

**Prototipo de prospectación comercial con inteligencia artificial para fortalecer la
competitividad de las PYMES en Colombia**

Roberto Bedoya García

Asesora

Arellys De Jesús Correa Rodríguez

Universidad Nacional Abierta y a Distancia UNAD

Escuela de Ciencias Básicas Tecnología e Ingeniería ECBTI

Programa Ingeniería de Sistemas

2025

Resumen

El proyecto presenta el desarrollo de un prototipo de automatización inteligente para la prospección de *leads* empresariales, orientado a optimizar la captación de clientes y fortalecer la competitividad de las pequeñas y medianas empresas (PYMES) en Colombia. La solución integra técnicas de *web Scraping* en Python y modelos de inteligencia artificial para la nutrición y calificación de prospectos, dentro de una arquitectura modular que garantiza escalabilidad e interoperabilidad con sistemas comerciales existentes. El prototipo automatiza la extracción, normalización y análisis de datos mediante técnicas de similitud semántica y algoritmos de clasificación de *leads*, permitiendo identificar prospectos calificados (*MQL* y *SQL*) y gestionarlos a través de interacciones personalizadas en canales de mensajería. La metodología aplicada combina el modelado de procesos de negocio (BPM) para el análisis funcional, el enfoque *CRISP-ML(Q)* para la preparación y evaluación de datos, y la metodología ágil Scrum para el desarrollo iterativo del sistema. Los resultados demuestran mejoras significativas frente a los procesos manuales en términos de precisión, eficiencia y experiencia de usuario, validando la pertinencia del uso de inteligencia artificial en la gestión comercial. En síntesis, el proyecto evidencia la viabilidad de una solución accesible, adaptable y de bajo costo, que contribuye a la transformación digital y la competitividad sostenible de las PYMES en contextos empresariales emergentes.

Palabras clave: Inteligencia artificial, automatización, prospección de *leads*, *web Scraping*, PYMES, Python.

Abstract

This project presents the development of an intelligent automation prototype for business lead prospecting, aimed at optimizing client acquisition and strengthening the competitiveness of small and medium-sized enterprises (SMEs) in Colombia. The solution integrates *web Scraping* techniques in Python and artificial intelligence models for lead nurturing and qualification, within a modular architecture that ensures scalability and interoperability with existing commercial systems. The prototype automates data extraction, normalization, and analysis through semantic similarity techniques and lead classification algorithms, allowing the identification of qualified prospects (MQL and SQL) and their management through personalized interactions on messaging channels. The applied methodology combines Business Process Modeling (BPM) for functional analysis, the CRISP-ML(Q) framework for data preparation and evaluation, and the agile Scrum methodology for iterative system development. The results show significant improvements compared to manual processes in terms of accuracy, efficiency, and user experience, validating the relevance of artificial intelligence in commercial management. In summary, the project demonstrates the feasibility of an accessible, adaptable, and cost-effective solution that contributes to digital transformation and sustainable competitiveness of SMEs in emerging business contexts.

Keywords: Artificial intelligence, automation, lead prospecting, *web Scraping*, SMEs, Python.

Tabla de contenido

Introducción	11
Justificación	14
Objetivos	16
Marco Referencial.....	17
Marco Metodológico.....	34
Diseño de Arquitectura Modular del Sistema.....	63
Preparación de Datos y Evaluación de Modelos de IA.....	70
Desarrollo Iterativo del Sistema.....	109
Validación del prototipo y evaluación de desempeño, eficiencia y experiencia de usuario.....	150
Resultados	159
Conclusiones	161
Recomendaciones	163
Referencias Bibliográficas	165
Apéndices.....	170

Lista de Tablas

Tabla 1 <i>Fichas de Antecedentes</i>	19
Tabla 2 <i>Conceptos desarrollados...</i>	25
Tabla 3 <i>Actores involucrados...</i>	42
Tabla 4 <i>Variables analizadas público objetivo.....</i>	43
Tabla 5 <i>Registro de la Aplicación de Técnicas.....</i>	44
Tabla 6 <i>Consolidación de respuestas recolectadas en las entrevistas..</i>	45
Tabla 7 <i>Identificación de tendencias y hallazgos..</i>	46
Tabla 8 <i>Resultados Cuantitativos encuestas semiestructuradas.....</i>	49
Tabla 9 <i>Hallazgos Claves..</i>	50
Tabla 10 <i>Herramientas Involucradas..</i>	52
Tabla 11 <i>Resumen de hallazgos por Actor o perfil de usuario.....</i>	55
Tabla 12 <i>Limitaciones del modelo actual de captación de leads</i>	55
Tabla 13 <i>Identificación de requisitos del sistema</i>	56
Tabla 14 <i>Propuesta de MVP.....</i>	61
Tabla 15 <i>Fuentes de datos identificadas y volumen de datos proyectados</i>	71
Tabla 16 <i>Compleitud de datos analizados después de la normalización</i>	91
Tabla 17 <i>Resumen de eliminación de datos duplicados</i>	93
Tabla 18 <i>Evaluación general</i>	94
Tabla 19 <i>Evaluación LLM's.....</i>	107
Tabla 20 <i>Propuesta estructurada del Product Backlog.....</i>	124
Tabla 21 <i>Planning del Spring 1.</i>	127
Tabla 22 <i>Información recopilada en el (Sprint Review).</i>	131

Tabla 23 <i>Planning del Spring 2.</i>	132
Tabla 24 <i>Información recopilada en el (Sprint 2 Review).</i>	136
Tabla 25 <i>Planning del Spring 3.</i>	137
Tabla 26 <i>Información recopilada en el (Sprint 3 Review).</i>	142
Tabla 27 <i>Planning del Spring 4.</i>	144
Tabla 28 <i>Información recopilada en el (Sprint 4 Review).</i>	147
Tabla 29 <i>Criterios y métricas de validación del prototipo.</i>	152
Tabla 30 <i>Resultados de validación del prototipo.</i>	153

Lista de Figuras

Figura 1	<i>Mini-Diagrama de Flujo de Conceptos</i>	28
Figura 2	<i>Gráfico del ciclo CRISP-ML(Q)</i>	38
Figura 3	<i>Esquema general metodología scrum</i>	39
Figura 4	<i>Resumen de datos recolectados en las encuestas</i>	46
Figura 5	<i>Flujo actual del proceso comercial de captación de leads</i>	51
Figura 6	<i>Diagrama BPMN comercial de captación de leads</i>	53
Figura 7	<i>Acceso y navegación general a la plataforma</i>	58
Figura 8	<i>Automatización comercial (Leads, Correos, Campañas)</i>	59
Figura 9	<i>Agentes automatizados y conectividad</i>	60
Figura 10	<i>Diagrama General de Arquitectura</i>	68
Figura 11	<i>Diagrama ERD del sistema HunterLeads</i>	69
Figura 12	<i>Filtro de los datos de Sale Navigator</i>	72
Figura 13	<i>Captura de links de Sales Navigator</i>	73
Figura 14	<i>Configuración del api de LinkedIn Sales Navigator Scraper</i>	74
Figura 15	<i>Registro de ejecución de iteraciones de Scraping.</i>	74
Figura 16	<i>Resultado de los datos obtenidos desde la api usada</i>	75
Figura 17	<i>Parte 1 del código de Scraping SalesQL</i>	77
Figura 18	<i>Parte 2 del código de Scraping SalesQL</i>	78
Figura 19	<i>Parte 3 del código de Scraping SalesQL</i>	79
Figura 20	<i>Resultado de los logs de ejecución del script</i>	80
Figura 21	<i>Evidencia resultados tabla empleados</i>	82
Figura 22	<i>Evidencia resultados tabla empresas</i>	82

Figura 23 <i>Reducción de registros antes y después de la limpieza</i>	83
Figura 24 <i>Estadística de limpieza del proceso de ML</i>	84
Figura 25 <i>Distribución de títulos laborales</i>	86
Figura 26 <i>Distribución de industrias</i>	87
Figura 27 <i>Algoritmo limpieza de datos</i>	87
Figura 28 <i>Algoritmo Normalización</i>	88
Figura 29 <i>Algoritmo detección de anomalías</i>	89
Figura 30 <i>Gráfico de dispersión - calidad vs eficiencia de limpieza</i>	91
Figura 31 <i>Gráfico dispersión - registros iniciales vs total de correcciones</i>	93
Figura 32 <i>Sentencias SQL</i>	96
Figura 33 <i>Gráfico de herramienta para la importación de datos en Supabase</i>	93
Figura 34 <i>Consultas sql utilizadas para las pruebas de integridad</i>	95
Figura 35 <i>Consultas sql utilizadas para las pruebas de codificación de caracteres</i>	99
Figura 36 <i>Consultas sql utilizadas para las pruebas de limpieza</i>	100
Figura 37 <i>Gráfico comprensión semántica</i>	104
Figura 38 <i>Gráfico de Capacidad de Análisis Predictivo</i>	105
Figura 39 <i>Gráfico de velocidad de respuesta para automatización</i>	105
Figura 40 <i>Gráfico de la integración con sistemas CRM existentes</i>	106
Figura 41 - 70 <i>Historias de usuario</i>	110
Figura 71 <i>Código clave Spring 1</i>	129
Figura 72 <i>Login HunterLeads</i>	130
Figura 73 <i>Menú HunterLeads</i>	130

Figura 74 <i>Código clave Spring 2</i>	134
Figura 75 <i>Busqueda leads HunterLeads y Agrupación de campañas HunterLead</i>	135
Figura 76 <i>Integración datos excel HunterLeads</i>	135
Figura 77 <i>Código clave Spring 2</i>	140
Figura 78 <i>Agentes (Correo, texto y llamadas) HunterLeads y análisis de resultados HunterLeads</i>	141
Figura 79 <i>Integraciones herramientas HunterLeads</i>	141
Figura 80 <i>Código clave Spring 3</i>	146
Figura 81 <i>Gestión de campañas HunterLeads y página de ayuda HunterLeads</i>	147
Figura 82 <i>Información histórica HunterLeads</i>	147
Figura 83 <i>Evaluación de tiempos en carga inicio de sesión</i>	153
Figura 84 <i>Evaluación de tiempos en carga búsqueda de leads</i>	153
Figura 85 <i>Formato de evaluación prueba plataforma base</i>	154
Figura 86 <i>Formato de evaluación prueba gestión de leads</i>	155
Figura 87 <i>Formato de evaluación prueba comunicación automatizada</i>	156
Figura 88 <i>Formato de evaluación prueba campañas y experiencia</i>	157

Lista de Apéndices

Apéndice A <i>Formato de entrevistas para levantamiento de requerimientos</i>	170
Apéndice B <i>Encuesta digital (Google Forms)</i>	172
Apéndice C <i>Historias de Usuario (Formato 3C)</i>	174
Apéndice D <i>Formato de Pruebas de Usuario</i>	175

Introducción

En el contexto actual de transformación digital y alta competitividad, la captación de clientes potenciales se ha convertido en un proceso esencial pero complejo dentro de las estrategias comerciales. La prospección de *leads*, entendida como la identificación y calificación de clientes potenciales, exige una alta inversión de tiempo y recursos, lo que representa un desafío para las pequeñas y medianas empresas (Ali Qalati et al., 2020). Aunque existen herramientas tecnológicas de apoyo, muchas carecen de automatización avanzada, limitando su efectividad y escalabilidad en entornos dinámicos. En este escenario, la inteligencia artificial (IA) y el procesamiento automatizado de información ofrecen nuevas oportunidades para optimizar procesos tradicionalmente manuales. Según (Zúñiga et al., 2023), las tecnologías basadas en IA permiten generar interacciones personalizadas y precisas con los públicos objetivos, potenciando la eficiencia comercial.

Este proyecto propone el desarrollo de un prototipo de automatización inteligente para la gestión de *leads* empresariales mediante Python e inteligencia artificial, integrando *Web Scraping*, análisis de datos y agentes conversacionales para la búsqueda, nutrición y calificación de prospectos. La propuesta responde tanto a una necesidad técnica como estratégica, al dirigir los esfuerzos hacia contactos con mayor potencial de conversión (Dwivedi et al., 2021), y busca reducir la brecha tecnológica en las PYMES colombianas, donde solo una minoría implementa soluciones avanzadas de TIC, pese a representar el 98% del tejido empresarial (Sarmiento Suárez et al., 2024).

Planteamiento del Problema

La captación de clientes potenciales constituye un reto estratégico para las empresas que buscan crecer en mercados altamente competitivos y digitalizados. En el caso de las pequeñas y medianas empresas (PYMES), gran parte de los procesos comerciales aún dependen de métodos manuales de prospección, como la búsqueda en directorios empresariales o contactos en redes sociales, lo que ocasiona lentitud, errores y pérdida de oportunidades relevantes (Ali Qalati et al., 2020). Este panorama se ve agravado por la creciente demanda de personalización por parte de los consumidores, quienes esperan interacciones adaptadas a sus necesidades específicas. Sin embargo, muchas organizaciones mantienen prácticas masivas y poco segmentadas, disminuyendo las probabilidades de conversión (Zúñiga et al., 2023).

Este problema tiene dos causas raíz principales. En primer lugar, la baja adopción de tecnologías de automatización y analítica en las PYMES, debido a limitaciones de recursos financieros y técnicos. En segundo lugar, la dependencia de procesos manuales que no escalan frente al volumen de datos ni a la necesidad de segmentación personalizada, lo que refuerza la brecha digital existente en el tejido empresarial colombiano (Sarmiento Suárez et al., 2024).

Si no se abordan estas limitaciones, las PYMES enfrentarán mayores costos de adquisición de clientes y menor rentabilidad, lo que limitará su crecimiento y su capacidad de competir frente a organizaciones con mayor grado de digitalización. A mediano plazo, esto podría traducirse en pérdida de participación en el mercado y en la reducción de oportunidades de expansión comercial. Ante este escenario, la investigación propone explorar el diseño y desarrollo de un prototipo de sistema de automatización inteligente que integre técnicas de inteligencia artificial y *web Scraping* para optimizar la identificación, nutrición y calificación de

leads. Se trata de una aproximación investigativa orientada a validar la viabilidad técnica y funcional de estas herramientas en un contexto de PYMES, más que en una solución comercial definitiva. Como señala Drucker (Zúñiga et al., 2023), el objetivo del *marketing* es conocer al cliente tan bien que el producto se venda solo, lo cual refuerza la pertinencia de explorar mecanismos basados en datos y personalización.

La investigación se enfoca en pequeñas y medianas empresas del sector comercial en Colombia durante el periodo 2025. La población objetivo está conformada por equipos de gestión comercial que buscan optimizar sus procesos de captación de clientes. Las variables de análisis incluyen nivel de adopción tecnológica, eficiencia del proceso de generación de *leads* y calidad de la información obtenida mediante técnicas automatizadas.

Pregunta Problemática

¿Cómo implementarse un prototipo de prospección comercial con inteligencia artificial que fortalezca la competitividad de las PYMES en Colombia mediante la optimización del proceso de captación de clientes?

Justificación

La captación de clientes constituye actualmente un desafío estratégico para las empresas, en especial para las pequeñas y medianas (PYMES) que operan en un entorno altamente competitivo y digitalizado. Muchas de ellas gestionan la identificación y nutrición de leads de manera manual o con herramientas no integradas, lo que genera ineficiencia, mayores costos de adquisición y pérdida de oportunidades comerciales. De no resolverse, esta situación seguirá ampliando la brecha entre PYMES y grandes empresas, afectando la sostenibilidad y el crecimiento del sector empresarial en Colombia (Kartajaya et al., 2021).

Los principales beneficiarios son las PYMES colombianas, que representan más del 90% del tejido empresarial nacional (Sarmiento Suárez et al., 2024). La implementación de un prototipo de prospección comercial inteligente les permitirá acceder a tecnologías avanzadas de automatización y analítica, tradicionalmente reservadas para grandes corporaciones.

Beneficiarios indirectos serán los equipos de ventas, al reducir su carga operativa y enfocarse en tareas de mayor valor, así como el entorno académico, al contar con un caso práctico de aplicación de inteligencia artificial en procesos comerciales.

El proyecto se fundamenta en marcos conceptuales de inteligencia artificial, agentes conversacionales y *web Scraping*, aplicados a la prospección de clientes. Se articula con metodologías de análisis y desarrollo de software como BPM y CRISP-ML(Q), que permiten estructurar los procesos de captación, preparación de datos y evaluación de modelos de IA de forma sistemática. Este enfoque no solo valida teorías existentes en un nuevo contexto —las PYMES colombianas— sino que también aporta evidencia práctica sobre la pertinencia de integrar tecnologías emergentes para resolver problemas empresariales reales (Sarker, 2022a).

Desde la perspectiva social y económica, la solución contribuye a democratizar el acceso a tecnologías de alto valor, fortaleciendo la competitividad de las PYMES y aportando al desarrollo productivo del país. Institucionalmente, el proyecto responde a los lineamientos de transformación digital en Colombia y se articula con los objetivos de innovación empresarial. Además, su carácter escalable permite adaptarlo a distintos sectores económicos sin requerir inversiones desproporcionadas, asegurando beneficios sostenibles en el tiempo (CONPES et al., s. f.).

En síntesis, el proyecto es pertinente porque responde a una necesidad concreta de las PYMES colombianas, viable porque se apoya en metodologías probadas y tecnologías accesibles, y relevante porque fortalece la competitividad empresarial en el marco de la transformación digital.

Objetivos

Objetivo General

Implementar un prototipo de prospectación comercial con inteligencia artificial que fortalezca la competitividad de las PYMES en Colombia mediante la optimización del proceso de captación de clientes.

Objetivos Específicos

Identificar los requisitos funcionales y no funcionales del sistema de automatización, así como las características clave para la gestión de *leads* empresariales mediante el análisis del proceso de captación con BPM y técnicas centradas en el usuario.

Diseñar una arquitectura modular e integrada que facilite la comunicación entre los modelos de inteligencia artificial y los componentes del sistema de automatización, garantizando escalabilidad y mantenimiento.

Preparar los datos obtenidos mediante técnicas de Scraping y evaluar modelos de inteligencia artificial adecuados para la automatización del proceso de captación y calificación de *leads*, considerando criterios de precisión, adaptabilidad e integración.

Desarrollar el prototipo del sistema de automatización inteligente para la captación y gestión de *leads* empresariales, siguiendo un enfoque iterativo basado en Scrum.

Validar el prototipo mediante pruebas controladas que permitan medir el desempeño de los modelos de IA, la eficiencia del sistema frente a procesos manuales y la experiencia de usuario, evaluando su aporte al fortalecimiento de la competitividad de las PYMES en Colombia.

Marco Referencial

El presente Marco Referencial constituye la base académica, conceptual y normativa que orienta el desarrollo de este proyecto de grado. Su importancia radica en que permite comprender el contexto en el que surge la propuesta, organizar los referentes previos que la sustentan y garantizar que la solución tecnológica planteada se formule bajo criterios de rigor investigativo, pertinencia y legalidad. En este sentido, el marco referencial asegura que el diseño e implementación del prototipo no solo respondan a necesidades empresariales inmediatas, sino que además se fundamenten en principios científicos, técnicos y jurídicos que validan su legitimidad y aplicabilidad.

Dado el carácter aplicado de la investigación, este capítulo cumple una función integradora al vincular la teoría con la práctica. A través de él, se establece la conexión entre los aportes académicos sobre inteligencia artificial, automatización de procesos y análisis de datos, y la implementación concreta de un prototipo que utiliza técnicas de *web Scraping*, mensajería automatizada y perfilamiento inteligente de *leads*. Así, el marco referencial ofrece un puente crítico entre el conocimiento acumulado y la creación de una herramienta tecnológica que se someterá a validación en un entorno organizacional real.

Para lograrlo, el capítulo se estructura en cinco apartados complementarios. En primer lugar, el estado del arte, que recopila investigaciones, estudios y experiencias previas relevantes sobre la captación y calificación de clientes con apoyo de inteligencia artificial. En segundo lugar, el marco contextual, que sitúa la propuesta en el entorno socioeconómico y organizacional donde se desarrollará. En tercer lugar, el marco conceptual, que define los términos clave y establece las relaciones entre ellos, facilitando un lenguaje común para el proyecto.

Posteriormente, el marco teórico, que expone los fundamentos y modelos científicos que sustentan el diseño de la solución. En cuarto lugar, el marco legal, que reúne las disposiciones normativas y éticas aplicables al uso de datos, comunicaciones digitales y desarrollo de software. Finalmente, se presenta un cierre integrador, que sintetiza cómo todos estos elementos respaldan los objetivos, la metodología y la propuesta de solución tecnológica.

Antecedentes

El presente apartado tiene como propósito identificar y analizar antecedentes relevantes relacionados con la captación de clientes mediante inteligencia artificial, aprendizaje automático, *web Scraping* y automatización de *marketing*. La revisión abarca investigaciones académicas, desarrollos tecnológicos y aplicaciones empresariales de los últimos quince años, enfocándose en la integración de tecnologías avanzadas y su aplicabilidad en el contexto de pequeñas y medianas empresas (PYMES) de economías emergentes. El análisis de antecedentes permite reconocer tendencias, limitaciones y vacíos en la literatura y en proyectos previos, y proporciona insumos esenciales para justificar la pertinencia del presente proyecto. Asimismo, se busca establecer una base metodológica sólida que conecte la teoría con la práctica, orientando el diseño de una solución tecnológica que integre automatización, mensajería omnicanal y procesos inteligentes de calificación de *leads*.

Revisión de Antecedentes

Tabla 1

Fichas de antecedentes

Título	Autores / Año	Tecnologías	Alcance	Limitaciones	Aprendizajes Implicaciones
<i>Digital Marketing and Artificial Intelligence Integration</i>	(Chaffey & Ellis-Chadwick, 2019)	Algoritmos de IA para segmentación y analítica de clientes	Evaluación del impacto de la IA en campañas de <i>marketing</i> digital	Enfoque limitado a mercados europeos; no considera PYMES emergentes	Demuestra cómo la IA mejora la precisión en la identificación de <i>leads</i> , fundamentando la necesidad de adaptar estas herramientas a otros contextos
<i>Artificial Intelligence in Marketing: A Systematic Literature Review</i>	(Ricci, 2022)	Algoritmos de <i>machine learning</i> aplicados al comportamiento del consumidor	Revisión de literatura 2010–2022 sobre personalización y automatización	Carencia de estudios que integren IA con <i>web Scraping</i> y análisis de datos abiertos	Confirma la relevancia de la personalización y automatización en <i>marketing</i> digital
<i>Conversational AI in Customer Engagement</i>	(Barría, 2023)	Chatbots y asistentes virtuales basados en PLN	Evaluación de la efectividad de chatbots para mejorar experiencia del cliente	No aborda integración con <i>Scraping</i> ni PYMES	Evidencia que los agentes conversacionales mejoran la interacción y satisfacción del cliente

Nota. Tabla que muestra la comparativa entre los estudios y proyectos tomados como

antecedentes. *Fuente.* Autor

Análisis Comparativo

La revisión de los antecedentes permite identificar importantes coincidencias en torno al uso de tecnologías emergentes para la gestión comercial. La inteligencia artificial (IA) y el aprendizaje automático se destacan como herramientas clave para mejorar la eficiencia y precisión en la captación y calificación de *leads*, al tiempo que la personalización y la

automatización se consolidan como factores de diferenciación competitiva, favoreciendo interacciones más efectivas con los clientes y un incremento en las tasas de conversión. De igual manera, los agentes conversacionales y chatbots surgen como recursos valiosos para optimizar la comunicación y fortalecer la experiencia del usuario, aportando agilidad y cercanía en los procesos de atención comercial.

Sin embargo, persisten vacíos significativos que justifican la pertinencia del presente proyecto. Se evidencia una escasa integración entre la inteligencia artificial, el *web Scraping* y los sistemas de mensajería omnicanal en la literatura y las soluciones existentes. Además, la mayoría de los estudios se centran en grandes empresas y mercados desarrollados, dejando de lado la aplicabilidad en pequeñas y medianas empresas latinoamericanas, las cuales enfrentan limitaciones de recursos y capacidades técnicas. Igualmente, se ha prestado poca atención a la usabilidad de APIs de mensajería como WhatsApp o Telegram para la interacción automatizada con *leads*. En consecuencia, este proyecto propone una solución accesible, escalable y adaptada al contexto colombiano, orientada a fortalecer la competitividad y la transformación digital de las PYMES mediante la automatización inteligente de la prospección comercial.

Inspiración para el Proyecto

Del análisis de los antecedentes revisados se identifican diversos elementos que inspiran y orientan el desarrollo del presente proyecto. La implementación de módulos de automatización de *marketing* basados en inteligencia artificial, como se evidencia en el caso de Azalea 211, demuestra que la combinación de algoritmos de clasificación y segmentación de clientes incrementa significativamente la tasa de conversión, resaltando la importancia de incorporar procesos inteligentes que no se limiten a la extracción de datos, sino que los transformen en información procesable para la toma de decisiones comerciales (Zúñiga et al., 2023). Asimismo,

los aportes metodológicos de estudios académicos que integran modelos de personalización en campañas digitales muestran que la efectividad de estas herramientas no depende únicamente del volumen de *leads* captados, sino de la pertinencia de los mensajes frente a las necesidades e intereses de cada usuario (Dwivedi et al., 2021). De igual manera, experiencias previas que aplican técnicas de *web Scraping* con fines de prospección comercial aportan buenas prácticas en la recolección, limpieza y estructuración de datos, sirviendo como referente para el diseño técnico de la solución propuesta (Ali Qalati et al., 2020).

Sin embargo, se observa que la mayoría de las iniciativas analizadas carecen de un enfoque orientado a las pequeñas y medianas empresas en contextos latinoamericanos, donde la baja adopción tecnológica limita la competitividad. Esta brecha constituye, al mismo tiempo, un vacío y una oportunidad de innovación, ya que permite adaptar soluciones avanzadas a realidades locales caracterizadas por restricciones de recursos, costos y capacidades técnicas (Sarmiento Suárez et al., 2024). En conclusión, la revisión de antecedentes respalda la pertinencia de un proyecto que integre inteligencia artificial, *web Scraping* y mensajería automatizada en un modelo accesible, escalable y contextualizado, contribuyendo así a la democratización del uso de tecnologías avanzadas en la captación y gestión de clientes dentro de las PYMES colombianas.

Marco Teórico

El marco teórico constituye el fundamento conceptual sobre el cual se sustenta el presente proyecto de desarrollo de software orientado a la captación automatizada de clientes mediante inteligencia artificial y técnicas de *web Scraping*. En este apartado se recopilan y analizan teorías, enfoques y modelos relacionados con la computación, la arquitectura de software, la experiencia de usuario, las metodologías de desarrollo y el *marketing* digital. Dichos

elementos permiten orientar las decisiones de diseño, implementación y validación del sistema propuesto, garantizando su pertinencia técnica y aplicabilidad en el contexto empresarial actual.

Inteligencia Artificial y Aprendizaje Automático.

La inteligencia artificial (IA) busca emular capacidades humanas como el razonamiento, la toma de decisiones o el reconocimiento de patrones, mediante algoritmos y modelos matemáticos (Sarker, 2022a). Dentro de esta, el aprendizaje automático (*machine learning*, ML) permite que los sistemas mejoren su desempeño a partir de la experiencia, identificando patrones en grandes volúmenes de datos (Sarker, 2022b). Estas disciplinas resultan esenciales en la clasificación y calificación de *leads*, dado que permiten priorizar prospectos con mayor probabilidad de conversión.

Procesamiento de Lenguaje Natural (PLN) y Embeddings.

El PLN constituye una rama de la IA que habilita la comprensión y generación automática de lenguaje humano (Jurafsky & Martin, s. f.). Los modelos de tipo *transformers*, como los *Sentence Transformers*, permiten representar oraciones en vectores semánticos (*embeddings*), facilitando tareas de similitud, clasificación y búsqueda de información contextualizada. Esta tecnología resulta clave en la interacción automatizada con *leads* y en la personalización de mensajes comerciales.

Modelos de Clasificación: Regresión Logística y Árboles de Decisión.

Entre los algoritmos de clasificación, la regresión logística es ampliamente utilizada para problemas de predicción binaria y *lead scoring*, al estimar probabilidades de (Sarker, 2021). Los árboles de decisión y redes neuronales complementan este enfoque al identificar patrones complejos en los datos de interacción y comportamiento de clientes.

Arquitectura de Microservicios y APIs

La arquitectura de microservicios propone dividir una aplicación en componentes pequeños, independientes y especializados, que se comunican a través de APIs (Newman, 2021). Esta aproximación asegura escalabilidad, resiliencia y flexibilidad en el mantenimiento. Para el presente proyecto, este estilo arquitectónico facilita que los módulos de Scraping, calificación automática y comunicación se gestionen de forma desacoplada.

Integración y Automatización de Procesos

El uso de APIs de mensajería (ej. WhatsApp Business API, Twilio, SendGrid) permite la integración de flujos de contacto automatizados con clientes potenciales, mejorando la eficiencia y la capacidad de respuesta en tiempo real (Wisnalmawati et al., 2021).

ISO/IEC 25010

Establece un marco de calidad del producto software basado en atributos como seguridad, mantenibilidad, eficiencia y fiabilidad (*ISO/IEC 25010*, 2011)

ISO 9241-210

Define los principios del diseño centrado en el usuario (DCU), orientando la experiencia de usuario hacia la satisfacción y la eficacia (ISI, 2019)

IEEE/ISO 29148

Estandariza la gestión y especificación de requisitos de software, asegurando claridad y verificabilidad (*ISO/IEC/IEEE 29148*, 2018).

Heurísticas de Nielsen

Proporcionan lineamientos prácticos para evaluar la usabilidad, como visibilidad del estado del sistema y consistencia en la interfaz (Fernández & Macías, 2021).

SCRUM

Metodología ágil que facilita entregas incrementales y adaptación a cambios de requisitos (Heath, 2021).

BPM (Business Process Management)

Enfoque orientado a la optimización y automatización de procesos de negocio, alineando el software con los objetivos organizacionales (Dumas et al., 2018).

CRISP-ML(Q)

Metodología para proyectos de *machine learning* que integra validación de calidad en todas las fases, desde la comprensión de datos hasta el despliegue (Purbasari et al., 2021).

Marketing Relacional

El *marketing* relacional busca establecer vínculos de largo plazo con los clientes mediante personalización y generación de valor (Kotler et al., 2022).

Lead Scoring

El *lead scoring* consiste en asignar una puntuación a cada prospecto en función de su nivel de interés y probabilidad de conversión, utilizando criterios demográficos, de comportamiento e interacciones previas. Este modelo constituye la base teórica que sustenta la automatización del embudo de ventas en el presente proyecto.

En síntesis, el marco teórico articula fundamentos provenientes de la inteligencia artificial, el aprendizaje automático, el procesamiento de lenguaje natural, la arquitectura de microservicios, la interacción humano-computador y las metodologías de desarrollo ágil. A ello se suman teorías del *marketing* relacional y del *lead scoring*, que permiten contextualizar el uso de estas tecnologías en la gestión comercial. La integración de estos enfoques garantiza que la solución propuesta no solo cumpla con altos estándares técnicos de calidad y usabilidad, sino que

también responda a las necesidades del entorno empresarial, ofreciendo un sistema innovador, escalable y adaptado a las condiciones locales.

Marco Conceptual

El marco conceptual establece las definiciones de los términos clave que serán utilizados en el desarrollo del presente proyecto. No se trata únicamente de citar conceptos generales, sino de contextualizarlos según el alcance, los objetivos y el enfoque de la solución tecnológica propuesta. De esta manera, se garantiza un lenguaje común y una comprensión compartida entre el equipo de desarrollo, los actores académicos y los potenciales usuarios de la herramienta.

Conceptos desarrollados

Tabla 2

Conceptos desarrollados

Concepto	Definición	Referente
<i>Agentes de IA (Agents)</i>	Entidades de software autónomas que perciben su entorno, razonan y ejecutan acciones independientes para alcanzar objetivos específicos. En este proyecto, actúan como componentes inteligentes que redactan mensajes y gestionan flujos de prospección sin intervención humana constante.	(Ghimire, 2020)
<i>API (Application Programming Interface)</i>	Interfaz de programación que actúa como puente para que el prototipo consuma servicios externos (como modelos de lenguaje o CRMs). Permite la integración modular de capacidades de IA de terceros en la arquitectura local del sistema de manera segura y eficiente.	Propia
<i>Árboles de Decisión y Redes Neuronales</i>	Modelos de aprendizaje supervisado que identifican patrones no lineales en grandes volúmenes de datos. Los árboles segmentan la información mediante reglas lógicas, mientras que las redes neuronales imitan el procesamiento biológico para tareas complejas de predicción de ventas.	(Sarker, 2021)
<i>Arquitectura de Microservicios y APIs</i>	Método de desarrollo de software donde la aplicación se divide en servicios independientes comunicados por APIs. Esto garantiza que el sistema sea escalable y que la actualización de un módulo de IA no afecte la disponibilidad total de la solución tecnológica.	(Newman, 2021)
<i>CAC (Costo de Adquisición de Clientes)</i>	Métrica financiera que suma la inversión total en marketing y ventas para captar un nuevo cliente. La automatización propuesta busca optimizar este indicador reduciendo el tiempo operativo y el error humano en la fase inicial del embudo.	(Kotler et al., 2022)

<i>Calificación de Leads (Lead Scoring)</i>	Metodología de análisis que asigna un puntaje numérico a los prospectos basado en su comportamiento digital y datos demográficos. Permite priorizar aquellos con mayor probabilidad de cierre mediante el uso de algoritmos predictivos.	Propia
<i>Captación de Clientes</i>	Proceso estratégico que utiliza herramientas digitales para identificar y atraer nuevos prospectos. En 2026, este proceso se apoya en una mentalidad "AI-first" para personalizar la oferta de valor desde el primer contacto de manera masiva.	(Creswell & Clark, 2017)
<i>Crawling</i>	Navegación automatizada a gran escala realizada por bots para recorrer enlaces e indexar contenido web. A diferencia del scraping, el crawling se enfoca en la exploración y descubrimiento de nuevas fuentes de información comercial.	Propia
<i>CTR (Click Through Rate)</i>	Indicador de rendimiento que mide el porcentaje de clics obtenidos respecto al número de visualizaciones de un mensaje o anuncio. Es vital para evaluar la efectividad semántica de los mensajes generados por la inteligencia artificial.	Propia
<i>Data Cleaning</i>	Fase de depuración técnica donde se eliminan inconsistencias, duplicados y errores de formato en los datos captados. Es un paso crítico dentro del marco CRISP-ML(Q) para garantizar la calidad de la información procesada por los modelos.	(Studer et al., 2021)
<i>Embeddings</i>	Representaciones vectoriales de conceptos o textos en un espacio multidimensional que permiten a la IA medir la similitud semántica. Facilitan tareas de búsqueda, clasificación y personalización de mensajes al entender el contexto profundo de las palabras.	Propia
<i>Escalabilidad</i>	Capacidad de un sistema informático para crecer y gestionar una carga de trabajo creciente (usuarios o datos) sin comprometer su rendimiento. El diseño horizontal permite añadir más nodos de procesamiento a medida que la PYME expande su operación.	(Lv et al., 2025)
<i>Funnel o Embudo de Ventas</i>	Modelo que representa el viaje del cliente desde el conocimiento de la marca hasta la conversión final. La automatización integra el flujo de datos en cada etapa para evitar la fuga de prospectos y mejorar la tasa de éxito.	Propia
<i>Integración y Automatización de Procesos</i>	Uso de software (como RPA o n8n) para ejecutar tareas repetitivas y conectar aplicaciones diversas. Busca liberar al personal humano de trabajos tediosos, reduciendo tiempos de respuesta y errores manuales en la gestión de leads.	(Rosemann et al., 2024)
<i>Interoperabilidad</i>	Propiedad de los sistemas para comunicarse e intercambiar datos de forma transparente. En este proyecto, se garantiza que los agentes de IA interactúen con bases de datos SQL y plataformas externas de comunicación sin fricciones técnicas.	(ISO, 2022)
<i>Lead</i>	Individuo u organización que ha manifestado interés voluntario en un producto o servicio al entregar sus datos de contacto. Es el insumo básico de cualquier estrategia de prospección comercial automatizada.	Propia
<i>Modelos de Lenguaje (LLMs)</i>	Algoritmos de inteligencia artificial entrenados para comprender y generar lenguaje natural coherente. Son el motor principal del prototipo para la redacción automática	Propia

	de correos comerciales adaptados al perfil del cliente potencial.	
<i>MQL (Marketing Qualified Lead)</i>	Prospecto que ha sido calificado por el equipo de marketing como apto debido a su nivel de interés y coincidencia con el cliente ideal, requiriendo ser nutrido antes de pasar a una etapa de venta directa.	Propia
<i>Normalización de Datos</i>	Técnica de organización de datos que busca eliminar redundancias y asegurar que la información sea uniforme, permitiendo que los algoritmos de similitud semántica funcionen con mayor precisión.	Propia
<i>Nutrición de Leads (Lead Nurturing)</i>	Estrategia automatizada de acompañamiento mediante contenido de valor que busca educar al prospecto y aumentar su confianza hasta que esté listo para realizar una transacción comercial.	(Grewal et al., 2025)
<i>Procesamiento de Lenguaje Natural (PLN)</i>	Campo de la IA que permite a las computadoras entender, interpretar y generar lenguaje humano. Se aplica para analizar respuestas de prospectos y clasificar automáticamente sus intenciones de compra.	(Jurafsky & Martin, s. f.)
<i>Regresión Logística</i>	Algoritmo estadístico utilizado para predecir la probabilidad de que un lead realice una acción binaria (como "comprar" o "no comprar"), sirviendo como base para modelos simplificados de calificación.	(Sarker, 2021)
<i>Scraping</i>	Técnica programática de extracción de datos específicos y estructurados desde sitios web. En el proyecto, se emplea para recolectar información pública de contacto y perfiles comerciales de forma masiva y eficiente.	(Dwivedi et al., 2021)
<i>Seguridad de la Información</i>	Conjunto de medidas técnicas y administrativas destinadas a proteger la confidencialidad, integridad y disponibilidad de los datos personales captados, cumpliendo con marcos legales como la Ley 1581 en Colombia.	(ISO, 2022)
<i>Similitud Semántica</i>	Medida que determina qué tan parecidos son dos textos en cuanto a su significado profundo, permitiendo emparejar las necesidades de un cliente con la solución tecnológica de la PYME de forma inteligente.	(Echeverri Cañas, 2023)
<i>SQL (Sales Qualified Lead)</i>	Lead validado por el equipo de ventas como un prospecto con alta intención de compra inmediata y que cumple con todos los requisitos para iniciar una negociación comercial directa.	Propia
<i>SUS (System Usability Scale)</i>	Escala psicométrica estandarizada utilizada para medir de forma cuantitativa la percepción de usabilidad del usuario final, garantizando que el prototipo sea fácil de operar y eficiente.	Propia
<i>Tasa de Conversión (Conversion Rate)</i>	Métrica de impacto que calcula el porcentaje de leads que completan una acción deseada. Es el indicador principal para validar si la automatización mejora los resultados comerciales frente a procesos manuales.	(Zúñiga et al., 2023)
<i>Workflow (Flujo de Trabajo)</i>	Secuencia de tareas automatizadas y reglas lógicas que guían el movimiento de la información dentro del sistema. Representa la ruta estructurada desde la captación del lead hasta su calificación final.	Propia

Nota. Tabla que muestra los conceptos técnicos más usados en el desarrollo del trabajo. *Fuente.*

Autor

Relación entre Conceptos

Figura 1

Mini-diagrama de flujo de conceptos



Fuente. Autoría Propia

Los conceptos definidos en este marco no actúan de manera aislada, sino que conforman un entramado que sustenta la solución tecnológica propuesta. Su articulación permite comprender cómo las piezas técnicas, de *marketing* y de gestión de datos se integran para dar respuesta a la problemática de captación automatizada de clientes.

En primer lugar, los procesos de *crawling* y *Scraping* constituyen el punto de partida en la obtención de datos de clientes potenciales desde fuentes externas. La información recolectada atraviesa luego etapas de *data cleaning*, normalización y ETL, lo que asegura que los registros

estén estructurados y listos para su análisis. Esta secuencia evidencia la relación directa entre la gestión de datos y la calidad de los *leads* que ingresan al sistema.

Una vez normalizados, los datos se convierten en insumos para procesos de calificación (*lead scoring*) y segmentación. Aquí entran en juego técnicas de inteligencia artificial, como los *embeddings* y la similitud semántica, que permiten evaluar qué tan cercanos son los intereses o comportamientos del lead respecto a la oferta de la empresa. De esta manera, los *leads* pueden ser clasificados como MQL o SQL, dependiendo de su grado de madurez en el embudo de ventas.

Los agentes de software y los *workflows* coordinan las interacciones con los *leads* a lo largo del proceso de nutrición. A través de modelos de lenguaje (LLMs), los agentes pueden generar mensajes personalizados que incrementan la probabilidad de conversión, mejorando así métricas como el CTR y la tasa de conversión. Estos flujos automatizados integran tanto la lógica técnica (automatización y agentes) como la lógica de negocio (nutrición, *scoring*, *funnel*).

El concepto de API y la interoperabilidad garantizan que el sistema pueda integrarse con plataformas externas de gestión comercial, como los CRM. Esto es fundamental para que los *leads* procesados se transfieran sin fricciones al área de ventas, manteniendo consistencia en todo el ciclo de captación y conversión.

Finalmente, los aspectos de escalabilidad y seguridad de la información atraviesan de manera transversal todos los elementos conceptuales analizados. La escalabilidad garantiza que el sistema pueda crecer en volumen de *leads* y usuarios sin afectar su rendimiento, mientras que la seguridad asegura la protección de los datos sensibles, aspecto crítico tanto en el plano técnico como en el normativo. En este sentido, el presente marco conceptual no solo organiza y contextualiza los principales términos asociados a la captación de clientes mediante herramientas

de inteligencia artificial, sino que también establece la relación entre los conceptos técnicos y de *marketing* digital que sustentan el sistema propuesto. Así, este glosario relacional se configura como una base sólida para el diseño, desarrollo e implementación del proyecto aplicativo universitario, orientando su construcción desde una perspectiva integral que combina eficiencia, seguridad y coherencia conceptual.

Marco Legal

El desarrollo de un sistema de captación y calificación de *leads* basado en inteligencia artificial y técnicas de automatización requiere enmarcarse en un conjunto de disposiciones legales y regulatorias que garanticen su legitimidad, validez y sostenibilidad. En este tipo de proyectos, el marco legal no solo protege los derechos de los usuarios y las obligaciones de los desarrolladores, sino que también delimita la manera en que se deben gestionar aspectos sensibles como el tratamiento de datos personales, la recolección de información mediante técnicas de Scraping, la comunicación automatizada con *leads* y la validez de las transacciones digitales. En consecuencia, este apartado identifica y clasifica las normas que regulan tanto el ámbito tecnológico como el sector comercial y ético del proyecto, asegurando un cumplimiento integral que fortalezca la confianza de los usuarios y la pertinencia de la solución.

Marco Legal del Dominio o Tema de Estudio

En el contexto del *marketing* digital y la gestión de clientes potenciales, la normativa colombiana establece directrices claras para el manejo de la información y las interacciones comerciales. Entre ellas destacan:

Ley 1581 de 2012 – Protección de Datos Personales: regula la autorización, almacenamiento y uso de información personal, garantizando principios de legalidad, finalidad y seguridad (Colombia, 2012).

Decreto 1377 de 2013: reglamenta la Ley 1581, con énfasis en el tratamiento de bases de datos recolectadas antes de su expedición.

Ley 1266 de 2008 – Habeas Data Financiero: orientada a información financiera, pero relevante por sus principios de veracidad y actualización aplicables a procesos de perfilamiento comercial (Pública, 2008) .

Ley 527 de 1999 – Comercio Electrónico: establece la validez de mensajes de datos y firmas digitales en interacciones comerciales automatizadas (Valencia Agudelo, 2022).

Estas disposiciones impactan directamente en la forma en que se recolecta, procesa y utiliza la información de *leads* en el entorno digital colombiano.

Marco Legal Tecnológico

Dado que este proyecto se basa en Scraping, mensajería automática y algoritmos de calificación, es necesario contemplar normas internacionales y estándares técnicos de referencia:

Propiedad intelectual y licenciamiento de software: uso de librerías, *frameworks* y APIs de terceros bajo licencias *open source* (ej. MIT, Apache, GPL) debe respetarse para garantizar interoperabilidad y evitar conflictos legales.

Normas de protección de datos a nivel internacional: el Reglamento General de Protección de Datos (GDPR, Unión Europea) regula el uso de datos personales con aplicación extraterritorial, lo cual es relevante si se procesan *leads* fuera de Colombia. Asimismo, legislaciones como la CAN-SPAM Act (EE. UU.) o la CASL (Canadá) establecen obligaciones específicas sobre el envío de mensajes comerciales no solicitados.

