

# **Implementación de ciencia de datos en MiPymes con déficit de información en Colombia**

Manuel Danilo Camargo Figueredo

Asesor

Luis Angel Anillo Arrieta

Universidad Nacional Abierta y a Distancia UNAD

Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI

Ciencia de Datos y Analítica

2024

## **Agradecimientos**

Quiero expresar mi más profundo agradecimiento a todas las personas e institución que contribuyeron a la realización de este trabajo de grado.

En primer lugar, a la Universidad Nacional Abierta y a Distancia (UNAD), por brindarme una formación académica de calidad por medio de los contenidos de los cursos realizados durante la especialización, especialmente a los profesores de las materias de Proyecto de Grado I y II, quienes con su orientación y conocimiento me permitieron llevar este proyecto a buen término. A mi asesor, Luis Ángel Anillo Arrieta, gracias por su valiosa guía, disponibilidad, paciencia y constante apoyo a lo largo de este proceso. Sus sugerencias y recomendaciones fueron fundamentales para el logro de los objetivos planteados.

A Michael D'chardi Hernández, CEO de Verde Mediterráneo y sus colaboradores, por la oportunidad de aplicar este proyecto en su empresa, y por creer en la capacidad transformadora de la ciencia de datos en las MiPymes rurales. Su colaboración fue crucial para la conexión entre la teoría y la práctica, y por su confianza en la implementación de las ideas en el campo empresarial, permitiendo que la ciencia de datos transforme la realidad una MiPyme en área rural de Cundinamarca, y por ello estoy profundamente agradecido.

Finalmente, a mi familia y amigos, por su apoyo incondicional, motivación y comprensión durante todo este proceso. Sin ustedes, este logro no hubiera sido posible.

## Resumen

Dentro de las opciones de grado para la especialización en ciencia de datos, para el desarrollo del presente trabajo, se escoge la opción de proyecto aplicado, el cual permite el diseño de proyectos para una transferencia social de conocimiento que contribuya de manera innovativa a la solución de problemas focalizados, en este caso específicamente en el desarrollo tecnológico, la cual tiene que ver con la línea de investigación de la cadena de formación en sistemas. (UNAD, Unad, 2020)

La temática identificada como necesidad o problemática la relaciono como opción de grado un proyecto aplicado específicamente en el desarrollo tecnológico, la cual tiene que ver con la línea de investigación de la cadena de formación en sistemas, ya que la temática seleccionada tendría que ver en un principio con el levantamiento de información, reconocimiento de los procesos de una empresa a la cual se le haya identificado un déficit en la captura de información y que mediante la implementación de herramientas existentes en el mercado, se permita registrar, almacenar, procesar y visualizar información para la toma de decisiones en alguno de los procesos dentro de una empresa.

**Palabras claves:** Datos, Análisis, Inteligencia de Negocios, MiPyme, Transformación Digital.

## **Abstract**

As part of the degree options for the Data Science specialization, this project adopts the applied project option, which enables the design of projects aimed at the social transfer of knowledge. This approach aims to innovatively address specific problems, in this case, focusing on technological development, aligned with the research line of the systems training pathway. (UNAD, Unad, 2020)

The identified issue or need is linked to an applied project specifically in technological development, which relates to the research line within the systems training chain. The chosen topic involves, at its core, the collection of information and the recognition of processes within a company that has been identified as having a deficiency in information capture. Through the implementation of existing tools on the market, the project aims to enable the recording, storage, processing, and visualization of information to support decision-making in one of the company's processes.

***Keywords:*** Data, Analysis, BI, SMEs, Digital Transformation.

## Tabla de Contenido

Introducción .....	8
Descripción del Problema .....	10
Justificación .....	12
Objetivos .....	14
Objetivo General .....	14
Objetivos Específicos .....	14
Marco de Referencia .....	15
Desarrollo .....	22
Identificación de Factores Principales que Impiden la Implementación de Ciencia de Datos en Colombia .....	22
Metodología .....	23
Diagnóstico de Registros Históricos .....	23
Caracterización de una Mipyme en Zona Rural de Cundinamarca para Implementar Ciencia de Datos.....	29
Conclusiones .....	39
Recomendaciones.....	42
Comentarios de Mejoramiento del Trabajo .....	46
Referencias Bibliográficas .....	47

## Lista de Tablas

<b>Tabla 1</b> <i>Variedad de Lechugas</i> .....	31
<b>Tabla 2</b> <i>Variables de la Base de Datos</i> .....	33
<b>Tabla 3</b> <i>Plan de Acción Modelos de Machine Learning</i> .....	43
<b>Tabla 4</b> <i>Metodología para Replicar en Otras MiPymes</i> .....	44

## Lista de Figuras

<b>Figura 1</b> <i>Causas - Efectos y su Relación con la Problemática Central</i> .....	22
<b>Figura 2</b> <i>Linea de Tiempo de Caracterización de la MiPyme</i> .....	23
<b>Figura 3</b> <i>Análisis Exploratorio de Datos</i> .....	23
<b>Figura 4</b> <i>Estado de Transformación de Empresas (Departamental)</i> .....	26
<b>Figura 5</b> <i>Estado de Transformación de Empresas (Cundinamarca)</i> .....	27
<b>Figura 6</b> <i>Indice Promedio de Acceso a Internet (Departamentos)</i> .....	28
<b>Figura 7</b> <i>Indice Promedio de Acceso a Internet (Municipios)</i> .....	28
<b>Figura 8</b> <i>Diagrama de Flujo Proceso Cultivo de Lechugas Hidropónicas</i> .....	32
<b>Figura 9</b> <i>Ezquema de la Aplicación</i> .....	35
<b>Figura 10</b> <i>Métricas Generadas a Partir de los Datos (Visualización Novedades)</i> .....	36
<b>Figura 11</b> <i>Métricas Generadas a Partir de los Datos (Visualización Seguimiento)</i> .....	37
<b>Figura 12</b> <i>Métricas Generadas a Partir de los Datos (Visualización Calidad de Datos)</i> .....	38

## Introducción

El presente trabajo tiene como objetivo principal demostrar cómo el uso de la ciencia de datos puede ser una herramienta clave para impulsar el crecimiento y sostenibilidad de las micro, pequeñas y medianas empresas (MiPymes) en zonas rurales de Cundinamarca. A lo largo del estudio, se identificó que una de las principales barreras que enfrentan estas empresas es la falta de información adecuada y oportuna para la toma de decisiones estratégicas, lo cual limita su capacidad para competir en un mercado cada vez más digital y dinámico.

Mediante el uso de técnicas de análisis de datos, este proyecto abordó el déficit de información, permitiendo a una de las MiPymes rurales acceder a datos valiosos principalmente sobre la parte productiva. Para ello, se implementó el uso de herramientas como el desarrollo de una aplicación para la captura de datos y la construcción de dashboards que facilitan la visualización y análisis de datos relevantes para la toma de decisiones.

El estudio se aplicó en una empresa seleccionada en la región de Cundinamarca, con el fin de validar la efectividad de estas herramientas tecnológicas en la realidad de las MiPymes rurales. Los resultados demostraron un impacto positivo en la gestión empresarial, evidenciando un incremento en la eficiencia operativa y una mejora en las decisiones comerciales. Específicamente, se observó una mejora en el proceso de identificación y registro de pérdidas, lo que se traduce en un mayor control sobre las mismas, generando así una sostenibilidad económica para la empresa al poder minimizar las pérdidas para incrementar la producción.

Desde un enfoque social, la implementación permitió fortalecer la cadena de valor agrícola al mejorar la toma de decisiones basadas en datos. Estos hallazgos refuerzan la importancia de la ciencia de datos como motor de transformación en este tipo de empresas, no

solo desde la perspectiva empresarial, sino también como catalizador del desarrollo socioeconómico en la región de Cundinamarca.

Lo que refuerza la importancia de la ciencia de datos como motor de transformación en este tipo de empresas, como por ejemplo el caso de implementación de la agricultura de precisión en Kenia, en donde pequeños agricultores en el país enfrentaban desafíos para maximizar sus rendimientos debido a la variabilidad del clima y la falta de información precisa sobre sus cultivos y que por medio de la implementación de soluciones tecnológicas como sensores, drones y análisis de datos, se desarrolló una plataforma que proporciona a los agricultores recomendaciones personalizadas sobre riego, fertilización y control de plagas. Esto ha llevado a un aumento significativo en la producción y la rentabilidad en estas MiPymes, siendo un caso de éxito sobre el alcance de la transformación de MiPymes en entornos rurales como lo es el sector agrícola. (Toro, 2024)

Este trabajo ofrece una base sólida para futuras investigaciones y aplicaciones prácticas de la ciencia de datos en áreas rurales, destacando su capacidad para reducir las brechas de información y contribuir al desarrollo económico local.