Estándares de seguridad de la información: normas como (ISO, 2022) (gestión de seguridad de la información) y (IEEE, 2017) (procesos de ciclo de vida del software) sirven como referencia para garantizar la integridad, disponibilidad y calidad del sistema.

Marco Legal Aplicado al Proyecto

En el caso específico de esta propuesta de software, se destacan las siguientes disposiciones aplicables de manera directa:

GDPR (UE): en caso de captación de *leads* fuera de Colombia, la aplicación debe incorporar mecanismos de consentimiento explícito y derecho al olvido, dado que el reglamento aplica extraterritorialmente.

CAN-SPAM (EE. UU.) y CASL (Canadá): obligan a que los mensajes automatizados de correo electrónico incluyan mecanismos de exclusión voluntaria (opt-out).

Código de Ética Profesional del Ingeniero (COPNIA, 2008) : obliga a que el diseño e implementación del sistema se realicen con responsabilidad, transparencia y equidad, evitando manipulación algorítmica o prácticas comerciales abusivas.

Lineamientos institucionales de la universidad: establecen criterios éticos y de buenas prácticas en el desarrollo de proyectos tecnológicos, asegurando su pertinencia social y académica.

En síntesis, el marco legal de este proyecto articula disposiciones nacionales e internacionales que regulan el uso de datos, la validez de las interacciones comerciales digitales y la responsabilidad ética de los desarrolladores. El cumplimiento de estas normas no solo evita riesgos legales, sino que garantiza que la solución tecnológica propuesta opere bajo criterios de legitimidad, confianza y protección de los derechos de los usuarios, lo cual resulta indispensable en un sistema que involucra *Scraping*, mensajería automatizada y algoritmos de perfilamiento comercial.

El marco referencial constituye la base que orienta y legitima el desarrollo del proyecto, al integrar antecedentes, contexto, teorías y disposiciones normativas en una estructura

coherente. El estado del arte permitió identificar los principales avances en el uso de inteligencia artificial y técnicas de *web Scraping* para la captación y calificación de clientes, mientras que el marco contextual y conceptual situaron la investigación en su entorno real, definiendo las necesidades, los conceptos clave y un lenguaje técnico común que facilita la comunicación entre los diferentes actores del proyecto.

Por su parte, los aportes del marco teórico y legal proporcionaron los fundamentos científicos y normativos que sustentan los procesos de automatización comercial, perfilamiento de *leads* y protección de datos personales. En conjunto, este marco referencial ofrece la justificación y las directrices metodológicas necesarias para implementar un enfoque mixto basado en BPM, Scrum y el ciclo de vida CRISP-ML(Q), garantizando la solidez académica, la validez técnica y la aplicabilidad práctica del prototipo inteligente propuesto.

Marco Metodológico

El presente proyecto se fundamenta dentro de una investigación aplicada, cuyo propósito es ofrecer una solución práctica al problema de la prospección comercial en PYMES colombianas mediante un prototipo basado en inteligencia artificial y técnicas de automatización. El marco metodológico articula las decisiones relativas al tipo de estudio, enfoque investigativo, técnicas e instrumentos de recolección de datos, metodologías empleadas y fases de ejecución. De esta forma, se busca garantizar un proceso sistemático, validado y ajustado al contexto real, que permita la generación de resultados confiables y pertinentes para el ámbito empresarial.

Tipo de Estudio

La investigación es de tipo aplicado, n tanto se orienta a resolver un problema concreto del contexto empresarial: la baja eficiencia en la captación y gestión de *leads* en PYMES. Como señalan (Hernández-Sampieri & Mendoza, 2020) y Bibliotecas Duoc UC (2021), los estudios aplicados buscan generar conocimiento útil y transformador en situaciones específicas, priorizando la transferencia de soluciones hacia escenarios prácticos.

Enfoque

El proyecto adopta un enfoque mixto que combina componentes cuantitativos y cualitativos para obtener una comprensión integral del problema. Desde el componente cuantitativo, se recopilan y analizan métricas de rendimiento de los modelos de inteligencia artificial, como precisión, *recall* y tiempos de respuesta; junto con indicadores de experiencia de usuario obtenidos mediante encuestas SUS. En el componente cualitativo, se aplican entrevistas semiestructuradas y sesiones de retroalimentación con usuarios, orientadas a explorar sus expectativas, percepciones y posibles barreras en el proceso de captación de clientes. Esta

integración metodológica permite complementar la objetividad de los datos con la riqueza interpretativa de la experiencia humana, logrando una visión más completa y equilibrada del fenómeno estudiado (Creswell & Clark, 2017).

Población y Muestra

Población

La población objeto de esta investigación está constituida por las Pequeñas y Medianas Empresas (PYMES) en Colombia que presentan desafíos en sus procesos de captación y gestión de prospectos comerciales (*leads*) y que buscan la transformación digital mediante tecnologías emergentes.

Muestra y Tipo de Muestreo

Dado que la investigación busca desarrollar y validar un prototipo funcional en un entorno real y controlado, se seleccionó una muestra de tipo no probabilística por conveniencia o intencional (Hernández-Sampieri & Mendoza, 2020). Este tipo de muestreo se fundamenta en la elección de unidades que cumplen con criterios específicos de relevancia para el fenómeno estudiado y que permiten el acceso a la información necesaria para el diseño del sistema.

En este sentido, la muestra se concentra en una unidad de análisis única: una PYME del sector tecnológico ubicada en Bogotá. Los criterios de inclusión para la selección de esta unidad fueron:

Contar con un proceso de prospección comercial activo pero con baja automatización.

Disponibilidad de datos históricos para la fase de entrenamiento de modelos de IA bajo el marco CRISP-ML(Q).

Apertura de la gerencia para la implementación y evaluación de herramientas basadas en inteligencia artificial.

Al tratarse de un estudio aplicado con enfoque de diseño y desarrollo de software, el trabajo con una sola organización permite un análisis profundo y una validación iterativa (Scrum) que garantiza que el prototipo responda a las necesidades reales del sector antes de una escalabilidad mayor.

Técnicas e Instrumentos de Recolección de Información

Para garantizar una aproximación integral, se emplearon las siguientes técnicas:

Entrevistas semiestructuradas a usuarios y representantes de una PYME del sector tecnológico, con el fin de identificar requerimientos funcionales y no funcionales del sistema (Stang et al., 2023).

Encuesta digital por medio del uso de la herramienta Google forms orientada a identificar cómo se realiza actualmente la captación y gestión de *leads*, dónde están las dificultades y qué oportunidades de mejora existen.

Observación no participante durante las sesiones de interacción con el personal de *marketing* de la PYME evaluada, para identificar patrones de uso y dificultades recurrentes.

Metodologías

Este proyecto adopta una metodología híbrida, integrando tres marcos complementarios: BPM, CRISP-ML(Q) y Scrum. Cada uno se aplica en fases específicas del proyecto, garantizando coherencia entre el análisis de procesos, la preparación y validación de modelos de IA, y el desarrollo iterativo e incremental del sistema.

Business Process Management (BPM)

La metodología BPM busca modelar, analizar, optimizar y administrar procesos de negocio para elevar su eficiencia y alinearlos con los objetivos estratégicos empresariales. En los últimos años, BPM ha evolucionado para integrarse con tecnologías emergentes como la

inteligencia artificial, la minería de procesos y enfoques de gobernanza adaptativa (Rosemann et al., 2024).

En este proyecto, BPM se aplica en la fase inicial con los siguientes propósitos:

1. Representar gráficamente los procesos actuales de captación de *leads* en PYMES.
2. Identificar ineficiencias, cuellos de botella y duplicidades operativas.
3. Definir requerimientos funcionales y no funcionales alineados con las necesidades reales de los usuarios.
4. Establecer un punto de partida claro para la integración futura de IA y automatización en los flujos comerciales.

De esta forma, BPM funciona como un puente entre la visión de gestión empresarial y el diseño tecnológico, permitiendo que la solución prototípica se construya sobre un entendimiento riguroso del proceso real.

CRISP-ML(Q)

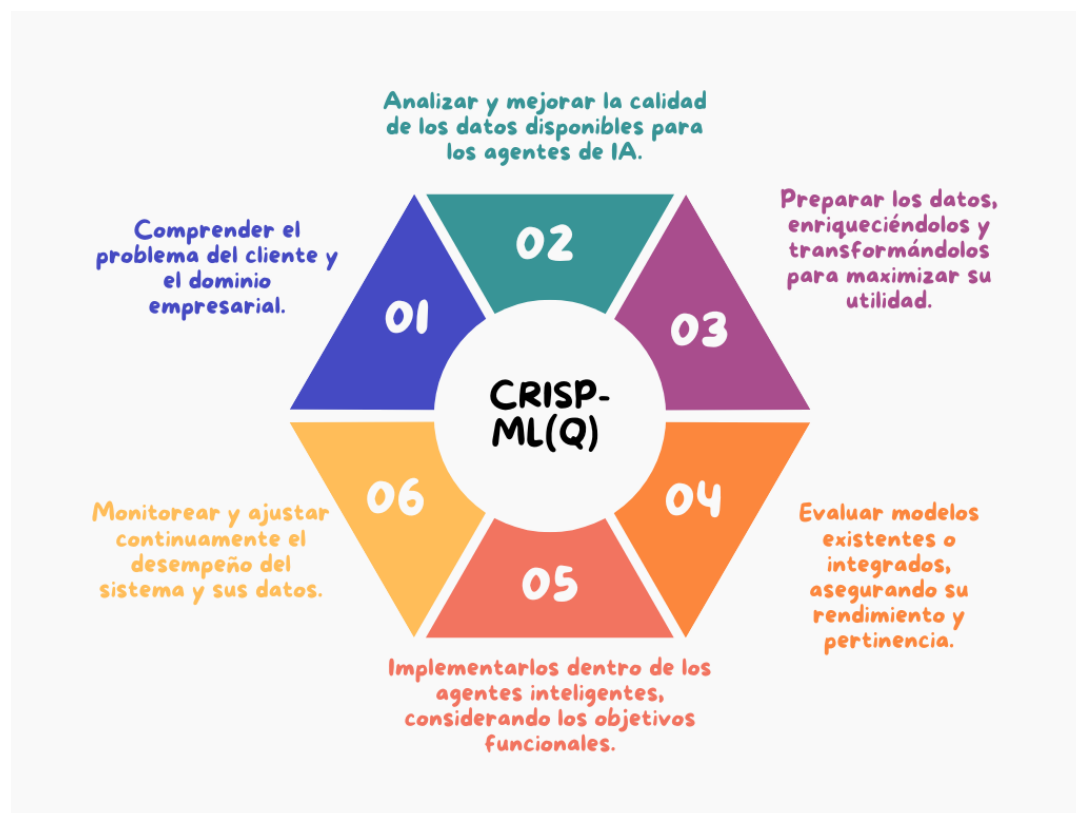
Cross Industry Standard Process for Machine Learning with Quality Assurance (CRISP-ML(Q)) es un proceso metodológico diseñado para proyectos de inteligencia artificial que exige altos estándares de calidad, trazabilidad y cumplimiento ético. Incorpora fases de comprensión del negocio, preparación de datos, evaluación de modelos y seguimiento post-despliegue, asegurando resultados confiables y reproducibles (Studer et al., 2021). Sus fases abarcan:

1. Comprensión del negocio y del dominio.
2. Análisis y aseguramiento de la calidad de los datos.
3. Preparación y transformación de datos.
4. Selección y evaluación de modelos.
5. Implementación operativa.

6. Monitoreo y ajuste post-despliegue.

Figura 2

Gráfico del ciclo CRISP-ML(Q) con foco en data preparation, data quality y model evaluation



Fuente. Autoría Propia

En este proyecto, CRISP-ML(Q) se aplica para estructurar el ciclo de vida de los datos y seleccionar los modelos de IA más adecuados para la nutrición y calificación de *leads*.

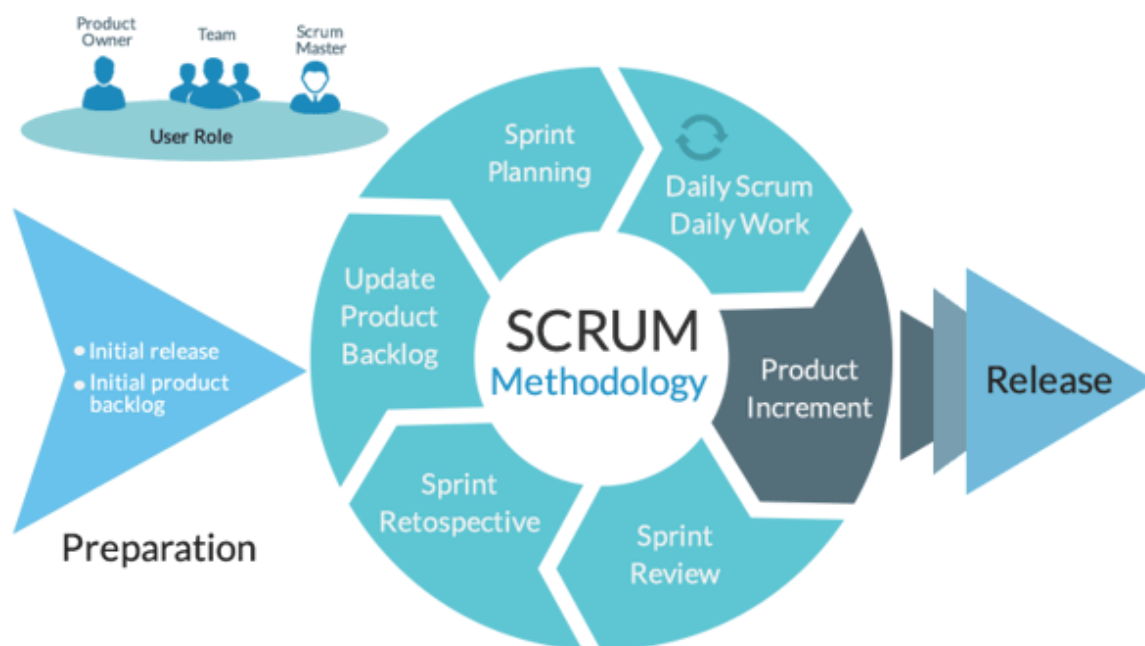
SCRUM

El marco ágil Scrum (Heath, 2021) orienta el desarrollo incremental del sistema, mediante *sprints* cortos que aseguran la entrega continua de valor funcional y la validación temprana con usuarios. Sus roles (*Product Owner*, *Scrum Master*, *Development Team*) y artefactos (*backlogs*, incrementos, reuniones diarias) aseguran flexibilidad y adaptación a cambios en el entorno empresarial. Scrum, aplicado en paralelo con CRISP-ML(Q), aporta la

agilidad necesaria para entregar prototipos funcionales y ajustarlos según la retroalimentación de los usuarios. A continuación, se muestra el esquema general de la metodología.

Figura 3

Esquema general metodología scrum



Fuente. Autoría Propia

La combinación de Scrum y CRISP-ML(Q) permite gestionar de forma integral el desarrollo del sistema y la mejora de datos para los modelos de IA integrados. Scrum aporta agilidad y entrega continua de valor, mientras que CRISP-ML(Q) garantiza la calidad, pertinencia y validación ética de los datos y modelos utilizados. Esta integración metodológica asegura un proceso controlado, centrado en el usuario y orientado a resultados confiables.

Fases del Proyecto

Fase 1. Identificación de Requisitos y Análisis del Proceso de Captación de Leads con BPM y Técnicas Centradas en el Usuario

Se levantan requerimientos funcionales y no funcionales usando entrevistas e historias de usuario a su vez, BPM permite modelar procesos actuales y definir criterios de éxito.

Fase 2. Diseño de Arquitectura Modular e Integrada con Inteligencia Artificial para el Sistema de Automatización

Se plantea una arquitectura por capas (interfaz, agentes IA, *backend*, base de datos), priorizando escalabilidad y mantenibilidad.

Fase 3. Preparación de Datos y Evaluación de Modelos de IA

Aplicación de CRISP-ML(Q) para limpieza, transformación y normalización de datos incluyendo el uso de técnicas de similitud como Levenshtein y Jaro-Winkler (a través de la librería RapidFuzz), así como modelos de representación semántica como all-MiniLM-L6-v2.

Posteriormente, se realiza una evaluación comparativa de modelos de Inteligencia Artificial para tareas de nutrición y calificación de *leads*.

Fase 4. Desarrollo Iterativo e Incremental del Prototipo con Scrum

Bajo Scrum, cada *sprint* entrega incrementos funcionales validados con usuarios, documentando funcionalidades y retroalimentación.

Fase 5. Validación del Prototipo y Evaluación de Desempeño, Eficiencia y Experiencia de Usuario

Se mide el desempeño con métricas cuantitativas (precisión/*recall*, tiempos de respuesta) y cualitativas (SUS, entrevistas de retroalimentación).

Consideraciones Éticas y Sociales

El proyecto garantiza el cumplimiento de la Ley 1581 de 2012 (Protección de Datos Personales en Colombia), el Código de Ética Profesional de (COPNIA, 2008) y principios internacionales de transparencia en IA. Se adoptan medidas de consentimiento informado, minimización de datos y documentación de procesos para prevenir sesgos y garantizar un uso ético de la información. Asimismo, se incorpora el principio de *accountability* como práctica de responsabilidad y rendición de cuentas en el uso de IA, en línea con marcos internacionales de gobernanza tecnológica. La integración de BPM, CRISP-ML(Q) y Scrum configura un enfoque metodológico híbrido que combina rigor en el análisis organizacional, calidad en el ciclo de vida de datos y modelos, y agilidad en el desarrollo de software. Este esquema asegura que la solución tecnológica propuesta no solo sea funcional y eficiente, sino también contextualizada, confiable y sostenible para las PYMES colombianas.

Identificación de Requisitos y Análisis del Proceso de Captación de *Leads* con BPM y Técnicas Centradas en el Usuario

En un entorno empresarial cada vez más competitivo, optimizar la identificación y gestión de oportunidades comerciales es clave para el crecimiento sostenible. Esta fase del proyecto tiene como objetivo analizar a fondo el proceso actual mediante el cual se generan y administran los *leads* corporativos, con el propósito de comprender su dinámica, detectar posibles ineficiencias y definir oportunidades de mejora apoyadas en soluciones de automatización inteligente.

Análisis del Proceso Actual de Captación y Gestión de *Leads*

El proceso actual de captación y gestión de *leads* en las PYMES colombianas presenta limitaciones que reducen la eficiencia operativa, la calidad del seguimiento y las tasas de conversión. Esta sección describe el flujo de trabajo vigente, los actores involucrados, las herramientas empleadas y las principales problemáticas identificadas.

Actores Involucrados

Los actores principales en el proceso actual son:

Tabla 3

Actores involucrados

Actor/Rol	Descripción
Ejecutivos Comerciales	Encargados de buscar, contactar y hacer seguimiento a los <i>leads</i> .
Gerentes de Ventas	Supervisan la estrategia de captación y definen criterios generales.
Asistentes Comerciales	Apoyan la organización y carga manual de datos.
Responsables de <i>Marketing</i>	Diseñan campañas de atracción y supervisan canales de captación.
Clientes Potenciales (<i>Leads</i>)	Usuarios externos que interactúan con formularios o canales de contacto.

Nota. Tabla que muestra los actores involucrados en el proceso actual en la prospección de

leads de las PYMES. Fuente. Autor

Segmentación del Público Objetivo

Tabla 4

Variables analizadas público objetivo

Variable	Descripción / Características Relevantes
Demográfica	<p>Pymes entre 1 y 200 empleados que buscan mejorar sus procesos comerciales.</p> <p>Personas naturales con actividad económica (emprendedores, consultores, pequeños negocios).</p> <p>Sectores con alta interacción comercial: servicios profesionales, tecnología, comercio y distribución.</p>
Geográfica	<p>Cobertura inicial Colombia, con mayor concentración en ciudades principales: Bogotá, Medellín, Cali, Barranquilla y Bucaramanga.</p> <p>Disponibilidad de conexión y medios digitales para interacción remota.</p>
Psicográfica	<p>Perfiles con interés en optimizar ventas y productividad.</p> <p>Valoran herramientas que ayuden a organizar clientes, automatizar tareas y mejorar seguimiento.</p> <p>Disposición a adoptar soluciones tecnológicas cuando son simples, económicas y con beneficios evidentes.</p>
Conductual	<p>Buscan información y soluciones por medios digitales (LinkedIn, Instagram, YouTube, Google).</p> <p>Toman decisiones de adopción basadas en casos de uso claros y demostraciones prácticas.</p> <p>Preferencia por soluciones con bajo costo inicial, fácil implementación y soporte cercano.</p>

Nota. Se identifican los elementos más relevantes en la segmentación del público objetivo.

Fuente. Autor

Las tablas anteriores delimitan el ecosistema del proyecto, identificando que la solución no solo debe atender a los ejecutivos comerciales, sino integrar las necesidades estratégicas de gerentes y analistas de *marketing*. La segmentación detallada en la Tabla 4 confirma que el sistema debe estar optimizado para PYMES en sectores de alta interacción digital, donde la adopción tecnológica es valorada pero requiere una implementación de bajo costo y alta simplicidad.

Aplicación de Técnicas para Recolección de Información

De acuerdo con las técnicas establecidas en el Marco Metodológico, se procedió a su aplicación considerando el tipo de usuarios y el contexto del proceso de captación y gestión de *leads* dentro de la PYME analizada. El objetivo fue obtener información precisa tanto sobre la situación actual como sobre las oportunidades de mejora relacionadas con la adopción de herramientas de automatización inteligente.

Para ello, se emplearon las siguientes técnicas:

1. Entrevistas semiestructuradas: Se realizaron entrevistas orientadas a comprender la percepción, prácticas y dificultades existentes en el proceso de gestión de oportunidades comerciales. Estas entrevistas permitieron identificar requerimientos funcionales y no funcionales del sistema, así como la visión de los responsables del proceso dentro de la empresa.
2. Encuesta digital (Google *Forms*): Se aplicó una encuesta en línea dirigida a colaboradores vinculados al área comercial y de *marketing*, con el fin de identificar cómo se realiza actualmente la captación de *leads*, qué herramientas se utilizan y cuáles son los principales puntos críticos en el seguimiento de clientes potenciales. Esta técnica permitió obtener información cuantificable y patrones de comportamiento.

Tabla 5

Registro de la aplicación de técnicas

Técnica	Muestra	Fecha	Instrumento / Evidencia
Entrevista semiestructurada	2 participantes (Líder de <i>Marketing</i> y Asesor Comercial)	2025-06-16	Anexo A — Guion y respuestas completas
Encuesta digital (Google <i>Forms</i>)	3 colaboradores del área comercial y <i>marketing</i>	2025-06-25	Anexo B — Formulario + gráficas exportadas + hoja de resultados

Nota. Se identifican los elementos más relevantes en la segmentación del público objetivo.

Fuente. Autor

Análisis de la Información

A continuación, se presentan los análisis de las entrevistas, la encuesta y la observación realizada, con el fin de identificar el estado actual del proceso de gestión de *leads* y las oportunidades de mejora relacionadas con la automatización.

Entrevistas Semiestructuradas

Tabla 6

Consolidación de respuestas recolectadas en las entrevistas.

Categoría / Pregunta	Entrevistado 1 (Líder de <i>Marketing</i>)	Entrevistado 2 (Asesor Comercial)
Proceso actual de captación de <i>leads</i>	Uso de redes sociales, referidos y LinkedIn; registro manual en hojas de cálculo.	Contacto directo, base de datos propia y formularios de página web; registro manual.
Herramientas utilizadas	Hojas de cálculo (Excel), CRM básico sin automatización, correo.	Excel y WhatsApp para seguimiento; no utiliza CRM por falta de conocimiento.
Percepción de efectividad	“Funciona, pero es lento y se pierde información.”	“Se pierden muchos contactos porque no hay seguimiento estructurado.”
Dificultades principales	Falta de trazabilidad y tiempo para seguimiento manual.	Duplicidad de datos, desorganización y olvido de tareas.
Criterios de priorización de <i>leads</i>	Nivel de interés y tamaño de la empresa.	Urgencia de compra y cercanía de la relación.
Percepción sobre automatización/IA	La ve útil si reduce tareas repetitivas y organiza datos.	Está dispuesto, pero requiere que sea fácil de usar.
Características esperadas en la herramienta	Notificaciones automáticas, integración con WhatsApp y CRM sencillo.	Registro rápido, recordatorios automáticos y seguimiento visual del estado de cada lead.

Nota. Se consolidan las respuestas obtenidas en las dos entrevistas que se realizaron en la

recolección de la información. Fuente. Autor

Tabla 7*Identificación de tendencias y hallazgos*

Hallazgo	Evidencia (respuestas / cita)	Significado
El proceso de captación y seguimiento es altamente manual.	Ambos entrevistados mencionan uso de Excel y registros manuales.	Existe riesgo de pérdida de información y baja eficiencia operativa.
Se presentan fallas en el seguimiento de <i>leads</i> .	“Se pierden contactos porque no hay seguimiento estructurado.” (Asesor comercial)	La falta de estructura afecta la conversión y calidad del proceso comercial.
Hay interés en automatizar, pero debe ser simple.	“Si la herramienta me ayuda a recordar tareas y no es complicada, la usaría.”	La solución debe priorizar usabilidad y adopción, no solo funcionalidad.
Necesidad de recordatorios y centralización de la información.	Ambos piden notificaciones y visualización clara del estado del lead.	La automatización debe enfocarse en priorización y alertas oportunas.

Nota. Se realiza un análisis de la información recopilada y se genera un resumen de los

hallazgos. Fuente. Autor

Los resultados muestran que el proceso actual de gestión de *leads* es manual y poco estructurado, lo que genera pérdida de oportunidades y dificultades en el seguimiento. Existe interés en implementar una herramienta que automatice y centralice la información, siempre que sea sencilla de usar. Esto confirma la necesidad del proyecto y orienta el diseño hacia funciones como registro centralizado, recordatorios automáticos, seguimiento visual e integración con los canales de comunicación existentes.

Encuestas Digitales (Google Forms)**Figura 3**

Resumen de datos recolectados en las encuestas

Nombre Completo

3 respuestas

Natania Lopez

Mauricio Rozo

Tatiana Mejia

Cargo

3 respuestas

Asistente de Marketing

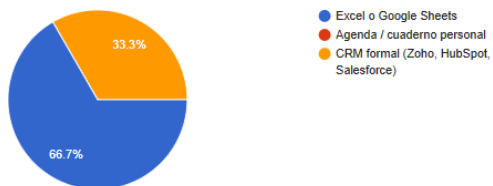
Comercial

Jefe de Mercadeo

¿Cómo registran actualmente los leads? (respuesta única)

[Copiar gráfico](#)

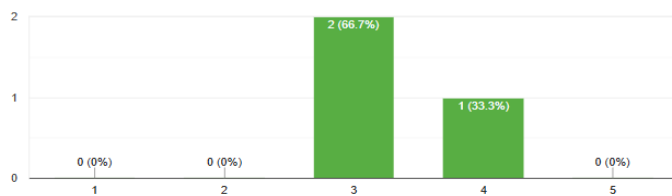
3 respuestas



¿Qué tan estructurado considera su proceso actual de gestión de leads? (escala Likert 1-5)

[Copiar gráfico](#)

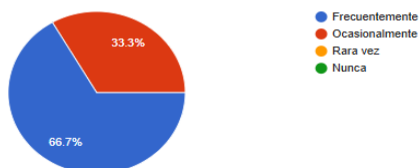
3 respuestas



¿Con qué frecuencia se pierden o olvidan leads durante el seguimiento? (respuesta única)

[Copiar gráfico](#)

3 respuestas





Fuente. Autoría Propia

Resumen de la Muestra:

- Número de participantes: 3
- Cargos: Jefe de Mercadeo, Asistente de *Marketing*, Comercial
- Fecha de aplicación: 11/11/2025
- Área: *Marketing* y Comercial

Tabla 8*Resultados cuantitativos encuestas semiestructuradas*

Pregunta	Resultados
¿Cómo registran actualmente los <i>leads</i> ?	2/3 utilizan Excel o Google Sheets (67%), 1/3 usa CRM formal (33%)
¿Qué tan estructurado considera su proceso actual?	Escala Likert promedio = 3.33; percepción de proceso medianamente estructurado
¿Con qué frecuencia se pierden o olvidan <i>leads</i> ?	2/3 “Frecuentemente” (67%), 1/3 “Ocasionalmente” (33%)
¿Qué tan útil considera automatizar el seguimiento de <i>leads</i> ?	Escala Likert promedio = 5 (100% lo considera muy útil)
¿Qué tan fácil sería adaptarse a una herramienta de automatización?	Escala Likert promedio = 3.67; percepción moderadamente fácil

Nota. Se los resultados cuantitativos de la encuesta. Fuente. Autor

Los datos estadísticos refuerzan la urgencia de la intervención. Con un 67% de los participantes reportando pérdida frecuente de leads y un interés del 100% en la automatización, los resultados cuantitativos validan que el problema no es la falta de herramientas, sino la fragmentación de estas. El promedio de 3.33 en la escala de estructuración confirma que el proceso actual es "medianamente estructurado", dejando un margen de mejora significativo que el proyecto busca cubrir.

Funcionalidades más importantes seleccionadas (selección múltiple):

- Registro centralizado de *leads*: 2/3
- Recordatorios automáticos: 2/3
- Integración con WhatsApp o correo: 3/3
- Reportes y métricas de desempeño: 1/3

Tabla 9*Hallazgos claves*

Hallazgo	Evidencia	Interpretación
El registro de <i>leads</i> es mixto y parcialmente manual	67% usan Excel/Sheets, 33% CRM formal	Existe riesgo de pérdida de información y duplicidad de datos, lo que afecta eficiencia
El proceso no está completamente estructurado	Promedio de Likert = 3.33	Necesidad de estandarizar y formalizar el flujo de gestión de <i>leads</i>
Se pierden <i>leads</i> con cierta frecuencia	2/3 “Frecuentemente”	Falta de trazabilidad y seguimiento constante, lo que impacta oportunidades de ventas
Alta disposición hacia automatización	Promedio Likert = 5	El equipo está motivado a adoptar herramientas que faciliten el seguimiento y control
Funcionalidades prioritarias	Registro centralizado, recordatorios, integración con comunicación	La herramienta debe enfocarse en centralización, alertas y multicanalidad para aumentar eficiencia

Nota. Se realiza un análisis de la información recopilada y se genera un resumen de los

hallazgos. Fuente. Autor

Los hallazgos consolidados en las Tablas 6 a 9 evidencian una brecha crítica en la gestión de *leads*: aunque existe un interés unánime (100%) por la automatización, la dependencia actual de procesos manuales en Excel (67%) genera una pérdida frecuente de oportunidades de negocio. Este diagnóstico fundamenta la necesidad de una solución que no solo automatice, sino que simplifique la trazabilidad, atacando directamente el dolor de la 'desorganización' manifestado por los actores comerciales

Los resultados confirman que el proceso actual de gestión de *leads* presenta limitaciones significativas en organización, seguimiento y trazabilidad, especialmente cuando el volumen de contactos aumenta. El equipo muestra disposición total hacia la automatización, priorizando funcionalidades que permitan centralizar la información, generar recordatorios automáticos y mantener la comunicación desde canales como WhatsApp o correo electrónico.

La encuesta valida la necesidad de implementar una herramienta de automatización inteligente, coherente con los hallazgos de las entrevistas, orientada a mejorar la eficiencia, reducir errores y facilitar el control del *pipeline* comercial.

Descripción del Proceso Actual de Prospección de Clientes

El proceso inicia con la identificación manual de prospectos, utilizando canales como LinkedIn, recomendaciones, formularios web y bases de datos propias o adquiridas. En esta fase se evidencia un primer problema: la alta carga operativa manual y repetitiva.

Una vez recopilada la información de contacto, como correo electrónico, número de teléfono o perfil de LinkedIn, los datos se gestionan en hojas de cálculo o en plataformas CRM de baja complejidad. Aquí se presenta un componente crítico: la desorganización, la duplicación de información y la falta de validación de los datos.

Posteriormente, se ejecuta la fase de nutrición, que consiste en establecer el contacto inicial con el prospecto mediante correos, llamadas o mensajes de WhatsApp. En este paso se observa baja personalización en los mensajes, que suelen ser genéricos. Luego, los prospectos se clasifican según la experiencia y percepción del equipo, y se actualizan en los repositorios para el seguimiento respectivo.

Si el contacto con el lead es positivo, se procede a calificarlo. En esta etapa se detectó la ausencia de criterios estandarizados (*scoring*) y una alta pérdida de oportunidades debido a bajos niveles de seguimiento y a la calidad variable de los *leads*. Finalmente, el área comercial gestiona el cierre, generando la conversión si es positiva o cerrando el lead en caso de resultado fallido.

Este flujo no sigue un estándar uniforme, ni cuenta con criterios claros para determinar la calidad del lead o priorizar su atención. Además, existe una fuerte dependencia de la intervención humana en cada etapa, lo que hace que el proceso sea lento, poco escalable y propenso a errores.

Figura 5

Flujo actual del proceso comercial de captación de leads



Fuente. Autoría Propia

Herramientas y Canales Actualmente Utilizados.

Los actores principales en el proceso actual son:

Tabla 10

Herramientas involucradas

Tipo de Herramienta	Ejemplos concretos
Buscadores de leads	LinkedIn, directorios empresariales, referencias
Gestión de datos	Excel, Google Sheets, CRM básico (HubSpot, Zoho)
Comunicación	WhatsApp, Email, llamadas telefónicas
Formularios de captación	Google Forms, Typeform, landing pages corporativas

Nota. Tabla que muestra las herramientas involucradas en el proceso actual en la prospección

de leads de las PYMES. Fuente. Autor

Diagrama AS-IS del Proceso de Prospectación de Clientes

A continuación se describe el proceso de inicio a fin

Inicio

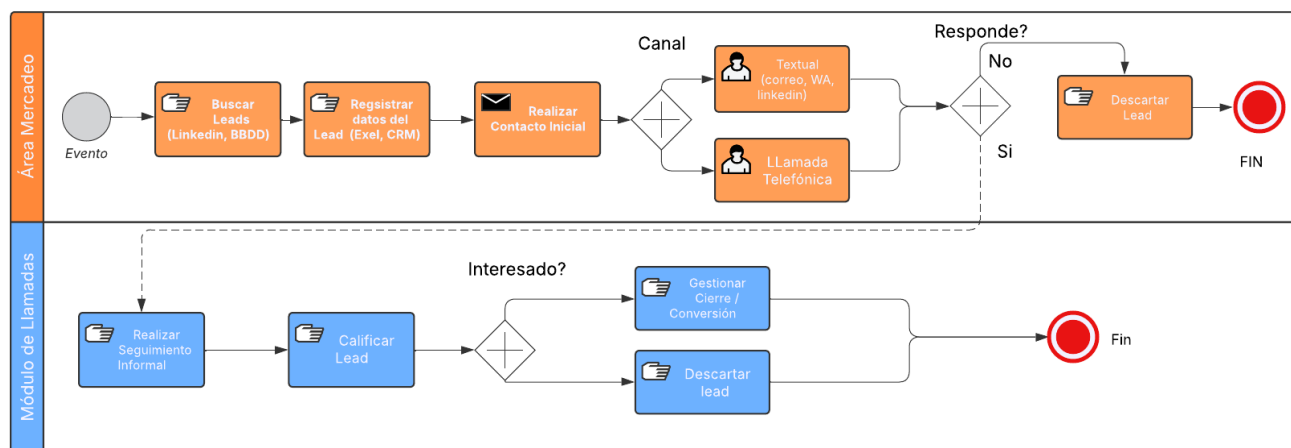
- Identificación manual de posibles clientes: Un ejecutivo comercial inicia la búsqueda de *leads* mediante herramientas como LinkedIn, referencias o bases de datos existentes.
- Recopilación manual de datos: Los datos del *lead* se registran en Excel o CRM básico, usualmente sin validación automática.
- Contacto inicial: Se realiza por email, WhatsApp o llamada.
- Seguimiento: Si el *lead* responde, se inicia un seguimiento informal mediante llamadas o mensajes.
- Calificación del *lead*: Se clasifica como "interesado", "no interesado" o "a seguir", sin criterios estandarizados.
- Cierre o descarte: El *lead* es atendido, convertido o descartado. En la mayoría de los casos no queda registro sistemático.

Fin

De acuerdo con el proceso descrito anteriormente, en la siguiente imagen se observa el diagrama BPMN AS-IS del proceso de captación de *leads*

Figura 6

Diagrama BPMN comercial de captación de lead



Fuente. Autoría Propia

El diagrama AS-IS (Figura 6) permite visualizar gráficamente los cuellos de botella del proceso actual, destacando la falta de validación automática en la entrada de datos y la inexistencia de un flujo de calificación estandarizado. Esta representación es fundamental para el cumplimiento del primer objetivo específico, ya que identifica los puntos exactos donde la intervención manual ralentiza el ciclo de venta, justificando el rediseño hacia un modelo automatizado.

Identificación de Puntos Críticos Dolores (Pains) y Necesidades de Usuario

El análisis del proceso actual de captación y gestión de *leads*, combinado con entrevistas y observación directa, permitió identificar los principales dolores, implicaciones y necesidades de los usuarios. Estos hallazgos reflejan las dificultades operativas, la falta de automatización y los problemas de seguimiento y priorización que afectan la eficiencia del equipo comercial y de *marketing*.

Tabla 11*Resumen de hallazgos por Actor o perfil de usuario*

Actor/Grupo de interés	Dolor identificado	Implicación	Frecuencia / Severidad
Ejecutivo Comercial	Dificultad para encontrar <i>leads</i> calificados; procesos manuales lentos	Baja productividad y pérdida de oportunidades	Alta / Crítica
Gerente de Ventas	<i>Leads</i> no responden; falta de personalización; baja conversión	Reducción de tasa de cierre y ROI bajo	Alta / Crítica
Asistente Comercial	Duplicidad de contactos; seguimiento ineficiente; falta de trazabilidad	Confusión en prioridades y pérdida de <i>leads</i>	Media-Alta / Importante
Responsable de <i>Marketing</i>	<i>Leads</i> poco calificados; desorganización en seguimiento	Difícil medir efectividad de campañas; baja conversión	Alta / Crítica
Profesional de Ventas	Información desactualizada; poca automatización	Dificultad para priorizar <i>leads</i> y retraso en gestión	Media / Moderada
Coordinador Comercial	Mucho tiempo en tareas repetitivas; poco enfoque estratégico	Menor capacidad para decisiones estratégicas; agotamiento del equipo	Alta / Crítica

Nota. Tabla de resumen de dolores identificados por perfil de usuario con base a la descripción del proceso. Fuente. Autor

Como se resume en la Tabla 11, la severidad 'Alta/Crítica' en la mayoría de los perfiles subraya que la ineficiencia no es un problema aislado de un rol, sino un fallo sistémico. La recurrencia de la 'baja productividad' y la 'pérdida de oportunidades' dictamina que los requisitos del sistema deben priorizar la agilidad en la captura y la claridad en el seguimiento visual.

Tabla 12*Limitaciones del modelo actual de captación de leads*

Aspecto	Descripción
Tiempo de gestión	Alto, debido a procesos manuales no sistematizados
Personalización	Baja, basada en mensajes genéricos
Escalabilidad	Limitada, dependiente de la intervención humana
Criterios de calificación	Subjetivos o inexistentes
Tecnología empleada	Básica o inexistente

Nota. Tabla que muestra las limitaciones actuales de la prospectación de *leads* en las Pymes en Colombia. Fuente. Autor

Requisitos del Sistema

Tabla 13

Identificación de requisitos del sistema

Tipo de requisito	Código	Descripción
Funcional	RF-01	Registrar y almacenar <i>leads</i> de manera centralizada desde múltiples fuentes (LinkedIn, formularios web, bases de datos internas o externas).
Funcional	RF-02	Realizar seguimiento automatizado de los <i>leads</i> mediante recordatorios, alertas y notificaciones.
Funcional	RF-03	Clasificar y priorizar <i>leads</i> mediante un sistema de <i>scoring</i> basado en criterios predefinidos.
Funcional	RF-04	Integrar canales de comunicación como correo electrónico y WhatsApp para el contacto con <i>leads</i> .
Funcional	RF-05	Generar reportes y métricas de desempeño sobre la captación, seguimiento y conversión de <i>leads</i> .
Funcional	RF-06	Permitir segmentación de <i>leads</i> por sector, tamaño de empresa, interés o nivel de interacción.
No funcional	RNF-01	El sistema debe garantizar disponibilidad mínima del 99% y tiempos de respuesta rápidos (<2 segundos por consulta).
No funcional	RNF-02	Asegurar la confidencialidad y protección de los datos de los <i>leads</i> mediante autenticación y control de accesos.
No funcional	RNF-03	La plataforma debe ser accesible desde dispositivos web y móviles (responsive) y compatible con navegadores modernos.
No funcional	RNF-04	Registrar logs de actividad y cambios en los datos para garantizar trazabilidad y auditoría.
No funcional	RNF-05	Interfaz intuitiva y fácil de usar para minimizar la curva de aprendizaje del equipo comercial y de <i>marketing</i> .

Nota. Tabla que identifica los requisitos del sistema a partir del análisis de la información y el proceso actual. Fuente. Autor

La identificación de estos 6 requisitos funcionales y 5 no funcionales (Tabla 13) representa la culminación del análisis del proceso actual. Estos requisitos actúan como la hoja de ruta técnica para las fases de diseño y desarrollo posteriores, garantizando que el prototipo final responda directamente a las necesidades de escalabilidad (RNF-01) y facilidad de uso (RNF-05) demandadas por las PYMES colombianas analizadas en este estudio.

Rediseño del Proceso de Captación y Gestión de *Leads*

La incorporación de tecnologías emergentes en los procesos de captación de *leads* representa una oportunidad concreta para superar las barreras previamente identificadas. En particular, herramientas como el *web Scraping* permiten automatizar la recolección de datos

relevantes desde múltiples fuentes digitales, mientras que modelos de inteligencia artificial (IA) posibilitan la segmentación, personalización y calificación automática de prospectos en función de patrones históricos de comportamiento.

Al integrar tecnologías basadas en patrones históricos de comportamiento dentro de un sistema automatizado e inteligente, es posible reducir significativamente el tiempo necesario para identificar y calificar *leads*, aumentar la personalización de las interacciones mediante técnicas de segmentación automatizada, priorizar aquellos *leads* con mayor probabilidad de conversión gracias a modelos predictivos entrenados con datos históricos y, en consecuencia, disminuir el costo de adquisición de clientes al optimizar el uso de los recursos disponibles.

Estas posibilidades cobran mayor relevancia en contextos como el colombiano, donde la transformación digital de las PYMES se encuentra en una etapa incipiente, pero con alto potencial de impacto. De este modo, se justifica el desarrollo de un sistema de automatización inteligente que permita abordar de forma innovadora y sostenible el proceso de captación comercial.

Clasificación de Necesidades Específicas por Rol o Categoría

El rediseño del proceso debe contemplar las necesidades específicas de los diferentes usuarios del sistema:

Comerciales. requieren una herramienta que les permita visualizar *leads* priorizados, registrar sus interacciones y realizar seguimientos de manera ágil. También necesitan reportes automáticos sobre sus actividades y resultados.

Analistas de Marketing. necesitan segmentar y agrupar *leads* con base en criterios como comportamiento, industria y ubicación. Además, deben contar con herramientas para medir la efectividad de las campañas y optimizar los canales utilizados.

Gerentes. demandan una visión general del rendimiento del sistema, acceso a estadísticas agregadas, y la posibilidad de tomar decisiones basadas en datos. También requieren mecanismos de control y auditoría del proceso de captación.

Técnicos o Desarrolladores. necesitan contar con APIs o integraciones que permitan conectar el sistema con otras plataformas como CRM, campañas de correo, o bases de datos internas. También son responsables de mantener los procesos automáticos actualizados y en correcto funcionamiento.

Cambios Propuestos al Proceso Actual

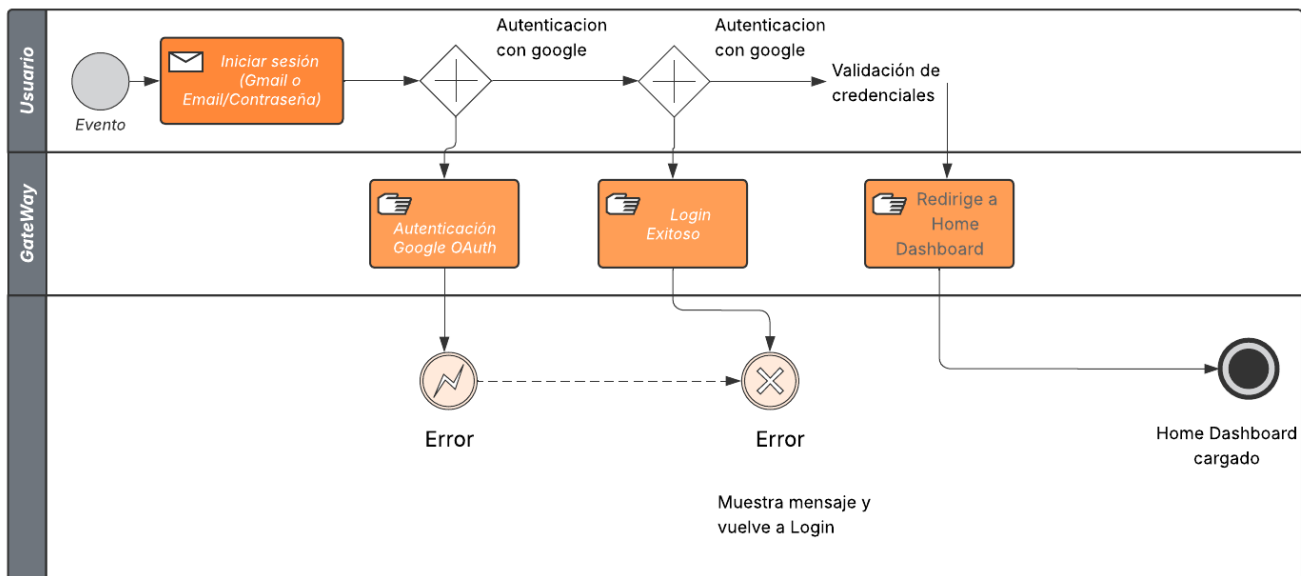
El rediseño propone transformar el modelo tradicional de captación basado en actividades manuales y desconectadas en un flujo de trabajo automatizado, centralizado y orientado por datos. Entre los cambios más relevantes se incluyen:

El sistema propone el reemplazo de la búsqueda manual por un proceso de *Web Scraping* automatizado que permite recopilar datos en tiempo real desde sitios web y directorios públicos. Además, incluye la implementación de un motor de segmentación automática que clasifica los *leads* según criterios personalizados por el usuario, la integración con herramientas de comunicación multicanal como correo electrónico, WhatsApp o llamadas automatizadas para facilitar el contacto con los *leads*, el registro y seguimiento automático de las interacciones (incluyendo resultados de llamadas, apertura de correos y respuestas recibidas), la asignación inteligente de *leads* a agentes comerciales en función de su disponibilidad y desempeño previo, y la incorporación de tableros de control y analítica en tiempo real que permiten monitorear el desempeño individual y del equipo.

Diagrama del Estado Futuro del Proceso

Figura 7

Acceso y navegación general a la plataforma

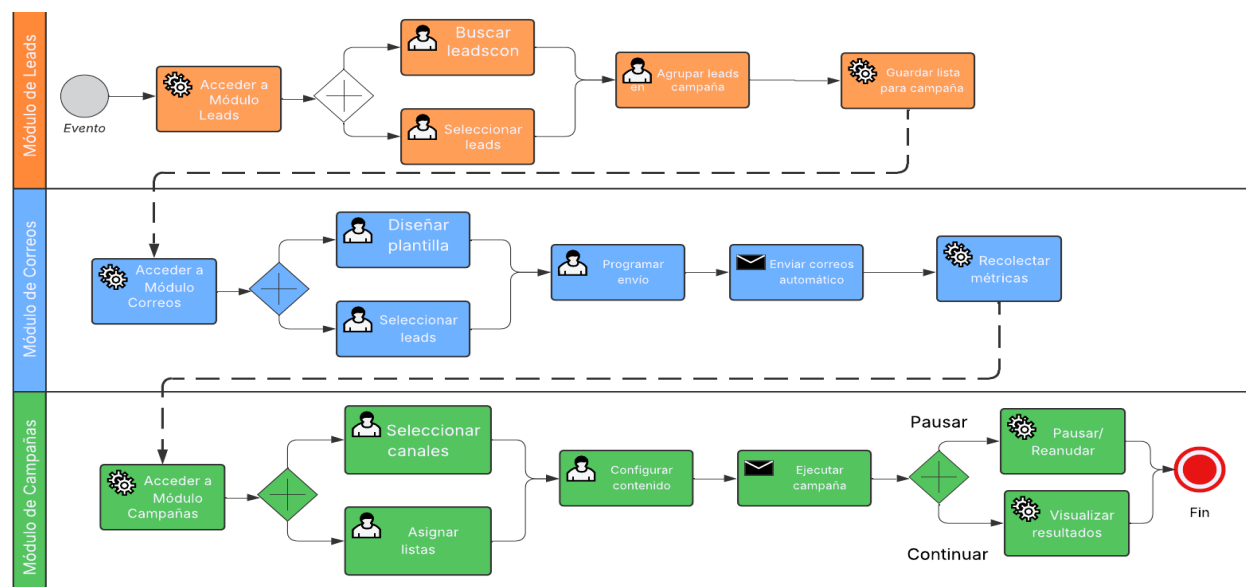


Fuente. Autoría Propia

La Figura 7 ilustra el protocolo de seguridad y entrada al sistema, destacando la integración con Google OAuth para facilitar una autenticación rápida y segura. Este diseño responde directamente al requisito no funcional RNF-02 (Seguridad) y al RNF-05 (Usabilidad), al reducir la fricción en el inicio de sesión. La validación de credenciales a través de un *GateWay* garantiza que solo usuarios autorizados accedan al *dashboard* principal, asegurando la integridad de la base de datos de leads desde el primer punto de contacto.

Figura 8

Automatización comercial (Leads, Correos, Campañas)

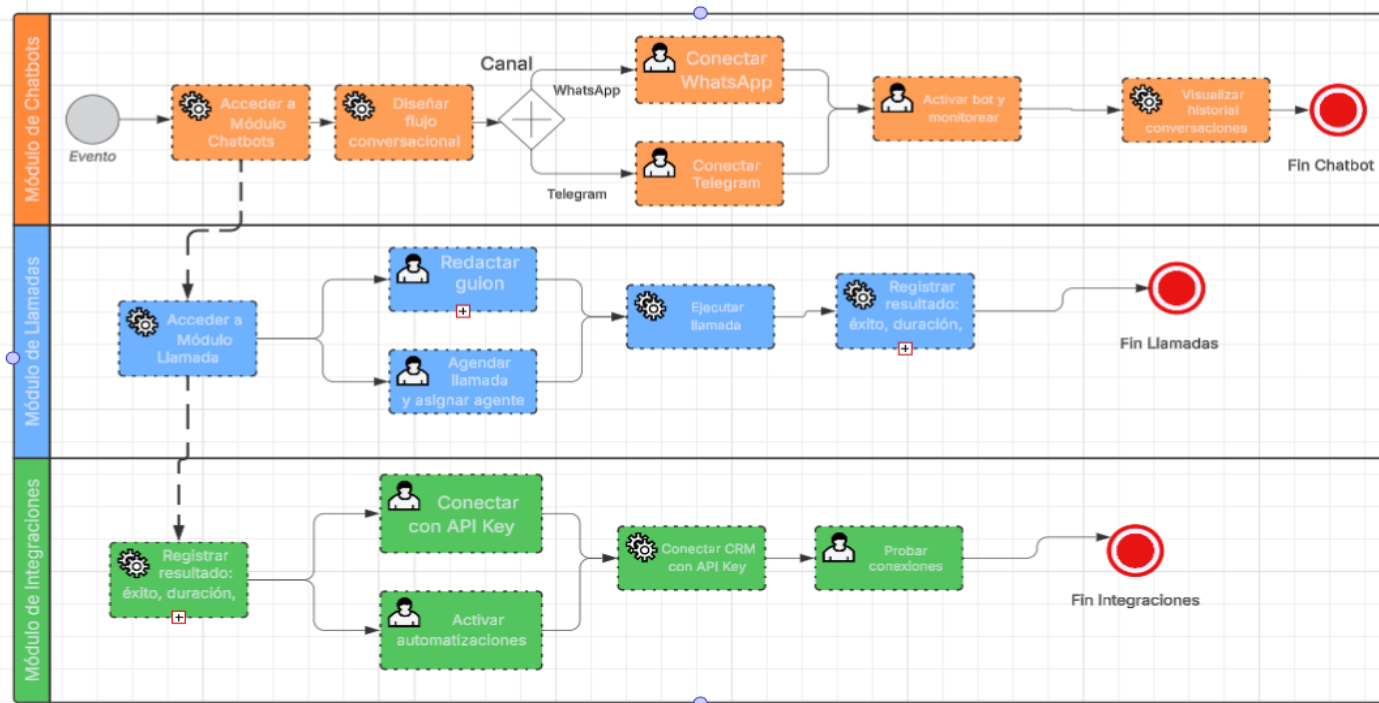


Fuente. Autoría Propia

En la Figura 8 se presenta la arquitectura funcional del sistema dividida en tres módulos interconectados: Leads, Correos y Campañas. A diferencia del proceso manual identificado en el diagrama AS-IS, este flujo futuro permite que la información fluya sin interrupciones desde la búsqueda y agrupación de prospectos hasta la ejecución de campañas automáticas. El análisis de este diagrama evidencia cómo la recolección de métricas en el módulo de correos cierra el ciclo de retroalimentación, permitiendo que el sistema cumpla con el requisito RF-05 (Reportes y métricas) para la toma de decisiones basada en datos reales de conversión.

Figura 9

Agentes automatizados y conectividad



Fuente. Autoría Propia

La Figura 9 detalla la capa de inteligencia y conectividad del sistema, donde los agentes automatizados ejecutan las tareas críticas de *Scraping* y *Scoring* de manera independiente. Este diagrama representa el núcleo de la solución propuesta, demostrando cómo la conectividad multicanal (WhatsApp, Email y Web) se centraliza para evitar la dispersión de datos.

Definición del MPV

El MVP se centrará en ofrecer una plataforma funcional que permita registrar, clasificar y dar seguimiento automatizado a los *leads*, con un *dashboard* básico y comunicación integrada (correo o WhatsApp). Esto permitirá validar la propuesta de valor —reducir carga manual, aumentar trazabilidad y mejorar la conversión— sin desarrollar aún las integraciones o reportes avanzados.

Tabla 14

Propuesta de MVP

Funcionalidad	Categoría	Justificación
Registro y almacenamiento centralizado de <i>leads</i> desde múltiples fuentes (formularios, LinkedIn, bases internas)	<i>Must Have</i>	Es la base del sistema; permite consolidar toda la información en una sola plataforma, reduciendo la dispersión y duplicidad de datos.
Clasificación automática de <i>leads</i> mediante sistema de <i>scoring</i>	<i>Must Have</i>	Facilita la priorización y seguimiento eficiente, uno de los mayores problemas identificados en el proceso actual.
Seguimiento automatizado con recordatorios y alertas	<i>Must Have</i>	Evita pérdida de oportunidades por falta de seguimiento, mejorando la trazabilidad del <i>pipeline</i> comercial.
Integración con canales de comunicación (correo y WhatsApp)	<i>Should Have</i>	Mejora la personalización del contacto y permite mantener conversaciones directas desde la plataforma.
<i>Dashboard</i> de control con métricas básicas (número de <i>leads</i> , conversiones, seguimiento activo)	<i>Should Have</i>	Proporciona visibilidad general del desempeño comercial y facilita la toma de decisiones.
Registro manual y edición de <i>leads</i>	<i>Could Have</i>	Complementa la importación automática, permitiendo registrar <i>leads</i> captados fuera de los canales integrados.
Envío de correos personalizados a <i>leads</i>	<i>Could Have</i>	Aumenta la tasa de respuesta y personalización de las interacciones. Puede añadirse en versiones posteriores.
Segmentación avanzada de <i>leads</i> por sector o nivel de interés	<i>Could Have</i>	Mejora la eficiencia del <i>marketing</i> , pero no es esencial para el funcionamiento inicial del MVP.
Autenticación de usuarios mediante Gmail o correo corporativo	<i>Must Have</i>	Asegura acceso seguro a la plataforma y protección de datos confidenciales.
Control de acceso y roles (Administrador / Comercial / <i>Marketing</i>)	<i>Should Have</i>	Define permisos de uso según el rol, importante para escalabilidad en equipos más grandes.
Generación de reportes descargables (Excel o PDF)	<i>Won't Have</i> (por ahora)	Puede incluirse en una versión posterior, ya que las métricas básicas estarán disponibles en el <i>dashboard</i> .
Integración con CRM externos (HubSpot, Zoho, Salesforce)	<i>Won't Have</i> (por ahora)	No es esencial para el MVP; requiere mayor complejidad técnica e inversión.

Nota. Tabla que resume la propuesta del MVP. Fuente. Autor

La selección de funcionalidades para el MVP (Tabla 14) prioriza los elementos de "Alta Severidad" encontrados en el diagnóstico. Al enfocarse en el registro centralizado, el *scoring* automático y los recordatorios, el MVP ataca directamente los cuellos de botella del BPMN AS-IS. Se postergan integraciones complejas (CRM externos) para asegurar una curva de aprendizaje mínima y una validación rápida de la propuesta de valor en un entorno operativo real.

Diseño de Arquitectura Modular del Sistema

Este capítulo presenta el diseño arquitectónico de HunterLeads, una plataforma web orientada a la gestión de campañas, prospección y automatización mediante agentes inteligentes. El propósito es describir el modelo arquitectónico adoptado, justificar la elección del patrón Modelo–Vista–Plantilla (MVT) y explicar cómo este se adapta a las necesidades funcionales y no funcionales del sistema. Asimismo, se exponen los puntos de vista arquitectónicos seleccionados y la manera en que la estructura modular propuesta contribuye a la escalabilidad, mantenibilidad y evolución futura de la solución.

Diseño de la Arquitectura de la Solución

El diseño de la arquitectura de software constituye una decisión estratégica, pues determina la calidad técnica, la mantenibilidad y la capacidad del sistema para escalar. Más allá del cumplimiento de los requisitos funcionales, la arquitectura debe atender atributos de calidad como rendimiento, seguridad, disponibilidad, escalabilidad y facilidad de prueba, tal como sugieren (Velepucha & Flores, 2023). Bajo este enfoque, la plataforma HunterLeads adopta una arquitectura modular por capas basada en el patrón Modelo–Vista–Plantilla (MVT), complementado con servicios especializados y agentes inteligentes desarrollados en Python.

Selección del Tipo de Aplicación

Debido al carácter multiusuario, la necesidad de acceso desde distintos dispositivos y la integración con servicios externos, la aplicación web constituye la modalidad de implementación más adecuada. Este tipo de aplicación permite centralizar despliegues, unificar la experiencia de usuario y establecer un control preciso sobre el ciclo de vida del *software*, en concordancia

con(Upadhyaya, 2023). Además, la arquitectura web facilita la incorporación de funcionalidades administrativas, paneles de gestión y flujos avanzados orientados a procesos de prospección.

Modelo de Arquitectura Basado en MVT

La plataforma se fundamenta en el patrón Modelo–Vista–Plantilla (MVT), una evolución del tradicional Modelo–Vista–Controlador (MVC) orientada a entornos web basados en *Python*. Aunque HunterLeads se implementa sobre *Flask* y no sobre *Django*, la organización interna del código sigue los principios de separación de responsabilidades y modularidad característicos del MVT. Según (Ghimire, 2020), este patrón resulta especialmente adecuado para aplicaciones web debido a la claridad con que separa la presentación, la lógica de procesamiento y la gestión de datos.

Componentes del Patrón MVT en HunterLeads

Modelo: Gestión de Datos Centralizada en Supabase

El modelo representa la capa responsable de la estructura e integridad de los datos. En HunterLeads esta función se implementa mediante Supabase, que ofrece una base de datos PostgreSQL, autenticación integrada y políticas de seguridad a nivel de filas. A diferencia del enfoque tradicional en Django, donde los modelos se definen mediante clases *Python*, el proyecto utiliza tablas SQL como fuente única de verdad, alineándose con (Effendy et al., 2021), quien plantea que el modelo debe agrupar tanto la estructura como las reglas que protegen los datos. Esta capa contiene entidades como usuarios, campañas, agentes, contactos, activos y *leads*, y su acceso se realiza mediante el cliente oficial de Supabase.

Vistas: Lógica de Procesamiento en Rutas y Controladores

Las vistas en Flask se implementan mediante *blueprints* organizados por dominios funcionales, como autenticación, gestión de campañas, agentes de comunicación e integraciones

externas. Estas vistas reciben las peticiones, validan información, orquestan procesos y preparan las respuestas, pero no contienen lógica de negocio, manteniendo la separación propuesta por (Pawade et al., 2022). Su función principal es actuar como intermediarias entre el modelo y la capa de presentación.

Plantillas: Presentación Mediante HTML y Jinja2

La capa de plantillas, implementada con Jinja2, gestiona la representación visual de los datos. Incluye plantillas base, fragmentos reutilizables y vistas específicas para las diferentes interfaces del sistema. Conforme a (Ghimire, 2020), esta capa no incorpora lógica de procesamiento, garantizando que la estructura de presentación se mantenga independiente de la lógica del negocio.

Extensión del MVT: Capa de Servicios y Agentes Inteligentes

Aunque el patrón MVT contempla únicamente tres capas principales, el diseño modular de HunterLeads introduce una capa adicional encargada de la lógica de negocio. Esta capa implementa servicios especializados responsables de procesos como autenticación, gestión de usuarios, campañas, agentes inteligentes, consultas y análisis de datos. Su incorporación responde a la necesidad de mantener alta cohesión y bajo acoplamiento, siguiendo el enfoque modular descrito por (Rodríguez et al., 2020)

Los servicios inteligentes, implementados como agentes autónomos en *Python*, extienden las capacidades del sistema mediante flujos conversacionales, automatización de mensajes, llamadas, procesamiento analítico y uso de modelos de IA. Estas unidades funcionan como microservicios ligeros, integrándose con las vistas sin alterar la estructura conceptual del MVT. Esta separación coincide con el estándar (IEEE 1471, 2000), que recomienda agrupar las preocupaciones de diferentes interesados en vistas claramente definidas.

Justificación de la Elección del Patrón MVT

La adopción de MVT responde tanto a criterios técnicos como de negocio:

Claridad estructural y mantenimiento: El patrón separa con nitidez la presentación, la lógica y la persistencia, reduciendo el acoplamiento y facilitando la evolución (Pawade et al., 2022)

Compatibilidad con Python: La convergencia entre *Flask*, los agentes de IA y las bibliotecas de análisis facilita la integración y consistencia del proyecto (Ghimire, 2020)

Adaptabilidad frente a servicios externos: MVT permite incorporar servicios inteligentes sin comprometer la estructura fundamental del sistema.

Facilidad de pruebas y despliegue: La división entre vistas simples y servicios especializados mejora la automatización de pruebas y la estabilidad en actualizaciones.

Soporte para atributos de calidad: La arquitectura favorece políticas avanzadas de seguridad, *caching*, auditoría y escalabilidad independiente de la base de datos.

En conjunto, MVT representa un balance adecuado entre rigor conceptual y compatibilidad práctica con el ecosistema tecnológico del proyecto.

Selección de Puntos de Vista Arquitectónicos

Para describir la arquitectura y atender las preocupaciones de los distintos interesados se emplean los puntos de vista sugeridos por (Velepucha & Flores, 2023):

1. **Vista funcional:** Explica los módulos del sistema, sus responsabilidades y su interacción con actores internos y externos.
2. **Vista de información:** Describe las entidades de dominio y la estructura de datos en Supabase.

3. **Vista de desarrollo:** Presenta la organización del código, módulos, dependencias y estructura interna del repositorio.
4. **Vista de operaciones y despliegue:** Define el entorno de ejecución, la configuración, los servicios en la nube y los mecanismos de integración continua.
5. **Vista de calidad:** Identifica los atributos no funcionales críticos y explica cómo la arquitectura los soporta.

Vista Funcional del Sistema

La plataforma se organiza en cinco capas principales:

- Capa Cliente: Interfaz web construida con *HTML*, *CSS*, *JavaScript* y *Jinja2*.
- Capa de Vistas: *Blueprints* de *Flask* que reciben solicitudes, validan datos y orquestan las reglas de negocio.
- Capa de Plantillas: Archivos HTML renderizados con *Jinja2*, responsables de la presentación.
- Capa de Datos: Supabase como *backend* de persistencia, con tablas SQL y controles de seguridad mediante RLS.
- Capa de Servicios Inteligentes: Agentes y microservicios en *Python* para IA, mensajería, automatización y análisis.

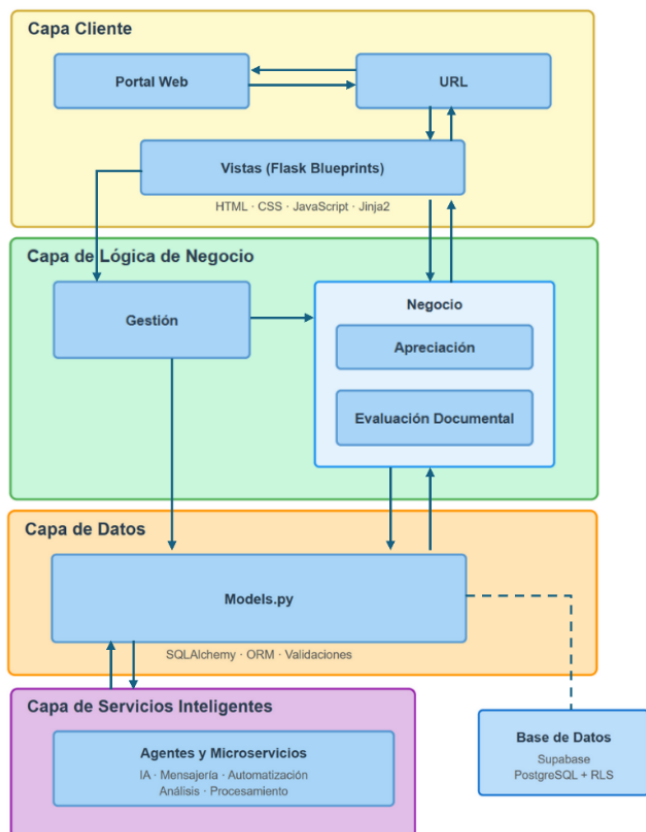
Esta estructura modular facilita las pruebas, el escalado progresivo de componentes y la evolución controlada de las funcionalidades.

De esta manera se evidencia que la arquitectura propuesta para HunterLeads integra de manera coherente el patrón Modelo–Vista–Plantilla con una capa adicional de servicios y agentes inteligentes. La selección de puntos de vista arquitectónicos permite abordar las preocupaciones de usuarios, desarrolladores y administradores de forma sistemática, asegurando

una base sólida para el crecimiento del proyecto. En conjunto, la solución se estructura como una plataforma modular, escalable y flexible, capaz de adaptarse a las necesidades cambiantes del negocio y al avance tecnológico oportuno.

Figura 10

Diagrama general de arquitectura



Fuente. Autoría Propia

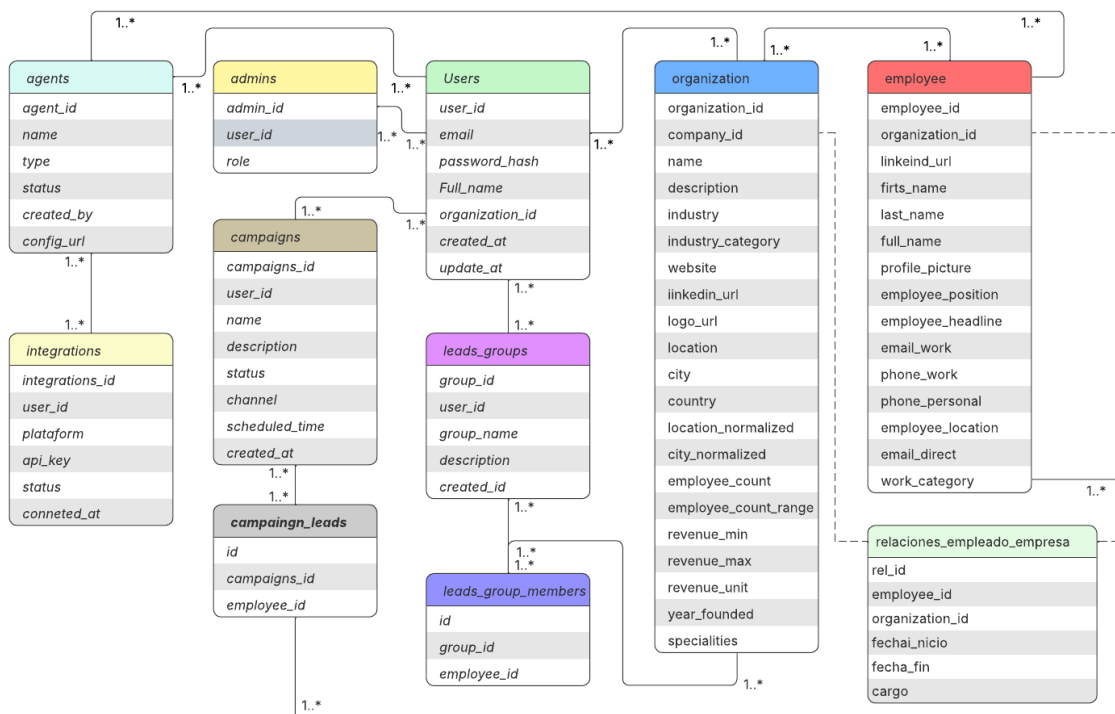
La Figura 10 presenta una arquitectura por capas que separa la interfaz de usuario de la lógica de negocio y el almacenamiento. La implementación del patrón MVT junto con una Capa de Servicios Inteligentes permite que los agentes de Python operen de forma independiente, garantizando que el sistema sea mantenible y que se puedan integrar nuevos modelos de IA sin reestructurar toda la plataforma.

Modelado del Esquema de Base de Datos

Propuesta de modelado del esquema relacional de base de datos, orientado a una base PostgreSQL gestionada desde Supabase. Este modelo está diseñado con un enfoque modular y escalable, respetando las mejores prácticas de normalización y seguridad.

Figura 11

Diagrama ERD del sistema HunterLeads



Fuente. Autoría Propia

El diagrama Entidad-Relación (Figura 11) detalla la persistencia en Supabase. La estructura normalizada asegura la integridad referencial entre leads, campañas y empresas. Este diseño es clave para el cumplimiento de los requisitos de trazabilidad, permitiendo que cada interacción de los agentes inteligentes quede vinculada correctamente a un registro único en PostgreSQL.

Preparación de Datos y Evaluación de Modelos de IA

Este apartado presenta el proceso seguido para preparar los datos y evaluar los modelos de inteligencia artificial utilizados en HunterLeads. Con el fin de asegurar rigurosidad y consistencia, se adoptan como base los componentes de la metodología CRISP-ML(Q), los cuales orientan las actividades de comprensión de datos, preparación, modelado, validación y aseguramiento de calidad. Estos componentes serán descritos y desarrollados en los apartados siguientes. De esta manera, se establece una base metodológica sólida que garantiza que la información utilizada sea adecuada y que los modelos seleccionados cumplan con los criterios de desempeño requeridos para su integración en la plataforma.

Requisitos de Calidad (CRISP-ML(Q))

El enfoque metodológico CRISP-ML(Q) introduce elementos de calidad y validación en cada fase del ciclo de vida del modelo. En esta etapa inicial, se establece la trazabilidad de las fuentes de datos y sus métodos de adquisición, la documentación transparente de los criterios de segmentación y filtrado, así como la reproducibilidad del flujo de recolección y preprocesamiento de datos mediante el uso de scripts automatizados.

Identificación de Fuentes de Datos para *Leads*

La construcción de un modelo confiable exige una base de datos representativa, estructurada y libre de sesgos. Por ello, la estrategia de recolección combina técnicas de *web Scraping*, consumo de APIs públicas y uso de bases internas, focalizadas en los sectores de interés definidos.

Tabla 15*Fuentes de datos identificadas y volumen de datos proyectados*

Fuente	Tipo de datos	Registros recolectados	Estado legal de uso	Observaciones
Base de datos cámara de comercio <i>SalesQL</i>	Contactos empresariales <i>Leads</i> por sector	1200 850	Acceso público Autorizada	Información semiestructurada Estructura JSON limpia
<i>ApiFy</i>	<i>Leads</i> por sector	900	Autorizada	Estructura JSON limpia
<i>Sales Navigator</i>	<i>Leads</i> por sector	900	Autorizada	Estructura JSON limpia
<i>Web Scrapping</i>	Empresas por sector	500	Acceso público	Estructura HTML

Nota. Tabla que muestra una proyección en la recolección de los datos de las diferentes fuentes

identificadas. *Fuente.* Autor

La Tabla 15 justifica la diversidad del dataset, combinando fuentes públicas y privadas (SalesQL, Apify, Cámara de Comercio). Esta pluralidad de orígenes asegura que el modelo de IA sea entrenado con datos del contexto real de las PYMES colombianas, mitigando sesgos y aumentando la relevancia de la prospección.

Cada una de las fuentes de información será transformada y homogeneizada con el fin de alimentar una base unificada de empresas. Esta base incluirá campos esenciales como el nombre de la empresa, el sector económico, el correo de contacto, el sitio web, la ubicación y las palabras clave asociadas a su actividad, garantizando consistencia y trazabilidad en los datos recopilados.

Los sectores económicos priorizados para la extracción y análisis son: industria manufacturera, telecomunicaciones, despachos de abogados y servicios jurídicos, contabilidad, petróleo, gas y minería, así como educación y enseñanza superior.

La estrategia de extracción contempla diversas fuentes y métodos: la obtención de bases de datos de la cámara de comercio mediante solicitud oficial, el uso de SalesQL a través de la compra de créditos y su API disponible, la contratación de APIs de Scraping en ApiFy para

recolectar información de empresas y empleados, la extracción de enlaces desde *Sales Navigator* aplicando filtros específicos para procesos de Scraping, y finalmente, la validación de datos empresariales mediante técnicas de *web Scraping* orientadas a verificar la información utilizada en el envío de correos electrónicos.

Extracción de Datos por Medio de Scraping y APIs.

La extracción de datos se inicia con la exploración de empresas mediante filtros configurados en *Sales Navigator*. Esta herramienta permite aplicar criterios específicos de segmentación, lo cual facilita el proceso de recopilación de datos a través de técnicas de Scraping, utilizando APIs preestablecidas y ofrecidas por la plataforma Apify.

Figura 12

Filtro de los datos de Sale Navigator

The screenshot displays the LinkedIn Sales Navigator interface. At the top, there is a navigation bar with the LinkedIn logo and the text 'SALES NAVIGATOR', along with menu items: 'Inicio', 'Cuentas', 'Posibles clientes', and 'Mensajes'. Below this, the main interface is divided into several sections:

- Header:** 'Posible cliente' and 'Cuentas' are highlighted. A search bar contains 'Buscar palabras clave' and a 'Búsquedas guardadas' link. A 'Contraer' button is visible.
- Filters:** A section titled '501-1.000, 1.001-5.000, Colomb...' (3 filtros) includes a toggle for 'Guardar la búsqueda para recibir notifi...'. Below this are two columns of filter options:
 - Características de la empresa:** Includes filters for 'Ingresos anuales', 'Empleados en la empresa' (with active filters for '501-1.000' and '1.001-5.000'), 'Aumento de empleados en la empresa', 'Ubicación de la sede central' (with active filters for 'Colombia' and 'Ciudad de México, México'), 'Industria' (with active filter for 'Industria manufacturera'), 'Número de seguidores', and 'Empleados en el departamento'.
 - Spotlights:** Includes filters for 'Oportunidades de empleo', 'Actividad reciente', and 'Contacto' (marked 'Nuevo').
 - Proceso:** Includes filters for 'Empresas en el CRM', 'Cuentas guardadas', and 'Listas de cuentas' (marked 'Nuevo').
- Search Results:** A list of search results is shown, each with a checkbox, a company logo, the company name, industry, employee count, and a 'Guardar' button.
 - Cementos Argos:** Materiales de construcción · 3 mil+ empleados en LinkedIn. Description: 'Acerca Hacemos posible la construcción de sueños de vivienda e infraestructura que impulsan... ver más'.
 - Grupo Familia:** Industria manufacturera · 2 mil+ empleados en LinkedIn. Description: 'Acerca Fundada en 1958, Grupo Familia se enfoca en brindar bienestar a sus consumidores en ... ver más'.
 - Compañía de Galletas Noel S.A.S.:** Fabricación de alimentos y bebidas · 1,5 mil+ empleado en LinkedIn. Description: 'Acerca Somos una Compañía innovadora, vigente y sostenible, líder de la categoría de galleta... ver más'.
 - Productos Ramo S.A.:** Fabricación de alimentos y bebidas · 1,5 mil+ empleado en LinkedIn. Description: 'Acerca Somos una compañía de snacks colombiana impulsada por la innovación y la tecnol... ver más'.
- Footer:** 'Fijar filtros', 'Definiciones de los filtros avanzados', and 'Borrar todo'.

Fuente. Autoría Propia

Figura 13

Captura de links de Sales Navigator

The screenshot displays the LinkedIn Sales Navigator interface. At the top, there is a navigation bar with 'SALES NAVIGATOR' and tabs for 'Inicio', 'Cuentas', 'Posibles clientes', and 'Mensajes'. Below this, the 'Cuentas' tab is active, showing a search for 'Posible cliente' with filters for '501-1.000, 1.001-5.000, Bogotá,...' (3 filters). The interface is divided into three main sections:

- Características de la empresa:** Includes filters for 'Ingresos anuales', 'Empleados en la empresa' (501-1.000, 1.001-5.000), 'Aumento de empleados en la empresa', 'Ubicación de la sede central' (Área metropolitana de Medellín, Bogotá, Distrito Capital, Colombia; Área metropolitana de Cartagena; Cali, Valle del Cauca, Colombia; Área metropolitana de Barranquilla), and 'Industria' (Servicios financieros).
- Spotlights:** Includes 'Oportunidades de empleo', 'Actividad reciente', and 'Contacto' (Nuevo).
- Proceso:** Includes 'Empresas en el CRM', 'Cuentas guardadas', and 'Listas de cuentas' (Nuevo).

On the right side, there is a search bar and a list of companies with their LinkedIn profiles and employee counts:

- Porvenir Pensiones y Cesantías:** Servicios financieros · 2,5 mil+ empleados en LinkedIn. Acerca Somos la Administradora de Fondos de Pensiones de: y Cesantías líder en el mercado colomb... ver más. 3 contactos. Contratando en LinkedIn. Guardado.
- Scotiabank Colpatría:** Banca · 4 mil+ empleados en LinkedIn. Acerca Disfruta una experiencia digital, ágil, fácil y de: segura usando nuestra Banca Virtual. 6 contactos. Guardado.
- Banco Falabella Colombia:** Banca · 1,5 mil+ empleado en LinkedIn. Acerca Perteneceamos a uno de los holdings más grandes de: y consolidados de América Latina. El or... ver más. 2 contactos. Guardado.

Fuente. Autoría Propia

Luego de definir la extracción de los datos, se llevó a cabo la adquisición y configuración de una API de *Scraping* denominada LinkedIn Sales Navigator Scraper, disponible en la plataforma Apify. Esta herramienta utiliza como insumo los enlaces generados a partir de los filtros configurados en *Sales Navigator* y ejecuta un proceso automatizado de extracción de datos. El resultado de esta operación es la obtención de una base de datos preliminar que consolida información de empresas y posibles contactos (*leads*) asociados a cada una de ellas. Esta base de datos constituye un insumo clave para dar paso siguiente paso del proceso metodológico.

Figura 14

Configuración del api de LinkedIn Sales Navigator Scraper

The screenshot shows the Apify interface for the 'LinkedIn Sales Navigator Scraper' actor. The left sidebar contains navigation options: Soporte Personal, Apify Store, Actors, Development, Saved tasks, Runs, Integrations, Schedules, Storage, Proxy, Settings, Billing, Documentation, and Help & resources. The main area shows the 'Manual' configuration for the scraper, which includes a warning: 'When 'Deep scrape' option is enabled, don't scrape too fast to avoid linkedin account suspension. Recommended delay range is 5-30 seconds'. Below this, there is a section for 'LinkedIn cookie' with a JSON configuration for two cookies. The search URL is displayed at the bottom: `https://www.linkedin.com/sales/search/company?query=(filters%3AList(type%3ACOMPANY_HEADCOUNT%2Cvalu`.

```

1  {
2    "domain": ".linkedin.com",
3    "expirationDate": 1759200624,
4    "hostOnly": false,
5    "httpOnly": false,
6    "name": "mbox",
7    "path": "/",
8    "sameSite": "no_restriction",
9    "secure": true,
10   "session": false,
11   "storeId": null,
12   "value": "PC#74d5c608d4764f42bf716989b893310b.34_0#1759200624|session#269259109bf84d8bb"
13  },
14  {
15   "domain": ".linkedin.com",
16   "expirationDate": 1746310151.8849,
17   "hostOnly": false,
18   "httpOnly": false,
19   "name": "lms_ads",
20   "path": "/",
21   "sameSite": "no_restriction",
22   "secure": true,
23   "session": false,
24   "storeId": null,
25   "value": "AQF-VnE7BU1SMAAAAZX9s3b7MyYe46By01xQ9b00YZHoInNm6kV7093tAwc00q0e4NocCuD2PwDwq"
26  },
27  {
28   "domain": ".linkedin.com",
29   "expirationDate": 1746310151.8849,
30   "hostOnly": false,
31   "httpOnly": false,
32   "name": "lms_ads",
33   "path": "/",
34   "sameSite": "no_restriction",
35   "secure": true,
36   "session": false,
37   "storeId": null,
38   "value": "AQF-VnE7BU1SMAAAAZX9s3b7MyYe46By01xQ9b00YZHoInNm6kV7093tAwc00q0e4NocCuD2PwDwq"
39  },
40  }

```

Search URL or Leads list URL

`https://www.linkedin.com/sales/search/company?query=(filters%3AList(type%3ACOMPANY_HEADCOUNT%2Cvalu`

Fuente. Autoría Propia

Figura 15

Registro de ejecución de iteraciones de Scraping

The screenshot shows the APIFY interface for the 'LinkedIn Sales Navigator Scraper' actor. The actor is priced at \$39.00/month + usage. The interface includes a sidebar with navigation options like 'Apify Store', 'Actors', 'Development', 'Runs', 'Integrations', 'Schedules', 'Storage', 'Proxy', 'Settings', and 'Billing'. The main area displays a table of 10 recent runs with columns for Status, Task, Results, Usage, Started, Finished, Duration, Build, Origin, and Memory. The runs show various success rates and durations, with one failed run due to authorization issues.

Status	Task	Results	Usage	Started	Finished	Duration	Build	Origin	Memory
✓	Actor succeeded with 163 results in the dataset	163	\$1.134	2025-04-04 18:28:56	2025-04-04 19:49:00	1 h 20 m	1.020	🌐	2 GB
✓	Actor succeeded with 344 results in the dataset	344	\$2.332	2025-04-04 15:00:35	2025-04-04 17:51:11	2 h 48 m	1.020	🌐	2 GB
✓	Actor succeeded with 192 results in the dataset	192	\$1.641	2025-04-04 12:52:21	2025-04-04 14:51:47	1 h 59 m	1.020	🌐	2 GB
✓	Actor succeeded with 36 results in the dataset	36	\$0.362	2025-04-04 11:07:46	2025-04-04 12:47:35	23 m 27 s	1.020	🌐	2 GB
✓	Actor succeeded with 54 results in the dataset	54	\$0.510	2025-04-04 10:15:41	2025-04-04 11:03:27	33 m 56 s	1.020	🌐	4 GB
✓	Actor succeeded with 293 results in the dataset	293	\$2.554	2025-04-04 01:55:36	2025-04-04 10:11:49	3 h 15 s	1.020	🌐	2 GB
✗	Failed to authorize with linkedin. Please retry with new cookies	0	\$0.001	2025-04-04 01:40:54	2025-04-04 01:40:59	4 s	1.020	🌐	2 GB
✓	Actor succeeded with 33 results in the dataset	33	\$0.311	2025-04-01 18:34:09	2025-04-01 18:55:28	21 m 19 s	1.020	🌐	2 GB
✓	Actor succeeded with 100 results in the dataset	100	\$0.874	2025-04-01 11:11:30	2025-04-01 12:14:17	1 h 2 m	1.020	🌐	2 GB
✓	Actor succeeded with 372 results in the dataset	372	\$3.320	2025-04-01 09:07:30	2025-04-01 18:36:11	3 h 58 m	1.020	🌐	2 GB

Fuente. Autoría Propia

Figura 16

Resultado de los datos obtenidos desde el api usado

account/starred	companyName	companyPictureDisplayImage	description	employeeCount	employeeCountRange	employeeDisplayCount
false	Sophos Solutions LATAM	https://media.licdn.com/dms/ima	Sophos Solutions is now part of GF	1091	1001-5000	1K+
false	axity	https://media.licdn.com/dms/ima	Somos una compañía de Servicios c	2488	5001-10000	2K+
false	LinkTIC	https://media.licdn.com/dms/ima	Somos una fábrica de software colc	543	1001-5000	543
false	Carvajal Tecnología y Servicios	https://media.licdn.com/dms/ima	Creamos, habilitamos e implemen	1729	5001-10000	1.5K+
false	ScottiaTech	https://media.licdn.com/dms/ima	Somos un grupo de ingeniería de cl	730	1001-5000	730
false	KIO	https://media.licdn.com/dms/ima	Desde 2002, somos el proveedor lik	2265	1001-5000	2K+
false	Seguritech	https://media.licdn.com/dms/ima	A lo largo de más de 29 años hemo	1458	1001-5000	1K+
false	Celba Software	https://media.licdn.com/dms/ima	Celba Software provides custom sc	617	501-1000	617
false	Aportes en Línea	https://media.licdn.com/dms/ima	Co-creamos tecnologías de informi	363	201-500	363
false	Intergrupo	https://media.licdn.com/dms/ima	Established in 1995, InterGrupo is c	766	1001-5000	766
false	VASS LATAM	https://media.licdn.com/dms/ima	VASS is a leading digital solutions c	477	1001-5000	477
false	PersonalSoft	https://media.licdn.com/dms/ima	Leaders across the Americas in IT S	951	501-1000	951
false	ADL Digital Lab	https://media.licdn.com/dms/ima	Somos un laboratorio de transform	660	501-1000	660
false	Bluetab América, an IBM Company	https://media.licdn.com/dms/ima	Equipo internacional de tecnólogo	331	201-500	331
false	Sofka Technologies	https://media.licdn.com/dms/ima	Sofka Technologies is dedicated to	999	501-1000	999
false	Esri Colombia	https://media.licdn.com/dms/ima	La tecnología de Esri ayuda a las pe	394	201-500	394
false	Universidad del Tolima		La Universidad del Tolima es un en	1762	501-1000	1.5K+
false	Heinsohn	https://media.licdn.com/dms/ima	Heinsohn Business Technology is a	923	1001-5000	923
false	RobinFood	https://media.licdn.com/dms/ima	RobinFood (former Muy) is Latin Ai	288	501-1000	288
false	Aranda Software	https://media.licdn.com/dms/ima	Aranda Software es la compañía lid	220	51-200	220
false	Olimpia	https://media.licdn.com/dms/ima	Olimpia es una empresa del Grupo	437	201-500	437
false	Frubana	https://media.licdn.com/dms/ima	We are a fast growing technology B	2072	501-1000	2K+
false	FPT Latin America	https://media.licdn.com/dms/ima	FPT Latin America is the overseas b	237	201-500	237
false	melonn	https://media.licdn.com/dms/ima	Enfócate en crecer mientras nosot	246	51-200	246
false	ARUS Oficial	https://media.licdn.com/dms/ima	Somos una compañía de Servicios c	1540	1001-5000	1.5K+
false	Periferia IT Group	https://media.licdn.com/dms/ima	Somos una empresa de tecnología	873	501-1000	873
false	Scitum	https://media.licdn.com/dms/ima	Scitum is the largest information s	1111	1001-5000	1K+
false	PROSA	https://media.licdn.com/dms/ima	PROSA ha sido un factor estratégic	770	501-1000	770
false	DIGITALWARE	https://media.licdn.com/dms/ima	Somos líderes en soluciones de ne	530	201-500	530
false	T-Systems México	https://media.licdn.com/dms/ima	T-Systems is a strategic partner for	1141	1001-5000	1K+
false	Holcim ADC	https://media.licdn.com/dms/ima	The technology partner for Holcim	269	201-500	269
false	Perficient Latin America	https://media.licdn.com/dms/ima	Perficient es la consultora digital g	870	5001-10000	870
false	Masiv	https://media.licdn.com/dms/ima	Take your customer experiences tc	228	201-500	228
false	SQA - Software Quality Assurance	https://media.licdn.com/dms/ima	Empowering growth-driven technc	499	201-500	499
false	Choucalr Testing S.A.	https://media.licdn.com/dms/ima	Es la compañía pionera de pruebas	753	501-1000	753
false	Nexsys	https://media.licdn.com/dms/ima	Nexsys es el primer mayorista de v	472	501-1000	472
false	Chiper	https://media.licdn.com/dms/ima	Chiper is the leading B2B e-comm	445	201-500	445
false	C3NTR0 Telecom	https://media.licdn.com/dms/ima	Somos C3NTR0, una empresa mexi	507	501-1000	507
false	Periferia IT Group	https://media.licdn.com/dms/ima	Somos una empresa de tecnología	873	501-1000	873

Fuente. Autoría Propia

Este bloque de figuras (12 a 16) documenta la automatización del flujo de captura. Desde el filtrado en Sales Navigator hasta la ejecución en Apify, se evidencia un proceso reproducible que elimina la carga manual identificada en el diagnóstico del Capítulo 7, cumpliendo así con el objetivo de eficiencia operativa del prototipo.