## Descripción del Problema

Con el paso del tiempo y el avance de la era digital, el volumen de datos generados a diario en todo el mundo ha crecido exponencialmente, lo que ha dado lugar a la necesidad de gestionar y analizar dicha información para tomar decisiones estratégicas. Sin embargo, en el caso de las micro, pequeñas y medianas empresas (MiPymes) en Colombia, existe un marcado déficit en el uso adecuado de los datos. A pesar de que muchas de estas empresas capturan información en sus procesos, esta no es procesada, analizada ni visualizada de forma que genere valor para la toma de decisiones empresariales. Esto implica que los datos, que podrían ser un activo valioso para resolver problemas, optimizar procesos y aumentar la rentabilidad, quedan desaprovechados.

Por otro lado, un problema aún más crítico en muchas MiPymes es que, en algunos casos, ni siquiera se lleva a cabo la captura de datos. En estas situaciones, la ausencia total de registros de información en los procesos empresariales impide el seguimiento y control de las actividades, lo que compromete gravemente la capacidad de la empresa para mejorar y optimizar sus operaciones. Sin datos registrados, los gerentes y propietarios de estas empresas carecen de las herramientas necesarias para implementar decisiones basadas en evidencia y, por ende, no pueden acceder a los beneficios que ofrece la tecnología y el análisis de datos para tomar decisiones informadas.

Este déficit de datos también refleja una falta de cultura empresarial orientada a la toma de decisiones basada en evidencia, una práctica cada vez más necesaria en un entorno empresarial globalizado y competitivo. Las MiPymes, especialmente en áreas rurales o menos urbanizadas, no logran integrar estas herramientas y, como resultado, pierden oportunidades de crecimiento, mejora de la eficiencia y, en última instancia, de maximización de sus

rentabilidades. La falta de análisis de datos limita la capacidad de estas empresas para adaptarse a los cambios del mercado, aprovechar tendencias emergentes y anticiparse a problemas antes de que escalen.

Asimismo, la falta de inversión en tecnología que facilite la captura y análisis de datos se traduce en una oportunidad perdida para implementar estrategias que aumenten la competitividad y reduzcan costos. El análisis de datos ofrece ventajas como la optimización de procesos, el análisis de comportamiento de clientes, la mejora en la gestión de inventarios y la identificación de oportunidades de crecimiento. Sin embargo, al no captar ni analizar esta información, las MiPymes en Colombia quedan rezagadas, reduciendo su capacidad para mejorar su rentabilidad y sostenibilidad a largo plazo.

En conclusión, la ausencia de una cultura de captura y análisis de datos en las MiPymes colombianas representa un obstáculo significativo para su crecimiento y adaptación al mundo digital. Este problema requiere una intervención urgente para que estas empresas puedan aprovechar el poder de los datos como un activo estratégico que mejore su competitividad y eficiencia.

## **Justificación**

En los últimos años, el acelerado avance de la transformación digital en Colombia ha evidenciado un incremento notable en la cantidad de datos generados y registrados por las empresas. Este fenómeno responde a la creciente necesidad de agilidad en los trámites y a la importancia de contar con información precisa para tomar decisiones estratégicas. Sin embargo, muchas micro, pequeñas y medianas empresas (MiPymes), especialmente en zonas rurales, aún no han adoptado estas prácticas. Esto crea una oportunidad significativa para aquellas que, al incorporar la digitalización y el registro de datos en sus operaciones, podrían obtener mayor control sobre sus procesos y mejorar su capacidad de toma de decisiones.

La transformación digital, además de optimizar procesos, permite un desarrollo empresarial más eficiente y sostenible. Las MiPymes que adopten estas tecnologías estarán mejor posicionadas para competir, no solo en el mercado local, sino también a nivel nacional e internacional. Para lograr dicha adopción de tecnologías, las condiciones para garantizar la escalabilidad de la metodología implementada, se ajustan con empresas en donde en un principio se logre identificar puntos de captura de datos en sus procesos, que permitan tomar decisiones basadas en datos, así como también potencial de crecimiento económico y de competitividad con la implementación de soluciones basadas en inteligencia de negocios, donde producto de ello se les permita controlar y orientar la toma de decisiones estratégicas en pro de optimizar sus procesos e incrementar su rentabilidad en cualquier sector económico. Con la constante evolución de tecnologías como la inteligencia artificial y el análisis de datos, es crucial que estas empresas empiecen a implementar mecanismos digitales que les permitan aprovechar el valor de los datos. De no hacerlo, estarían en desventaja frente a competidores que ya utilizan estas herramientas para mejorar su productividad y competitividad.