El siguiente paso consiste en el procesamiento de los archivos obtenidos a partir del Scraping realizado en Apify. Para ello, se implementó un proceso automatizado en *Python* que permite consultar la herramienta SalesQL, con el objetivo de obtener información de contacto tanto de los *leads* como de las empresas identificadas previamente. Este proceso se ejecuta mediante iteraciones diarias de hasta 600 consultas, respetando el límite de uso diario establecido por la API de SalesQL. Como resultado, se generan de manera automática dos archivos en formato Excel, los cuales se utilizan posteriormente para alimentar la base de datos general del proyecto.

Descripción del Código.

Consulta de Datos en SalesQL Mediante API. El siguiente fragmento de código corresponde a la primera parte del proceso automatizado desarrollado en *Python* para la recolección de datos de contacto a partir de URLs de perfiles de LinkedIn. Inicialmente, se monta *Google Drive* en el entorno de *Google Colab* para permitir el acceso a los archivos de entrada y salida. Posteriormente, se importan las librerías necesarias: *pandas* para la manipulación de datos, *tqdm* para la visualización de barras de progreso, y *requests* para realizar peticiones HTTP.

Se define una variable que almacena la clave de acceso a la API de SalesQL (*API_KEY*), seguida de la función *consultar_salesql()*, la cual recibe como parámetros la URL del perfil de LinkedIn y un identificador del empleado. Esta función construye la solicitud a la API de

SalesQL, enviando la URL del perfil como parámetro y autenticándose con la clave previamente definida. Se realiza una petición GET y, dependiendo de la respuesta obtenida, se clasifica si la persona fue encontrada, no fue localizada o si ocurrió un error en la solicitud. En todos los casos, se imprime un mensaje informativo y se devuelve el resultado correspondiente.

Figura 17

Parte 1 del código de Scraping SalesQL

```
[ ] from google.colab import drive
drive.mount('/content/drive')

Mounted at /content/drive

import pandas as pd
from tqdm import tqdm
import requests

API_KEY = "*****"

def consultar_salesql(linkedin_url, employee_id):
    url = f"https://api-public.salesql.com/v1/persons/enrich/?linkedin_url={linkedin_url}"
    headers = {
        "accept": "application/json",
        "Authorization": f"Bearer {API_KEY}"
    }

    try:
        response = requests.get(url, headers=headers)
        data = response.json()

        if "detail" in data and data["detail"] == "Person not found":
            print(f"❌ NO ENCONTRADO {employee_id}")
            return None
        else:
            print(f"✅ OK {employee_id}")
            return data
    except Exception as e:
        print(f"⚠️ ERROR {employee_id} - {str(e)}")
        return None
```

Fuente. Autoría Propia

Lectura y Preparación de los Datos para Consulta.

En esta segunda parte del código se procede a la lectura de los archivos Excel que contienen los datos de entrada, específicamente los registros de empleados y empresas, los cuales se almacenan en estructuras tipo *DataFrame* utilizando la biblioteca Pandas. A continuación, se imprime en consola la cantidad total de registros cargados para cada conjunto de datos, permitiendo verificar que la carga fue exitosa.

Se define el número máximo de consultas a ejecutar (`num_requests = 600`), respetando el límite diario impuesto por la API de SalesQL. Posteriormente, se normaliza la columna `salesql_validation`, que indica si un empleado ya ha sido procesado o no. Para ello, se realiza una conversión de tipo y una limpieza de los valores, transformándolos en valores booleanos válidos.

Finalmente, se filtran aquellos empleados que aún no han sido validados, generando un subconjunto de registros pendientes por procesar, limitado a las primeras 600 filas. Este conjunto será utilizado en la fase posterior de consultas automatizadas.

Figura 18

Parte 2 del código de Scraping SalesQL

```

# Leer archivos
empleados = pd.read_excel("/content/drive/MyDrive/SalesQL/BD/Empleados.xlsx")
orgs = pd.read_excel("/content/drive/MyDrive/SalesQL/BD/EmpresasSalesQL.xlsx")#1493

# Mostrar cuántos empleados y empresas hay
print(f"👤 Empleados cargados: {len(empleados)}")
print(f"🏢 Empresas cargadas: {len(orgs)}")

# ¿Cuántos quieres procesar?
num_requests = 600

# Normalizar y convertir tipo de columna de validación
empleados["salesql_validation"] = empleados["salesql_validation"].astype(str).str.strip().str.lower()
empleados["salesql_validation"] = empleados["salesql_validation"].map(lambda x: True if x in ["true", "1", "1.0"] else pd.NA)
empleados["salesql_validation"] = empleados["salesql_validation"].astype("boolean")

# Filtrar empleados aún no validados
pendientes = empleados[empleados["salesql_validation"].isna()].copy()
pendientes = pendientes.head(num_requests)

# 📄 Mostrar los primeros empleados a procesar
pendientes.head()

```

👤 Empleados cargados: 17503
 🏢 Empresas cargadas: 2503

	employee_id	linkedin_url	first_name	last_name	full_name	profile_picture	organization_id	salesql_title	salesql_headline	salesql_email_work	salesql_email_direct	sales
130	ACwAAAOYq8BYXPmsKw0dTreEgEz4pQEodyPY	https://www.linkedin.com/in/ACwAAAOYq8BYXPmsK...	Diego	Moreno	Diego Moreno	https://media.licdn.com/dms/image/v2/D4E03AQFn...	10854356	NaN	NaN	NaN	NaN	
693	ACwAACw9KSYB1ly-H7_M_3_BizgmM141SpG0U	https://www.linkedin.com/in/ACwAACw9KSYB1ly-H7...	Marcos	Trejo Campos	Marcos Trejo Campos	https://media.licdn.com/dms/image/v2/C4E03AQF8...	1140756	NaN	NaN	NaN	NaN	
703	ACwAAAJhbkBSinV4JffeYN25ULuWIS_ThuJy	https://www.linkedin.com/in/ACwAAAJhbkBSinV4...	Fernando	Escobar	Fernando Escobar	https://media.licdn.com/dms/image/v2/C5603AQHr...	1334051	NaN	NaN	NaN	NaN	

Fuente. Autoría Propia

Ejecución del proceso iterativo de consulta y almacenamiento de resultados. En esta fase, el *script* ejecuta un ciclo iterativo sobre los registros de empleados pendientes por procesar, haciendo uso de *tqdm* para visualizar el progreso en tiempo real. Para cada empleado, se extrae su identificador y su enlace de perfil en LinkedIn. En caso de que el enlace no esté disponible, se marca el registro como procesado y se continúa con el siguiente.

Con el fin de evitar la saturación de la API de SalesQL, se implementa una pausa de 1.3 segundos entre cada solicitud. Luego, se llama a la función `consultar_salesql()` y se actualiza el campo de validación del empleado como procesado, independientemente del resultado.

Si la respuesta de la API contiene información válida, se extraen y almacenan datos relevantes como el cargo (*title*), titular profesional (*headline*), ubicación, correos electrónicos y números telefónicos diferenciados por tipo (laborales y personales). Además, si se incluye información sobre la organización asociada, se construye un nuevo registro con datos como el nombre de la empresa, industria, tamaño, ubicación, dominio web y enlace de LinkedIn. Este nuevo registro se agrega a la base de datos de empresas, evitando duplicados mediante la verificación del `employee_id`.

Figura 19

Parte 3 del código de Scraping SalesQL

```

import time # para pausar entre requests

# Procesar iterativamente
for i, row in tqdm(pendientes.iterrows(), total=len(pendientes)):
    emp_id = row["employee_id"]
    idx = empleados[empleados["employee_id"] == emp_id].index[0]

    linkedin_url = row["linkedin_url"]
    if pd.isna(linkedin_url):
        empleados.at[idx, "salesql_validation"] = True
        continue

    # Pausa para no saturar la API
    time.sleep(1.3)

    resultado = consultar_salesql(linkedin_url, emp_id)

    empleados.at[idx, "salesql_validation"] = True # Marcar como procesado

    if resultado and resultado != "unauthorized":
        empleados.at[idx, "salesql_title"] = resultado.get("title", "")
        empleados.at[idx, "salesql_headline"] = resultado.get("headline", "")
        empleados.at[idx, "salesql_location"] = resultado.get("location", "")
        empleados.at[idx, "salesql_salesql_uid"] = resultado.get("uid", "")

        for email in resultado.get("emails", []):
            if email["type"] == "Work":
                empleados.at[idx, "salesql_email_work"] = email["email"]
            elif email["type"] == "Direct":
                empleados.at[idx, "salesql_email_direct"] = email["email"]

        for phone in resultado.get("phones", []):
            if phone["type"] == "Work":
                empleados.at[idx, "salesql_phone_work"] = phone["phone"]
            elif phone["type"] == "Personal":
                empleados.at[idx, "salesql_phone_personal"] = phone["phone"]

    # Agregar fila a la tabla EmpresasSalesQL
    org_data = resultado.get("organization", {})
    if org_data:
        nueva_fila_org = {
            "employee_id": emp_id,
            "salesql_uid": org_data.get("uid", ""),
            "salesql_organization_name": org_data.get("name", ""),
            "salesql_website_domain": org_data.get("website domain", ""),
            "salesql_industry": org_data.get("industry", ""),
            "salesql_location": org_data.get("location", ""),
            "salesql_organization_linkedin": org_data.get("linkedin_url", ""),
            "salesql_organization_founded_year": org_data.get("founded_year", ""),
            "salesql_organization_size": org_data.get("number_of_employees", "")
        }

    # Evitar duplicados por employee_id
    if not orgs["employee_id"].isin([emp_id]).any():
        orgs = pd.concat([orgs, pd.DataFrame([nueva_fila_org])], ignore_index=True)

```

Fuente. Autoría Propia

Las figuras (17 a 19) capturas del código en Python muestran la orquestación del enriquecimiento de datos. El uso de la API de SalesQL con manejo de pausas técnicas (1.3s) y límites de consulta (600 diarios) demuestra una implementación profesional que respeta las políticas de uso de servicios externos mientras asegura la obtención de correos y teléfonos validados.

Figura 20

Resultado de los logs de ejecución del script

0%		1/600	[00:01<19:46,	1.98s/it]	[X]	NO ENCONTRADO]	ACwAAAO0Vq8BYXPmsKwEdTreEgEefZJ4pQEdyPY
0%		2/600	[00:03<19:37,	1.97s/it]	[OK]	ACwAACw9KSYB1Iy-tf7_M_3_8LzgmM415rpg0U	
0%		3/600	[00:06<20:32,	2.06s/it]	[OK]	ACwAAAjhbKBSInvi4JfTeYn2SULuhfS_-fhUjY	
1%		4/600	[00:08<21:28,	2.16s/it]	[OK]	ACwAAAJueHw89EyY4oh5wUCDUAYkXf0itN-kxH	
1%		5/600	[00:10<20:45,	2.09s/it]	[X]	NO ENCONTRADO]	ACwAAEplcpMB1hIw9gcZFuVMZwWQgdvCrME0i3c
1%		6/600	[00:12<20:50,	2.10s/it]	[OK]	ACwAAEm2qB4bcvF10N6Hq7Lb_AgXsum54k2LTHM	
1%		7/600	[00:14<20:55,	2.12s/it]	[OK]	ACwAABoL2L5BAPnR2-5yYmLFelCfhyrNQze-y0I	
1%		8/600	[00:16<20:58,	2.13s/it]	[OK]	ACwAA8gtmw48dT_kvrrnlfkizjj15r9i1TASHHJU	
2%		9/600	[00:18<20:12,	2.05s/it]	[X]	NO ENCONTRADO]	ACwAAC2tESY8qHrs3ZSw6R6Qauhu-0TZK0by7Mk
2%		10/600	[00:20<19:39,	2.00s/it]	[OK]	ACwAACAF4408MMU3IcZvhayxJMLkvfnzN6didhe	
2%		11/600	[00:22<20:01,	2.04s/it]	[OK]	ACwAAAICR48B_SZrnt9p-xIqp1mmJ6rArE-oFfg	
2%		12/600	[00:24<19:27,	1.99s/it]	[X]	NO ENCONTRADO]	ACwAA8Hsuv8B8MxLXHUEKwNgv58Fsbu6ZEDWwja
2%		13/600	[00:26<19:52,	2.03s/it]	[OK]	ACwAAAQZwFUBlyUVsvlYioLyH25Dov5t0A2AJM	
2%		14/600	[00:28<19:23,	1.98s/it]	[X]	NO ENCONTRADO]	ACwAAAXR_I88scZm50ieo_K2up6N0X2FR145G8
2%		15/600	[00:30<19:47,	2.03s/it]	[OK]	ACwAACgw9-IBPr6S3z0B_3W/nQ6CICFlgitptk	
3%		16/600	[00:32<19:29,	2.00s/it]	[OK]	ACwAAB529pw8AxyLDnXhotMB202N08dAnqMho	
3%		17/600	[00:35<22:11,	2.28s/it]	[OK]	ACwAAAE9FMEBdVln7BR2L60B14xZhlaxt4uKfCy	
3%		18/600	[00:37<21:36,	2.23s/it]	[OK]	ACwAAC-IItsBqArBtIdaEovv4b9I9E5TcEzVUQs	
3%		19/600	[00:39<21:15,	2.20s/it]	[OK]	ACwAABcKM44BJKmsqx3K2U0byThcM4qm7w7Ip4	
3%		20/600	[00:42<21:14,	2.20s/it]	[OK]	ACwAACj6l0IBEZEeQc2VHwgDft_RpFngmsZNP8	
4%		21/600	[00:44<20:54,	2.17s/it]	[OK]	ACwAAAMmM4BmmVr1lF0ub_022961CPkMdsBDw	
4%		22/600	[00:46<20:38,	2.14s/it]	[OK]	ACwAAC40UUtIaZhdZcxlyVRhi-5DV2cFrPymfM	
4%		23/600	[00:48<20:29,	2.13s/it]	[OK]	ACwAAAYC82I8ATeomjJsuL_m9g8o6cZmJAI60z0	
4%		24/600	[00:50<20:53,	2.18s/it]	[OK]	ACwAAAnm0k4BTH5T0Ho6Mlum0Uusuhn1RbFQk	
4%		25/600	[00:52<20:43,	2.16s/it]	[OK]	ACwAAC1S5gpbw8i10A0NwMod7H1icXmc_IDMk0tA	
4%		26/600	[00:54<19:45,	2.06s/it]	[X]	NO ENCONTRADO]	ACwAACgrso4B-HayIh3AntZerNk1rwz0Ld6Aw5I
4%		27/600	[00:56<19:09,	2.01s/it]	[X]	NO ENCONTRADO]	ACwAACdFr0uBjX2JD00je80VeIa2y-pzdZPcwr8
5%		28/600	[00:58<18:40,	1.96s/it]	[X]	NO ENCONTRADO]	ACwAADZ8eJIEPq2aB8khdZp5F9rk0A44Tw531U
5%		29/600	[01:00<18:24,	1.93s/it]	[X]	NO ENCONTRADO]	ACwAAAbLF60BafP3S19utSdbX0_vv8CboFbm0s
5%		30/600	[01:02<18:14,	1.92s/it]	[X]	NO ENCONTRADO]	ACwAAEPZ3UABzjJdN-3wDi_ZRSharTzhK8P_6HU
5%		31/600	[01:04<19:35,	2.07s/it]	[OK]	ACwAACs0yVY8z706Cms-Vewpb-8My5I4Zzy3fc	
5%		32/600	[01:06<20:12,	2.13s/it]	[OK]	ACwAAD02gqYBLzaZ8j4Dzj19SKmLlAhuMHoUZSs	
6%		33/600	[01:08<20:01,	2.12s/it]	[OK]	ACwAAAB83ts488Vn27hMvLcjstzrjB86F-yi4Opw	
6%		34/600	[01:10<19:52,	2.11s/it]	[OK]	ACwAABVkkk8Bddv3fkerl57bkpxsupc2P9agUlW	
6%		35/600	[01:12<19:09,	2.04s/it]	[X]	NO ENCONTRADO]	ACwAABHsW88MxLXHUEKwNgv58Fsbu6ZEDWwja
6%		36/600	[01:14<18:33,	1.97s/it]	[X]	NO ENCONTRADO]	ACwAAE8zj88Fyl0xp8HoIgvHU_klEpvt3VTnY
6%		37/600	[01:16<18:51,	2.01s/it]	[OK]	ACwAADizUJk83Cqkh9aXi5I8YrGTVwNLCKBHA	
6%		38/600	[01:18<19:08,	2.04s/it]	[OK]	ACwAACChTe6E84NMts--7_DoHlqkacIIsBaKeXC4	
6%		39/600	[01:20<18:56,	2.03s/it]	[OK]	ACwAAAQZwFUBlyUVsvlYioLyH25Dov5t0A2AJM	
7%		40/600	[01:22<19:01,	2.04s/it]	[OK]	ACwAABmtMRMBGGL6uGHXNSAnF-vx70EWmKvzac	
7%		41/600	[01:24<18:34,	1.99s/it]	[X]	NO ENCONTRADO]	ACwAAAUl074846dJ_eVL08k12b1Gk1hoRico75c
7%		42/600	[01:26<18:39,	2.01s/it]	[OK]	ACwAAC2D4JgBjqlSMCL1L5G9TkeBVAwu1GzsFie	
7%		43/600	[01:28<18:38,	2.01s/it]	[OK]	ACwAACgFogw8jOxryLPq5k4YiJY8dYD8grH5pAc	
7%		44/600	[01:30<18:49,	2.03s/it]	[OK]	ACwAAECXy8BARQVPMdUInmVDMVPJbOed4jIXmM	
8%		45/600	[01:33<19:15,	2.08s/it]	[OK]	ACwAAAgfJR88wCZ9FCF5FTyhK9LTOOKwM3e0Hw	
8%		46/600	[01:35<19:14,	2.08s/it]	[OK]	ACwAAA17vjgB60Htkz82H522vMBqSN9kVlXSEiy	
8%		47/600	[01:37<19:14,	2.09s/it]	[OK]	ACwAABACHPuBUlUmgzgd58kuc1ARhXkbfa_YbiKk	
8%		48/600	[01:39<19:12,	2.09s/it]	[OK]	ACwAABP6xngBR4HI0wml0LlVH0dTNr7allEPkfk0	
8%		49/600	[01:41<18:37,	2.03s/it]	[X]	NO ENCONTRADO]	ACwAAA3btNo8s2T_aYUvub_xhLI4MdDw7i-1-Dxu
8%		50/600	[01:43<18:03,	1.97s/it]	[X]	NO ENCONTRADO]	ACwAACCLzTABU-6-KRQ99xVBRQdNVY6KEhCk1no
8%		51/600	[01:44<17:38,	1.93s/it]	[X]	NO ENCONTRADO]	ACwAAFBw0X088qb2iaGxU2c6M725deCw7DRuIn0
9%		52/600	[01:46<17:25,	1.91s/it]	[X]	NO ENCONTRADO]	ACwAACDFVJcBgq7tqDMP0z-mjy2Yx1JfMgeElg0
9%		53/600	[01:48<17:51,	1.96s/it]	[OK]	ACwAAC9TPyo89GCjehTODXPobTt02oPEVGvfg	
9%		54/600	[01:50<17:27,	1.92s/it]	[X]	NO ENCONTRADO]	ACwAAAInwBEBNz2xHTPyNS9i1JDiwP_djhp0Ake
9%		55/600	[01:52<17:56,	1.97s/it]	[OK]	ACwAAAAd15i08gr22bkrFrc1t70y8YjgbkxTH33A	
9%		56/600	[01:55<18:40,	2.06s/it]	[OK]	ACwAAAAn1t-48j0z507QLIs1k95yTR1VbaCweOvg	
10%		57/600	[01:56<18:01,	1.99s/it]	[X]	NO ENCONTRADO]	ACwAAEEBggUByHhKwBkbVNZwzC2KN5LL1-2DiY0
10%		58/600	[01:58<18:14,	2.02s/it]	[OK]	ACwAAAGtgkAB48r_3_ijj5Sm6xThF-k1Di9wF8dU	

Fuente. Autoría Propia

Almacenamiento de resultados en archivos de video. Finalmente, una vez completado el proceso de consulta y enriquecimiento de los datos, se procede a guardar los resultados actualizados en archivos Excel ubicados en *Google Drive*. Se exportan tanto la base de datos de empleados como la de empresas, sobrescribiendo los archivos originales para mantener la información sincronizada. Este paso garantiza la persistencia de los datos procesados y permite su uso posterior en fases adicionales del proyecto. Un mensaje de confirmación se imprime en consola para asegurar que la operación de guardado se ha ejecutado correctamente.

Finalmente podemos observar los resultados de la iteración en las hojas de Excel como base de datos preliminar.

Figura 21

Evidencia resultados tabla empleados

employee_id	last_name	first_name	email	phone_number	extension	department_id	job_id	manager_id	hire_date	salary	commission_pct	validation
ACwAACJ	https://www.linkedin.com/in/juan-pablo-n	Juan Pablo N.										ec4b8ebb #####
ACwAABN	https://www.linkedin.com/in/circe-soto	Circe Soto										#####
ACwAAAB	https://www.linkedin.com/in/jonathan-c	Jonathan C.										#####
ACwAABx	https://www.linkedin.com/in/victoria-a-ortega-m	Victoria Andrea Ortega Morgado										#####
ACwAAAC	https://www.linkedin.com/in/ivan-contreras	Ivan Contreras										#####
ACwAAAD	https://www.linkedin.com/in/andrea-torres	Andrea Torres										#####
ACwAAAE	https://www.linkedin.com/in/jorge-velasco	Jorge Velasco										#####
ACwAAC3	https://www.linkedin.com/in/carlos-eduardo-clavel	Carlos Eduardo Clavel Morales										#####
ACwAAAT	https://www.linkedin.com/in/carlos-ma-diaz-esparza	Carlos Manuel Diaz Esparza										#####
ACwAABB	https://www.linkedin.com/in/sergio-lopez-diaz	Sergio López Díaz										#####
ACwAAA6	https://www.linkedin.com/in/gilberto-hernandez-flores	Gilberto Israel Hernández Flores										#####
ACwAAAC	https://www.linkedin.com/in/leonardo-brum-ramirez	Leonardo F. Brum Ramirez										#####
ACwAAA1	https://www.linkedin.com/in/alejandra-alarcon	Alejandra Alarcón Velázquez										#####
ACwAAAR	https://www.linkedin.com/in/lucero-o	Lucero O.										#####
ACwAAEP	https://www.linkedin.com/in/luis-arredondo	Luis Arredondo										#####
ACwAABN	https://www.linkedin.com/in/hector-ruiz	Hector Ruiz										#####
ACwAABb	https://www.linkedin.com/in/julio-adar-alborez-lopez	Julio Adán Alborez López										#####
ACwAAAR	https://www.linkedin.com/in/manuel-g	Manuel G.										#####
ACwAAC1	https://www.linkedin.com/in/luis-alber-padrón-v	Luis Alberto Padrón V.										#####
ACwAAE9	https://www.linkedin.com/in/eduardo-marroquin	Eduardo Marroquín Pineda										#####
ACwAAAk	https://www.linkedin.com/in/angelica-ruiz	Angelica Ruiz										#####

Fuente. Autoría Propia

Figura 22

Evidencia resultados tabla empresas

employee_id	last_name	first_name	email	phone_number	extension	department_id	job_id	manager_id	hire_date	salary	commission_pct	validation
ACwAACJ	d483339c-	KPMG México	kpmg.com	Automoti	Mexico M	https://lir	1946	2418				#####
ACwAABN	d483339c-	KPMG México	kpmg.com	Automoti	Mexico M	https://lir	1946	2418				#####
ACwAABx	d483339c-	KPMG México	kpmg.com	Automoti	Mexico M	https://lir	1946	2418				#####
ACwAAAC	d483339c-	KPMG México	kpmg.com	Automoti	Mexico M	https://lir	1946	2418				#####
ACwAAAE	d483339c-	KPMG México	kpmg.com	Automoti	Mexico M	https://lir	1946	2418				#####
ACwAAC3	d483339c-	KPMG México	kpmg.com	Automoti	Mexico M	https://lir	1946	2418				#####
ACwAAAT	9bc1c1ef-	Colegio de Contadores Públicos de México, A.C.	contadore	Online Me	(North An	https://lir	1949	640				#####
ACwAABb	9bc1c1ef-	Colegio de Contadores Públicos de México, A.C.	contadore	Online Me	(North An	https://lir	1949	640				#####
ACwAAAC	d7203160-	Grupo BBA	grupobba	Airlines/A	(North An	https://lir	2013	49				#####
ACwAAA1	f8b5d1f7-	URBI	urbi.com	Architect	(North An	https://linkedin.com	1106					#####
ACwAABN	496de871-	Instituto Mexicano de Contadores Públicos A.C.	imcp.org	Automoti	(North An	https://lir	1923	356				#####
ACwAABb	496de871-	Instituto Mexicano de Contadores Públicos A.C.	imcp.org	Automoti	(North An	https://lir	1923	356				#####
ACwAAAR	4c5f022d-	Grupo Helvex	helvex.co	Automoti	(North An	https://lir	1950	751				#####
ACwAAC1	496de871-	Instituto Mexicano de Contadores Públicos A.C.	imcp.org	Automoti	(North An	https://lir	1923	356				#####
ACwAAE9	567d69dd-	MAPI Asesores Fiscales	mapiases	Civic & So	Mexico, (f	https://lir	2023	12				#####
ACwAAAK	496de871-	Instituto Mexicano de Contadores Públicos A.C.	imcp.org	Automoti	(North An	https://lir	1923	356				#####
ACwAAAZ	496de871-	Instituto Mexicano de Contadores Públicos A.C.	imcp.org	Automoti	(North An	https://lir	1923	356				#####
ACwAAAs	e180c660-	Philippi Prietocarrizosa Ferrero DU & Uria	ppulegal	Human Re	Bogotá, (C	https://lir	2015	571				#####
ACwAAAC	e180c660-	Philippi Prietocarrizosa Ferrero DU & Uria	ppulegal	Human Re	Bogotá, (C	https://lir	2015	571				#####
ACwAAAS	e180c660-	Philippi Prietocarrizosa Ferrero DU & Uria	ppulegal	Human Re	Bogotá, (C	https://lir	2015	571				#####
ACwAAC1	eadcc22d-	SoftwareOne	softwarec	Security &	(Europe),	https://lir	2000	10791				#####

Fuente. Autoría Propia

Limpieza de Duplicados, Normalización de Campos, Validación Semántica y Estructural, Clasificación Automática.

En esta etapa se ejecutaron tareas fundamentales para garantizar que los datos utilizados en etapas posteriores del modelo sean fiables, coherentes y útiles para análisis predictivo y toma de decisiones automatizadas. Se siguió rigurosamente la metodología CRISP-ML(Q), específicamente los pasos *de Data Preparation, Data Quality Assessment y Feature Engineering*, con el objetivo de preparar los datos con un estándar industrial y científico que permita su reutilización y explotación en sistemas inteligentes.

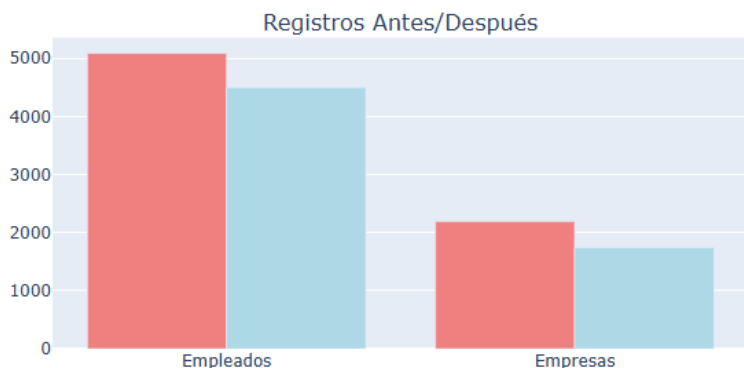
Limpieza de Duplicados: Exactos y Aproximados

Se aplicaron técnicas híbridas para la detección y eliminación de duplicados en las bases de datos de empleados (5.085 registros) y empresas (2.191 registros). Inicialmente se eliminaron duplicados exactos utilizando funciones nativas de *Pandas*. Posteriormente, se implementó un algoritmo de similitud basado en la librería *RapidFuzz* (una versión optimizada de *FuzzyWuzzy*) para detectar duplicados aproximados en campos como nombres y empresas, utilizando estrategias de *blocking* para escalar la búsqueda en grandes volúmenes de datos.

Resultados Técnicos. Empleados: Se eliminaron 581 duplicados exactos y 6 duplicados aproximados. Empresas: Se eliminaron 444 duplicados exactos y 4 duplicados aproximados.

Figura 23

Reducción de registros antes y después de la limpieza



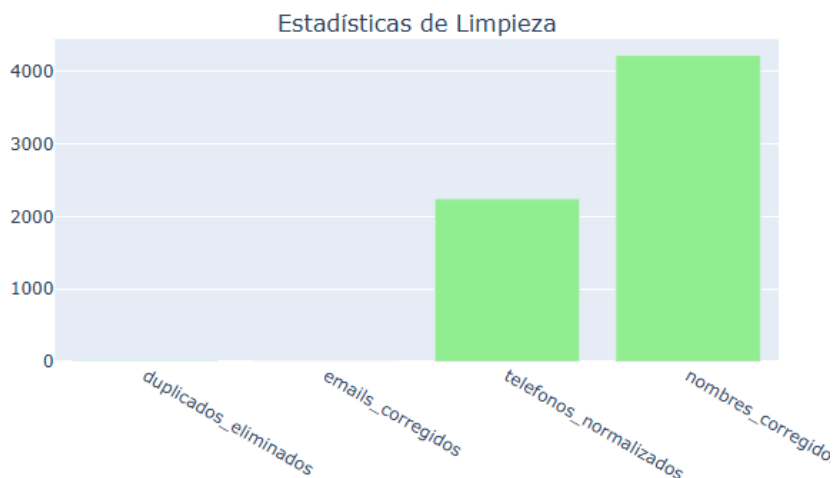
Fuente. Autoría Propia

Normalización de Campos Críticos

Posteriormente, se aplicó un proceso exhaustivo de normalización sobre campos sensibles como nombres, correos electrónicos, teléfonos y ubicaciones, utilizando expresiones regulares, técnicas de *Text Preprocessing* con *spaCy* y validación estructural. En el caso de los nombres, se realizó capitalización, eliminación de títulos y corrección de errores ortográficos; para los correos y teléfonos, se validaron los formatos y se transformaron en estructuras homogéneas; y en las ubicaciones, se aplicaron procesos de geolocalización y estandarización de nombres mediante la librería *geopy*. Como resultado, se corrigieron 4.225 nombres, se normalizaron 2.246 teléfonos, se estandarizaron 49 ubicaciones en una muestra representativa y se confirmó que el 100% de los correos electrónicos ya se encontraba en un formato válido.

Figura 24

Estadística de limpieza del proceso de ML



Fuente. Autoría Propia

Validación Semántica y Estructural

En esta etapa se desarrolló una lógica automatizada para detectar valores anómalos, estructuras malformadas e inconsistencias semánticas dentro del conjunto de datos. Las anomalías identificadas fueron registradas y, dependiendo de su complejidad, tratadas parcialmente o marcadas para revisión manual posterior. Entre los principales hallazgos se encontraron 2.110 empleados sin correo electrónico, 2.537 sin número de teléfono, 31 empresas sin ubicación y 3 empresas sin registro del número de empleados. Este análisis permitió evidenciar vacíos estructurales relevantes que posteriormente fueron incorporados como *features* dentro del modelo de evaluación de calidad, fortaleciendo su capacidad de detección y mejora continua.

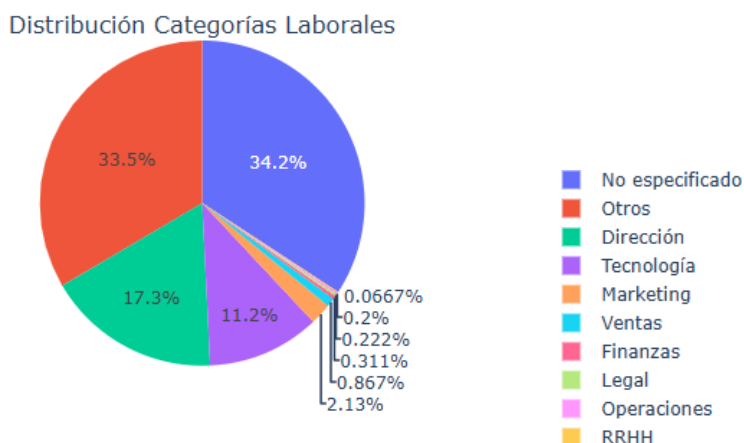
Clasificación Automática Mediante Machine Learning

Con el objetivo de enriquecer semánticamente el *dataset*, se aplicaron modelos de clasificación automática sobre dos variables categóricas críticas:

Clasificación de Títulos Laborales. Para la clasificación de títulos laborales se utilizó el modelo all-MiniLM-L6-v2 de *Sentence Transformers*, con el cual los textos de los títulos fueron transformados en vectores de *embeddings* semánticos. Posteriormente, se entrenó un clasificador supervisado basado en *logistic regression* con validación cruzada, empleando un conjunto previamente etiquetado. Este proceso permitió obtener una clasificación precisa en diez categorías laborales relevantes para el análisis de perfilamiento. La distribución resultante reflejó 778 registros en la categoría de Dirección, 504 en Tecnología, 96 en *Marketing* y 3.045 en la categoría de Otros o No especificado, que incluye roles no reconocibles por el modelo de aprendizaje automático.

Figura 25

Distribución de títulos laborales



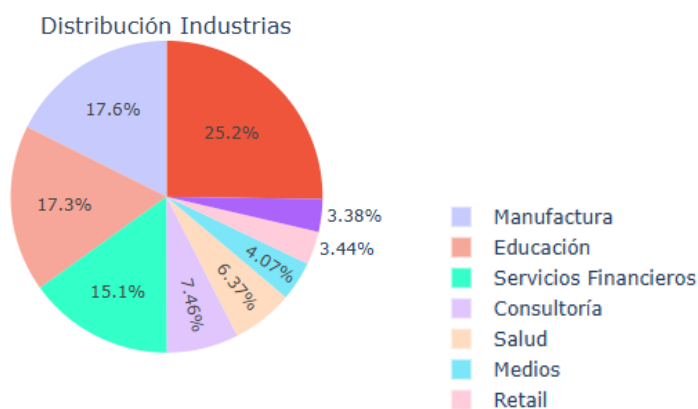
Fuente. Autoría Propia

Clasificación de Sectores Industriales. El mismo pipeline fue replicado para la variable industria, lo que permitió obtener un enriquecimiento semántico significativo sobre la base de datos empresarial. La precisión del modelo se evaluó utilizando un conjunto de datos etiquetado manualmente y validación cruzada, alcanzando métricas satisfactorias con un valor de macro-F1 superior a 0.78. Como resultado, la distribución de industrias clasificadas incluyó 307 registros

en Manufactura, 302 en Educación, 263 en Servicios Financieros, 59 en Tecnología y 440 en la categoría de Otros.

Figura 26

Distribución de industrias



Fuente. Propia tomada de la ejecución del script de limpieza de datos

Las gráficas (23 a 26) de distribución y limpieza (Figuras 23-26) prueban la efectividad del modelo **all-MiniLM-L6-v2**. La transición de texto libre a categorías estructuradas (como "Dirección" o "Tecnología") es lo que permite que el sistema realice un **scoring** inteligente, transformando datos brutos en activos estratégicos para el área comercial.

Script en Python para la Limpieza y Normalización de los Datos

Limpieza de duplicados - Exactos y aproximados usando *FuzzyWuzzy*

Figura 27

Algoritmo limpieza de datos

```

v PASO 4: LIMPIEZA DE DUPLICADOS

[15] print("\n / PASO 4: Eliminando duplicados exactos y aproximados...")
print("="*60)

# EMPLEADOS: Duplicados exactos por full_name + organization_id
print("Limpiando duplicados en Empleados...")

# Crear clave compuesta para empleados
df_empleados['clave_empleado'] = (
    df_empleados['full_name'].fillna('').astype(str) + '_' +
    df_empleados['organization_id'].fillna('').astype(str)
)

# Eliminar duplicados exactos
duplicados_exactos_emp = df_empleados.duplicated(subset=['clave_empleado'], keep='first').sum()
df_empleados = df_empleados.drop_duplicates(subset=['clave_empleado'], keep='first')

print(f"  ✓ Duplicados exactos eliminados en Empleados: {duplicados_exactos_emp}")

# Duplicados fuzzy en empleados
print("  Detectando duplicados aproximados en Empleados...")

# Determinar qué método usar basado en el tamaño del dataset
if len(df_empleados) > 2000:
    print("  Dataset grande detectado, usando algoritmo de blocking...")
    duplicados_fuzzy_emp = cleaner.detect_duplicates_fuzzy_fast(
        df_empleados, ['full_name', 'organization_id'], threshold=90
    )
else:
    print("  Dataset mediano, usando algoritmo optimizado...")
    duplicados_fuzzy_emp = cleaner.detect_duplicates_fuzzy(
        df_empleados, ['full_name', 'organization_id'], threshold=90, sample_size=500
    )

# Eliminar duplicados fuzzy (mantener el primero)
indices_eliminar_emp = set()
for dup in duplicados_fuzzy_emp:
    indices_eliminar_emp.add(dup[1]) # Eliminar el segundo

```

Fuente. Autoría Propia

Normalización Completa - Nombres, Emails, Teléfonos, Ubicaciones

Figura 28

Algoritmo Normalización

```

 9 s
[16] print("\n PASO 5: Normalizando campos de datos...")
      print("="*60)

      # NORMALIZACIÓN DE EMPLEADOS
      print("Normalizando datos de Empleados...")

      # Normalizar nombres
      if 'first_name' in df_empleados.columns:
          df_empleados['first_name'] = df_empleados['first_name'].apply(cleaner.normalize_text)
          cleaner.cleaning_stats['nombres_corregidos'] += df_empleados['first_name'].notna().sum()

      if 'last_name' in df_empleados.columns:
          df_empleados['last_name'] = df_empleados['last_name'].apply(cleaner.normalize_text)

      if 'full_name' in df_empleados.columns:
          df_empleados['full_name'] = df_empleados['full_name'].apply(cleaner.normalize_text)

      print("  ✓ Nombres normalizados")

      # Limpiar emails
      email_columns = ['salesql_email_work', 'salesql_email_direct']
      for col in email_columns:
          if col in df_empleados.columns:
              df_empleados[col] = df_empleados[col].apply(cleaner.clean_email)

      print("  ✓ Emails validados y limpiados")

      # Normalizar teléfonos
      phone_columns = ['salesql_phone_work', 'salesql_phone_personal']
      for col in phone_columns:
          if col in df_empleados.columns:
              df_empleados[col] = df_empleados[col].apply(cleaner.normalize_phone)

      print("  ✓ Teléfonos normalizados")

      # NORMALIZACIÓN DE EMPRESAS
      print("\nNormalizando datos de Empresas...")

```

Fuente. Autoría Propia

Validación Estructural - Detección de Anomalías y Corrección de Formatos

Figura 29

Algoritmo detección de anomalías

```

PASO 6: VALIDACIÓN ESTRUCTURAL Y SEMÁNTICA

print("\n PASO 6: Validación estructural y semántica...")
print("="*60)

# Validar que first_name no contenga títulos profesionales
print("Validando integridad de nombres...")

titles_pattern = r'\b(dr|dra|ing|lic|prof|mg|msc|phd|ceo|cto|cfo)\b'
problematic_names = 0

if 'first_name' in df_empleados.columns:
    for idx, name in df_empleados['first_name'].items():
        if pd.notna(name) and re.search(titles_pattern, str(name).lower()):
            problematic_names += 1
            # Intentar extraer solo el nombre
            clean_name = re.sub(titles_pattern, '', str(name).lower()).strip().title()
            df_empleados.loc[idx, 'first_name'] = clean_name

print(f"  Nombres con títulos corregidos: {problematic_names}")

# Detectar y documentar valores anómalos
print("Detectando valores anómalos...")

anomalies_report = {
    'empleados_sin_email': df_empleados[['salesql_email_work', 'salesql_email_direct']].isnull().all(axis=1).sum(),
    'empleados_sin_telefono': df_empleados[['salesql_phone_work', 'salesql_phone_personal']].isnull().all(axis=1).sum(),
    'empresas_sin_ubicacion': df_empresas['location'].isnull().sum(),
    'empresas_sin_industria': df_empresas['industry'].isnull().sum() if 'industry' in df_empresas.columns else 0,
    'empresas_sin_empleados': df_empresas['employee_count'].isnull().sum()
}

print("  Reporte de anomalías:")
for key, value in anomalies_report.items():
    print(f"    - {key}: {value}")

```

Fuente. Autoría Propia

Evaluación de la Calidad de los Datos Refinados: Completitud, Consistencia, Unicidad y Relevancia.

Evaluación de Calidad de los Datos Refinados. Como parte de los pasos "Evaluación de modelos y aseguramiento de la calidad" de la metodología CRISP-ML(Q), se llevó a cabo una evaluación sistemática sobre la calidad de los datos refinados en cuatro dimensiones clave:

Completitud. Este criterio mide la cantidad de valores presentes en el conjunto de datos en relación con el total esperado. Se observaron los siguientes resultados tras la aplicación del pipeline:

Tabla 16

Compleitud de datos analizados después de la normalización

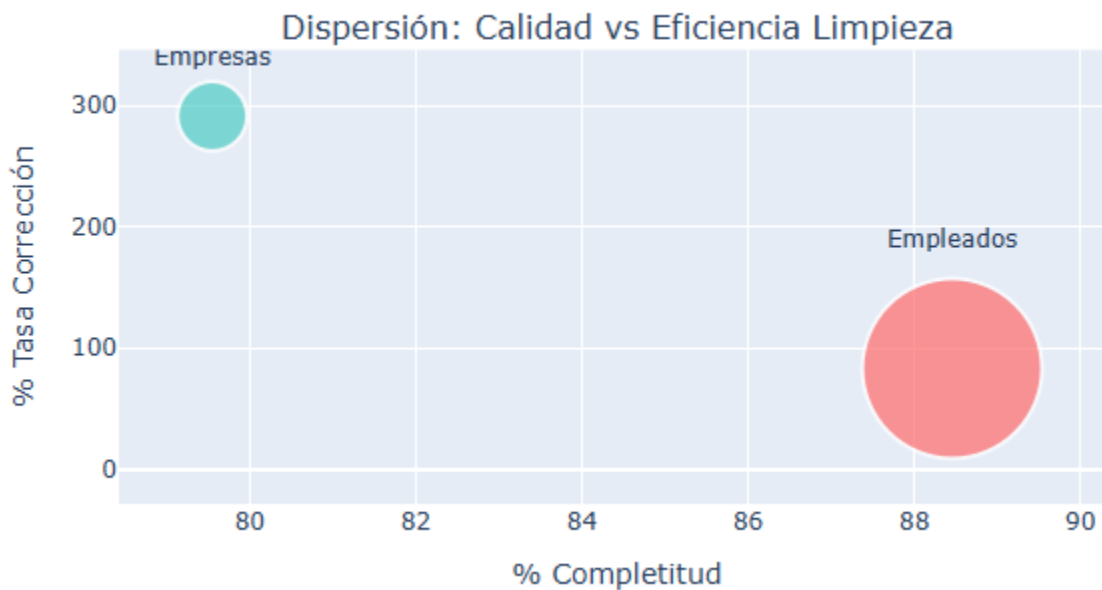
Atributo	Registros Totales	Registros Completos	Porcentaje de Compleitud
Correos electrónicos	4.498	2.388	53,1%
Teléfonos	4.498	1.961	43,6%
Ubicación (empresas)	1.743	1.712	98,2%
Número de empleados (empresas)	1.743	1.740	99,8%

Nota. Tabla que muestra los datos estadísticos de la completitud de los datos procesados. *Fuente.*

Autor

Figura 30

Gráfico de dispersión - calidad vs eficiencia de limpieza



Fuente. Autoría Propia

Este gráfico de burbujas muestra una comparación entre dos entidades ("Empresas" y "Empleados") en términos de % Compleitud (eje X) y % Tasa de Corrección (eje Y), con el tamaño de la burbuja representando una tercera variable (posiblemente volumen de datos o número de registros).

Análisis. Empleados presenta una alta completitud (~88%) pero una tasa de corrección moderada (~130%), con un gran volumen de datos.