Este trabajo de grado parte de la premisa de que muchas MiPymes rurales en Colombia carecen de sistemas de registro digital de información, lo que representa una barrera para su crecimiento y desarrollo. La implementación de un ciclo completo de ciencia de datos, desde la captura y procesamiento de información hasta su análisis y visualización, ofrece una solución tangible a esta problemática. Mediante la identificación de una MiPyme que carezca de estas prácticas, se busca implementar un modelo que le permita no solo optimizar sus operaciones, sino también mejorar la toma de decisiones basadas en datos, lo que se traducirá en una mejora continua y en el fortalecimiento de su competitividad.

La justificación de este proyecto radica en la necesidad de brindar a las MiPymes rurales herramientas que les permitan aprovechar los avances tecnológicos y adaptarse a las nuevas exigencias del mercado. Este enfoque contribuirá al crecimiento económico de estas empresas, al mismo tiempo que fomentará su capacidad para integrar tecnologías emergentes, potenciando así su desarrollo en un entorno cada vez más digital y competitivo.

## **Objetivos**

### **Objetivo General**

Implementar ciencia de datos en una MiPyme con déficit de datos en área rural del departamento de Cundinamarca.

### **Objetivos Específicos**

Identificar los factores principales que impiden la implementación de ciencia de datos en Colombia.

Caracterizar una MiPyme en zona rural de Cundinamarca donde se pueda implementar ciencia de datos.

Aplicar una metodología que permita la recolección, análisis exploratorio, visualización y modelado de datos.

## Marco de Referencia

En la era en la que nos encontramos es común hablar sobre tecnología como una adopción que se debe tener en las empresas a nivel global, ya que los procesos tienden a ser digitalizados y con el pasar el tiempo automatizados con el fin de mejorar la forma de hacer las cosas, además de mantener un control y minimizar errores en los procesos. Lo anterior hace parte de la cuarta revolución industrial, también llamada Industria 4.0 o Revolución 4.0, que describe la creciente digitalización de toda la cadena de valor y la interconexión resultante de personas, objetos y sistemas a través del intercambio de datos en tiempo real. Esta revolución modificará fundamentalmente la forma en que vivimos, trabajamos y nos relacionamos, su naturaleza global afectará a todos los países, economías, sectores y personas. (Jiménez, 2021)

Es importante que la empresa y la academia se involucren en la mencionada cuarta revolución industrial, ya que, en la actualidad con el crecimiento de datos en las empresas, se necesita contar con talento que se encuentre en la capacidad de administrar, procesar y convertir los datos a información de manera tangible para que los líderes de los procesos y gerentes puedan tomar decisiones basados en datos.

En la actualidad, se han alcanzado niveles considerables de inmersión y absorción tecnológica en todos los sectores clave de la sociedad. Dentro de estos sectores, las universidades como instituciones rectoras de procesos de transformación social han tenido que adoptar un nuevo modelo educativo, lo cual incluye innovación en procesos de enseñanza-aprendizaje y construcción colectiva del conocimiento mediante el uso de Internet. Esta visión colectiva del conocimiento, ligada al desarrollo de aplicaciones web, compromete a las universidades a empoderar digitalmente a comunidades educativas en términos de accesibilidad, integración digital, equidad o igualdad, y gestión para la inserción laboral.

(Iris Agustina Jiménez-Pitre, Raúl José Martelo Gómez, Juannys Chiquillo Rodelo, Denny de Jesús Lloreda Gracia, Maribel Sofía Morales Camacho, 2017)

Las empresas en Colombia en los últimos años han adoptado la transformación digital en sus procesos, capturando datos y convirtiéndolos en información para la toma de decisiones, aplicando la inteligencia de negocios e invirtiendo en recursos profesionales en el área y recursos tecnológicos, con el fin de digitalizar y optimizar los procesos, obteniendo grandes beneficios en las empresas y los clientes. Pero, por otra parte, especialmente las MiPymes en Colombia, apenas están ejecutando la transformación digital, o en el peor de los casos, no han comenzado a montarse en la cuarta revolución industrial, por tal motivo no nace la iniciativa de inversión en la misma, tal vez por desconocimiento de las oportunidades que brinda el aprovechamiento de los datos, o por que definitivamente no tienen acceso a recursos tecnológicos, como lo es un computador, celular o internet. Para ello es importante involucrar las MiPymes en la transformación digital, dándoles a conocer las técnicas, herramientas y ventajas que trae consigo la implementación de la ciencia de datos como parte de sus procesos.