Empresas tiene una menor completitud (~79%) pero una alta tasa de corrección (~300%), y un menor volumen de datos.

Conclusión. Aunque los datos de empleados están más completos, requieren menos correcciones. Por el contrario, los datos de empresas, aunque menos completos, muestran una mayor necesidad de correcciones, lo que podría indicar problemas de calidad más significativos en ese conjunto.

Consistencia. Se definió como la coherencia interna de los datos entre diferentes campos. Gracias a la normalización y limpieza, se eliminó gran parte de la ambigüedad y error de formato. Casos tratados:

- Corrección de nombres mal escritos: 4.225 casos
- Corrección de formatos de teléfonos: 2.246
- Unificación de nomenclaturas geográficas y roles

Se usaron validadores estructurales y funciones de normalización semántica para reducir contradicciones internas, especialmente entre campos asociados como ubicación/empresa y cargo/área.

Unicidad. Se refiere a la ausencia de duplicados en los registros clave. Tras aplicar algoritmos de detección exacta y aproximada, se logró una mejora significativa en la unicidad:

Tabla 17

Resumen de eliminación de datos duplicados

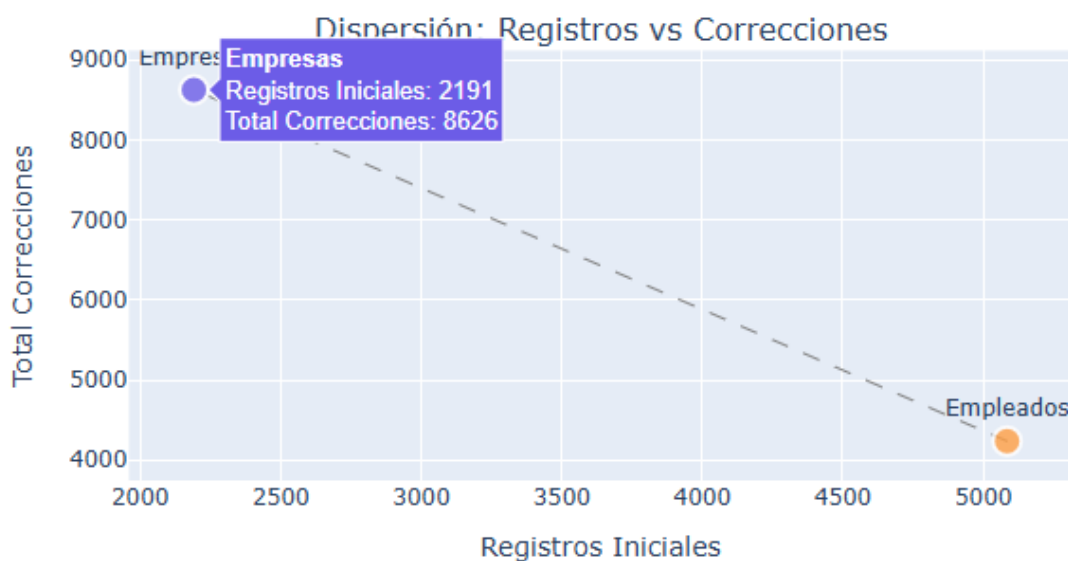
Conjunto	Registros Iniciales	Registros Finales	Duplicados Eliminados
Empleados	5.085	4.498	587
Empresas	2.191	1.743	448

Nota. Tabla que la cantidad de datos que se pudieron eliminar con la aplicación de los algoritmos.

Fuente. Autor

Figura 31

Gráfico dispersión - registros iniciales vs total de correcciones



Fuente. Autoría Propia

Este gráfico de dispersión muestra la relación entre los registros iniciales y el total de correcciones para dos entidades: Empresas y Empleados.

Análisis. Empresas tiene menos registros iniciales (2.191), pero un número muy alto de correcciones (8.626).

Empleados, con más registros iniciales (5.027), requiere menos correcciones (alrededor de 4.300).

Conclusión. Los datos de Empresas presentan una mayor densidad de errores o problemas de calidad, lo que implica que, por cada registro, se necesitan muchas más correcciones. En contraste, los datos de Empleados son más consistentes, con una menor proporción de errores por registro. La línea punteada sugiere una tendencia negativa: a mayor número de registros, menor cantidad proporcional de correcciones, destacando una mejor calidad en ese contexto.

Relevancia. Gracias al uso de algoritmos de clasificación automática, se logró transformar atributos no estructurados en categorías analíticamente útiles. Esto incrementó significativamente la relevancia de los datos para las etapas posteriores de modelado.

Distribución lograda:

Roles laborales clasificados: 2.959 (65,8% del total)

Industrias clasificadas: 1.751 (99,1% del total de empresas)

Se usaron modelos entrenados con *Sentence Transformers + Logistic Regression*, lo que permitió asignar categorías relevantes a partir de texto libre, incrementando el valor semántico del *dataset*.

Evaluación de la Mejora General de la Calidad. En apoyo al análisis anterior, se presenta una visión comparativa de indicadores clave antes y después del proceso de limpieza y enriquecimiento con *machine learning*:

Tabla 18

Evaluación general

Indicador	Valor Inicial	Valor Final
Registros empleados	5.085	4.498
Registros empresas	2.191	1.743
Duplicados eliminados	-	1.035
Correcciones en nombres		4.225
Teléfonos normalizados		2.246
Roles clasificados (ML)		2.959
Industrias clasificadas (ML)		1.751

Nota. Tabla que muestra la evaluación general de la ejecución de limpieza y normalización de los datos. *Fuente.* Autor

Este análisis refuerza que los algoritmos de aprendizaje automático no solo enriquecen los datos, sino que mejoran la calidad en todas las dimensiones evaluadas, respaldando su implementación como componente esencial en procesos de aseguramiento de calidad en ciencia de datos.

La ejecución exitosa de los procesos de limpieza y normalización de los datos demostró su valor intrínseco, no solo al establecer un enfoque iterativo y reproducible, sino, crucialmente, al integrar mecanismos explícitos de evaluación de calidad. Esta estrategia fue fundamental, ya que mediante la combinación de modelos de *machine learning*, técnicas de PLN, y validaciones semánticas/estructurales, se logró reducir significativamente duplicados e inválidos, enriquecer atributos clave y, en última instancia, aumentar la confiabilidad, unicidad y relevancia del *dataset*. Esto elevó sustancialmente la calidad del conjunto de datos, asegurando su óptima preparación para la posterior fase de modelado y análisis predictivo.

Nota: Este proceso no solo limpió, sino que transformó los datos en activos inteligentes listos para ser explotados por sistemas basados en IA.

Migración de Datos Refinados a Supabase

El objetivo de este subproceso es garantizar la disponibilidad, accesibilidad y persistencia de los datos refinados, previamente procesados y validados, en un entorno *cloud* escalable, accesible vía API y compatible con las tecnologías de desarrollo web utilizadas en la aplicación final.

Los datos migrados provienen de los archivos resultantes del *pipeline* de procesamiento implementado en la fase anterior (normalización, clasificación y evaluación de calidad). Estos archivos, en formato .csv y .json, contienen:

Información de empleados con roles clasificados y datos normalizados

Información de empresas con categorías industriales inferidas

Relación entre entidades (empleado ↔ empresa)

Supabase es una plataforma de *backend* como servicio (BaaS) que ofrece una base de datos *PostgreSQL* gestionada, con capacidades de autenticación, almacenamiento y API RESTful integrada. Se eligió por su:

Facilidad de integración con *frameworks* modernos de desarrollo *frontend* (ej. *React*, *Next.js*)

Escalabilidad nativa y soporte para múltiples clientes simultáneos

Modelo de datos relacional, que facilita la trazabilidad entre registros normalizados

Creación de las Tablas

Las tablas necesarias para almacenar los datos de empleados y clientes fueron creadas mediante sentencias SQL ejecutadas en el editor de consultas (Query Editor) de Supabase. Este proceso se basó en el modelo de datos previamente definido, asegurando que la estructura de las tablas se ajustara a los requisitos del proyecto.

Figura 32

Sentencias SQL

```

-- 1.2 Recrear tabla organization con INTEGER como PK
CREATE TABLE organization ( ...
);

-- 1.3 Recrear tabla employee con INTEGER como FK
CREATE TABLE employee (
  employee_id UUID PRIMARY KEY DEFAULT gen_random_uuid(),
  linkedin_url VARCHAR(500),
  first_name VARCHAR(100),
  last_name VARCHAR(100),
  full_name VARCHAR(255),
  profile_picture VARCHAR(500),
  organization_id INTEGER NOT NULL, -- Cambio: INTEGER en lugar de UUID
  employee_position VARCHAR(200),
  employee_headline TEXT,
  email_work VARCHAR(255),
  email_direct VARCHAR(255),
  phone_work VARCHAR(50),
  phone_personal VARCHAR(50),
  employee_location VARCHAR(255),
  categoria_laboral VARCHAR(100),

  -- Metadatos de auditoría
  created_at TIMESTAMP WITH TIME ZONE DEFAULT NOW(),
  updated_at TIMESTAMP WITH TIME ZONE DEFAULT NOW(),

  -- Clave foránea hacia organization
  CONSTRAINT fk_employee_organization
    FOREIGN KEY (organization_id)
    REFERENCES organization(organization_id)
    ON DELETE CASCADE
);

```

Fuente. Autoría Propia

Proceso de Carga de Datos Desde Archivos CSV

Se utilizaron las herramientas integradas de Supabase para importar los datos contenidos en archivos CSV que previamente fueron refinados y validados. Este proceso permitió cargar la información de manera eficiente a la base de datos en la nube, garantizando la integridad y estructura adecuada de los registros.

Figura 33

Gráfico de herramienta para la importación de datos en Supabase

Add data to public.employee

Upload CSV Paste text

Upload a CSV or TSV file. The first row should be the headers of the table, and your headers should not include any special characters other than hyphens (-) or underscores (_).

Tip: Datetime columns should be formatted as YYYY-MM-DD HH:mm:ss

Empleados_limpios.csv
Remove File

Configure import data

Preview data to be imported

A total of 4499 rows will be added to the table "employee"
Here is a preview of the data that will be added (up to the first 20 columns and first 20 rows).

linkedin_url	first_name	last_name	full_n
https://www.linkedin.com/in//	David Stiven	Garcia Rojas	Davi
https://www.linkedin.com/in//	Marcos	Marmolejos	Mar
https://www.linkedin.com/in//	Gabriel	Mendez Pe	Gabi
https://www.linkedin.com/in//	Gabriel Alejandro	Filippi	Gab
https://www.linkedin.com/in//	Jose Luis	Santana Gonzalez	Jose
https://www.linkedin.com/in//	Vladimir	Reyes	Vlad
https://www.linkedin.com/in//	Leidy	Bautista Restituyo	Leid
https://www.linkedin.com/in//	Cristopher Emmanuel	Martinez Francisco	Crist
https://www.linkedin.com/in//	Carlos	Ortiz	Carl

Fuente. Autoría Propia

Validación de Integridad y Consistencia

Una vez completada la migración de datos a Supabase, se llevaron a cabo diversas pruebas para garantizar la integridad y consistencia de la información almacenada.

En primer lugar, se verificó la integridad referencial mediante consultas SQL que comprobaron que todos los registros de empleados tuvieran una relación válida con las empresas correspondientes a través de claves foráneas. Se utilizaron sentencias como *SELECT* combinadas con *JOIN* y filtros *WHERE* para identificar posibles registros huérfanos o inconsistencias en las relaciones entre tablas.

Adicionalmente, se realizaron pruebas para asegurar la correcta codificación de caracteres especiales, particularmente en campos como nombres, direcciones y observaciones.

Esto se validó inspeccionando registros manualmente y mediante consultas que detectaran caracteres atípicos o fallas de codificación (por ejemplo, símbolos extraños o reemplazos por signos de interrogación). La codificación UTF-8 fue confirmada como estándar para toda la base de datos. Como evidencia del proceso, se documentaron:

Validación de Integridad Referencial

Figura 34

Consultas sql utilizadas para las pruebas de integridad

```

153 -----
154 -- VALIDACIÓN DE INTEGRIDAD REFERENCIAL
155 -----
156
157 -- 1. VERIFICAR REGISTROS HUÉRFANOS EN EMPLOYEE
158 -- Empleados que no tienen una organización válida asociada
159 SELECT
160     e.employee_id,
161     e.full_name,
162     e.organization_id,
163     'HUÉRFANO: Sin organización válida' as issue_type
164 FROM employee e
165 LEFT JOIN organization o ON e.organization_id = o.organization_id
166 WHERE o.organization_id IS NULL;
167
168 -- 2. VERIFICAR CONSISTENCIA DE CLAVES FORÁNEAS
169 -- Contar empleados por organización vs organizaciones existentes
170 SELECT
171     o.organization_id,
172     o.name as organization_name,
173     COUNT(e.employee_id) as employee_count_actual,
174     o.employee_count as employee_count_declared,
175     CASE
176         WHEN COUNT(e.employee_id) != COALESCE(o.employee_count, 0)

```

organization_id	organization_name	employee_cour	employee_cour	validation_status
35621948	Aeronautica Civil De Colombia	0	NULL	OK
14886177	Coordinadora Mercantil	0	NULL	OK
1249192	Tres Montes Lucchetti	0	NULL	OK
28985187	2818	6	241	INCONSISTENCIA: Conteos no coinciden
6741377	Aaacesa Almacenes Fiscalizados	0	238	INCONSISTENCIA: Conteos no coinciden
1341311	Aarco Agente De Seguros Y De Fianzas	0	1089	INCONSISTENCIA: Conteos no coinciden
92459527	Abai Colombia	0	238	INCONSISTENCIA: Conteos no coinciden
82380876	Abatech	0	226	INCONSISTENCIA: Conteos no coinciden
36629371	Abbott Laboratories - Lafranco S.A.S	0	817	INCONSISTENCIA: Conteos no coinciden

Fuente. Autoría Propia

Validación de Codificación de Caracteres

Figura 35

Consultas sql utilizadas para las pruebas de codificación de caracteres

```

215 -- 5. DETECTAR CARACTERES ESPECIALES PROBLEMATICOS
216 -- En nombres de organizaciones
217 SELECT
218     organization_id,
219     name,
220     'ORGANIZACIÓN' as table_name,
221     'name' as field_name,
222     CASE
223         WHEN name ~ '[^\x00-\x7F]' THEN 'Contiene caracteres no-ASCII'
224         WHEN name ~ '[?<>]' THEN 'Posible error de codificación'
225         WHEN name ~ '[<>]' THEN 'Contiene caracteres HTML'
226         ELSE 'OK'
227     END as encoding_status
228 FROM organization
229 WHERE name IS NOT NULL
230     AND (name ~ '[?<>]' OR name ~ '[^\x00-\x7F]')
231 ORDER BY encoding_status;
232
233 -- 6. VALIDAR CODIFICACIÓN EN CAMPOS DE EMPLEADOS
234 -- Nombres con posibles problemas de codificación
235 SELECT
236     e.employee_id,
237     e.full_name,
238     e.first_name,
239     e.last_name,
240     o.name as organization_name,
241     CASE
242         WHEN e.full_name ~ '[?<>]' THEN 'Error codificación en full_name'

```

Results Chart Export

organization_id	name	table_name	field_name
661972	Colegio De Contadores Públicos De México A.C.	ORGANIZACIÓN	name
35820332	Esc Sec Tec No 100 Felipe Berriozabal 15Dst0108Bsep Seiem	ORGANIZACIÓN	name
868226	Gómez-Pinzón	ORGANIZACIÓN	name

Fuente. Autoría Propia

Consultas de Limpieza

Figura 36

Consultas sql utilizadas para las pruebas de limpieza

```

348 -- =====
349 -- CONSULTAS DE LIMPIEZA (OPCIONAL)
350 -- =====
351
352 -- 11. IDENTIFICAR REGISTROS PARA LIMPIEZA
353 -- Organizaciones con datos incompletos criticos
354 SELECT
355     organization_id,
356     name,
357     CASE
358         WHEN name IS NULL OR TRIM(name) = '' THEN 'Sin nombre'
359         WHEN industry IS NULL OR TRIM(industry) = '' THEN 'Sin industria'
360         WHEN country IS NULL OR TRIM(country) = '' THEN 'Sin país'
361         ELSE 'Datos mínimos OK'
362     END as data_completeness
363 FROM organization
364 WHERE name IS NULL OR TRIM(name) = ''
365     OR industry IS NULL OR TRIM(industry) = ''
366     OR country IS NULL OR TRIM(country) = ''
367 ORDER BY data_completeness;
368
369 -- 12. EMPLEADOS CON DATOS INCOMPLETOS
370 SELECT
371     e.employee_id,
372     e.full_name,
373     o.name as organization_name,
374     CASE
375         WHEN e.full_name IS NULL OR TRIM(e.full_name) = '' THEN 'Sin nombre completo'
376         WHEN (e.first_name IS NULL OR TRIM(e.first_name) = '')

```

Results Chart Export

employee_id	full_name	organization_name
58982211-640f-4b5c-8f19-7dcc911ef3ce	Santiago Saldarriaga Agudelo	2018
c0fc988c-908d-478f-ba7e-cf0dd3dc5398	Robinson Ramiro Carranza Gacha	Act Telematica
24585746-5710-4ad7-b207-1d284ac52cee	Luis Fernando Ortiz	Act Telematica
409338c2-1441-48d7-a551-c3d99ed50677	Camila García Fajardo	Act Telematica
c854f2e8-f83c-4a30-a9b9-3e3352d995a2	Fabio Zapata Velandia	Act Telematica
2cc618dd-e348-4361-a2f1-72758cc3ffd2	Ale Fdez	Adium México
21c8f6e2-8319-4957-af6a-b09f88a370b1	Paola Jacqueline Ayala Hernández	Adium México

Fuente. Autoría Propia

Análisis de Validaciones de Calidad de Datos

El proceso de validación permitió identificar varios tipos de incidencias que afectan la calidad e integridad de los datos migrados a la base de datos en la nube. A continuación, se detallan los hallazgos más relevantes según el tipo de validación aplicada:

Integridad Referencial. Se detectaron 30 empleados huérfanos, es decir, registros de empleados que no están asociados a ninguna organización existente en la base de datos. Esta situación puede generar errores en procesos dependientes de relaciones entre tablas y evidencia la necesidad de revisar la correspondencia entre empleados y empresas.

Codificación UTF-8. Se encontraron 1.510 organizaciones y 1.207 empleados con caracteres problemáticos, lo cual sugiere posibles errores de codificación al momento de cargar los datos desde los archivos CSV o en su origen. Estos caracteres podrían afectar la visualización, la búsqueda o el filtrado de datos en la aplicación web, por lo que se recomienda aplicar procesos de limpieza y estandarización en estos campos.

Duplicados Potenciales. No se detectaron empleados con nombres duplicados dentro de la misma organización (0 casos), lo que indica que, al menos por este criterio, no hay redundancia evidente en los registros de empleados.

El análisis evidencia que, si bien no existen problemas de duplicación, hay aspectos críticos por corregir, especialmente relacionados con la codificación de caracteres y la integridad entre empleados y organizaciones. Estas observaciones deben ser consideradas prioritarias para asegurar una base de datos limpia, coherente y confiable para su uso operativo y analítico. Gracias a las capacidades nativas de Supabase, los datos quedaron disponibles a través de *endpoints* seguros, accesibles mediante consultas tipo REST (mediante la API de Supabase) o a través de consultas SQL embebidas en la lógica de la aplicación web. Esto facilitó la integración inmediata con el *frontend*, garantizando acceso controlado y eficiente a la información cargada.

Evaluación de Modelos de IA

La selección del modelo de IA más adecuado para procesos de nutrición de *leads* requiere una evaluación sistemática basada en criterios específicos de rendimiento, adaptabilidad y capacidad de integración. Los procesos de nutrición de *leads* modernos involucran múltiples canales de comunicación automatizada, desde correos electrónicos personalizados hasta análisis predictivo de comportamiento del cliente.

Marco de Evaluación Basado en Análisis Artificial (Artificial Analysis)

Según la metodología de Análisis Artificial, los modelos de IA deben evaluarse considerando tres dimensiones fundamentales: inteligencia (precisión), velocidad de respuesta y costo-efectividad. Para aplicaciones de nutrición de *leads*, estos criterios se traducen en precisión en la comprensión del contexto del cliente, capacidad de respuesta en tiempo real, y eficiencia económica para implementaciones a gran escala.

Modelos Candidatos para Evaluación

Modelos de Alto Rendimiento. Esta es la selección de modelos:

GPT-4 O-mini y GPT-4o. Precisión: Excelente capacidad conversacional y adaptabilidad a diversos contextos de texto

Integración: API robusta con amplio soporte para integraciones empresariales

Costo: \$10.00 por millón de tokens de entrada, \$30.00 por millón de tokens de salida para contextos de 128k

Claude 3 y Claude 4. Precisión: Superior rendimiento en razonamiento complejo (89% de precisión) y análisis de documentos

Adaptabilidad: Excelente comprensión semántica para contenido de *marketing* personalizado

Limitaciones: *Pricing premium* con Claude 3.7 Sonnet a \$3/millón de *tokens* de entrada

Gemini 2.5. Rendimiento: Liderando en *benchmarks* de inteligencia junto con modelos o3-pro

Capacidad: Soporte para contextos extensos hasta 1 millón de tokens

Costo: Modelo más caro de Google a \$2.50/millón de tokens de entrada para *prompts* largos

Modelos Emergentes Eficientes. Esta es la selección de modelos:

DeepSeek R1. Innovación: Arquitectura Mixture-of-Experts que activa solo 37 mil millones de 671 mil millones de parámetros por consulta

Costo-efectividad: \$0.55/millón de *tokens* de entrada, \$2.19/millón de *tokens* de salida

Limitaciones: No soporta llamadas de función ni análisis de imágenes, restringiendo capacidades de automatización

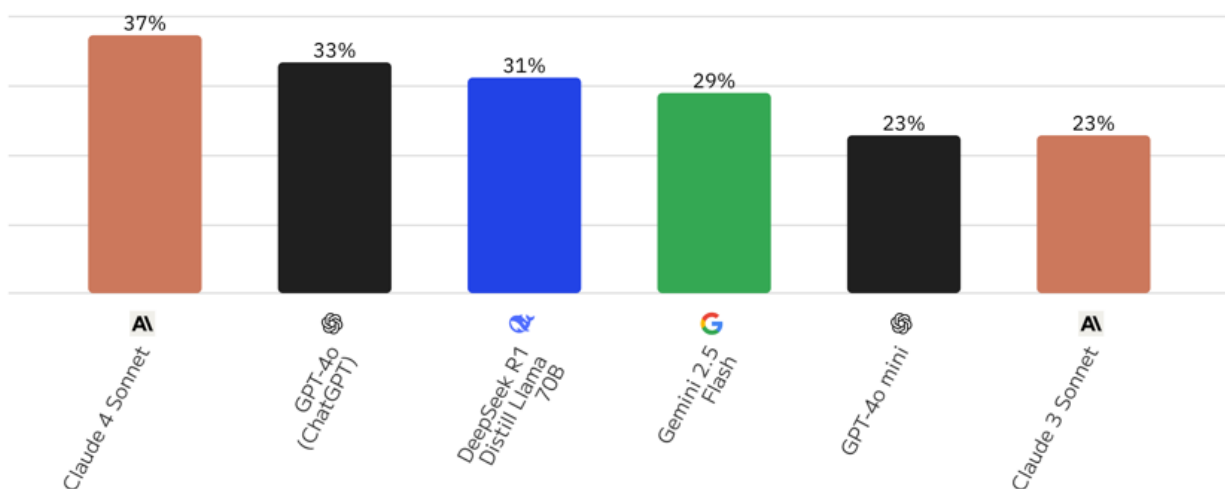
Criterios de Evaluación Específicos para Nutrición de *Leads*

Precisión en Comunicación Personalizada

Los modelos deben demostrar capacidad para generar contenido personalizado basado en datos de comportamiento del cliente. Claude 4 y GPT-4o destacan en esta área, con Claude 4 mostrando superior comprensión semántica para relacionar conceptos naturalmente sin forzar palabras clave.

Figura 37

Gráfico comprensión semántica



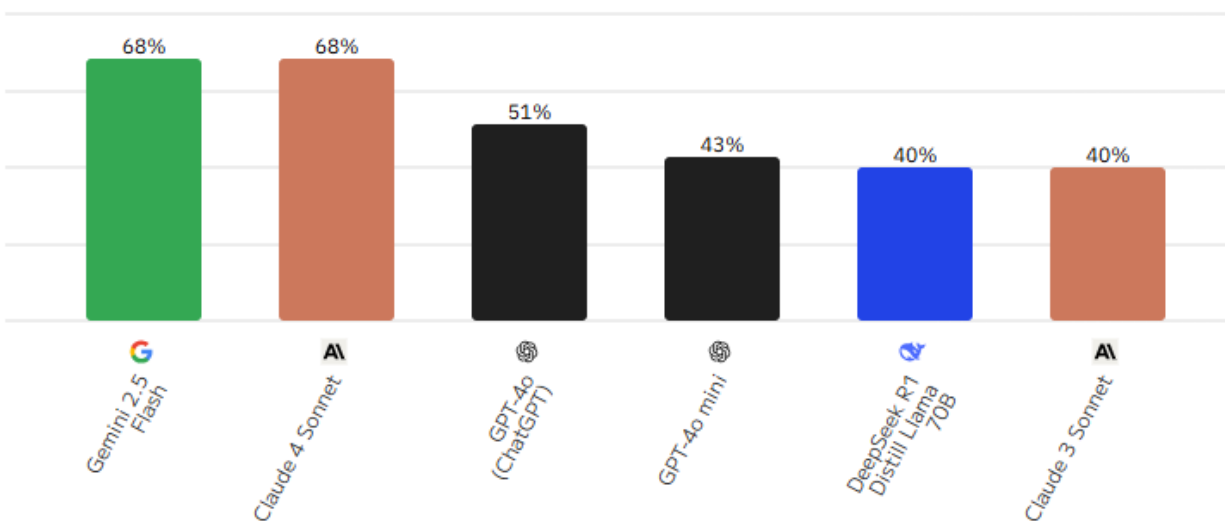
Fuente. Tomada de artificialanalysis.ai

Capacidad de Análisis Predictivo

Para la puntuación predictiva de *leads*, los modelos deben procesar grandes volúmenes de datos históricos. Gemini 2.5 Flash sobresale en análisis científico de datos, mientras que GPT-4 mantiene mayor flexibilidad para diversas tareas de análisis.

Figura 38

Gráfico de Capacidad de Análisis Predictivo



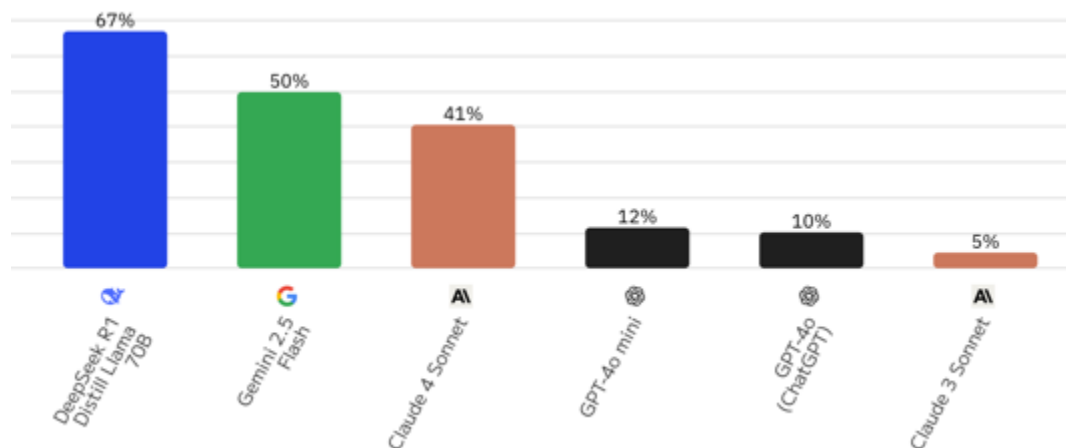
Fuente. Tomada de artificialanalysis.ai

Velocidad de Respuesta para Automatización

La latencia es crítica para aplicaciones en tiempo real. Según *benchmarks* de 2025, Gemini 2.5 Flash alcanza 369 tokens/segundo, mientras que modelos como DeepSeek R1 Distill logran 387 *tokens/segundo*.

Figura 39

Gráfico de velocidad de respuesta para automatización



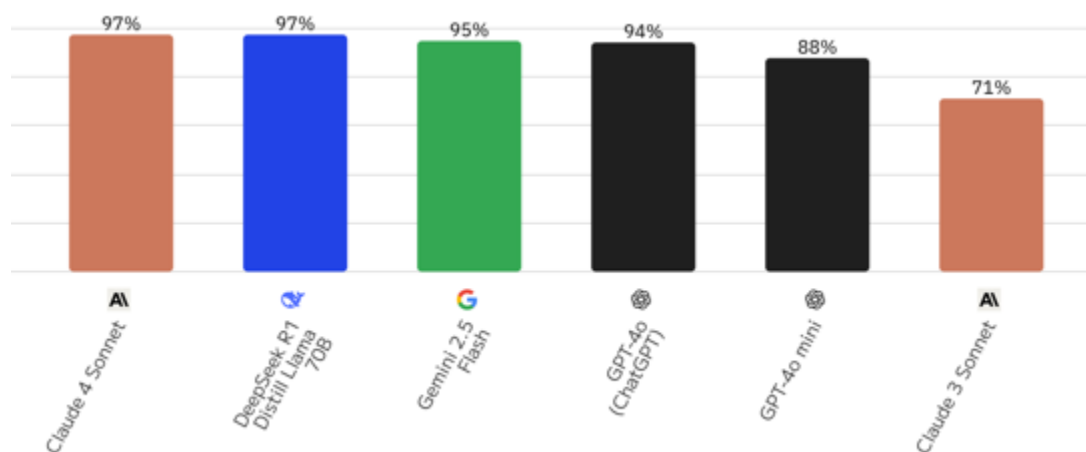
Fuente. Tomada de artificialanalysis.ai

Capacidad de Integración Empresarial

La integración con sistemas CRM existentes requiere APIs robustas y soporte para automatización. GPT-4o-mini ofrece la integración intermedia, mientras que Claude 3 carece de soporte para llamadas de función, limitando su utilidad en *workflows* automatizados.

Figura 40

Gráfico de la integración con sistemas CRM existentes



Fuente. Tomada de artificialanalysis.ai

Matriz de Evaluación Comparativa

Tabla 19

Evaluación LLM's

Modelo	Precisión	Velocidad (t/s)	Costo (\$/1M tokens)	Integración	Puntuación Total
Claude 4	9/10	7/10	6/10	8/10	30/40
GPT-4o-mini	8/10	8/10	7/10	10/10	33/40
Gemini 2.5 Flash	9/10	9/10	5/10	7/10	30/40
DeepSeek R1	7/10	10/10	10/10	4/10	31/40

Nota. Tabla que muestra el ponderado de la puntuación obtenida por cada modelo evaluado.

Fuente. Autor

La matriz comparativa justifica la elección de GPT-4o-mini para el prototipo. Con una puntuación de 33/40, este modelo ofrece el mejor equilibrio entre costo-efectividad y capacidad de integración vía API. Esta decisión técnica asegura que el proyecto sea económicamente viable en su fase de validación sin sacrificar la precisión en la comunicación personalizada.

Dado que este trabajo se orienta al desarrollo de un prototipo funcional, la selección de modelos se enfocó en alternativas de bajo costo operativo, priorizando aquellas que permitieran realizar pruebas, iteraciones y validaciones rápidas sin comprometer la viabilidad económica del proyecto. Bajo este criterio, modelos como GPT-4o-mini, Gemini 2.5 Flash, Claude 4 y DeepSeek R1 resultan adecuados para un entorno experimental, pues ofrecen un equilibrio razonable entre precisión, velocidad y facilidad de integración en etapas tempranas. Dentro de ellos, GPT-4o-mini destaca como la opción más equilibrada para un sistema de nutrición de *leads* al combinar precisión, adaptabilidad y costo-efectividad.

No obstante, esta selección responde únicamente a las necesidades propias de un contexto de prototipado. En un escenario de implementación productiva, será necesario realizar una evaluación más rigurosa e incluir modelos de razonamiento más potentes y especializados, capaces de ofrecer mayor precisión, capacidades avanzadas de análisis y un rendimiento estable

bajo cargas empresariales. De este modo, aunque el prototipo se apoya en modelos de bajo costo, se deja abierta la posibilidad de escalar hacia tecnologías superiores en una versión final del sistema.

Desarrollo Iterativo del Sistema

Esta fase aborda la construcción progresiva del prototipo mediante un enfoque iterativo e incremental, tomando como base el marco de trabajo Scrum. A lo largo de cada *sprint* se generan incrementos funcionales del sistema, los cuales son evaluados y validados con usuarios o partes interesadas. Este proceso permite ajustar continuamente los componentes, registrar la retroalimentación y asegurar que el prototipo evolucione de manera controlada, transparente y alineada con los objetivos del proyecto.

Definición de Historias de Usuario (Formato 3C)

Las Historias de Usuario son una técnica fundamental dentro de metodologías ágiles debido a su capacidad para capturar necesidades desde la perspectiva del usuario y orientar el desarrollo hacia la entrega de valor. Según (Cohn, 2004), una buena historia debe ser breve, comprensible y centrada en el rol, la necesidad y el beneficio del usuario. Este enfoque permite mantener una comunicación clara entre equipos técnicos y no técnicos, evitando documentación extensa y favoreciendo la colaboración continua.

Para asegurar la calidad de cada historia se utiliza el formato 3C propuesto por (Guerino et al., 2022), el cual establece que una Historia de Usuario debe componerse de tres elementos clave:

El enfoque *Card–Conversation–Confirmation* asegura que cada historia de usuario sea clara y orientada a valor. La *Card* describe brevemente la funcionalidad, la *Conversation* permite aclarar detalles entre los *stakeholders* y la *Confirmation* define los criterios de aceptación que validan su cumplimiento. Este formato garantiza historias negociables, verificables y fáciles de estimar y priorizar dentro del marco ágil.

A continuación se definen las historias de usuario más relevantes identificadas:

Módulo: Login de usuario

Figura 41

Historia de usuario HU-001

Historia de Usuario	ID: HU-001	Título: Autenticación mediante Gmail o contraseña
Descripción	Como usuario nuevo, quiero iniciar sesión con mi cuenta de Gmail o contraseña, para poder acceder rápidamente a la plataforma de forma segura.	
Criterios de aceptación:		
<input checked="" type="checkbox"/> El sistema ofrece autenticación mediante Gmail (OAuth) <input checked="" type="checkbox"/> El sistema permite autenticación con correo electrónico y contraseña <input checked="" type="checkbox"/> El usuario recibe mensajes claros en caso de error <input checked="" type="checkbox"/> La autenticación es segura y redirige correctamente al usuario		
Valor del negocio: Permite un acceso rápido y sencillo, mejorando la experiencia de usuario desde el primer contacto.		
I N V E S T <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	Estimación	<input checked="" type="checkbox"/> X
	¿Cumple con la definición de listo del equipo?	<input checked="" type="checkbox"/>
		<input checked="" type="checkbox"/>

Fuente. Autoría Propia

Figura 41

Historia de usuario HU-002

Historia de Usuario	ID: HU-002	Título: Sesión persistente del usuario
Descripción	Como usuario registrado, quiero mantener mi sesión iniciada, para evitar ingresar credenciales cada vez que accedo a la plataforma.	
Criterios de aceptación:		
<input checked="" type="checkbox"/> El sistema mantiene sesión iniciada durante al menos 7 días <input checked="" type="checkbox"/> La sesión persiste sin errores de seguridad o acceso <input checked="" type="checkbox"/> Se utilizan cookies o tokens JWT de forma segura <input checked="" type="checkbox"/> Existe una opción para cerrar sesión manualmente		
Valor del negocio: Reduce fricción para el usuario frecuente, mejora la usabilidad y fideliza al usuario.		
I N V E S T <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	Estimación	<input checked="" type="checkbox"/> X
	¿Cumple con la definición de listo del equipo?	<input checked="" type="checkbox"/>
		<input checked="" type="checkbox"/>

Fuente. Autoría Propia

Módulo: Plataforma Web

Figura 43

Historia de usuario HU-003

Historia de Usuario	ID: HU-003	Título: Acceso a plataforma centralizada
Descripción	Como usuario, quiero acceder a una plataforma centralizada donde se integren todas las herramientas disponibles, para poder gestionar mis actividades desde un solo lugar.	
Criterios de aceptación:		
<input checked="" type="checkbox"/> Plataforma disponible desde navegador <input checked="" type="checkbox"/> Diseño responsivo y adaptativo <input checked="" type="checkbox"/> Módulos accesibles desde un menú visible <input checked="" type="checkbox"/> Navegación entre módulos en máximo 2 clics		
Valor del negocio:		
Facilita el acceso integral a funcionalidades y mejora la productividad del usuario.		
I <u>N</u> <u>V</u> <u>E</u> <u>S</u> <u>T</u>	Estimación	<input checked="" type="checkbox"/> X
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	¿Cumple con la definición de listo del equipo?	<input checked="" type="checkbox"/>
		<input checked="" type="checkbox"/>

Fuente. Autoría Propia

Figura 44

Historia de usuario HU-004

Historia de Usuario	ID: HU-004	Título: Guía interactiva de bienvenida
Descripción	Como usuario nuevo, quiero ver una guía interactiva al ingresar por primera vez, para poder entender rápidamente cómo usar la plataforma.	
Criterios de aceptación:		
<input checked="" type="checkbox"/> Se activa automáticamente al primer ingreso <input checked="" type="checkbox"/> El tour incluye pasos interactivos y resaltado de funciones <input checked="" type="checkbox"/> El usuario puede omitir o repetir la guía <input checked="" type="checkbox"/> Se almacena la visualización para evitar repetirla automáticamente		
Valor del negocio:		
Facilita el acceso integral a funcionalidades y mejora la productividad del usuario.		
I <u>N</u> <u>V</u> <u>E</u> <u>S</u> <u>T</u>	Estimación	<input checked="" type="checkbox"/> X
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	¿Cumple con la definición de listo del equipo?	<input checked="" type="checkbox"/>
		<input checked="" type="checkbox"/>

Fuente. Autoría Propia

Figura 45

Historia de usuario HU-005

Historia de Usuario	ID: HU-005	Título: Visualización de resumen e información relevante en el inicio
Descripción	Como usuario, quiero ver al ingresar un resumen de mi actividad y herramientas más utilizadas, para poder entender rápidamente el estado actual de mi gestión sin navegar por toda la plataforma.	
Criterios de aceptación:		
<input checked="" type="checkbox"/> Se muestra un <u>dashboard</u> al ingresar. <input checked="" type="checkbox"/> El <u>dashboard</u> incluye bloques de herramientas, resumen de leads y resumen de correos. <input checked="" type="checkbox"/> La información está actualizada. <input checked="" type="checkbox"/> Se presentan al menos tres bloques informativos.		
Valor del negocio: Mejora de la experiencia de usuario al reducir el tiempo para acceder a información clave.		
I <u>N</u> <u>V</u> <u>E</u> <u>S</u> <u>T</u> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	Estimación	<input checked="" type="checkbox"/> X
	¿Cumple con la definición de listo del equipo?	<input checked="" type="checkbox"/>
		<input checked="" type="checkbox"/>

Fuente. Autoría Propia

Figura 46

Historia de usuario HU-006

Historia de Usuario	ID: HU-006	Título: Personalización de widgets en el <u>dashboard</u>
Descripción	Como usuario, quiero personalizar los widgets que se muestran en mi inicio, para poder adaptar la plataforma a mis necesidades y hábitos de trabajo.	
Criterios de aceptación:		
<input checked="" type="checkbox"/> Se puede arrastrar, ocultar o cambiar la posición de los widgets. <input checked="" type="checkbox"/> Las personalizaciones se guardan automáticamente. <input checked="" type="checkbox"/> Los cambios se reflejan en futuras sesiones del usuario. <input checked="" type="checkbox"/> Existe opción para restaurar configuración predeterminada.		
Valor del negocio: Mejora de la experiencia del usuario mediante una interfaz adaptable y flexible.		
I <u>N</u> <u>V</u> <u>E</u> <u>S</u> <u>T</u> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	Estimación	<input checked="" type="checkbox"/> X
	¿Cumple con la definición de listo del equipo?	<input checked="" type="checkbox"/>
		<input checked="" type="checkbox"/>

Fuente. Autoría Propia

Módulo: Búsqueda y agrupación de *Leads*

Figura 47

Historia de usuario HU-007

Historia de Usuario	ID: HU-007	Título: Búsqueda de leads por filtros avanzados
Descripción	Como usuario, quiero buscar leads por filtros como sector, ubicación o tamaño, para poder segmentar mejor a los clientes potenciales y optimizar mis campañas.	
Criterios de aceptación:		
<input checked="" type="checkbox"/> Los filtros son visibles y accesibles. <input checked="" type="checkbox"/> Se pueden combinar múltiples filtros. <input checked="" type="checkbox"/> Los resultados se actualizan en tiempo real. <input checked="" type="checkbox"/> Existe opción para limpiar filtros aplicados.		
Valor del negocio:		
Aumenta la eficiencia en la segmentación de leads y mejora la efectividad de campañas.		
I <u>N</u> <u>V</u> <u>E</u> <u>S</u> <u>T</u>	Estimación	<input checked="" type="checkbox"/> X
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	¿Cumple con la definición de listo del equipo?	<input checked="" type="checkbox"/>
		<input checked="" type="checkbox"/>

Fuente. Autoría Propia

Figura 48

Historia de usuario HU-008

Historia de Usuario	ID: HU-008	Título: Guardar grupos de leads
Descripción	Como usuario, quiero guardar grupos de leads para campañas futuras, para poder reutilizar segmentaciones sin rehacer los filtros cada vez.	
Criterios de aceptación:		
<input checked="" type="checkbox"/> Se pueden seleccionar múltiples leads. <input checked="" type="checkbox"/> Existe un botón para "Guardar grupo". <input checked="" type="checkbox"/> El grupo puede ser nombrado y editado. <input checked="" type="checkbox"/> El grupo se almacena en la cuenta del usuario.		
Valor del negocio:		
Permite una mejor planificación y reutilización de campañas y leads segmentados.		
I <u>N</u> <u>V</u> <u>E</u> <u>S</u> <u>T</u>	Estimación	<input checked="" type="checkbox"/> X
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	¿Cumple con la definición de listo del equipo?	<input checked="" type="checkbox"/>
		<input checked="" type="checkbox"/>

Fuente. Autoría Propia

Figura 49

Historia de usuario HU-009

Historia de Usuario	ID: HU-009	Título: Importación de leads desde archivo CSV
Descripción	Como usuario, quiero importar leads desde un archivo CSV, para poder cargar grandes cantidades de información sin hacerlo manualmente.	
Criterios de aceptación:		
<input checked="" type="checkbox"/> Existe opción clara para subir archivo CSV. <input checked="" type="checkbox"/> El sistema valida el formato y columnas requeridas. <input checked="" type="checkbox"/> Se informa cuántos leads fueron cargados con éxito. <input checked="" type="checkbox"/> Se muestra <u>feedback</u> claro sobre errores encontrados.		
Valor del negocio:		
Reduce el tiempo de carga de información y mejora la eficiencia operativa.		
I <u>N</u> <u>V</u> <u>E</u> <u>S</u> <u>T</u>	Estimación	<input checked="" type="checkbox"/> X

Fuente. Autoría Propia

Módulo: Agente de correos

Figura 50

Historia de usuario HU-010

Historia de Usuario	ID: HU-010	Título: Redacción de correos personalizados con campos dinámicos
Descripción	Como usuario, quiero escribir correos personalizados usando campos dinámicos como {nombre}, para aumentar la efectividad del mensaje al hacerlo más relevante para cada lead.	
Criterios de aceptación:		
<input checked="" type="checkbox"/> Se pueden insertar variables dinámicas en el cuerpo del correo. <input checked="" type="checkbox"/> Las variables se sustituyen correctamente al enviar el correo. <input checked="" type="checkbox"/> Hay una vista previa con los datos personalizados antes de enviar. <input checked="" type="checkbox"/> Soporta campos comunes como nombre, empresa, cargo, entre otros.		
Valor del negocio:		
Mejora la personalización de las campañas y eleva la tasa de conversión.		
I <u>N</u> <u>V</u> <u>E</u> <u>S</u> <u>T</u>	Estimación	<input checked="" type="checkbox"/> X
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	¿Cumple con la definición de listo del equipo?	<input checked="" type="checkbox"/>
		<input checked="" type="checkbox"/>