La minería de datos se concibe como el proceso mediante el cual es posible descubrir información no trivial a partir de grandes volúmenes de datos, utilizando técnicas de inteligencia artificial y estadística. En una revisión sobre las aplicaciones de la minería de datos, se puede encontrar gran variedad de áreas como el sector salud, finanzas, banca, educación, biología, entre otras. Estas aplicaciones son realizadas en su gran mayoría sobre datos estructurados, es decir, datos organizados en bases de datos. De igual manera, las metodologías y plataformas para minería han centrado sus esfuerzos sobre minería de datos estructurados; de hecho, se puede encontrar diversidad de opciones de acceso libre y licenciado.

(Efraín Alberto Oviedo Carrascal, Ana Isabel Oviedo Carrascal, Gloria Liliana Velez Saldarriaga, 2016)

Para lo anterior es importante tener en cuenta los recursos mínimos necesarios para implementar la ciencia de datos en una MiPyme, que realmente resulta algo tan sencillo como un hardware y software, así como conectividad a internet.

Cuando se habla de las tic, al primer dispositivo al que se remite es al computador; representación que se da de manera general. Este dispositivo está inscrito dentro del conjunto de las tic cuando se asocia a internet; es decir que es condición necesaria que un computador tenga acceso a internet para que pueda cumplir con los propósitos que los sujetos le asignan. En el contexto urbano, en donde hay más acceso a internet en términos de cobertura y velocidad de conexión y en donde los sujetos tienen mayor acceso a los dispositivos, los actores también inscriben en el conjunto de las tics dispositivos como teléfonos inteligentes y tabletas digitales, debido a la posibilidad que tienen de conectarse a la red.

(Eliaana Esperanza Ordóñez Gómez, Nicolás Penagos Muñetón, 2015)

En Colombia se han evidenciado varios casos de éxito en diferentes sectores, una de ellas que tiene que ver con investigación sobre el alto volumen de datos, el tiempo de procesamiento y la visualización de información que son problemas que hoy en día deben afrontar las organizaciones en el sector de tecnología. En esta investigación se desarrolló un modelo para la toma de decisiones basado en la interacción de los criterios y las etapas del ciclo de ventas. Se usó una herramienta especializada en inteligencia de negocios que facilitó el manejo de grandes volúmenes de datos, su procesamiento y la visualización de la información. Dentro de los resultados conseguidos se encontró una significativa reducción en el tiempo de obtención de la información, pasando de horas a minutos.

(Diego Armando Vanegas, Giovanni Mauricio Tarazona-Bermudez, Luz Andrea Rodríguez-Rojas, 2020)

Otra iniciativa ejecutada en el país aportando a la transformación digital, denominada como Kioscos Digitales, son centros que usan plataformas de conectividad digital para promover el cambio en la comunidad desde diferentes alternativas de inclusión social en el contexto cultural desde lo que implican los procesos de conectividad digital; así mismo, delinean desde el reconocimiento del capital humano y los aprendizajes colaborativos en lugares rurales de difícil acceso tecnológico, pedagógico y con fallas en el acceso a servicios digitales, la comunicación eficiente para la promoción de hábitos y estilos de vida saludables. (Beltran, 2016)

Por otra parte, en el sector automotriz, se destaca un proyecto que permitió la transformación tecnológica en procesos del negocio en la unidad HONDA Motocicletas de FANALCA, aumentando el nivel de satisfacción de servicio al cliente gracias a procesos organizacionales eficientes y rápidos que permiten brindar un servicio integral a los usuarios de la marca HONDA. Para el cual se diseñaron integraciones tomando como origen la ensambladora en la creación de productos y como fuente la Red de concesionarios. Una vez realizadas las integraciones expuestas a través de un bus de datos o a través de Web-Services, se definió un procedimiento de manejo de errores y de excepciones creado entre la Ensambladora y la Red de Concesionarios que define tanto el proceso en los sistemas de información, como el proceso de negocio. El proceso de transformación generó un esfuerzo continuo de capacitación al nuevo personal y refuerzo en las mejores prácticas tanto en FANALCA como en la red de distribución, generando por esta vía, la transformación de la Red de concesionarios que incluyó una revisión y mejora de los procesos de mercadeo, ventas y posventa.

El personal de FANALCA se vio beneficiado al poder acceder a capacitaciones, logrando una alta cultura de autogestión apoyada en los sistemas de información, reduciendo tareas operativas y aumentando la confiabilidad de la información.

(Mónica Vélez O., Hermann Stuart Fuquen Gonzalez, Claudia Marcela Sánchez, 2020)

A pesar de que en las zonas rurales aún no se cuenta con total conectividad y acceso a recursos tecnológicos, es posible aportar a la transformación tecnológica por medio de la ciencia de datos, destacando un proyecto aplicado en el sector agrícola, resaltando que las herramientas desarrolladas para la agricultura han demostrado ser efectivas, ya que permiten realizar el monitoreo constante tanto de los cultivos como de sus procesos de producción.

Todo lo anterior se traduce en la búsqueda de la obtención de información que sea oportuna, verás, confiable y respaldada, con la finalidad de mejorar todos los procesos involucrados y de esta manera alcanzar mayores ganancias. Algunas de las tecnologías implementadas son: Softwares webs que brindan todo tipo de información, realizan el monitoreo de variables agrícolas, entre otros; teledetecciones con drones que permiten hacer el análisis de los cultivos con rapidez en grandes superficies, estudios topográficos, controles de temperatura, etc, y sistemas de sensores con la finalidad de monitorizar los parámetros determinantes para contar con una la mejor producción.

Los aplicativos móviles, por su parte, se han convertido en tecnologías ampliamente utilizadas por los caficultores y agricultores en general debido a su facilidad de portabilidad, además de permitirles tener en tiempo real los datos correspondientes al monitoreo de sus cultivos y procesos de producción para la toma de decisiones temprana y acertada. (Yurley T. Tovar, Andrés F. Calvo, Arley Bejarano, 2022)

Si bien la analítica de datos posibilita acometer estudios desde diferentes disciplinas, el acervo documental disponible en las bases de datos especializadas demuestra que las áreas de Bioquímica, Ingeniería y Farmacología son las más recurrentes en el uso y aplicación de la analítica, por su parte, en la literatura disponible se pueden encontrar estudios soportados en la analítica de datos particularmente en temas asociados al sector público, tales como: efectos socioeconómicos potenciales derivados del COVID, estudios de migraciones, análisis espacial para la localización de puntos de atención médicos, diseño de eco ciudades y ciudades, entre otros; demostrando la existencia de un interés creciente, en el uso de los datos más allá de las disciplinas tradicionales y por supuesto, se elucidan las ventajas de usar los grandes conjuntos de datos y la analítica como apoyo en la toma de decisiones en el sector público; no obstante, en la literatura se percibe algo de escepticismo, debido a las limitaciones a nivel de infraestructura y en mayor parte a la falta de más científicos de datos. Ciertamente, dicho escepticismo guarda relación con la poca disponibilidad de nuevos profesionales y científicos de datos un profesional con capacidades analíticas que lo ubican hoy, con un rol relevante en la llamada sociedad digital.

(Daniela Rivero-Riqueme, Luis Felipe Ortiz-Clavijo, 2021)

Finalmente es importante resaltar la importancia de la transformación digital y la ciencia de datos para mejorar la competitividad de las empresas, específicamente en el contexto colombiano, ya que como región, debemos adoptar las tecnologías que actualmente el mercado nos brinda para poder implementarlas al interior de los procesos, especialmente en las MiPymes que se encuentran en gran desventaja respecto a grandes empresas que inclusive ya utilizan la inteligencia artificial para la optimización de sus procesos. De tal manera que, si en las pequeñas empresas se comienza con la implementación de la transformación digital, poco a poco entrarían

a ser mucho más competitivas si toman decisiones basados en datos generados en sus propios procesos.