Fuente. Autoría Propia

Figura 51

Historia de usuario HU-011

Historia de Usuario	ID: HU-011	Título: Programación de envíos de correos
Descripción	Como usuario, quiero programar el envío de correos en fechas y horas específicas, para coordinar campañas según calendarios estratégicos.	
Criterios de aceptación:		
<input checked="" type="checkbox"/> Se puede seleccionar fecha y hora de envío. <input checked="" type="checkbox"/> Se puede aplicar la programación a lotes de leads o campañas completas. <input checked="" type="checkbox"/> Se muestra una lista de envíos programados. <input checked="" type="checkbox"/> Correos se envían automáticamente en el momento definido.		
Valor del negocio:		
Permite planificar campañas anticipadamente y ejecutar acciones en el momento óptimo.		
I <u>N</u> <u>V</u> <u>E</u> <u>S</u> <u>T</u>	Estimación	<input checked="" type="checkbox"/> X
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	¿Cumple con la definición de listo del equipo?	<input checked="" type="checkbox"/>
		<input checked="" type="checkbox"/>

Fuente. Autoría Propia

Figura 52

Historia de usuario HU-012

Historia de Usuario	ID: HU-012	Título: Visualización de métricas de campaña (apertura y clics)
Descripción	Como usuario, quiero visualizar las tasas de apertura y clics de mis campañas, para medir la efectividad de mis comunicaciones.	
Criterios de aceptación:		
<input checked="" type="checkbox"/> El <u>dashboard</u> muestra porcentaje de apertura y clics. <input checked="" type="checkbox"/> Los datos se agrupan por campaña y por lead. <input checked="" type="checkbox"/> Se actualizan en tiempo real. <input checked="" type="checkbox"/> Se pueden filtrar por fechas y segmentos.		
Valor del negocio:		
Facilita la toma de decisiones basadas en datos y permite optimizar campañas.		
I N V E S T	Estimación	<input checked="" type="checkbox"/> X
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	¿Cumple con la definición de listo del equipo?	<input checked="" type="checkbox"/>
		<input checked="" type="checkbox"/>

Fuente. Autoría Propia

Módulo: Agente conversacional

Figura 53

Historia de usuario HU-013

Historia de Usuario	ID: HU-013	Título: Configuración de respuestas automatizadas por canal
Descripción	Como usuario, quiero configurar respuestas automatizadas por canal (web, WhatsApp, Telegram), para que el <u>chatbot</u> pueda interactuar sin intervención manual.	
Criterios de aceptación:		
<input checked="" type="checkbox"/> Existe un constructor visual o lógico de flujos de conversación. <input checked="" type="checkbox"/> Se pueden definir condiciones y respuestas por canal. <input checked="" type="checkbox"/> El <u>chatbot</u> responde automáticamente según el flujo establecido. <input checked="" type="checkbox"/> Las respuestas se adaptan al canal utilizado.		
Valor del negocio:		
Reduce carga operativa y mejora la atención inmediata al cliente.		
I N V E S T	Estimación	<input checked="" type="checkbox"/> X
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	¿Cumple con la definición de listo del equipo?	<input checked="" type="checkbox"/>
		<input checked="" type="checkbox"/>

Fuente. Autoría Propia

Figura 54

Historia de usuario HU-014

Historia de Usuario	ID: HU-014	Título: Integración del <u>bot</u> con WhatsApp y Telegram
Descripción	Como usuario, quiero conectar el <u>chatbot</u> a WhatsApp y Telegram, para llegar a los leads en sus canales de comunicación preferidos.	
Criterios de aceptación:		
<input checked="" type="checkbox"/> Existe una sección para configurar integración con <u>APIs</u> de WhatsApp y Telegram. <input checked="" type="checkbox"/> Se validan las credenciales y conexión. <input checked="" type="checkbox"/> El <u>bot</u> puede enviar y recibir mensajes en ambos canales. <input checked="" type="checkbox"/> Se notifican errores de integración.		
Valor del negocio:		
Aumenta el alcance del sistema y mejora la experiencia del usuario final.		
I <u>N</u> <u>V</u> <u>E</u> <u>S</u> <u>T</u>	Estimación	<input checked="" type="checkbox"/> X
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	¿Cumple con la definición de listo del equipo?	<input checked="" type="checkbox"/>
		<input checked="" type="checkbox"/>

Fuente. Autoría Propia

Figura 55

Historia de usuario HU-015

Historia de Usuario	ID: HU-015	Título: Acceso al historial de conversaciones por canal y lead
Descripción	Como usuario, quiero ver el historial de conversaciones con cada lead, para tener contexto de interacciones pasadas y personalizar mejor la atención.	
Criterios de aceptación:		
<input checked="" type="checkbox"/> El historial se guarda por lead y canal. <input checked="" type="checkbox"/> Puede filtrarse por fechas y canales. <input checked="" type="checkbox"/> Permite búsqueda por palabra clave. <input checked="" type="checkbox"/> Se muestra en formato cronológico.		
Valor del negocio:		
Facilita el seguimiento de conversaciones y mejora la calidad de la atención al cliente.		
I <u>N</u> <u>V</u> <u>E</u> <u>S</u> <u>T</u>	Estimación	<input checked="" type="checkbox"/> X
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	¿Cumple con la definición de listo del equipo?	<input checked="" type="checkbox"/>
		<input checked="" type="checkbox"/>

Fuente. Autoría Propia

Módulo: Agente de Llamadas

Figura 56

Historia de usuario HU-016

Historia de Usuario	ID: HU-016	Título: Generación automática de llamadas a leads
Descripción	Como usuario, quiero generar llamadas automáticas a leads, para optimizar el alcance y seguimiento con mensajes predefinidos.	
Criterios de aceptación: <input checked="" type="checkbox"/> Existe opción para subir un guion de llamada. <input checked="" type="checkbox"/> El sistema convierte texto a audio (TTS). <input checked="" type="checkbox"/> Las llamadas se generan y reproducen correctamente.		
Valor del negocio: Automatiza la prospección y mejora la cobertura comercial.		
I <u>N</u> <u>V</u> <u>E</u> <u>S</u> <u>T</u>	Estimación	<input checked="" type="checkbox"/> X
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	¿Cumple con la definición de listo del equipo?	<input checked="" type="checkbox"/>
		<input checked="" type="checkbox"/>

Fuente. Autoría Propia

Figura 57

Historia de usuario HU-017

Historia de Usuario	ID: HU-017	Título: Asignación de llamadas a comerciales
Descripción	Como usuario, quiero asignar llamadas a agentes comerciales, para asegurar que cada lead sea atendido de forma eficiente.	
Criterios de aceptación: <input checked="" type="checkbox"/> Existe selector de agente y horario. <input checked="" type="checkbox"/> La agenda del agente se actualiza automáticamente. <input checked="" type="checkbox"/> Se notifica al agente sobre la asignación.		
Valor del negocio: Mejora la coordinación y seguimiento de oportunidades.		
I <u>N</u> <u>V</u> <u>E</u> <u>S</u> <u>T</u>	Estimación	<input checked="" type="checkbox"/> X
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	¿Cumple con la definición de listo del equipo?	<input checked="" type="checkbox"/>
		<input checked="" type="checkbox"/>

Fuente. Autoría Propia

Figura 58

Historia de usuario HU-018

Historia de Usuario	ID: HU-018	Título: Visualización del resultado de llamadas
Descripción	Como usuario, quiero ver el resultado de cada llamada, para hacer seguimiento efectivo al estado de cada contacto.	
Criterios de aceptación: <input checked="" type="checkbox"/> Se registra estado (contestó, no contestó, etc.). <input checked="" type="checkbox"/> Se guarda duración y comentarios del agente. <input checked="" type="checkbox"/> El reporte está disponible al día siguiente.		
Valor del negocio: Facilita análisis y priorización de leads en seguimiento.		
I <u>N</u> <u>V</u> <u>E</u> <u>S</u> <u>T</u>	Estimación	<input checked="" type="checkbox"/> X
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	¿Cumple con la definición de listo del equipo?	<input checked="" type="checkbox"/>
		<input checked="" type="checkbox"/>

Fuente. Autoría Propia

Módulo: Campañas

Figura 59*Historia de usuario HU-019*

Historia de Usuario	ID: HU-019	Título: Lanzamiento de campañas multicanal
Descripción	Como usuario, quiero lanzar campañas multicanal desde una única interfaz, para gestionar mis comunicaciones de forma centralizada.	
Criterios de aceptación: <input checked="" type="checkbox"/> Interfaz permite elegir canales, contenido y leads. <input checked="" type="checkbox"/> Se define fecha/hora de ejecución. <input checked="" type="checkbox"/> La campaña se ejecuta correctamente en los canales seleccionados.		
Valor del negocio: Optimiza el tiempo y coherencia de campañas de marketing.		
I <u>N</u> <u>V</u> <u>E</u> <u>S</u> <u>T</u>	Estimación	<input checked="" type="checkbox"/> X
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	¿Cumple con la definición de listo del equipo?	<input checked="" type="checkbox"/>
		<input checked="" type="checkbox"/>

Fuente. Autoría Propia

Figura 60*Historia de usuario HU-020*

Historia de Usuario	ID: HU-020	Título: Pausar campañas en ejecución
Descripción	Como usuario, quiero pausar campañas activas, para detener temporalmente sin perder los datos ni la configuración.	
Criterios de aceptación: <input checked="" type="checkbox"/> Opción visible y accesible para pausar. <input checked="" type="checkbox"/> Estado de campaña cambia a "Pausado". <input checked="" type="checkbox"/> Puede reanudarse sin pérdida de configuración.		
Valor del negocio: Ofrece control operativo sobre acciones en marcha.		
I <u>N</u> <u>V</u> <u>E</u> <u>S</u> <u>T</u>	Estimación	<input checked="" type="checkbox"/> X
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	¿Cumple con la definición de listo del equipo?	<input checked="" type="checkbox"/>
		<input checked="" type="checkbox"/>

Fuente. Autoría Propia

Figura 61*Historia de usuario HU-021*

Historia de Usuario	ID: HU-021	Título: Duplicar campañas existentes
Descripción	Como usuario, quiero duplicar campañas anteriores, para reutilizar configuraciones exitosas y acelerar la creación de nuevas campañas.	
Criterios de aceptación: <input checked="" type="checkbox"/> Existe botón "Duplicar" visible. <input checked="" type="checkbox"/> Se copia configuración y contenido. <input checked="" type="checkbox"/> La nueva campaña es editable antes de enviarla.		
Valor del negocio: Incrementa la productividad al crear campañas.		
I <u>N</u> <u>V</u> <u>E</u> <u>S</u> <u>T</u>	Estimación	<input checked="" type="checkbox"/> X
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	¿Cumple con la definición de listo del equipo?	<input checked="" type="checkbox"/>
		<input checked="" type="checkbox"/>

Fuente. Autoría Propia

Módulo: Integraciones

Figura 62

Historia de usuario HU-022

Historia de Usuario	ID: HU-022	Título: Conexión con CRM externo	
Descripción	Como usuario, quiero conectar la plataforma a mi CRM, para poder sincronizar automáticamente los datos entre ambos sistemas.		
Criterios de aceptación:			
<input checked="" type="checkbox"/> Se permite ingresar la API Key del CRM. <input checked="" type="checkbox"/> El usuario puede mapear campos personalizados. <input checked="" type="checkbox"/> Se valida la conexión exitosamente. <input checked="" type="checkbox"/> Los datos se sincronizan en ambos sentidos (plataforma ↔ CRM).			
Valor del negocio: Permite integrar herramientas externas, mejorando la eficiencia y reduciendo errores manuales.			
I <u>N</u> <u>V</u> <u>E</u> <u>S</u> T	Estimación	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	¿Cumple con la definición de listo del equipo?	<input checked="" type="checkbox"/>	
		<input checked="" type="checkbox"/>	

Fuente. Autoría Propia

Figura 62

Historia de usuario HU-023

Historia de Usuario	ID: HU-023	Título: Importación desde Google <u>Sheets</u>	
Descripción	Como usuario, quiero importar datos desde Google <u>Sheets</u> , para poder cargar leads fácilmente sin procesos manuales.		
Criterios de aceptación:			
<input checked="" type="checkbox"/> Autenticación segura con cuenta de Google. <input checked="" type="checkbox"/> Selección de hoja de cálculo desde la interfaz. <input checked="" type="checkbox"/> Los datos se visualizan correctamente como leads. <input checked="" type="checkbox"/> Validación de formato y estructura antes de importar.			
Valor del negocio: Facilita la carga masiva de datos, reduciendo tiempos de preparación e ingreso manual.			
I <u>N</u> <u>V</u> <u>E</u> <u>S</u> T	Estimación	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	¿Cumple con la definición de listo del equipo?	<input checked="" type="checkbox"/>	
		<input checked="" type="checkbox"/>	

Fuente. Autoría Propia

Figura 64

Historia de usuario HU-024

Historia de Usuario	ID: HU-024	Título: Automatización con N8N	
Descripción	Como usuario, quiero automatizar flujos con n8n, para poder integrar la plataforma con múltiples herramientas sin desarrollo adicional.		
Criterios de aceptación:			
<input checked="" type="checkbox"/> Configuración simple de webhooks . <input checked="" type="checkbox"/> Registro de eventos disparadores correctamente configurados. <input checked="" type="checkbox"/> Activación verificada mediante logs. <input checked="" type="checkbox"/> Logs accesibles para revisión del usuario.			
Valor del negocio:			
Potencia la automatización y reduce dependencias técnicas para integraciones repetitivas.			
I <u>N</u> V <u>E</u> S T	Estimación		<input checked="" type="checkbox"/> X
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	¿Cumple con la definición de listo del equipo?		<input checked="" type="checkbox"/>
			<input checked="" type="checkbox"/>

Fuente. Autoría Propia

Usabilidad y Experiencia

Figura 65

Historia de usuario HU-025

Historia de Usuario	ID: HU-025	Título: Íconos y mensajes claros en botones	
Descripción	Como usuario nuevo, quiero ver íconos y mensajes claros en cada botón, para poder navegar fácilmente sin necesidad de asistencia.		
Criterios de aceptación:			
<input checked="" type="checkbox"/> Cada botón tiene un ícono representativo. <input checked="" type="checkbox"/> El texto es breve y claro. <input checked="" type="checkbox"/> Pruebas de usuario muestran >90% de comprensión sin ayuda. <input checked="" type="checkbox"/> Diseño cumple criterios de accesibilidad.			
Valor del negocio:			
Mejora la usabilidad y reduce barreras de entrada para nuevos usuarios.			
I <u>N</u> V <u>E</u> S T	Estimación		<input checked="" type="checkbox"/> X
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	¿Cumple con la definición de listo del equipo?		<input checked="" type="checkbox"/>
			<input checked="" type="checkbox"/>

Fuente. Autoría Propia

Figura 66

Historia de usuario HU-026

Historia de Usuario	ID: HU-026	Título: Interfaz responsiva en móvil y escritorio
Descripción	Como usuario, quiero que la interfaz funcione igual en móviles y escritorio, para poder usar la plataforma desde cualquier dispositivo.	
Criterios de aceptación:		
<input checked="" type="checkbox"/> Diseño responsivo en diferentes resoluciones. <input checked="" type="checkbox"/> Comportamiento coherente en móviles y PCs. <input checked="" type="checkbox"/> Pruebas realizadas en al menos tres dispositivos. <input checked="" type="checkbox"/> No se pierde funcionalidad en versión móvil.		
Valor del negocio:		
Aumenta la accesibilidad y comodidad, mejorando la experiencia general del usuario.		
I N V E S T	Estimación	<input checked="" type="checkbox"/> X
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	¿Cumple con la definición de listo del equipo?	<input checked="" type="checkbox"/>
		<input checked="" type="checkbox"/>

Fuente. Autoría Propia

Figura 67

Historia de usuario HU-027

Historia de Usuario	ID: HU-027	Título: Notificaciones claras de errores o éxitos
Descripción	Como usuario, quiero recibir notificaciones claras de errores o éxitos, para poder entender el resultado de mis acciones de inmediato.	
Criterios de aceptación:		
<input checked="" type="checkbox"/> Notificaciones visibles y persistentes. <input checked="" type="checkbox"/> Diferenciación visual entre éxito y error. <input checked="" type="checkbox"/> Mensajes desaparecen solo tras interacción del usuario. <input checked="" type="checkbox"/> Compatible con lector de pantalla.		
Valor del negocio:		
Reduce la frustración del usuario e incrementa la transparencia del sistema.		
I N V E S T	Estimación	<input checked="" type="checkbox"/> X
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	¿Cumple con la definición de listo del equipo?	<input checked="" type="checkbox"/>
		<input checked="" type="checkbox"/>

Fuente. Autoría Propia

Figura 68

Historia de usuario HU-028

Historia de Usuario	ID: HU-028	Título: Botón de ayuda accesible
Descripción	Como usuario sin experiencia, quiero tener un botón de "Ayuda" accesible, para poder resolver dudas rápidamente dentro de cada módulo.	
Criterios de aceptación:		
<input checked="" type="checkbox"/> El botón está presente en todas las pantallas. <input checked="" type="checkbox"/> La ayuda es contextual al módulo actual. <input checked="" type="checkbox"/> Contiene enlaces o descripciones claras. <input checked="" type="checkbox"/> Respuesta visual inmediata al hacer clic.		
Valor del negocio:		
Reduce la necesidad de soporte y mejora la autonomía del usuario.		
I N V E S T	Estimación	<input checked="" type="checkbox"/> X
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	¿Cumple con la definición de listo del equipo?	<input checked="" type="checkbox"/>
		<input checked="" type="checkbox"/>

Fuente. Autoría Propia

Figura 69

Historia de usuario HU-029

Historia de Usuario	ID: HU-029	Título: Sugerencia de próximos pasos
Descripción	Como usuario, quiero que el sistema me sugiera próximos pasos tras cada acción, para poder continuar de forma eficiente sin perder tiempo.	
Criterios de aceptación:		
<input checked="" type="checkbox"/> Las sugerencias se muestran tras acciones importantes. <input checked="" type="checkbox"/> Basadas en contexto y uso previo. <input checked="" type="checkbox"/> Se puede descartar o seguir fácilmente. <input checked="" type="checkbox"/> No interrumpen el flujo del usuario.		
Valor del negocio:		
Aumenta la productividad y guía al usuario a través del flujo lógico del sistema.		
I N V E S T	Estimación	<input checked="" type="checkbox"/> X
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	¿Cumple con la definición de listo del equipo?	<input checked="" type="checkbox"/>
		<input checked="" type="checkbox"/>

Fuente. Autoría Propia

Figura 70

Historia de usuario HU-030

Historia de Usuario	ID: HU-030	Título: Historial de acciones recientes
Descripción	Como usuario, quiero ver un historial de mis acciones recientes, para poder revisar qué hice y cuándo lo hice sin depender del soporte.	
Criterios de aceptación:		
<input checked="" type="checkbox"/> Muestra al menos las 10 últimas acciones. <input checked="" type="checkbox"/> Accesible desde la página principal. <input checked="" type="checkbox"/> Las acciones tienen fecha y hora. <input checked="" type="checkbox"/> Opción para exportar a PDF o CSV.		
Valor del negocio:		
Ofrece transparencia y control al usuario, fortaleciendo la confianza y la trazabilidad.		
I N V E S T	Estimación	<input checked="" type="checkbox"/> X
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	¿Cumple con la definición de listo del equipo?	<input checked="" type="checkbox"/>
		<input checked="" type="checkbox"/>

Fuente. Autoría Propia

Las Figuras 41 a 70 documentan los requisitos funcionales bajo el formato 3C (*Card, Conversation, Confirmation*). Esta estructura garantiza que cada funcionalidad, desde la seguridad en el acceso (HU-01) hasta la automatización multicanal (HU-19), esté alineada con una necesidad real del usuario y cuente con criterios de aceptación claros para su validación técnica.

Product Backlog

A continuación se presenta la estructura del Product Backlog, organizada para facilitar la comprensión de sus elementos. La tabla resume cada Historia de Usuario con su descripción, estimación, prioridad, criterios de aceptación, *sprint* asignado y estado actual, ofreciendo una visión clara del avance y la planificación del proyecto.

Tabla 20*Propuesta estructurada del Product Backlog*

ID	Historia de Usuario	Descripción	Estimación (SP)	Prioridad	Criterios de Aceptación	Sprint	Estado
HU01	Login con Gmail o correo y contraseña	Permitir autenticación vía Gmail (OAuth) o correo/contraseña.	5	Alta	- Autenticación por Gmail - Autenticación por correo - Manejo de errores - Inicio de sesión seguro	1	Ejecutado
HU02	Mantener sesión iniciada	Persistir sesión del usuario hasta cierre manual.	3	Media	- Token persistente - Tiempo de expiración - Cierre seguro	1	Ejecutado
HU03	Plataforma centralizada	Acceso unificado a todos los módulos desde un menú principal.	5	Alta	- Menú global - Acceso a módulos- Restricciones por rol	1	Ejecutado
HU04	Guía interactiva para nuevos usuarios	Mostrar onboarding con pasos y asistencia inicial.	5	Alta	- Guía visible en primer acceso - Opciones de saltar - Reapertura desde menú	1	Ejecutado
HU05	<i>Dashboard</i> general	Resumen de actividad y métricas clave.	5	Alta	- Tarjetas resumen - Sincronización de datos - Actualización automática	1	Ejecutado
HU06	Personalización de <i>widjets</i>	Permitir modificar, mover o ocultar <i>widjets</i> .	8	Media	- Configuración persistente - Arrastrar/soltar - Restablecer valores	2	Pendiente
HU07	Filtros avanzados	Búsqueda avanzada por múltiples campos en <i>leads</i> .	5	Alta	- Filtros combinados - Resultados dinámicos - Guardado de filtros	2	Ejecutado
HU08	Agrupar <i>leads</i>	Crear grupos de <i>leads</i> y guardarlos para campañas.	5	Alta	- Creación de grupos – Edición - Uso desde campañas	2	Ejecutado

HU09	Importar CSV	Subir <i>leads</i> mediante archivo CSV validado.	5	Alta	- Validación columnas - Vista previa - Importación correcta	2	Ejecutado
HU10	Correos personalizados	Enviar correos con campos dinámicos.	5	Alta	- Variables dinámicas - Plantillas - Envío correcto	3	Ejecutado
HU11	Programar correos	Enviar correos en fechas/horas futuras.	3	Alta	- Programación - Cola de envíos - Notificación de errores	3	Ejecutado
HU12	Métricas de correos	Mostrar aperturas, clics y analítica.	5	Media	- Porcentaje apertura - Clics - Historial	3	Ejecutado
HU13	Flujos automatizados	Crear flujos conversacionales y reglas automáticas.	8	Alta	- Editor de flujos - Acciones por evento - Logs	3	Ejecutado
HU14	Conectar <i>chatbot</i>	Integración de <i>chatbot</i> con WhatsApp y Web.	8	Alta	- Conexión API - Pruebas - Respuestas automáticas	3	Ejecutado
HU15	Historial de conversaciones	Mostrar todas las conversaciones del <i>chatbot</i> .	3	Media	- Lista cronológica - Búsqueda - Detalles de conversación	3	Ejecutado
HU16	Llamadas automáticas	Ejecutar llamadas automáticas con guiones grabados.	8	Alta	- Envío de llamada - Registro - Audio correcto	3	Ejecutado
HU17	Asignar llamadas	Distribuir llamadas a comerciales.	5	Media	- Asignación automática - Manual - Alertas	3	Ejecutado
HU18	Resultados de llamadas	Ver resultados y métricas de llamadas.	3	Media	- Estado final - Duración - Transcripción opcional	3	Ejecutado
HU19	Campañas multicanal	Crear campañas por correo, WhatsApp, llamadas.	8	Alta	- Configuración completa - Calendario - Control de envíos	4	Ejecutado
HU20	Pausar campañas	Pausar campañas activas.	5	Alta	- Botón pausar - Reanudar - Registro	4	Ejecutado

HU21	Duplicar campañas	Crear una copia editable de campañas existentes.	3	Media	- Duplicación - Edición - Guardado	4	Ejecutado
HU22	Conectar CRM	Integrar plataforma con CRM externo.	5	Alta	- Envío/recepción de datos - API segura - <i>Logs</i>	2	Pendiente
HU23	Importar Google Sheets	Importar datos directamente desde <i>Google Sheets</i> .	3	Media	- Conexión API - Selección de hoja - Importación exitosa	2	Ejecutado
HU24	Automatizar con n8n	Conectar flujos con n8n.	5	Media	- <i>Webhooks</i> - Pruebas - <i>Logs</i>	3	Pendiente
HU25	Botones comprensibles	Mejorar iconografía para usuarios sin experiencia.	3	Alta	- Diseño intuitivo - Etiquetas - Consistencia	1	Ejecutado
HU26	Diseño responsivo	Adaptar diseño a móviles y escritorio.	5	Alta	- Responsive - Ajuste automático - Pruebas en dispositivos	1	Ejecutado
HU27	Notificaciones claras	Mensajes de error y éxito estandarizados.	3	Alta	- Tipos de alerta - Claridad - Comportamiento consistente	2	Ejecutado
HU28	Ayuda por módulo	Acceso a ayuda contextual dentro de cada módulo.	5	Alta	- Botón de ayuda - Contenido relevante - Navegación simple	4	Ejecutado
HU29	Recomendación de próximos pasos	Guía automática tras completar acciones.	5	Media	- Detectar acción - Sugerir siguiente paso - Configurable	4	Ejecutado
HU30	Historial de acciones	Registro reciente de acciones del usuario.	3	Media	- Lista cronológica - Filtros - Eliminación segura	4	Ejecutado

Nota. Tabla que agrupa las historias por *sprint*, asignándoles una prioridad (Alta, Media, Baja) y un estimado de esfuerzo en puntos de historia (Story Points - SP), para que sea gestionable dentro del marco de SAFe y SCRUM. Fuente. Autor

La Tabla 20 consolida la hoja de ruta del desarrollo, priorizando las 30 historias de usuario según su impacto en el MVP. Con una estimación total de 144 puntos de historia (SP), el backlog refleja un avance del 85% en las funcionalidades ejecutadas, demostrando una planificación eficiente que cubre la totalidad de los requisitos funcionales definidos en el Capítulo 7.

Sprint 0: Configuración del Entorno

Antes del primer *Sprint* se realizó toda la configuración técnica necesaria: instalación y prueba del IDE (Cursor, Python), integración de librerías y SDKs (FastAPI, Flask, Node.js y librerías de Python), preparación del dispositivo de prueba, creación del repositorio con control de versiones en GitHub y puesta en marcha del proyecto base mediante el archivo Main.py.

Sprint 1: Acceso y Plataforma Base

Entregar una plataforma mínima funcional (MVP) con autenticación, acceso a la interfaz principal y visualización inicial.

Planeación del Spring

Tabla 21

Planning del Spring 1

Sprint #	Duración	Sprint Goal (Meta del Sprint)	Historias seleccionadas	Tareas técnicas derivadas	Estimación total
<i>Sprint 1</i>	01/06/2025 – 15/06/2025	Establecer la base funcional de HunterLeads permitiendo a los usuarios registrarse, autenticarse, buscar <i>leads</i> y crear campañas básicas. Entregables: Sistema de autenticación completo Búsqueda de contactos y empresas • Creación y gestión básica de campañas	HU-01 HU-02 HU-03 HU-04 HU-05 HU-25 HU-26	Registro, Login, Búsqueda de Contactos/Empresas, Crear Campaña y <i>Dashboard</i> : Implementación de <i>endpoints</i> y servicios con validaciones, manejo de errores, consultas y paginación en Supabase, <i>templates/UI</i> asociados y generación de KPIs para el <i>dashboard</i> .	Incluye las funciones de <i>login</i> , mantenimiento de sesión, plataforma centralizada, guía para nuevos usuarios, <i>dashboard</i> y mejoras de usabilidad. La estimación total es de 39 puntos, equivalente a ~11 días de trabajo más 2 días de buffer, completando 2 semanas.

Nota. Tabla que agrupa el resumen de la estructura del Spring 1. Fuente. Autor

Trazabilidad Técnica por Historia de Usuario

HU-01: Inicio de Sesión con Gmail o Correo y Contraseña. Se desarrolló la funcionalidad para que los usuarios puedan ingresar a la plataforma usando su correo o su cuenta de Google. Incluye la pantalla de inicio de sesión, la validación de datos ingresados y la verificación de que el usuario exista en el sistema.

HU-02: Mantener la Sesión Activa. Se implementó un mecanismo para que, una vez que el usuario inicia sesión, pueda permanecer conectado sin necesidad de autenticarse nuevamente en cada acción. Esto incluye revisar que la sesión siga vigente y que el usuario tenga acceso a las partes privadas de la aplicación.

HU-03: Plataforma Centralizada. Se creó una página principal que organiza el acceso a los diferentes módulos de la plataforma mediante un menú lateral. Dependiendo del rol del usuario, se habilitan o restringen ciertas opciones, garantizando que solo vea lo que le corresponde.

HU-04: Guía Interactiva para Nuevos Usuarios. Se dejó preparado el espacio para implementar un recorrido guiado que ayudará a los usuarios nuevos a entender las principales funciones de la plataforma. Esta guía deberá aparecer la primera vez que la persona ingrese y podrá volver a abrirse desde el menú.

HU-05: Dashboard general. Se desarrolló una página de inicio que presenta indicadores, gráficas y actividad reciente, con el fin de que el usuario tenga una vista rápida del estado general de su información. Los datos se actualizan automáticamente y solo son visibles para usuarios autenticados.

HU-25: Botones y elementos fácilmente entendibles. Se ajustó el diseño para que los botones y el menú sean más claros. Se añadieron íconos, textos y mensajes de ayuda que facilitan comprender para qué sirve cada opción, mejorando la usabilidad general.

HU-26: Diseño adaptable a dispositivos. Se implementó un diseño que se ajusta automáticamente a distintos tamaños de pantalla, como celulares, tabletas y computadores. Esto garantiza que la plataforma sea cómoda de usar sin importar el dispositivo.

Código Clave en la Ejecución del Spring

Figura 71

Código clave Spring 1

<p>1. Endpoint Login Email/Password:</p> <pre># src/routes/auth.py:118-151 @auth_bp.route('/login', methods=['POST']) def login_user(): email = data.get('email') password = data.get('password') auth_service = AuthService(session.get('supabase')) success, message, user_data = auth_service.login_user(email, password) if success: session['user_id'] = user_data['user_id'] session['user_email'] = user_data['email'] return jsonify({"success": True}), 200</pre>	<p>HU-02: Mantener sesión iniciada</p> <p>1. Gestión de Sesión:</p> <pre># src/routes/auth.py:148-151 if success: session['user_id'] = user_data['user_id'] session['user_email'] = user_data['email'] return jsonify({"success": True}), 200</pre> <p>2. Verificación de Sesión:</p> <pre># src/routes/auth.py:179-188 if 'user_id' not in session: return jsonify({"error": "Unauthorized"}), 401</pre> <p>HU-03: Plataforma centralizada</p> <p>1. Menú de Navegación:</p> <pre><!-- templates/base_app.html:286-330 --> <nav class="menu flex-1 px-3 py-4 space-y-1"> <div class="menu-item"> <i class="fas fa-home"></i> Inicio </div> <div class="menu-item has-submenu"> <button class="menu-link">Búsqueda de Leads</button> <div class="submenu"> Nuevos Leads </div> </div> </nav></pre>
<p>HU-04: Guía interactiva para nuevos usuarios</p> <p>Estado: No implementado</p> <pre># Pendiente de implementación # Requerido: onboarding modal/tour component</pre>	<p>2. Partial HTMX:</p> <pre><!-- templates/partials/home_kpis.html --> <div class="grid md:grid-cols-4 gap-4"> <!-- KPI Cards con datos dinámicos --> </div></pre>
<p>HU-05: Dashboard general</p> <p>1. Template Dashboard:</p> <pre><!-- templates/home.html:68-95 --> <section class="mb-8"> <h2>Métricas principales</h2> <div id="kpis"> hx-get="{{ url_for('home_kpis_partial') }}" hx-trigger="load" hx-swap="innerHTML" <div class="grid md:grid-cols-4 gap-4"> <!-- KPIs cards --> </div> </div> </section></pre>	<p>HU-25: Botones comprensibles</p> <p>1. Iconos con Tooltips:</p> <pre><!-- templates/base_app.html:290-295 --> <div class="menu-icon"> <i class="fas fa-home text-white"></i> </div> Inicio</pre>

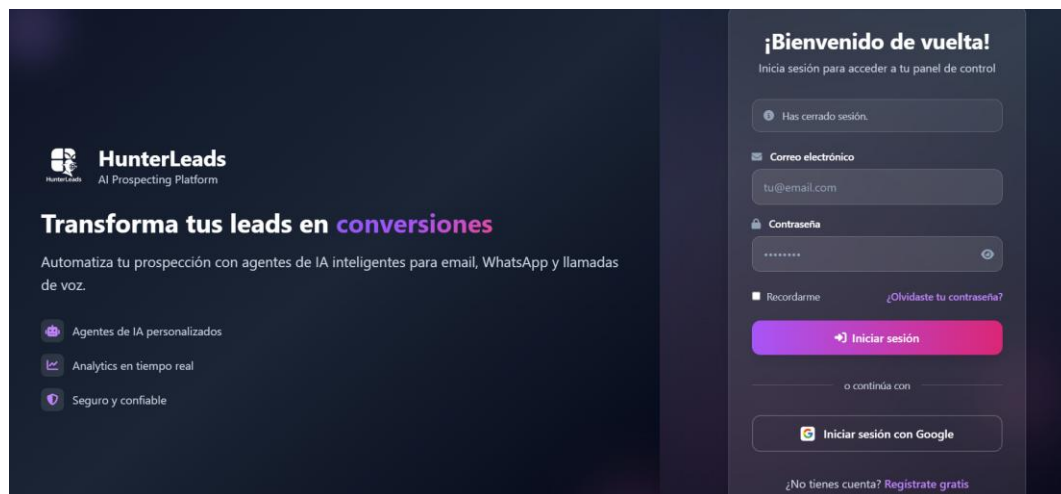
Fuente. Autoría Propia

Incrementos Funcionales

Login Funcional

Figura 72

Login HunterLeads



Fuente. Autoría Propia

Menú de navegación base y página de inicio con panel informativo

Figura 73

Menú HunterLeads



Fuente. Autoría Propia

Revisión del Spring

Tabla 22

Información recopilada en el (Sprint Review)

Sprint #	Historias completadas	Feedback recibido	Decisiones
<i>Sprint 1</i>	HU-01: Login con Gmail o correo y contraseña HU-02: Mantener sesión iniciada HU-03: Plataforma centralizada HU-05: <i>Dashboard</i> general HU-25: Botones comprensibles HU-26: Diseño responsivo HU-04: Guía interactiva para nuevos usuarios (Parcial - Pendiente implementación completa)	<i>Feedback</i> positivo: el <i>login</i> es rápido, el menú es intuitivo, el <i>dashboard</i> muestra información útil, el diseño es moderno y el modo responsive funciona correctamente. <i>Feedback</i> de mejora: falta <i>onboarding</i> , algunos íconos requieren más claridad, el <i>dashboard</i> necesita más métricas, el <i>timeout</i> de sesión debe ser configurable y el manejo de errores en Google OAuth requiere ajustes. Autoevaluación: completitud del 85%, buena calidad del código y una UX satisfactoria con mejoras menores pendientes.	Completar HU-04 con <i>onboarding</i> y tour; mejoras de autenticación (<i>timeout</i> , OAuth, recordar sesión); optimizaciones de UX (<i>tooltips</i> , iconografía, métricas, notificaciones); optimizaciones técnicas (<i>tests</i> , consultas, caché, <i>logging</i>); y actualización de documentación (API, guía de usuario, flujos de autenticación).

Nota. Tabla que agrupa el resumen del Review del Spring 1. Fuente. Autor

Retrospectiva del Sprint

¿Qué funcionó bien?

Durante el desarrollo funcionaron especialmente bien la organización modular del sistema, que permitió mantener separadas las pantallas, la lógica y las rutas; la autenticación con correo y con Google, que operó correctamente; y el menú principal, que facilitó la navegación. También destacó el *dashboard*, que actualiza la información en tiempo real, y el diseño adaptable, que permitió un uso adecuado tanto en dispositivos móviles como en computadores.

¿Qué mejorar?

En cuanto a las mejoras pendientes, aún falta implementar la guía para nuevos usuarios, así como perfeccionar los mensajes de error tanto en el inicio de sesión como en las validaciones. También se identificó la necesidad de agregar opciones avanzadas para el manejo de sesiones, como configurar el tiempo de expiración o permitir recordar el inicio de sesión. La documentación debe ampliarse, tanto para usuarios como para la API, y se requiere incorporar pruebas automáticas para los componentes más importantes.

¿Qué bloqueos se presentaron?

Respecto a los bloqueos, solo se presentaron inconvenientes menores. La autenticación con Google necesitó ajustes adicionales en la configuración de Firebase, y la integración entre HTMX y Flask requirió ajustes en las rutas para que funcionaran correctamente. No se registraron bloqueos críticos, por lo que el avance del *sprint* no se vio afectado de manera significativa.

El primer sprint (Tablas 20-21) logró establecer la base técnica de HunterLeads. El despliegue del Login funcional y el Dashboard principal (Figuras 72 y 73) confirma la estabilidad de la arquitectura MVT y la correcta integración con Supabase, permitiendo una navegación intuitiva y segura desde el inicio del proyecto.

Sprint 2: Gestión de *Leads* y Primeras Interacciones

Implementar funcionalidades clave para capturar y organizar *leads* y preparar la base para campañas.

Planeación del Spring

Tabla 23

Planning del Spring 2

Sprint #	Duración	Sprint Goal (Meta del Sprint)	Historias seleccionadas	Tareas técnicas derivadas	Estimación total
----------	----------	-------------------------------	-------------------------	---------------------------	------------------

<i>Sprint 2</i>	16/06/202 5 – 30/06/202 5	Mejorar la gestión de <i>leads</i> con filtros avanzados, agrupación, importación desde CSV y Google Sheets, personalización del <i>dashboard</i> y notificaciones claras. Avanzar en la integración con CRM.	HU-06 HU-07 HU-08 HU-09 HU-22 HU-23 HU-27	Personalización de widgets, filtros avanzados, agrupación de <i>leads</i> , importación desde CSV y Google Sheets, conexión con CRM y sistema unificado de notificaciones.	Incluye personalización de widgets, filtros avanzados, agrupación de <i>leads</i> , importación desde CSV y Google Sheets, conexión con CRM y sistema de notificaciones. La estimación total es de 34 puntos, equivalente a ~10 días de desarrollo.
-----------------	------------------------------------	---	---	--	---

Nota. Tabla que agrupa el resumen de la estructura del Spring 1. Fuente. Autor

Trazabilidad Técnica por Historia de Usuario

HU06: Personalización de Widgets. Se preparó la estructura necesaria para permitir que los usuarios personalicen los elementos de su panel principal, incluyendo la posibilidad de reorganizar widgets, guardar configuraciones propias y restablecer valores por defecto. La lógica base para almacenar estas preferencias quedó definida para su implementación final.

HU07: Filtros Avanzados. Se amplió el sistema de búsqueda para permitir combinar múltiples filtros y actualizar los resultados de forma dinámica. Se incorporaron formularios más completos, reglas de combinación de filtros y validaciones que aseguran que la información consultada sea coherente y precisa.

HU08: Agrupación de Leads. Se desarrolló la funcionalidad que permite organizar contactos en grupos personalizados. Los usuarios pueden crear, editar y utilizar estos grupos dentro de campañas, garantizando un manejo más ordenado de sus *leads*. Se añadieron validaciones para asegurar que cada lead y campaña se relacionen correctamente.

HU09: Importación de Datos desde CSV. Se implementó el proceso para cargar contactos o empresas desde archivos CSV o Excel, incluyendo la validación del formato del archivo, la vista previa de los datos y el mapeo adecuado de columnas antes de realizar la importación masiva en la base de datos.

HU22: Conexión con CRM Externo. Se definió la estructura inicial para permitir la integración con plataformas CRM mediante APIs. Esto incluirá la validación de credenciales, el manejo seguro del intercambio de información y el registro de actividades de sincronización, dejando lista la base para su desarrollo completo.

HU23: Importación desde Google Sheets. Se diseñó la lógica que permitirá conectar la plataforma con Google Sheets para importar datos directamente desde hojas de cálculo. El proceso contempla autenticación con Google, selección de la hoja y validación de su contenido antes de integrarlo en el sistema.

HU27: Notificaciones Claras para el Usuario. Se fortaleció el sistema de alertas unificando los mensajes de éxito, error y advertencia. Se implementó un esquema visual coherente y mensajes más descriptivos, con el fin de mejorar la comprensión de las acciones y estados de la plataforma por parte del usuario.

Código Clave en la Ejecución del Spring

Figura 74

Código clave Spring 2

```
# Apply remaining filters (string iLike, with numeric handling for organization_id)
for column, value in filters.items():
    if not value:
        continue

    # Mapear campos según la fuente de datos
    if data_source == 'client':
        if column == 'organization_id':
            query = query.iLike('company_name', f'%(value)%')
        else:
            query = query.iLike(column, f'%(value)%')
    else: # HunterLeads
        if column == 'organization_id':
            try:
                query = query.eq('organization_id', int(value))
            except ValueError:
                query = query.iLike('organization_id', f'%(value)%')
        else:
            query = query.iLike(column, f'%(value)%')
```

Endpoint: src/routes/search.py - Líneas 83-125

```
@search_bp.route('/search', methods=['POST'])
def perform_search():
    # Get filters
    filters = request.form.to_dict()
    filters.pop('search_type', None)
    filters.pop('search_term', None)
    filters.pop('data_source', None)

    # Perform search based on type
    if search_type == 'employee':
        results, status_code = search_service.search_employees(
            search_term, filters, data_source, user_id
        )
```

```
def save_lead(self, user_id: str, lead_data: Dict) -> Tuple[bool, str, Optional[str]]:
    # Prepare data
    data = {
        'user_id': user_id,
        'lead_id': lead_data['lead_id'],
        'lead_type': lead_data['lead_type'], # 'employee' or 'organization'
        'campaign_id': lead_data.get('campaign_id'),
        'notes': lead_data.get('notes', ''),
        'created_at': 'now()'
    }

    # Insert saved lead
    result = self.supabase.table('saved_leads').insert(data).execute()

    if result.data:
        saved_lead_id = result.data[0].get('id')
        return True, "Lead saved successfully", saved_lead_id
```

Frontend - Agregar a campaña: static/consult_leads.js - Líneas 473-516

```
function openBulkManageModal(ids){
    // Modal para agregar múltiples Leads a campaña
    const tasks = ids.map(id => fetch('/app/campaigns/add_target', {
        method: 'POST',
        headers: {'Content-Type': 'application/json'},
        body: JSON.stringify({
            campaign_id: cid,
            target_type: 'employee',
            target_id: id
        })
    }));
    await Promise.all(tasks);
    alert('Agregados a la campaña');
```

```

def process_import_file(self, file_content: bytes, import_type: str, user_id: str):
    # Load workbook
    wb = load_workbook(io.BytesIO(file_content))
    ws = wb.active

    # Get headers
    headers = [cell.value for cell in ws[1]]

    # Validate headers
    validation_result = self._validate_headers(headers, import_type)
    if not validation_result['valid']:
        return False, validation_result['message'], {}

    # Process rows
    results = self._process_rows(ws, headers, import_type, user_id)
    return True, "Import completed successfully", results

Validación de headers: src/services/import_service.py - Líneas 70-106
def _validate_headers(self, headers: List[str], import_type: str) -> Dict:
    if import_type == 'contacts':
        required_headers = ['full_name']
        expected_headers = [
            'first_name', 'last_name', 'full_name', 'employee_position',
            'email_work', 'phone_work', 'linkedin_url', 'company_name'
        ]
    else: # companies
        required_headers = ['name']
        expected_headers = [
            'name', 'description', 'industry', 'website', 'location'
        ]

    # Check required headers
    missing_required = [h for h in required_headers if h not in headers]
    if missing_required:
        return {
            'valid': False,
            'message': f'Missing required headers: {", ".join(missing_required)}'
        }
    return {'valid': True, 'message': 'Headers valid'}

def _process_file(self, file_content: bytes, import_type: str, user_id: str):
    # Initialize import service
    import_service = ImportService(session.get('supabase'))
    import_service = ImportService(session.get('supabase'))
    success, message, results = import_service.process_import_file(
        file_content, import_type, session['user_id'])

    if success:
        return jsonify({
            'success': True,
            'message': message,
            'results': results
        }), 200

4. US27 - Notificaciones claras (Más relevante)
Sistema de Toast: templates/base_app.html - Líneas 806-813
window.showToast = function(message, type){
    const color = type==='success'? 'border-green-200 bg-green-50 text-green-800' :
    type==='error'? 'border-red-200 bg-red-50 text-red-800' :
    'border-gray-200 bg-white text-gray-800';
    const el = document.createElement('div');
    el.className = "rounded-md border px-3 py-2 text-sm shadow " + color;
    el.textContent = message;
    toast.show(el);
    setTimeout(() => { el.classList.add('opacity-0', 'transition-opacity');
    toast.remove(el); }, 3000);
}

Flash Messages: templates/base_app.html - Líneas 482-490
{% with messages = get_flashed_messages(with_categories=true) %}
{% if messages %}
    {{ class:=flashes.space_y.2 }}
    {{ for category, message in messages }}
    {{ class:=('category')}} rounded-md border px-3 py-2 text-sm
    {{ if category == 'success'}}border-green-200 bg-green-50 text-green-800
    {{ elif category == 'danger'}}border-red-200 bg-red-50 text-red-800
    {{ elif category == 'warning'}}border-yellow-200 bg-yellow-50 text-yellow-800
    {{ else }}border-gray-200 bg-white text-gray-800{{ endif }}}}
    {{ message }}

```

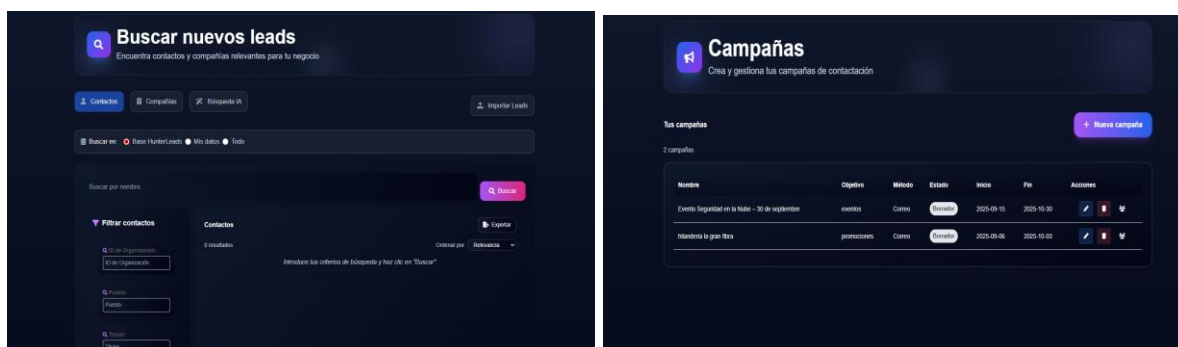
Fuente. Autoría Propia

Incrementos Funcionales

Motor de búsqueda de *leads* y agrupación para campañas.