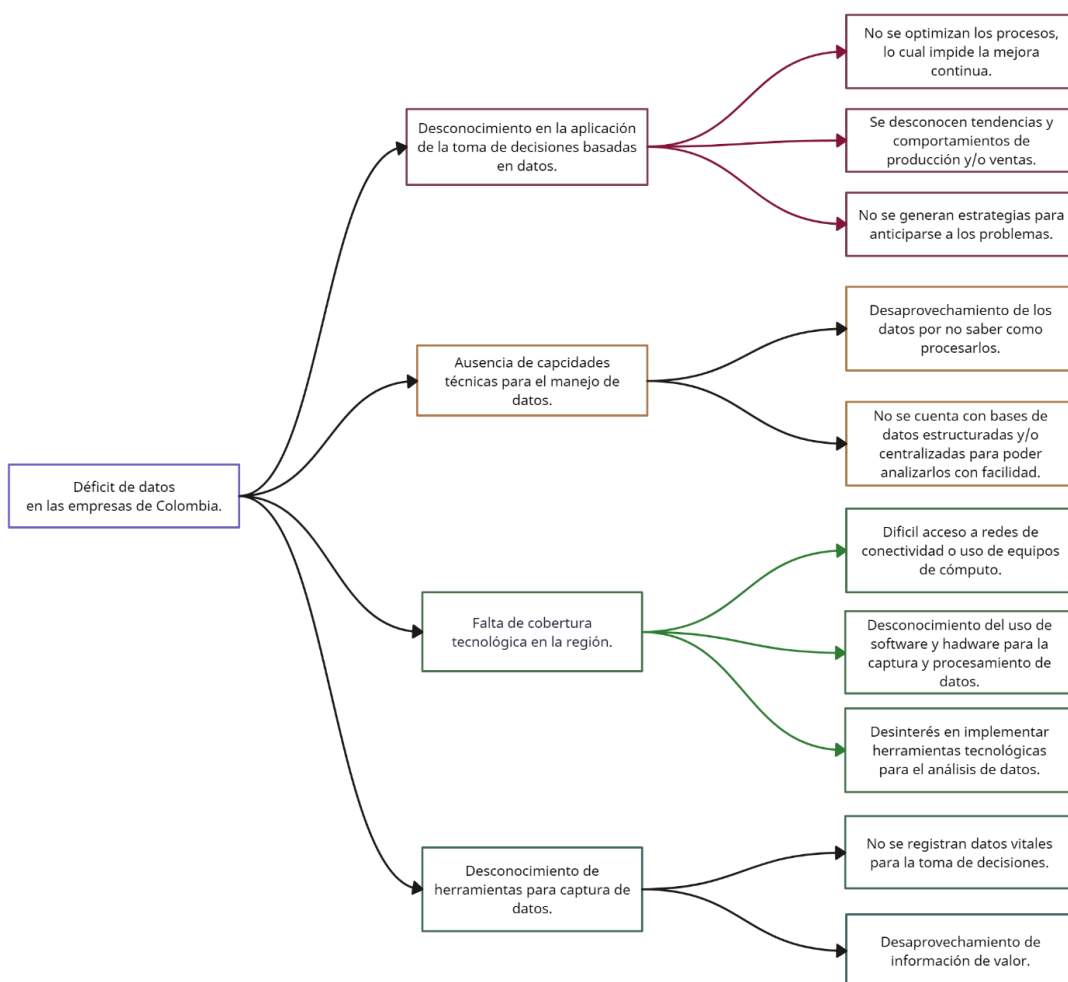
## Desarrollo

### Identificación de Factores Principales que Impiden la Implementación de Ciencia de Datos en Colombia

Para poder abordar la problemática principal sobre el déficit de datos en las empresas en Colombia, se plantearon cuatro causas principales a las cuales probablemente se deba dicho déficit de datos y sus respectivos efectos, como se muestra en el siguiente figura:

**Figura 1**

*Causas - Efectos y su Relación con la Problemática Central*



## Metodología

**Figura 2**

### *Línea de Tiempo de Caracterización de la MiPyme*



## Diagnóstico de Registros Históricos

**Figura 3**

### *Análisis Exploratorio de Datos*



Se inició con un proceso de recolección de información para obtener bases de datos que facilitaran la identificación de factores relacionados con una de las causas mencionadas en la figura 1: la falta de cobertura tecnológica en la región. Como parte de esta investigación, se consultó la página web de datos abiertos de Colombia, donde se descargaron inicialmente catorce archivos en formatos Excel y CSV. Tras una exploración preliminar de su contenido y estructura, se seleccionaron dos archivos que, una vez procesados, podrían proporcionar información relevante para comprender la falta de cobertura tecnológica en los municipios del país. Este análisis permitirá también identificar, con base en datos, un municipio específico donde implementar este proyecto de grado aplicado en una MiPyme.

Posteriormente, se llevó a cabo un análisis exploratorio, que consistió en la limpieza y transformación de los datos obtenidos de los dos archivos seleccionados. En una primera etapa, se utilizó Excel para realizar ajustes preliminares, como la revisión inicial de la calidad de los datos, detección de valores nulos y errores básicos de formato. Posteriormente, se empleó Visual Studio Code como entorno de desarrollo para implementar procesos más avanzados con el lenguaje de programación Python, utilizando la biblioteca Pandas. Este enfoque permitió realizar ajustes más detallados, como la conversión de formatos, la normalización y limpieza de textos, la eliminación de registros irrelevantes, y la selección de columnas clave relacionadas con la cobertura tecnológica. Una vez completado este proceso, los resultados se consolidaron y exportaron a formatos accesibles, como tablas de Excel y Google Sheets, para facilitar su análisis en etapas posteriores.