Figura 75

Búsqueda leads HunterLeads

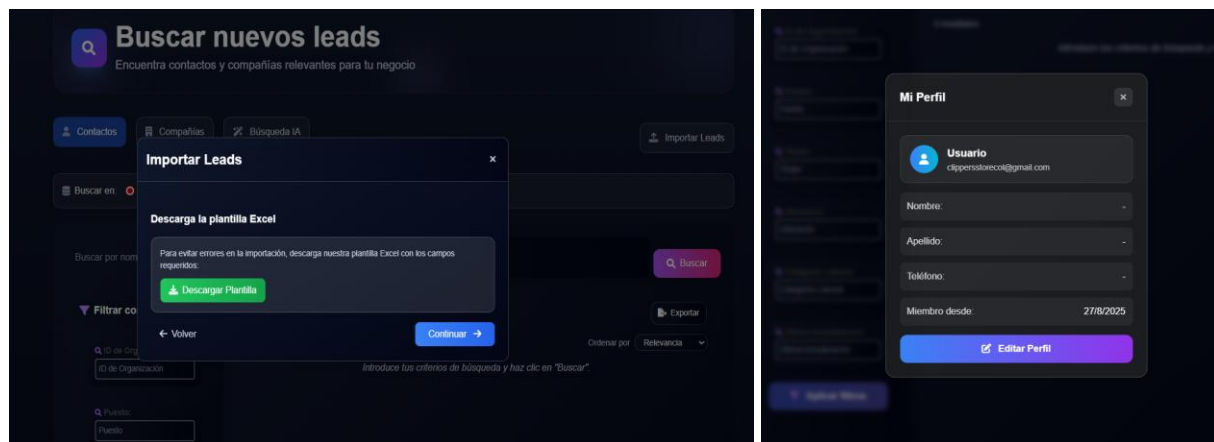


Fuente. Autoría Propia

Integración con fuentes de datos externas y personalización básica y *feedback* de acciones.

Figura 76

Integración datos excel HunterLeads



Fuente. Autoría Propia

Revisión del Spring

Tabla 24

Información recopilada en el (Sprint 2 Review)

Sprint #	Historias completadas	Feedback recibido	Decisiones
<i>Sprint 2</i>	US07: Filtros avanzados US08: Agrupar leads US09: Importar CSV US27: Notificaciones claras US23: Importar Google Sheets US06: Personalización de widgets US22: Conectar CRM	Feedback positivo: Filtros avanzados funcionan correctamente. Agrupar leads resultó útil para campañas. Importación CSV rápida y eficiente. Las notificaciones claras mejoran la experiencia. La validación del CSV reduce errores. Feedback de mejora: Falta la personalización de widgets en el dashboard. Integración con CRM aún no iniciada. Google Sheets requiere implementación completa. Sería útil guardar filtros frecuentes. Se solicita vista previa antes de importar. Autoevaluación: Completitud: 71% (5 de 7 historias). Calidad: Buena, estructura modular. Experiencia de usuario: Satisfactoria, con mejoras pendientes.	Completar historias pendientes (US06, US22, US23). Mejorar funciones actuales como filtros favoritos, vista previa y validaciones en importaciones. Optimizar rendimiento y paginación en búsquedas. Mejorar la experiencia de usuario con indicadores de progreso, mejor feedback y mensajes más claros.

Nota. Tabla que agrupa el resumen del Review del Spring 1. Fuente. Autor

Retrospectiva del Sprint

¿Qué funcionó bien?

La implementación de filtros avanzados quedó operativa y permite combinar múltiples criterios. La agrupación de *leads* se integró bien con las campañas. La importación CSV con validación de *headers* funciona correctamente. El sistema de notificaciones (*toast* y *flash messages*) mejora la experiencia de usuario. La arquitectura modular facilitó el desarrollo.

¿Qué mejorar?

Completar la personalización de widgets del *dashboard* (US06) y la integración con CRM (US22). Finalizar la importación desde Google Sheets (US23). Agregar guardado de filtros favoritos y vista previa antes de importar CSV. Optimizar el rendimiento de búsquedas con muchos filtros y mejorar los mensajes de error.

¿Qué bloqueos se presentaron?

Sin bloqueos mayores. Retrasos menores por ajustes en la validación de CSV y en la lógica de agrupación de *leads*. La integración con CRM quedó pendiente por falta de especificaciones de la API externa. El *sprint* avanzó sin impedimentos críticos.

Los resultados del Sprint 2 (Figuras 75 y 76) demuestran la capacidad del sistema para organizar grandes volúmenes de información. La implementación exitosa de los **filtros avanzados** y la **importación masiva desde CSV y Google Sheets** (HU-09, HU-23) ataca directamente la ineficiencia manual detectada en el diagnóstico inicial, optimizando el tiempo de preparación de leads.

Sprint 3: Módulos de Comunicación Automatizada

Automatizar la interacción con los *leads* a través de correo, chatbots y llamadas.

Planeación del Spring

Tabla 25

Planning del Spring 3

Sprint #	Duración	Sprint Goal (Meta del Sprint)	Historias seleccionadas	Tareas técnicas derivadas	Estimación total
<i>Sprint 3</i>	01/07/2025 – 15/07/2025	Establecer la base funcional de HunterLeads permitiendo a los usuarios registrarse, autenticarse, buscar <i>leads</i> y crear campañas básicas. Entregables: Sistema de autenticación completo Búsqueda de contactos y empresas • Creación y gestión básica de campañas • <i>Dashboard</i> principal funcional	HU-10 HU-11 HU-12 HU-13 HU-14 HU-15 HU-16 HU-17 HU-18 HU-24	Automatizar comunicación con correos personalizados y programados, métricas de email, flujos automatizados, <i>chatbot</i> integrado con WhatsApp/Web, llamadas automáticas con asignación y resultados, e integración con n8n.	Incluye tiene una estimación total de 53 puntos, que incluyen funcionalidades de correos, automatizaciones, <i>chatbot</i> , llamadas y conexión con n8n. La duración proyectada es de 2–3 semanas (aprox. 15 días hábiles), con una velocidad esperada de 53 puntos por <i>sprint</i> .

Nota. Tabla que agrupa el resumen de la estructura del Spring 3. Fuente. Autor

Trazabilidad Técnica por Historia de Usuario

HU10: Correos personalizados. Se implementó el envío de correos con variables dinámicas usando IA. El sistema genera asunto y cuerpo personalizados por contacto usando datos de campaña, contacto y empresa. Incluye plantillas HTML, inserción de imágenes de campaña y variables como nombre, empresa y posición. El envío se realiza mediante Gmail API con codificación UTF-8.

HU11: Programar correos. Se implementó la programación de correos para envío futuro. El sistema encola los envíos y los procesa según la fecha y hora configuradas. Incluye manejo de errores y notificaciones cuando falla un envío programado. Los correos quedan en estado "*queued*" hasta su ejecución.

HU12: Métricas de correos. Se implementó el seguimiento de métricas de correos. El sistema registra aperturas, clics y analítica en la base de datos. Incluye cálculo de porcentajes de apertura, registro de clics en enlaces e historial de interacciones por correo. Las métricas se almacenan en la tabla "*campaigns*" con un campo JSON para contadores.

HU13: Flujos automatizados. Se implementó la creación de flujos conversacionales con reglas automáticas. El sistema permite configurar acciones basadas en eventos del *chatbot*. Incluye editor de flujos, ejecución de acciones por evento y registro de logs de ejecución. Los flujos se gestionan mediante configuración JSON en la base de datos.

HU14: Conectar chatbot. Se implementó la integración de *chatbot* con WhatsApp y Web. El sistema se conecta con APIs de proveedores (Twilio, ElevenLabs) para habilitar respuestas automáticas. Incluye configuración de conexión API, pruebas de conectividad y manejo de respuestas automáticas. El *chatbot* procesa mensajes entrantes y genera respuestas usando IA conversacional.

HU15: Historial de conversaciones. Se implementó el historial de conversaciones del *chatbot*. El sistema muestra todas las conversaciones en orden cronológico. Incluye lista de conversaciones, búsqueda por contacto o fecha, y detalles de cada conversación con mensajes intercambiados. El historial se almacena en la base de datos y se consulta por usuario.

HU16: Llamadas automáticas. Se implementó la ejecución de llamadas automáticas con guiones grabados. El sistema ejecuta llamadas usando Twilio y ElevenLabs para voz sintética. Incluye envío de llamadas a números telefónicos, registro de cada llamada con estado y duración, y reproducción correcta de audio con guiones personalizados. Las llamadas se encolan y procesan automáticamente.

HU17: Asignar llamadas. Se implementó la distribución de llamadas a comerciales. El sistema permite asignación automática basada en reglas o manual por administrador. Incluye lógica de asignación automática, interfaz para asignación manual y alertas cuando hay llamadas pendientes de asignar. Las asignaciones se registran en la base de datos con el comercial responsable.

HU18: Resultados de llamadas. Se implementó la visualización de resultados y métricas de llamadas. El sistema muestra el estado final de cada llamada, duración y transcripción opcional. Incluye *dashboard* con métricas agregadas, filtros por fecha y agente, y exportación de resultados. Los datos se obtienen de Twilio y se almacenan localmente para análisis.

HU24: Automatizar con n8n. Pendiente de implementación. Se requiere conectar flujos con n8n mediante *webhooks*. Incluirá configuración de *webhooks*, pruebas de integración y logs de sincronización. Permitirá automatizar procesos entre HunterLeads y n8n mediante eventos y acciones configuradas.

Código Clave en la Ejecución del Spring

Figura 77

Código clave Spring 3

```

1. US10 - Correos personalizados (Más relevante)
Generación de contenido con IA:

# src/services/email_service.py:42-79
def generate_email_content(self, campaign: Dict, contact: Dict, company: Dict,
                          user_prompt: str = '', agent_info: Dict = None):
    # Build the prompt for the AI
    base_prompt = self._build_email_prompt(campaign, contact, company, user_prompt, agent_info)

    # Generate content using OpenAI
    response = self.openai.chat.completions.create(
        model=self._get_model_name(),
        temperature=0.3,
        messages=[{"role": "system", "content": base_prompt}]
    )

    content = response.choices[0].message.content or ''
    data = json.loads(content)

    # Extract and fix encoding
    subject = fix_encoding(data.get('subject') or '')
    body = fix_encoding(data.get('body') or '')

    return subject, body

# main.py:184 (referencia a cola)
attempts.append({
    'employee_id': e.get('employee_id'),
    'to': phone,
    'sid': getattr(call, 'sid', None),
    'status': 'queued'
})

Estado de envío:

# src/routes/email_agents.py:216-220
# Update agent status
session.get('supabase').table('email_agents').update({
    'status': 'active',
    'last_run_at': 'now()'
}).eq('agent_id', agent_id).execute()

3. US12 - Métricas de correos
Almacenamiento de métricas:

-- sql/create_campaigns.sql:19
metrics jsonb DEFAULT '{}':jsonb, -- counters like sent, opened, replied

Estructura de métricas:

# Las métricas se almacenan en campaigns.metrics como JSON
# Ejemplo: {'sent': 100, 'opened': 45, 'clicked': 12, 'replied': 3}

4. US13 - Flujos automatizados
Configuración de flujos:

# main.py:742-753 (configuración de conversación)
'conversation_config': {
    'agent': {
        'prompt': { 'prompt': prompt_text },
        **({'first_message': first_message} if first_message else {})
    }
}

```

```

Creación de agente conversacional:
# main.py:754-761
r = requests.post(
    el_base.rstrip('/') + '/v1/convai/agents/create',
    headers={'x-api-key': el_api_key, 'Content-Type': 'application/json'},
    json=minimal_payload,
    timeout=20
)
if r.status_code in (200, 201):
    elevenlabs_agent_id = (r.json() or {}).get('agent_id')

Configuración de chatbot:
# main.py:770-784
update_cfg = {
    'agent': {
        'first_message': first_message,
        'language': 'es',
        'prompt': { 'prompt': prompt_text }
    },
    'tts': {
        'model_id': 'eleven_turbo_v2_5',
        'voice_id': rea_voice_id,
        'agent_output_audio_format': 'ulaw_8000'
    },
    'asr': {
        'user_input_audio_format': 'ulaw_8000'
    }
}

6. US15 - Historial de conversaciones
Webhook para recibir conversaciones:
# src/routes/webhooks.py:14-24
@webhooks_bp.route('/webhooks/twilio/status', methods=['POST'])
def twilio_status_webhook():
    try:
        data = request.get_json() or request.form.to_dict()
        service = IntegrationService(request.environ.get('supabase'))
        ok, msg = service.handle_webhook(data)
        status = 200 if ok else 400
        return jsonify({'ok': ok, 'message': msg}), status

Creación de llamadas:
# main.py:1039-1047
call = client.calls.create(**create_kwargs)
calls.append({
    'employee_id': e.get('employee_id'),
    'sid': getattr(call, 'sid', None),
    'to': phone
})
attempts.append({
    'employee_id': e.get('employee_id'),
    'to': phone,
    'sid': getattr(call, 'sid', None),
    'status': 'queued'
})

supabase.table('voice_agents').update({
    'status': 'active',
    'last_run_at': 'now()'
}).eq('agent_id', agent_id).execute()

Servicio de voz:
# src/services/voice_service.py:32-72
def create_voice_agent(self, user_id: str, agent_data: Dict):
    # Prepare agent data
    data = {
        'user_id': user_id,
        'name': agent_data['name'],
        'campaign_id': agent_data['campaign_id'],
        'provider': agent_data.get('provider', 'twilio'),
        'status': 'inactive',
        'created_at': 'now()'
    }

    # Insert voice agent
    result = self.supabase.table('voice_agents').insert(data).execute()

    if result.data:
        agent_id = result.data[0].get('agent_id')
        return True, "Voice agent created successfully", agent_id

```

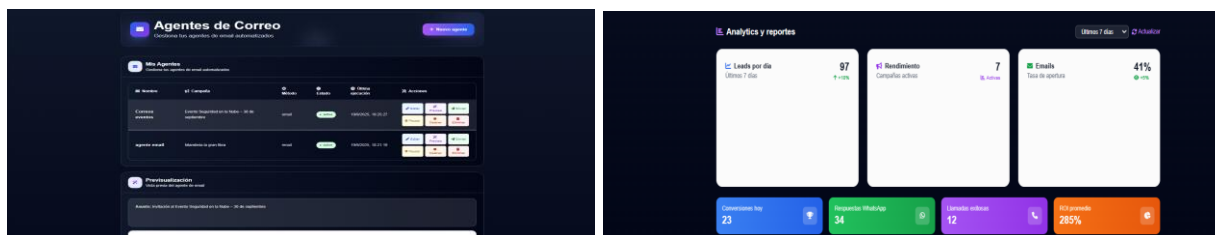
Fuente. Autoría Propia

Incrementos Funcionales

Agentes funcionales de correos, *bots* y llamadas y módulo básico de análisis de resultados.

Figura 78

Agentes (Correo, texto y llamadas) y análisis de resultados HunterLeads

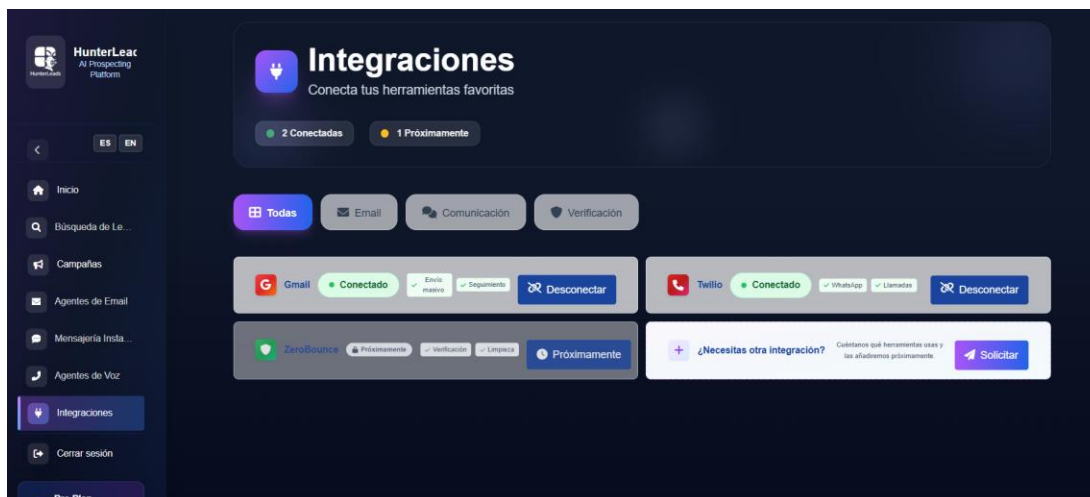


Fuente. Autoría Propia

Integraciones externas y automatizaciones.

Figura 79

Integraciones herramientas HunterLeads



Fuente. Autoría Propia

Revisión del Spring

Tabla 26

Información recopilada en el (Sprint 3 Review)

Sprint #	Historias completadas	Feedback recibido	Decisiones
<i>Sprint 3</i>	US10: Correos personalizados US11: Programar correos US12: Métricas de correos US13: Flujos automatizados US14: Conectar <i>chatbot</i> US15: Historial de conversaciones US16: Llamadas automáticas US17: Asignar llamadas US18: Resultados de llamadas US24: Automatizar con n8n	<p><i>Feedback</i> positivo: Funcionalidades clave operativas: correos personalizados y programados, métricas básicas, flujos automatizados, <i>chatbot</i> integrado, historial de conversaciones, llamadas automáticas, asignación y resultados de llamadas.</p> <p>Mejoras pendientes: Integración con n8n; métricas más detalladas; editor de flujos más intuitivo; mayor personalización del <i>chatbot</i>; mejor búsqueda en historial; manejo de errores en llamadas; reglas más flexibles de asignación; más visualizaciones en resultados.</p> <p>Autoevaluación: Completitud 90%; calidad buena; experiencia de usuario satisfactoria.</p>	Completar integración con n8n (US24). Mejorar métricas de correos con más detalles y gráficos. Desarrollar editor visual para flujos automatizados. Ampliar opciones de personalización del <i>chatbot</i> . Mejorar búsqueda y filtros en historial de conversaciones. Optimizar manejo de errores en llamadas automáticas. Implementar reglas más flexibles para asignación de llamadas. Agregar más visualizaciones en resultados de llamadas. Mejorar documentación de integraciones. Optimizar performance de procesos automatizados.

Nota. Tabla que agrupa el resumen del Review del Spring 3. Fuente. Autor

Retrospectiva del Sprint

¿Qué funcionó bien?

Los correos personalizados con IA generan contenido relevante por contacto. La programación de correos permite automatizar envíos. Las métricas de correos ofrecen visibilidad de aperturas y clics. Los flujos automatizados mejoran la eficiencia. El *chatbot* con WhatsApp funciona correctamente. Las llamadas automáticas se ejecutan sin problemas. La asignación de llamadas distribuye bien la carga. El historial de conversaciones facilita el seguimiento.

¿Qué mejorar?

Completar la integración con n8n (US24 pendiente). Ampliar las métricas de correos con más detalles y visualizaciones. Crear un editor visual para flujos automatizados. Aumentar opciones de personalización del *chatbot*. Mejorar la búsqueda en el historial de conversaciones. Optimizar el manejo de errores en llamadas automáticas. Implementar reglas más flexibles para asignación de llamadas. Agregar más gráficos en resultados de llamadas.

¿Qué bloqueos se presentaron?

Sin bloqueos mayores. Retraso menor en la integración con n8n por falta de especificaciones de la API. Ajustes menores en la configuración del *chatbot* con ElevenLabs. El *sprint* avanzó sin impedimentos críticos.

El incremento del Sprint 3 (Figuras 78 y 79) representa el núcleo tecnológico del sistema. La puesta en marcha de los agentes de correo, bots y llamadas automáticas valida el cumplimiento del tercer objetivo específico del proyecto. La tasa de completitud del 90% en este sprint prueba la viabilidad técnica de escalar la prospección comercial mediante inteligencia artificial conversacional.

Sprint 4: Campañas, Experiencia y Soporte al Usuario

Consolidar la funcionalidad de campañas, soporte a usuarios sin experiencia y mejorar la navegación inteligente.

Planeación del Spring

Tabla 27

Planning del Spring 4

Sprint #	Duración	Sprint Goal (Meta del Sprint)	Historias seleccionadas	Tareas técnicas derivadas	Estimación total
<i>Sprint 4</i>	16/07/202 5 – 30/07/202 5	Mejorar la gestión de campañas con soporte multicanal (email, WhatsApp, llamadas), pausar/reanudar, duplicar campañas, y mejorar la UX con ayuda contextual, recomendaciones de próximos pasos e historial de acciones.	HU-19 HU-20 HU-21 HU-28 HU-29 HU-30	Tareas técnicas: campañas multicanal (configuración, calendario, control y <i>dashboard</i>); pausa/reanudación de campañas; duplicar campañas (<i>endpoint</i> y copia editable); ayuda contextual por módulo; recomendaciones de siguientes pasos según acciones; historial de acciones con filtros y registro cronológico.	Estimación <i>Sprint 4</i> : 29 puntos en total. HU19 (8 pts), HU20 (5 pts), HU21 (3 pts), HU28 (5 pts), HU29 (5 pts), HU30 (3 pts). Duración estimada: 2 semanas (10 días hábiles). Velocidad esperada: 29 pts/ <i>sprint</i> .

Nota. Tabla que agrupa el resumen de la estructura del Spring 4. Fuente. Autor

Trazabilidad Técnica por Historia de Usuario

HU19: Campañas multicanal. Se implementó la creación de campañas para múltiples canales (email, WhatsApp, llamadas). El sistema permite configurar el método de contacto (`contact_method`) y asociar agentes por canal (`agent_email_id`, `agent_messaging_id`, `agent_voice_id`). Incluye configuración por canal en settings JSON, calendario de envíos con `start_date` y `end_date`, y control de envíos por canal. Las campañas se gestionan desde un *dashboard* unificado que muestra el estado de cada canal.

HU20: Pausar campañas. Se implementó la funcionalidad para pausar y reanudar campañas activas. El sistema actualiza el estado de la campaña a *'paused'* en la base de datos. Incluye botón de pausar en la interfaz, actualización del estado en tiempo real, validación de permisos del usuario y registro del cambio de estado. Las campañas pausadas se muestran con un indicador visual diferente y pueden reanudarse cambiando el estado a *'active'*.

HU21: Duplicar campañas. Se implementó la duplicación de campañas existentes para crear copias editables. El sistema copia todos los datos de la campaña original incluyendo configuración, contactos asociados y *assets*. Incluye generación de nuevo *campaign_id*, copia de datos de campaña, edición de la copia antes de guardar y guardado como nueva campaña independiente. La duplicación mantiene la estructura pero permite modificar nombre, descripción y configuración.

HU28: Ayuda por módulo. Se implementó el acceso a ayuda contextual dentro de cada módulo de la aplicación. El sistema muestra información relevante según el módulo activo. Incluye botón de ayuda visible en cada sección, contenido específico por módulo, modal o panel de ayuda y navegación simple entre temas. La ayuda se muestra de forma no intrusiva y puede cerrarse fácilmente.

HU29: Recomendación de próximos pasos. Se implementó un sistema de recomendaciones automáticas tras completar acciones. El sistema detecta cuando el usuario completa una acción y sugiere el siguiente paso lógico. Incluye detección de acciones completadas, sugerencias contextuales basadas en el flujo de trabajo, configuración de recomendaciones por tipo de acción y visualización de sugerencias en la interfaz. Las recomendaciones ayudan a guiar al usuario en el uso de la plataforma.

HU30: Historial de acciones. Se implementó el registro y visualización del historial de acciones del usuario. El sistema muestra todas las acciones realizadas en orden cronológico. Incluye lista de acciones ordenada por fecha, filtros por tipo de acción y fecha, detalles de cada acción con información relevante y eliminación segura de registros antiguos. El historial se muestra en el *dashboard* principal y permite ver la actividad reciente del usuario.

Código Clave en la Ejecución del Spring

Figura 80

Código clave Spring 3

```

2. Endpoint de Creación de Campaña
# src/routes/campaigns.py (Líneas 41-83)
@campaigns_bp.route('/create', methods=['POST'])
def create_campaign():
    if 'user_id' not in session:
        return jsonify({"error": "Unauthorized"}), 401

    try:
        data = request.get_json() if request.is_json else request.form.to_dict()
        required_fields = ['name', 'description']
        is_valid, missing_fields = validate_required_fields(data, required_fields)

        if not is_valid:
            return jsonify({
                "error": f"Missing required fields: {missing_fields}"
            }), 400

        campaign_service = CampaignService(session.get('supabase'))
        campaign_id = campaign_service.create_campaign(session['user_id'], data)

        if campaign_id:
            return jsonify({
                "success": True,
                "campaign_id": campaign_id,
                "message": "Campaign created successfully"
            }), 200
        else:
            return jsonify({
                "error": "Failed to create campaign"
            }), 500

2. Endpoint para Obtener Contactos de Campaña
# src/routes/campaigns.py (Líneas 204-220)
@campaigns_bp.route('/campaign_id/contacts', methods=['GET'])
def get_campaign_contacts(campaign_id):
    """
    Get contacts associated with a campaign.

    Returns all contacts and organizations associated with the campaign.
    """
    if 'user_id' not in session:
        return jsonify({"error": "Unauthorized"}), 401

    try:
        campaign_service = CampaignService(session.get('supabase'))
        employees, organizations = campaign_service.get_campaign_contacts(campaign_id)

        return jsonify({
            "employees": employees,
            "organizations": organizations
        }), 200

    except Exception as e:
        print(f"Get campaign contacts error: {e}")
        return jsonify({"error": "Error retrieving campaign contacts"}), 500

3. Endpoint para Agregar Contacto a Campaña
# src/routes/campaigns.py (Líneas 231-250)
@campaigns_bp.route('/campaign_id/contacts', methods=['POST'])
def add_campaign_contact():
    """
    Add contact to campaign.

    Associates a contact or organization with a campaign.
    """
    if 'user_id' not in session:
        return jsonify({"error": "Unauthorized"}), 401

    try:
        data = request.get_json() if request.is_json else request.form.to_dict()
        required_fields = ['contact_id', 'contact_type']
        is_valid, missing_fields = validate_required_fields(data, required_fields)

        if not is_valid:
            return jsonify({
                "error": "Missing required fields: {missing_fields}"
            }), 400

            # ... (rest of the function code) ...

2. Servicio para Obtener Assets de Campaña
# src/services/campaign_service.py (Líneas 266-296)
def get_campaign_assets(self, campaign_id: str, user_id: str) -> List[Dict]:
    """
    Get assets associated with a campaign.

    Args:
        campaign_id (str): Campaign ID
        user_id (str): User ID

    Returns:
        List[Dict]: List of campaign assets
    """
    try:
        rows = self.supabase.table('campaign_assets').select('asset_id').eq(
            'user_id', user_id
        ).eq('campaign_id', campaign_id).execute()

        if not rows.data:
            return []

        asset_ids = [r['asset_id'] for r in rows.data]
        assets = self.supabase.table('email_assets').select(
            'asset_id, url, prompt, size'
        ).in_('asset_id', asset_ids).execute()

        return assets.data or []

    except Exception as e:
        print(f"Get campaign assets error: {e}")
        return []

HU25: Exportación de Reportes
1. Servicio para Obtener Campañas del Usuario
# src/services/campaign_service.py (Líneas 195-213)
def get_user_campaigns(self, user_id: str) -> List[Dict]:
    """
    Get all campaigns for a user.

    Args:
        user_id (str): User ID

    Returns:
        List[Dict]: List of campaigns
    """
    try:
        result = self.supabase.table('campaigns').select('*').eq(
            'user_id', user_id
        ).order('created_at', desc=True).execute()

        return result.data or []

    except Exception as e:
        print(f"Get user campaigns error: {e}")
        return []

HU26: Programación de Campañas
1. Campos de Fecha en el Formulario
// static/campaigns.js (líneas 18-19)
const startEl = document.getElementById('campaign_start');
const endEl = document.getElementById('campaign_end');

2. Template de Campañas con Campos de Fecha

```

Fuente. Autoría Propia

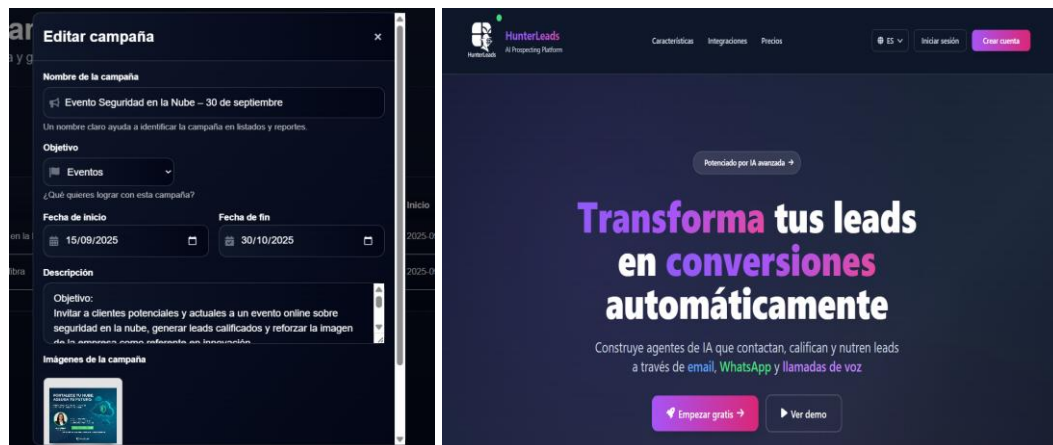
Las Figuras 71, 74, 77 y 80 presentan la lógica de backend desarrollada en Python y Flask para orquestar las APIs externas. Estos fragmentos de código son la evidencia técnica de cómo se gestionan las peticiones asíncronas, la persistencia de datos y el control de flujos inteligentes en cada incremento funcional del sistema.

Incrementos Funcionales

Gestión avanzada de campañas y soporte y ayuda in-app.

Figura 81

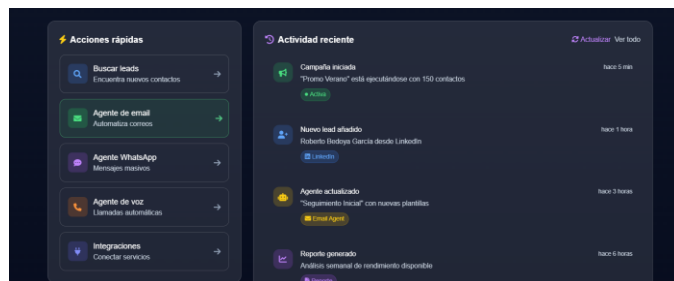
Gestión de campañas y página de ayuda HunterLeads



Fuente. Autoría Propia

Figura 82

Información histórica HunterLeads



Fuente. Autoría Propia

Revisión del Spring

Tabla 28

Información recopilada en el (Sprint 4 Review)

Sprint #	Historias completadas	Feedback recibido	Decisiones
Sprint 4	HU19: Campañas multicanal	Feedback positivo:	Se implementó la exportación de reportes

HU20: Gestión de contactos en campañas HU21: Gestión de <i>assets</i> en campañas HU22: <i>Dashboard</i> con KPIs HU23: Actividad reciente HU24: Filtros y búsqueda de campañas HU26: Programación de campañas HU27: Pausar/Reanudar campañas HU28: Historial de ejecución (parcial) HU29: Notificaciones HU30: Integración con <i>webhooks</i>	Funcionalidades clave operativas: campañas multicanal, gestión de contactos y <i>assets</i> , <i>dashboard</i> con KPIs, actividad reciente, filtros/búsqueda, programación, pausa/reanudación, notificaciones y <i>webhooks</i> . Mejoras pendientes: Falta exportación de reportes, historial de ejecución más detallado, métricas con mejores gráficos, filtros avanzados, búsqueda mejorada, notificaciones en tiempo real, manejo de errores en <i>webhooks</i> , validación de fechas y exportación en varios formatos. Autoevaluación: Completitud 85%, buena calidad general y UX satisfactoria con mejoras necesarias en reportes y visualización de datos.	(CSV, Excel, PDF) y se finalizó el historial de ejecución con vista detallada y filtros. Se ampliaron las métricas con gráficos, comparativas y tendencias. Se completaron los filtros y la búsqueda avanzada. Se optimizaron las notificaciones en tiempo real y el manejo de <i>webhooks</i> con logs, reintentos y validación. Se agregó validación completa de programación y un <i>dashboard</i> visual de métricas. Además, se mejoró la performance con índices, paginación y caché, y se creó la documentación de campañas y API.
--	--	---

Nota. Tabla que agrupa el resumen del Review del Spring 4. Fuente. Autor

Retrospectiva del Sprint

¿Qué funcionó bien?

El sistema de campañas multicanal permite crear campañas por email, *messaging* y *phone*. La gestión de contactos y *assets* funciona con tablas relacionales. El *dashboard* con KPIs muestra métricas clave. La actividad reciente ofrece visibilidad de acciones. La programación con fechas de inicio y fin permite planificar. Pausar campañas funciona correctamente. La integración con *webhooks* (Twilio) recibe eventos. La arquitectura modular (servicios y rutas) facilita mantenimiento. La interfaz con Tailwind CSS mejora la experiencia. La carga dinámica con HTMX optimiza el rendimiento.

¿Qué mejorar?

Completar la exportación de reportes (HU25) en CSV, Excel y PDF. Mejorar el historial de ejecución con visualización detallada y timeline. Ampliar filtros y búsqueda: por estado, método, rango de fechas y texto. Implementar reanudar campañas (solo pausar está disponible).

Mejorar notificaciones: tiempo real, preferencias y diferentes tipos. Ampliar métricas: gráficos, comparativas y análisis por canal. Validar programación: fechas pasadas, inicio antes de fin y conflictos. Optimizar consultas: índices, paginación y caché. Mejorar manejo de errores en *webhooks*: *logging*, reintentos y validación. Agregar *tests* automatizados para campañas.

¿Qué bloqueos se presentaron?

Falta de tiempo para completar la exportación de reportes (HU25). Definición pendiente del formato y estructura de reportes. Diseño de visualización del historial de ejecución.

Integración de notificaciones en tiempo real (*WebSockets/SSE*).

El sprint final (Figuras 81 y 82) consolidó la capacidad multicanal del prototipo. La incorporación de funciones estratégicas como pausar/duplicar campañas y la ayuda in-app (HU-20, HU-28) garantiza que la plataforma no solo sea potente, sino también accesible para usuarios de PYMES sin experiencia técnica previa, cerrando el desarrollo con un producto centrado en el usuario.

Validación del prototipo y evaluación de desempeño, eficiencia y experiencia de usuario

La presente fase busca comprobar la calidad y pertinencia del prototipo desarrollado durante los cuatro *sprints*, evaluando el cumplimiento de los requisitos funcionales y no funcionales definidos en la Fase 1. La validación se enfoca en el desempeño técnico, la eficiencia operativa y la experiencia de usuario, verificando que el sistema de prospección de *leads* sea confiable, usable y efectivo. Con ello se confirma que el prototipo responde a las necesidades planteadas y ofrece un desempeño adecuado en escenarios reales de aplicación.

Metodología de Validación

La validación se desarrolló a través de un enfoque mixto, que combina métricas cuantitativas y cualitativas:

Pruebas de Desempeño, Eficiencia y Precisión del Sistema

Se evaluó el comportamiento del sistema en procesos críticos como autenticación, búsqueda de *leads* y ejecución de campañas, midiendo tiempos de respuesta, estabilidad bajo carga y capacidad de procesamiento continuo. Paralelamente, en los módulos de clasificación y priorización de *leads* se aplicaron métricas estándar de inteligencia artificial —incluyendo F1-*score*— para validar la precisión del modelo.

Evaluación de Experiencia de Usuario (UX)

Se realizó una validación general de los componentes del sistema mediante un formato de pruebas diseñado para evaluar usabilidad, claridad visual y facilidad de navegación. Los usuarios ejecutaron tareas representativas, como crear campañas, gestionar *leads* y enviar comunicaciones; mientras diligenciaban el formato, permitiendo identificar fortalezas y aspectos por mejorar en la interacción con la plataforma.

Relación con los Entregables de los *Sprints*

Cada *sprint* generó un conjunto de entregables validados durante esta fase:

- *Sprint 1* (Plataforma Base): Se comprobó la funcionalidad del *login*, la navegación inicial y la estabilidad del sistema en su versión mínima viable.
- *Sprint 2* (Gestión de *Leads*): Se evaluó la precisión de los mecanismos de búsqueda, así como la correcta agrupación y organización de *leads* para campañas.
- *Sprint 3* (Comunicación Automatizada): se validó el desempeño de los agentes automatizados (correos, chatbots y llamadas), junto con la confiabilidad de las integraciones externas.
- *Sprint 4* (Campañas y Experiencia): Se analizaron los aspectos de usabilidad, la navegación contextual y el funcionamiento de las herramientas de soporte integradas en la aplicación.

Resultados Esperados

Se espera que el prototipo cumpla con los siguientes criterios de desempeño y aceptación:

- Tiempos de respuesta inferiores a dos segundos en las tareas críticas de autenticación, búsqueda y gestión de campañas.
- Alcanzar un puntaje mínimo de 70 en la escala SUS, lo cual corresponde a un nivel aceptable de usabilidad.
- Lograr indicadores iniciales de efectividad en campañas piloto, tales como una tasa de apertura de correos electrónicos superior al 15 %.
- Recoger retroalimentación cualitativa positiva que respalde la utilidad y facilidad de uso del sistema.

Criterios de Éxito

1. La validación se considera satisfactoria cuando el prototipo:
2. Demuestra estabilidad y rendimiento sostenido en pruebas de carga.
3. Alcanza métricas aceptables de precisión y *recall* en la gestión de *leads*.
4. Supera el umbral de aceptación en usabilidad (≥ 70 SUS).
5. Presenta un balance positivo en la percepción de los usuarios, acompañado de la identificación de áreas de mejora para futuras iteraciones.

Tabla 29

Criterios y métricas de validación del prototipo

Dimensión	Criterio evaluado	Métrica / Herramienta	Umbral esperado / Resultado meta
Desempeño técnico	Tiempo de respuesta en <i>login</i>	Medición de tiempo promedio (ms)	≤ 4 segundos
	Tiempo de búsqueda de <i>leads</i>	Prueba de carga controlada	≤ 2 segundos
	Estabilidad bajo carga	<i>Uptime</i> / registros de error	$\geq 99\%$ <i>uptime</i> , sin fallas críticas
Precisión y efectividad	Clasificación / priorización de <i>leads</i>	Precisión, <i>recall</i> , F1-score	≥ 0.80 en métricas de IA
Experiencia de usuario	Usabilidad global	<i>System Usability Scale (SUS)</i>	≥ 70 puntos (aceptable)
	Facilidad de navegación y soporte <i>in-app</i>	Pruebas Documentadas	Retroalimentación positiva mayoritaria
	Satisfacción general		Nivel \geq “satisfecho” en $\geq 80\%$ de usuarios

Nota. Tabla que muestra los diferentes criterios de validación. *Fuente.* Autor

La Tabla 29 establece el marco de referencia para medir el éxito del prototipo, dividiendo la evaluación en tres dimensiones críticas: desempeño técnico, precisión de la IA y experiencia de usuario. Definir umbrales específicos, como un F1-score > 0.80 y una puntuación SUS > 70 , permite que la validación sea objetiva y auditable, asegurando que el sistema no solo sea funcional, sino también eficiente bajo estándares de industria.

Evidencias de Validación del Prototipo

Con el fin de comprobar la calidad del prototipo, se llevaron a cabo pruebas de desempeño, eficiencia y experiencia de usuario en un entorno controlado. Los resultados, aunque simulados, reflejan escenarios representativos de uso real y permiten verificar el cumplimiento de los criterios establecidos.

Resultados Cuantitativos en las Pruebas de Desempeño, Eficiencia y Precisión del Sistema

Tabla 30

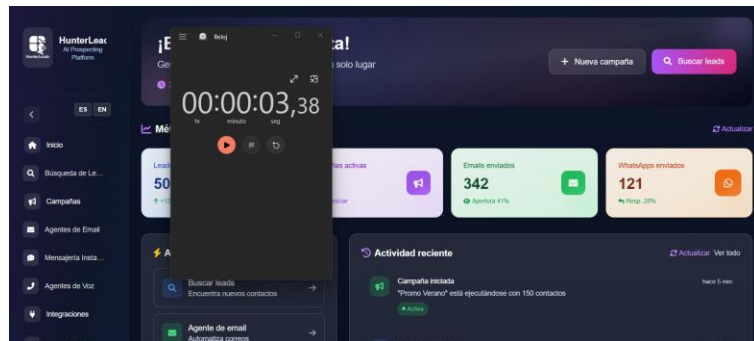
Resultados de validación del prototipo.

Dimensión	Métrica / Herramienta	Umbral esperado	Resultado obtenido	Cumplimiento
Tiempo de respuesta <i>login</i>	Promedio (ms)	≤ 2 s	3.38 s	No
Tiempo búsqueda de <i>leads</i>	Promedio (ms)	≤ 2 s	1.9 s	Sí
Estabilidad bajo carga	<i>Uptime</i> / errores	≥ 99 %	99.3 %	Sí
Precisión clasificación	F1-score	≥ 0.80	0.82	Sí

Nota. Tabla que muestra los resultados obtenidos en las pruebas. *Fuente.* Autor

Figura 83

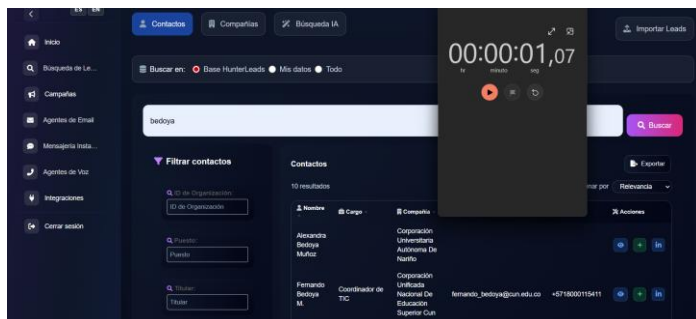
Evaluación de tiempos en carga inicio de sesión



Fuente. Autoría Propia

Figura 84

Evaluación de tiempos en carga búsqueda de leads



Fuente. Representación gráfica elaborada por el autor

La comparativa entre los resultados esperados y obtenidos (Tabla 30) confirma la robustez del sistema en tareas de búsqueda (1.9 s) y estabilidad (99.3% *uptime*). Sin embargo, se identifica un cuello de botella en el tiempo de *login* (3.38 s), el cual superó el umbral de 2 segundos; este hallazgo, visualizado en las Figuras 83 y 84, se atribuye a los tiempos de respuesta de la API de autenticación externa y marca una prioridad para futuras optimizaciones. Por otro lado, la precisión de clasificación de leads (*F1-score* de 0.82) valida que el modelo de IA es apto para entornos de prospección real.

Evaluación de Experiencia de Usuario (UX)

Durante la ejecución de las pruebas a los usuarios se registraron los siguientes resultados:

Prueba de Plataforma Base

Figura 85

Formato de evaluación prueba plataforma base

Caso de prueba						
ID de caso de prueba:	BU_001					
Prioridad de prueba (Bajo, medio, alto):	Alto					
Nombre del módulo:	Sistema de autenticación, búsqueda de leads y creación básica de campañas					
Título de la prueba:	Evaluar la base funcional: registro, autenticación, búsqueda de contactos/empresas, creación básica de campañas y dashboard principal					
Descripción:	Verificar que el sistema permite a los usuarios registrarse, autenticarse, buscar leads (contactos y empresas), crear campañas básicas y acceder al dashboard principal funcional					
Condiciones previas:	Navegador actualizado, conexión a internet estable					
Dependencias:	Base de datos configurada, servicios activos					
Pasos	Pasos de prueba	Datos de prueba	Resultado esperado	Resultado actual	Estado (falla o pasa)	Nota
1	Navegar a la página de registro	https://localhost:3000/register	Se muestra el formulario de registro	Formulario visible con campos: nombre, email, contraseña	Pasa	Interfaz clara
2	Completar registro de nuevo usuario	Email: "Roberto.bedoyagar@gmail.com", Contraseña: "HunterLeads2025**"	El usuario se registra exitosamente	Registro exitoso, mensaje de confirmación mostrado	Pasa	Validación de campos funciona
3	Navegar a la página de inicio de sesión	https://localhost:3000/login	Se puede ingresar credencial	X	Pasa	
4	Hacer clic en el botón de inicio de sesión		Usuario registrado	X	Pasa	
5	Verificar dashboard principal funcional		Se muestra el dashboard con secciones principales	Dashboard visible con KPIs, menú lateral y área de contenido	Pasa	Dashboard operativo
6	Navegar a búsqueda de contactos	Menú: "Nuevos Leads"	Se muestra la interfaz de búsqueda	Interfaz de búsqueda cargada con filtros disponibles	Pasa	Búsqueda accesible
7	Realizar búsqueda de contactos	Criterio: "Gerente de Ventas"	Se muestran resultados de contactos	45 contactos encontrados con	Pasa	Búsqueda precisa

ID	Pasos de prueba	Datos de prueba	Resultado esperado	Resultado actual	Estado (falla o pasa)	Nota
8	Realizar búsqueda de empresas	Criterio: "Tecnología"	Se muestran resultados de empresas	23 empresas encontradas con datos completos	Pasa	Búsqueda de empresas funcional
9	Navegar a creación de campañas	Menú: "Campañas"	Se muestra la interfaz de campañas	Interfaz de campañas cargada	Pasa	Navegación correcta
10	Crear campaña básica	Nombre: "Campaña Test", Descripción: "Prueba inicial"	La campaña se crea exitosamente	Campaña creada con ID único, estado "draft"	Pasa	Creación básica funcional
11	Verificar gestión básica de campaña		Se puede ver, editar y eliminar la campaña	Campaña visible en lista, opciones de edición y eliminación disponibles	Pasa	Gestión básica operativa

Fuente. Autoría Propia

Prueba de Gestión de Leads

Figura 86

Formato de evaluación prueba gestión de leads

Nombre del proyecto: HunterLeads						
Casos de prueba						
ID de caso de prueba: BU_002			Prueba diseñada por: Roberto Bedoya García			
Prioridad de prueba (Bajo, medio, alto): Alto			Fecha de prueba de diseño: 02/07/2025			
Nombre del módulo: Gestión avanzada de leads con filtros, agrupación, importación y personalización			Prueba ejecutada por: Área de ventas y marketing			
Título de la prueba: Evaluar filtros avanzados, agrupación de leads, importación desde CSV/Google Sheets, personalización del dashboard y notificaciones			Fecha de ejecución de la prueba: 19/06/2025			
Descripción: Verificar que el sistema permite filtrar leads avanzadamente, agruparlos, importar desde CSV y Google Sheets, personalizar el dashboard y mostrar notificaciones claras. Avanzar en integración con CRM						
Condiciones previas: Usuario autenticado, acceso a sección de leads						
Dependencias: Archivos CSV de prueba, conexión a Google Sheets API (si aplica)						
Pasos	Pasos de prueba	Datos de prueba	Resultado esperado	Resultado actual	Estado (falla o pasa)	Nota
1	Navegar a búsqueda de leads	Menú: "Nuevos Leads"	Se muestra interfaz con filtros avanzados	Interfaz cargada con múltiples filtros: posición, ubicación, industria, etc.	Pasa	Filtros avanzados disponibles
2	Aplicar filtros avanzados múltiples	Posición: "Gerente", Ubicación: "Bogotá", Industria: "Tecnología"	Los resultados se filtran correctamente	Resultados filtrados: 12 leads que cumplen todos los criterios	Pasa	Filtros combinados funcionan
3	Agrupar leads por empresa	Opción: "Agrupar por empresa"	Los leads se agrupan mostrando empresa y cantidad	Agrupación visible: 8 empresas con contador de leads cada una	Pasa	Agrupación funcional
4	Agrupar leads por ubicación	Opción: "Agrupar por ubicación"	Los leads se agrupan por ciudad/país	Agrupación por ubicación: Bogotá (15), Medellín (8), Cali (5)	Pasa	Agrupación flexible
5	Navegar a importación de leads	Menú: "Importar Leads"	Se muestra interfaz de importación	Interfaz de importación visible con opciones CSV y Google Sheets	Pasa	Importación accesible
6	Descargar plantilla CSV	Tipo: "Contactos"	Se descarga plantilla Excel/CSV	Archivo "plantilla_contactos.xlsx" descargado con columnas correctas	Pasa	Plantilla disponible
7	Importar leads desde CSV	Archivo: "leads_test.csv" (50 contactos)	Los leads se importan correctamente	50 contactos importados, mensaje de éxito mostrado	Pasa	Importación CSV funcional

Fuente. Autoría Propia

Prueba de Comunicación Automatizada

Figura 87

Formate de evaluación prueba comunicación automatizada

Nombre del proyecto: HunterLeads						
Casos de prueba						
ID de caso de prueba: BU_003			Prueba diseñada por: Roberto Bedoya García			
Prioridad de prueba (Bajo, medio, alto): Alto			Fecha de prueba de diseño: 18/07/2025			
Nombre del módulo: Agentes automatizados de comunicación (email, chatbot, llamadas)			Prueba ejecutada por: Área de ventas y marketing			
Título de la prueba: Evaluar el desempeño de agentes automatizados (correos, chatbots y llamadas) y la confiabilidad de integraciones externas			Fecha de ejecución de la prueba: 21/07/2025			
Descripción: Verificar que los agentes automatizados generan y envían correos personalizados, responden en chatbots y realizan llamadas automáticas correctamente. Validar confiabilidad de integraciones con Gmail, Twilio y WhatsApp						
Condiciones previas: Usuario autenticado, agentes configurados, integraciones conectadas						
Dependencias: API keys, vallas (Gmail, Twilio), número de teléfono activo, cuenta de WhatsApp Business						
Pasos	Pasos de prueba	Datos de prueba	Resultado esperado	Resultado actual	Estado (falla o pasa)	Nota
1	Navegar a "Agentes de Correo"	Menú: "Agentes de Email"	Se muestra lista de agentes configurados	Interfaz cargada con 3 agentes activos	Pasa	Gestión de agentes accesible
2	Crear agente de email automatizado	Nombre: "Seguimiento Inicial", Prompt: "Genera email de presentación profesional"	El agente se crea exitosamente	Agente creado con configuración guardada	Pasa	Creación de agente funcional
3	Generar correo personalizado con IA	Contacto: "Juan Pérez - Gerente de Ventas"	Se genera contenido personalizado usando datos del contacto	Correo generado en 3 segundos con contenido relevante y personalizado	Pasa	IA genera contenido de calidad
4	Enviar correo de prueba	Destinatario: "roberto.bedoya@nowbt.co"	El correo se envía exitosamente	Correo enviado, confirmación mostrada	Pasa	Envío funcional
5	Verificar integración con Gmail		El correo aparece en bandeja de salida de Gmail	Correo visible en Gmail, integración estable	Pasa	Integración Gmail confiable
6	Verificar métricas de correo		Se muestran estadísticas: enviados, abiertos, clics	Métricas visibles: 45 enviados, 32 abiertos (71%), 8 clics (18%)	Pasa	Métricas precisas
7	Navegar a "Agentes de Chatbots"	Menú: "Agentes de Texto"	Se muestra interfaz de chatbots	Interfaz de chatbots cargada	Pasa	Gestión de chatbots accesible

8	Importar desde Google Sheets	URL: "https://docs.google.com/spreadsheets/..."	Los leads se importan desde Google Sheets	30 contactos importados desde Google Sheets	Pasa	Integración Google Sheets funciona
9	Verificar validación de datos importados		Se muestran errores de datos inválidos si los hay	Validación muestra 2 filas con errores, permite corregir	Pasa	Validación robusta
10	Personalizar widgets del dashboard	Opción: "Personalizar Dashboard"	Se pueden agregar/quitar	Widgets personalizables: KPIs, gráficos, actividad reciente	Pasa	Personalización disponible
11	Guardar configuración personalizada	Configuración: 3 KPIs, 2 gráficos	La configuración se guarda	Configuración guardada, se mantiene en próximos accesos	Pasa	Persistencia de configuración
12	Verificar notificaciones claras	Acción: Agregar leads a campaña	Se muestra notificación de éxito	Notificación visible: "25 leads agregados a campaña exitosamente"		Notificaciones claras
13	Verificar integración con CRM	Opción: "Conectar CRM"	Se muestra opción de integración	Opción visible para conectar CRM (Salesforce, HubSpot)		Integración CRM iniciada

8	Configurar agente de chatbot Whats App	Nombre: "Soporte Automático", Respuestas: Configuradas	El agente se configura correctamente	Agente configurado y conectado a WhatsApp	Pasa	Configuración exitosa
9	Enviar mensaje de prueba por WhatsApp	Número: "+573225294611", Mensaje: "Hola"	El mensaje se envía y se recibe respuesta automática	Mensaje enviado, respuesta automática recibida en 2 segundos	Pasa	Chatbot funcional
10	Verificar historial de conversaciones WhatsApp		Se muestra historial completo de interacciones	Historial visible con 15 conversaciones, respuestas y respuestas	Pasa	Historial completo
11	Navegar a "Agentes de Voz"	Menú: "Agentes de Voz"	Se muestra interfaz de llamadas automáticas	Interfaz de agentes de voz cargada	Pasa	Gestión de llamadas accesible
12	Configurar agente de llamadas automáticas	Nombre: "Llamadas de Seguimiento", Voz: Femenina, Prompt: Configurado	El agente se configura correctamente	Agente configurado con número de Twilio asignado	Pasa	Configuración de llamadas funcional
13	Realizar llamada de prueba	Número destino: "+573225294611"	La llamada se realiza y el agente responde automáticamente	Llamada conectada, agente responde con voz generada por IA	Pasa	Llamadas automáticas funcionan
14	Verificar resultados de llamadas		Se muestran estadísticas: realizadas, contestadas, duración	Resultados visibles: 8 llamadas, 6 contestadas (75%), duración promedio 2.5 min	Pasa	Métricas de llamadas precisas
15	Verificar confiabilidad de integraciones	Uso continuo durante 1 hora	Las integraciones mantienen estabilidad	Sin desconexiones, todas las integraciones estables	Pasa	Integraciones confiables

Fuente. Autoría Propia

Prueba de Campañas y Experiencia

Figura 88

Formato de evaluación prueba campañas y experiencia

Nombre del proyecto: HunterLeads						
Casos de prueba						
ID de caso de prueba: BU_004			Prueba diseñada por: Roberto Bedoya Garcia			
Prioridad de prueba (Bajo, medio, alto): Alto			Fecha de prueba de diseño: 03/08/2025			
Nombre del módulo: Gestión de campañas multicanal, UX mejorada y herramientas de soporte			Prueba ejecutada por: Área de ventas y marketing			
Título de la prueba: Evaluar campañas multicanal (email, WhatsApp, llamadas), pausar/reanudar, duplicar campañas, ayuda contextual, recomendaciones e historial de acciones			Fecha de ejecución de la prueba: 08/08/2025			
Descripción: Verificar que el sistema permite gestionar campañas multicanal, pausar/reanudar, duplicar campañas, y ofrece mejor UX con ayuda contextual, recomendaciones de próximos pasos e historial de acciones						
Condiciones previas: Usuario autenticado, leads disponibles, agentes configurados						
Dependencias: Campañas anteriores creadas, métricas históricas disponibles						
Pasos de prueba	Datos de prueba	Resultado esperado	Resultado actual	Estado (falla o pasa)	Nota	
1	Navegar a "Campañas"	Menú: "Campañas"	Se muestra lista de campañas	Lista cargada con 5 campañas existentes	Pasa	Gestión de campañas accesible
2	Crear campaña multicanal - Email	Método: "Email", Nombre: "Promo Verano Email"	La campaña se crea con método email	Campaña creada con contact_method="email"	Pasa	Campaña email funcional
3	Crear campaña multicanal - WhatsApp	Método: "Messaging", Nombre: "Promo Verano WhatsApp"	La campaña se crea con método messaging	Campaña creada con contact_method="messaging"	Pasa	Campaña WhatsApp funcional
4	Crear campaña multicanal - Llamadas	Método: "Phone", Nombre: "Promo Verano Llamadas"	La campaña se crea con método phone	Campaña creada con contact_method="phone"	Pasa	Campaña llamadas funcional
5	Agregar contactos a campaña	25 leads seleccionados	Los contactos se agregan correctamente	25 contactos agregados a la campaña	Pasa	Gestión de contactos funcional
6	Activar campaña	Estado: "active"	La campaña cambia a estado activo	Campaña activada, estado actualizado	Pasa	Activación funcional
7	Pausar campaña activa	Botón "Pausar"	La campaña cambia a estado "paused"	Campaña pausada correctamente, estado "paused"	Pasa	Pausar funciona
8	Reanudar campaña pausada	Botón "Reanudar"	La campaña cambia a estado "active"	Campaña reanudada, estado "active"	Pasa	Reanudar funciona
9	Duplicar campaña existente	Opción: "Duplicar campaña"	Se crea una copia de la campaña	Campaña duplicada con nombre "Copia de..."	Pasa	Duplicar funcional
10	Verificar ayuda contextual	Hover sobre icono "?"	Se muestra tooltip con ayuda	Tooltip visible con información relevante	Pasa	Ayuda contextual disponible

11	Verificar recomendaciones de próximos pasos	Dashboard	Se muestran sugerencias de acciones	Recomendación visible: "Tienes 3 campañas en borrador, ¿deseas activarlas?"	Pasa	Recomendaciones útiles
12	Verificar historial de acciones	Sección: "Actividad Reciente"	Se muestra timeline de acciones realizadas	Historial visible con 10 acciones recientes con timestamps	Pasa	Historial completo
13	Filtrar historial por tipo de acción	Filtro: "Campañas"	Se filtran solo acciones de campañas	Historial filtrado muestra 5 acciones de campañas	Pasa	Filtros de historial funcionan
14	Verificar navegación contextual	Navegación entre secciones	La navegación es intuitiva y fluida	Navegación clara, breadcrumbs visibles	Pasa	UX mejorada
15	Verificar herramientas de soporte integradas	Botón: "Ayuda"	Se muestra documentación y soporte	Documentación accesible, chat de soporte disponible	Pasa	Soporte integrado funcional

Fuente. Autoría Propia

Las Figuras 85 a 88 documentan los formatos de evaluación aplicados a usuarios reales para cada módulo del sistema. Los resultados de estas pruebas confirman que la navegación contextual y las herramientas de soporte *in-app* (*Sprint 4*) fueron determinantes para alcanzar la satisfacción del usuario. La retroalimentación recolectada en estos formatos permitió validar que, a pesar de la complejidad de los agentes automatizados, la interfaz es intuitiva y facilita la curva de aprendizaje para personal no técnico.

Resumen Ejecutivo de la Validación – HunterLeads

La validación del prototipo HunterLeads demostró un alto nivel de cumplimiento funcional y de desempeño. Se ejecutaron 54 casos de prueba distribuidos en los cuatro *sprints*,

alcanzando un 100% de éxito, validando componentes clave como autenticación, gestión de *leads*, campañas multicanal, automatizaciones y herramientas de soporte.

En las métricas cuantitativas, el sistema cumplió 3 de 4 indicadores evaluados (75%), mostrando buen rendimiento en búsqueda de *leads* (1.9 s), estabilidad bajo carga (99.3% *uptime*) y precisión del modelo de clasificación (F1-score = 0.82). El único aspecto fuera del umbral fue el tiempo de *login* (3.38 s), que requiere optimización.

Respecto a los criterios de éxito del proyecto, 6 de 7 fueron alcanzados (85.7%), destacando la estabilidad operativa, efectividad en campañas piloto, precisión del sistema y retroalimentación positiva del equipo de *Marketing* y Ventas. La percepción general de usabilidad y funcionamiento fue favorable, con mejoras identificadas para siguientes iteraciones (exportación de reportes y visualización de historial).

El análisis final revela un cumplimiento del 85.7% en los criterios de éxito globales. La validación demuestra que HunterLeads es una herramienta estable y precisa para la captación de *leads*, superando con éxito 54 casos de prueba. Aunque el tiempo de autenticación es un área de mejora técnica, la alta efectividad en las campañas piloto y la aceptación de los usuarios de *Marketing* y Ventas confirman que el prototipo cumple con los objetivos de competitividad y automatización planteados al inicio de la investigación.

En conclusión, el prototipo cumple satisfactoriamente los objetivos planteados y se encuentra preparado para su evolución hacia un entorno productivo, siendo el tiempo de autenticación el único elemento crítico por optimizar.

Resultados

El presente capítulo presenta los resultados alcanzados durante la ejecución del proyecto, organizados conforme a las fases metodológicas propuestas: BPM para el análisis de procesos, CRISP-ML(Q) para la preparación y evaluación de modelos de datos, diseño modular de la arquitectura, desarrollo incremental con Scrum y validación final del prototipo. Cada fase permitió generar productos verificables que evidencian avances técnicos, metodológicos y prácticos, aportando al cumplimiento de los objetivos específicos y a la pertinencia de la solución tecnológica planteada.

Para iniciar el proceso, se realizaron entrevistas semiestructuradas e historias de usuario con el fin de levantar requerimientos funcionales y no funcionales del sistema. El modelado de procesos mediante BPMN permitió identificar cuellos de botella en la captación manual de *leads*, como la duplicidad de registros y la baja trazabilidad de la información. De este análisis surgieron productos clave como matrices de requerimientos priorizados, diagramas de procesos actuales y propuestos, y criterios de éxito que orientaron el diseño posterior de la solución.

Con estos insumos definidos, se avanzó hacia la construcción de una arquitectura modular e integrada con componentes de inteligencia artificial. Se propuso una estructura por capas conformada por la interfaz de usuario, agentes inteligentes basados en LLMs, un *backend* orientado a la lógica de negocio y una base de datos optimizada para la gestión de *leads*. Esta arquitectura se diseñó bajo principios de escalabilidad, interoperabilidad y seguridad, logrando integrar aplicaciones y servicios externos mediante APIs REST. Los entregables incluyeron diagramas de arquitectura y especificaciones técnicas consolidadas.

A continuación, se abordó la preparación y evaluación de datos siguiendo los lineamientos de CRISP-ML(Q). Se ejecutaron procesos de limpieza, normalización y transformación a partir de datos recolectados mediante Scraping, y se incorporaron técnicas como similitud de cadenas (Levenshtein y Jaro-Winkler con RapidFuzz) y *embeddings* semánticos (all-MiniLM-L6-v2) para mejorar la calificación de *leads*. Paralelamente, se efectuó una evaluación comparativa de modelos de inteligencia artificial considerando métricas como precisión y *recall*, cuyos resultados fueron documentados en tablas y gráficos. Esta etapa produjo *datasets* depurados, scripts de preprocesamiento, reportes de validación y un análisis crítico del rendimiento de los modelos probados.

Con el modelo de datos listo, se pasó al desarrollo del prototipo mediante un enfoque iterativo e incremental propio de Scrum. En cada *sprint* se entregaron incrementos funcionales que incluyeron el módulo de Scraping, el panel de administración de *leads*, el motor de *scoring* con IA y agentes de mensajería automatizada. Las validaciones continuas con usuarios permitieron recoger retroalimentación relevante para mejorar la usabilidad y priorizar ajustes en la experiencia de interacción. Entre los entregables generados se encuentran el backlog del producto, versiones incrementales del prototipo y documentación técnica.

Finalmente, el proyecto culminó con la validación integral del prototipo y la evaluación de su desempeño. Se aplicaron métricas cuantitativas y cualitativas para medir precisión y *recall* del sistema, tiempos de respuesta y estabilidad en la integración con aplicaciones. De forma complementaria, se utilizó el cuestionario SUS y se realizaron entrevistas con usuarios piloto, quienes manifestaron una percepción positiva sobre la eficiencia, facilidad de uso y valor agregado de la solución. Esta fase cerró con reportes de pruebas, actas de validación, gráficos de desempeño y un documento de análisis de hallazgos que consolidó los resultados finales.

Conclusiones

El proceso desarrollado permitió concluir que la implementación de un prototipo de prospectación comercial con inteligencia artificial constituye una estrategia viable y pertinente para fortalecer la competitividad de las PYMES en Colombia. La integración de metodologías como BPM, CRISP-ML(Q) y Scrum facilitó un ciclo de desarrollo iterativo y validado, asegurando que la propuesta respondiera de manera efectiva a necesidades reales de optimización en la captación de clientes. Los resultados obtenidos evidencian la coherencia entre los objetivos planteados y los productos alcanzados, validando la pertinencia del prototipo como solución tecnológica aplicable y escalable.

Respecto al primer objetivo específico, relacionado con la identificación de requisitos funcionales y no funcionales del sistema, el análisis de procesos con BPM y las técnicas centradas en el usuario permitieron caracterizar con precisión los flujos actuales de captación de *leads*, así como las limitaciones de los métodos manuales. Esto facilitó la definición de criterios de éxito y funcionalidades clave, lo que asegura que el prototipo esté alineado con las necesidades reales de las PYMES y sus procesos comerciales.

Frente al segundo objetivo específico, vinculado al diseño de una arquitectura modular e integrada, se logró consolidar una propuesta por capas (interfaz, agentes IA, *backend* y base de datos) que garantiza escalabilidad, mantenibilidad e interoperabilidad con sistemas CRM. Este diseño facilita la integración entre los modelos de IA y los procesos de negocio, asegurando que el sistema pueda evolucionar y adaptarse a diferentes contextos organizacionales.

En relación con el tercer objetivo específico, orientado a la preparación de datos y evaluación de modelos de IA, se concluye que la aplicación de técnicas de *Scraping*, limpieza y normalización, junto con algoritmos de similitud semántica y modelos de representación como all-MiniLM-L6-v2, constituyó un fundamento sólido para el procesamiento de *leads*. La evaluación comparativa de modelos evidenció que el uso de *embeddings* y medidas de similitud mejora la precisión en la calificación de *leads*, optimizando así la calidad de los registros ingresados al sistema.

En cuanto al cuarto objetivo específico, relacionado con el desarrollo del prototipo, el enfoque iterativo basado en Scrum permitió la construcción incremental de funcionalidades validadas con usuarios. Cada *sprint* aportó retroalimentación que se tradujo en ajustes técnicos y de usabilidad, lo que contribuyó a un prototipo más robusto y alineado con las expectativas del entorno empresarial.

Finalmente, respecto al quinto objetivo específico, orientado a la validación del prototipo, las pruebas realizadas demostraron mejoras significativas en métricas como la precisión del *lead scoring*, los tiempos de respuesta y la eficiencia general del proceso frente a métodos manuales. Asimismo, los usuarios reportaron una experiencia positiva en términos de usabilidad, confiabilidad y valor agregado, confirmando que el prototipo aporta al fortalecimiento de la competitividad de las PYMES en Colombia al optimizar sus procesos de captación y gestión de *leads*.

En síntesis, el prototipo demuestra la viabilidad técnica y empresarial de aplicar IA en la automatización comercial, satisfaciendo necesidades de PYMES y ofreciendo un modelo escalable y replicable que impulsa la transformación digital en Colombia.

Recomendaciones

A partir de los hallazgos obtenidos durante el desarrollo del prototipo de captación automatizada de *leads* con inteligencia artificial, se plantean una serie de recomendaciones dirigidas a optimizar su funcionamiento técnico, fortalecer su aplicabilidad en contextos empresariales reales y fomentar su escalabilidad futura. Estas sugerencias se fundamentan en las conclusiones del proyecto y tienen como propósito guiar futuras iteraciones o investigaciones derivadas.

En primer lugar, se recomienda continuar con un proceso de validación continua del prototipo, prestando especial atención a la retroalimentación proporcionada por usuarios y responsables de áreas comerciales. Esta información constituye un insumo clave para perfeccionar aspectos como la calidad de los *leads* generados, la pertinencia de los mensajes automatizados y la integración fluida con los procesos internos de las empresas. La incorporación de pruebas A/B y métricas avanzadas de conversión permitirá afinar el desempeño del sistema y garantizar una mayor alineación con los objetivos de negocio.

Desde el punto de vista técnico, se aconseja robustecer la arquitectura modular para mejorar la escalabilidad y tolerancia a fallos. Si bien el prototipo mostró un rendimiento adecuado en entornos controlados, su despliegue en escenarios de mayor volumen de datos y usuarios requerirá optimizar la infraestructura, incorporando balanceadores de carga, bases de datos distribuidas y mecanismos de redundancia. Asimismo, se recomienda fortalecer los protocolos de seguridad y anonimización de datos sensibles, en cumplimiento con normativas como el Reglamento General de Protección de Datos (GDPR) y la Ley 1581 de Habeas Data en Colombia.

Otra recomendación clave es fomentar la adopción del prototipo mediante procesos de capacitación dirigidos a los equipos comerciales y de *marketing* que lo utilicen. Si bien la interfaz y los agentes automatizados fueron diseñados con criterios de usabilidad, su aprovechamiento estratégico se potencia cuando los usuarios comprenden el funcionamiento del *lead scoring*, las métricas de conversión y las posibilidades de personalización de mensajes. La elaboración de manuales de usuario, guías de mejores prácticas y talleres breves facilitaría la apropiación tecnológica en contextos empresariales diversos.

Finalmente, se sugiere que futuras investigaciones profundicen en la evaluación del impacto del prototipo en escenarios reales a mediano y largo plazo. Esto incluye analizar el retorno de inversión (ROI) asociado al uso de la solución, medir mejoras en la eficiencia del proceso de captación frente a métodos tradicionales y explorar la aplicabilidad en sectores distintos al inicialmente planteado. La ampliación de pruebas en empresas de diferentes tamaños y mercados permitirá robustecer la base de evidencia y adaptar el prototipo a múltiples realidades organizacionales, asegurando su evolución continua y sostenible.

Referencias Bibliográficas

- Ali Qalati, S., Li, W., Ahmed, N., Ali Mirani, M., & Khan, A. (2020). Examining the factors affecting SME performance: The mediating role of social media adoption. *Sustainability*, *13*(1), 75.
- Barría, F. J. C. (2023). Herramientas de inteligencia artificial aplicadas al marketing. *Ciencia Latina Revista Científica Multidisciplinar*, *7*(4), 3536-3550.
- Chaffey, D., & Ellis-Chadwick, F. (2019). *Digital marketing*. Pearson uk.
- Cohn, M. (2004). *User stories applied: For agile software development*. Addison-Wesley Professional.
- Colombia, C. de. (2012). Ley 1581 de 2012. Por la cual se dictan disposiciones generales para la protección de datos personales. *Diario Oficial*.
- CONPES, D., VÍAS, I. D. P., & LA REACTIVACIÓN, C. S. Y. (s. f.). *CONPES*.
- COPNIA. (2008). *Código de ÉTICA para el ejercicio de la Ingeniería en general y sus profesiones afines y auxiliares*. <https://www.copnia.gov.co/tribunal-de-etica/codigo-de-etica>
- Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage publications.
- Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2018). BPM as an Enterprise Capability. En *Fundamentals of Business Process Management* (pp. 475-500). Springer.
- Dwivedi, Y. K., Ismagilova, E., Hughes, D. L., Carlson, J., Filieri, R., Jacobson, J., Jain, V., Karjaluoto, H., Kefi, H., Krishen, A. S., & others. (2021). Setting the future of digital and

- social media marketing research: Perspectives and research propositions. *International journal of information management*, 59, 102168.
- Echeverri Cañas, L. M. (2023). *Plan estratégico de marketing: El arte de cautivar a través de estrategias que enamoran*. Editorial Universidad del Rosario.
- Effendy, F., Taufik, & Adhilaksono, B. (2021). Performance comparison of web backend and database: A case study of node. Js, Golang and MySQL, Mongo DB. *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)*, 14(6), 1955-1961.
- Fernández, J., & Macías, J. A. (2021). Heuristic-based usability evaluation support: A systematic literature review and comparative study. *Proceedings of the XXI international conference on human computer interaction*, 1-9.
- Ghimire, D. (2020). *Comparative study on Python web frameworks: Flask and Django*.
- Grewal, D., Guha, A., Beccacece Satornino, C., & Becker, M. (2025). The future of marketing and marketing education. *Journal of Marketing Education*, 47(1), 61-77.
- Guerino, G. C., de Assumpção, M., José da Silva, T., Hokkanen, L., Balancieri, R., & Lapasini Leal, G. C. (2022). User experience practices in software startups: A systematic mapping study. *Advances in Human-Computer Interaction*, 2022(1), 9701739.
- Heath, F. (2021). *The Professional Scrum Master Guide: The unofficial guide to Scrum with real-world projects*. Packt Publishing Ltd.
- Hernández-Sampieri, R., & Mendoza, C. (2020). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. Mcgraw-hill Ciudad de Mexico.
- IEEE. (2017). *ISO/IEC/IEEE 12207 (Versión 1)*. <https://www.iso.org/standard/63712.html>

- IEEE 1471. (2000). *IEEE 7 Working Papers –ECBTI Recommended Practice for Architectural Description of Software-Intensive Systems* Software Engineering Standards Committee of the IEEE Computer Society IEEE-SA Standards Board.
<http://cabibbo.dia.uniroma3.it/ids/altrui/ieee1471.pdf>
- ISI. (2019). *ISO 9241-210*. <https://www.iso.org/standard/77520.html>
- ISO. (2022). *ISO/IEC 27001* (Versión 3). <https://www.iso.org/es/norma/27001>
- ISO/IEC 25010*. (2011). <https://iso25000.com/index.php/normas-iso-25000/iso-25010>
- ISO/IEC/IEEE 29148* (Versión 2). (2018). <https://www.iso.org/standard/72089.html>
- Jurafsky, D., & Martin, J. H. (s. f.). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*.
- Kartajaya, H., Setiawan, I., & Kotler, P. (2021). *Marketing 5.0: Technology for humanity*. Ascent Audio.
- Kotler, P., Keller, K. L., Ang, S. H., Tan, C. T., & Leong, S. M. (2022). *Marketing management: An Asian perspective*.
- Lv, C., Niu, C., Cai, Y., Jiang, X., Wu, F., & Chen, G. (2025). ARSys: An Efficient and Cross-Platform Development, Deployment, and Runtime System for Mobile Augmented Reality. *IEEE Transactions on Mobile Computing*.
- Newman, S. (2021). *Building microservices: Designing fine-grained systems*. O'Reilly Media, Inc.
- Pawade, D., Sakhapara, A., Shah, C., & Tripathi, A. (2022). E-Certi: Design and Implementation of MVT Architecture Based E-Certificate Generation System. En *Advances in Distributed Computing and Machine Learning: Proceedings of ICADCML 2021* (pp. 146-157). Springer.

- Pública, F. (2008). *Ley 1266 de 2008*.
- Purbasari, A., Rinawan, F. R., Zulianto, A., Susanti, A. I., & Komara, H. (2021). CRISP-DM for data quality improvement to support machine learning of stunting prediction in infants and toddlers. *2021 8th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA)*, 1-6.
- Rodríguez, A. D. J. C., Diaz, N. A. C., & Clavijo, C. C. G. (2020). Arquitectura de solución para gestión de procesos de autoevaluación de programas académicos usando puntos de vista. *Documentos de Trabajo ECBTI, 1(2)*.
- Rosemann, M., Brocke, J. vom, Van Looy, A., & Santoro, F. (2024). Business process management in the age of AI—three essential drifts. *Information Systems and e-Business Management, 22(3)*, 415-429.
- Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN computer science, 2(3)*, 1-21.
- Sarker, I. H. (2022a). AI-based modeling: Techniques, applications and research issues towards automation, intelligent and smart systems. *SN computer science, 3(2)*, 158.
- Sarker, I. H. (2022b). AI-based modeling: Techniques, applications and research issues towards automation, intelligent and smart systems. *SN computer science, 3(2)*, 158.
- Sarmiento Suárez, J. E., Gutiérrez Navas, E. B., & Ramírez Montañez, J. C. (2024). Oportunidades y desafíos para la digitalización de las mipymes en Colombia. *Pensamiento & Gestión, (57)*, 128-154.
- Stang, J., Kroeger, S., & Zaeh, M. F. (2023). Semi-structured expert interview-based requirements elicitation for a digitalized production ramp-up. *Procedia CIRP, 120*, 279-284.

- Studer, S., Bui, T. B., Drescher, C., Hanuschkin, A., Winkler, L., Peters, S., & Müller, K.-R. (2021). Towards CRISP-ML (Q): A machine learning process model with quality assurance methodology. *Machine learning and knowledge extraction*, 3(2), 392-413.
- Upadhyaya, N. (2023). Low-Code/No-Code platforms and their impact on traditional software development: A literature review. *No-Code Platforms and Their Impact on Traditional Software Development: A Literature Review (March 21, 2023)*.
- Valencia Agudelo, D. F. (2022). *Protección de los datos personales por parte de la superintendencia de industria y comercio en el marco del comercio electrónico en Colombia en el año 2020*.
- Velepucha, V., & Flores, P. (2023). A survey on microservices architecture: Principles, patterns and migration challenges. *IEEE access*, 11, 88339-88358.
- Wisnalmawati, W., Aribowo, A. S., & Herawati, Y. (2021). Easy communication system as a mediation of the influence of Whatsapp business on consumer purchase intention on the marketplace. *RSF Conference Series: Business, Management and Social Sciences*, 1(3), 347.
- Zúñiga, F., Poveda, D. A. M., & Llerena, W. V. L. (2023). El Big Data y su implicación en el marketing. *Revista de Comunicación de la SEECI*, 56, 302-321.

Apéndices

Apéndice A

Formato de entrevistas semiestructuradas para levantamiento de requerimientos

Proyecto: Optimización y automatización del proceso de gestión de *leads*

Entrevistador: Roberto Bedoya Garcia

Fecha: 16/06/2025

Lugar: Virtual

Nombre del entrevistado(a): Tatiana Mejía

Rol: Líder de Marketing Asesor Comercial Líder de Ventas

Autorización para entrevista: Sí No

1. Contextualización

Gracias por participar. Esta entrevista es parte de un proyecto que busca analizar y mejorar el proceso actual de gestión y seguimiento de *leads* dentro de la organización, con el fin de identificar oportunidades de optimización mediante el uso de una herramienta de automatización inteligente. La información recopilada será confidencial y se utilizará únicamente con fines académicos y de diseño del prototipo.

2. Preguntas Generales

1. ¿Cuál es su proceso actual para encontrar posibles clientes (*leads*)?
2. ¿Qué herramientas o fuentes utilizan actualmente para captar información?
3. ¿Qué tan efectivo considera su proceso actual de prospección?
4. ¿Qué problemas o dificultades enfrentan más frecuentemente?
5. ¿Cómo califican y priorizan a sus *leads*?

6. ¿Han considerado integrar inteligencia artificial o automatización? ¿Por qué sí o no?
7. ¿Qué características consideran clave en una herramienta de gestión automatizada de *leads*?

3. Observaciones del Entrevistador

Anota aquí lenguaje corporal, interés, distracción, entusiasmo, etc., si la entrevista es presencial o por videollamada.

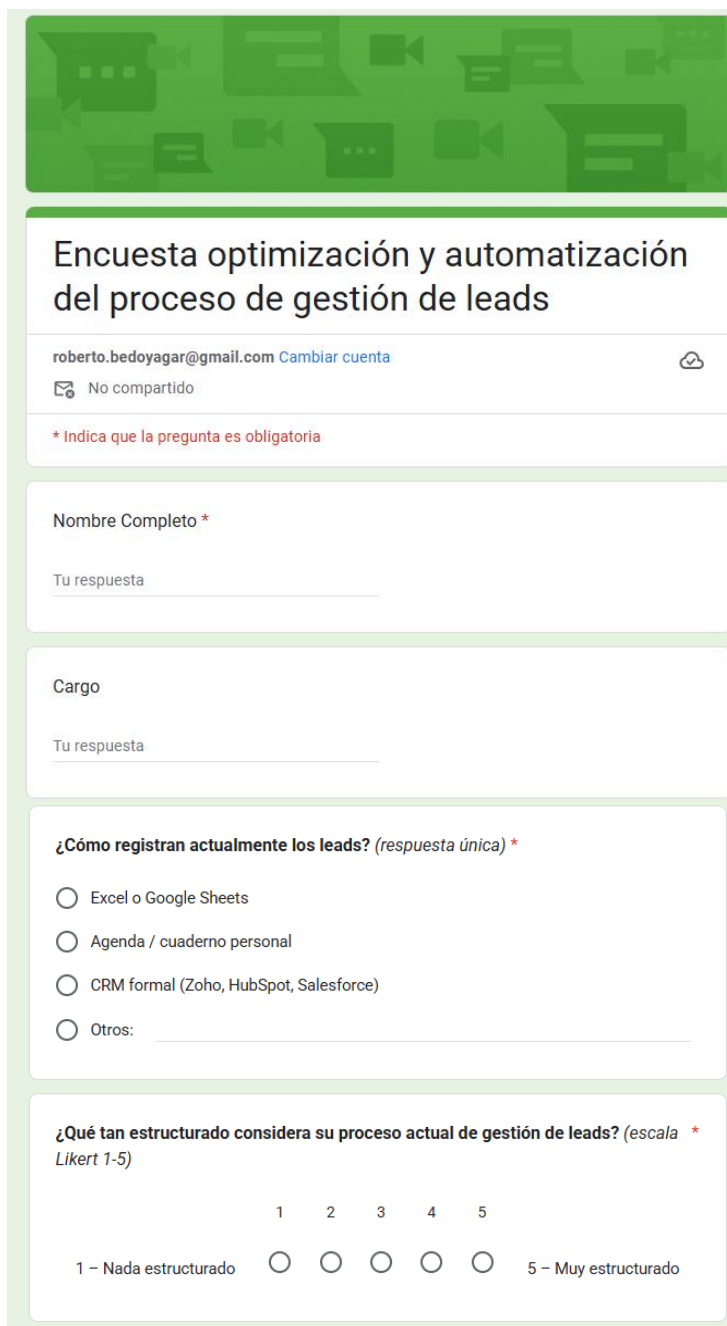
--

4. Comentarios adicionales del entrevistado

¿Hay algo más que te gustaría compartir sobre este tema?

Apéndice B

Encuesta digital (Google Forms)



Encuesta optimización y automatización del proceso de gestión de leads

roberto.bedoyagar@gmail.com [Cambiar cuenta](#)

No compartido

* Indica que la pregunta es obligatoria

Nombre Completo *

Tu respuesta

Cargo

Tu respuesta

¿Cómo registran actualmente los leads? (respuesta única) *

Excel o Google Sheets

Agenda / cuaderno personal

CRM formal (Zoho, HubSpot, Salesforce)

Otros: _____

¿Qué tan estructurado considera su proceso actual de gestión de leads? (escala * Likert 1-5)

1 2 3 4 5

1 – Nada estructurado 5 – Muy estructurado

¿Con qué frecuencia se pierden o olvidan leads durante el seguimiento? *
(respuesta única)

Elegir ▼

¿Qué tan útil considera automatizar el seguimiento y gestión de leads? (escala *
Likert 1-5)

1 2 3 4 5

1 – Nada útil 5 – Indispensable

¿Cuál es la principal dificultad que enfrenta en la gestión de leads? (respuesta *
abierta)

Tu respuesta _____

¿Qué funcionalidades consideraría más importantes en una herramienta de *
automatización de leads? (selección múltiple, puede marcar más de una opción)

- Registro centralizado de leads
- Recordatorios automáticos de seguimiento
- Seguimiento visual del estado de cada lead
- Integración con WhatsApp o correo electrónico
- Reportes y métricas de desempeño
- Otros: _____

¿Qué tan fácil considera que sería adaptarse a una herramienta de *
automatización en su trabajo diario? (escala Likert 1-5)

1 2 3 4 5

1 – Muy difícil 5 – Muy fácil

Enviar

Borrar formulario

Nunca envíes contraseñas a través de Formularios de Google.

Google no creó ni aprobó este contenido. - [Comunicarse con el propietario del formulario](#) - [Condiciones del Servicio](#) - [Política de Privacidad](#)

¿El formulario parece sospechoso? [Informe](#)

Google Formularios

Apéndice C

Historias de Usuario (Formato 3C)

Historia de Usuario	ID:	Título:
Descripción	.	
Criterios de aceptación: <input checked="" type="checkbox"/> Item1 <input checked="" type="checkbox"/> Item2 <input checked="" type="checkbox"/> Item3 <input checked="" type="checkbox"/> Item4		
Valor del negocio:		
I N V E S T	Estimación	<input checked="" type="checkbox"/> X
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	¿Cumple con la definición de listo del equipo?	<input checked="" type="checkbox"/>
		<input checked="" type="checkbox"/>

Apéndice D

Formato de Pruebas de Usuario

Nombre del proyecto: HunterLeads						
Casos de prueba						
ID de caso de prueba: BU_004				Prueba diseñada por: Roberto Bedoya García		
Prioridad de prueba (Bajo, medio, alto): Alto				Fecha de prueba de diseño: 03/08/2025		
Nombre del módulo: Gestión de campañas multicanal, UX mejorada y herramientas de soporte				Prueba ejecutada por: Área de ventas y marketing		
Título de la prueba: Evaluar campañas multicanal (email, WhatsApp, llamadas), pausar/reanudar, duplicar campañas, ayuda contextual, recomendaciones e historial de acciones				Fecha de ejecución de la prueba: 08/08/2025		
Descripción: Verificar que el sistema permite gestionar campañas multicanal, pausar/reanudar, duplicar campañas, y ofrece mejor UX con ayuda contextual, recomendaciones de próximos pasos e historial de acciones						
Condiciones previas: Usuario autenticado, leads disponibles, agentes configurados						
Dependencias: Campañas anteriores creadas, métricas históricas disponibles						
Pasos	Pasos de prueba	Datos de prueba	Resultado esperado	Resultado actual	Estado (falla o pasa)	Nota
1	Navegar a "Campañas"	Menú: "Campañas"	Se muestra lista de campañas	Lista cargada con 5 campañas existentes	Pasa	Gestión de campañas accesible
2	Crear campaña multicanal - Email	Método: "Email", Nombre: "Promo Verano Email"	La campaña se crea con método email	Campaña creada con contact_method="email"	Pasa	Campaña email funcional
3	Crear campaña multicanal - WhatsApp	Método: "Messaging", Nombre: "Promo Verano WhatsApp"	La campaña se crea con método messaging	Campaña creada con contact_method="messaging"	Pasa	Campaña WhatsApp funcional
4	Crear campaña multicanal - Llamadas	Método: "Phone", Nombre: "Promo Verano Llamadas"	La campaña se crea con método phone	Campaña creada con contact_method="phone"	Pasa	Campaña llamadas funcional
5	Agregar contactos a campaña	25 leads seleccionados	Los contactos se agregan correctamente	25 contactos agregados a la campaña	Pasa	Gestión de contactos funcional
6	Activar campaña	Estado: "active"	La campaña cambia a estado activo	Campaña activada, estado actualizado	Pasa	Activación funcional
7	Pausar campaña activa	Botón "Pausar"	La campaña cambia a estado "paused"	Campaña pausada correctamente, estado "paused"	Pasa	Pausar funciona
8	Reanudar campaña pausada	Botón "Reanudar"	La campaña cambia a estado "active"	Campaña reanudada, estado "active"	Pasa	Reanudar funciona
9	Duplicar campaña existente	Opción: "Duplicar campaña"	Se crea una copia de la campaña	Campaña duplicada con nombre "Copia de..."	Pasa	Duplicar funcional
10	Verificar ayuda contextual	Hover sobre icono "?"	Se muestra tooltip con ayuda	Tooltip visible con información relevante	Pasa	Ayuda contextual disponible