Con las tablas procesadas, se desarrollaron diversas estrategias de análisis. En Excel, se calcularon métricas clave y se generaron histogramas para visualizar la distribución y tendencias de los datos. Estas herramientas fueron fundamentales para identificar patrones generales y

realizar un análisis estadístico descriptivo inicial. Paralelamente, las tablas subidas a Google Sheets se integraron en Looker Studio para diseñar un dashboard interactivo. Este dashboard permitió una visualización más dinámica y detallada de los datos, proporcionando una perspectiva más clara sobre la relación entre las variables y la falta de cobertura tecnológica en los municipios analizados.

El uso combinado de estas herramientas no solo facilitó el análisis, sino que también garantizó la coherencia y la calidad de la información procesada. Los insights obtenidos en esta etapa fueron esenciales para entender las causas subyacentes de la problemática y permitieron identificar un municipio específico que cumpla con los criterios necesarios. A partir de esta selección, se pudo avanzar en la identificación de una MiPyme relevante dentro del municipio elegido, donde se implementará el proyecto de grado, asegurando que las decisiones tomadas estén fundamentadas en datos sólidos y confiables.

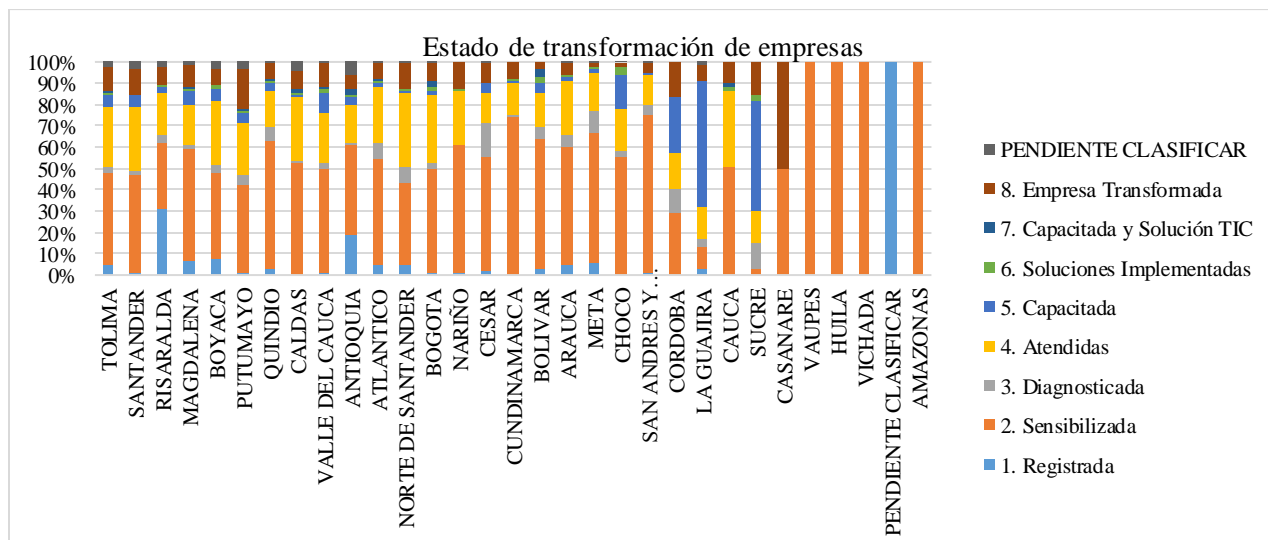
La primera base de datos que se consultó y se analizó fue sobre los centros de transformación digital en Colombia suministrado por el ministerio de tecnologías de la información y las comunicaciones (MINTIC) (Comunicaciones, Datos, 2024), en la cual se registran las entidades acreditadas por el gobierno colombiano, como por ejemplo las cámaras de comercio, para apoyar las pequeñas y medianas empresas en su proceso de transformación digital. Dichos centros proporcionan servicios de acompañamiento a las empresas para que adopten tecnologías digitales, mejorando así su competitividad y optimicen los procesos productivos mediante la incorporación de herramientas digitales, lo cual sería útil para esta parte de la investigación.

El conjunto de datos está con fecha de corte al 21 de mayo de 2021 y contiene 70.325 filas y 19 columnas. Luego de limpiar y transformar los datos usando Excel, las variables de

interés fueron las siguientes: ciudad, departamento, estado transformación, MiPyme Tipo de empresa, obteniendo los siguientes resultados:

#### Figura 4

##### *Estado de Transformación de Empresas (Departamental)*

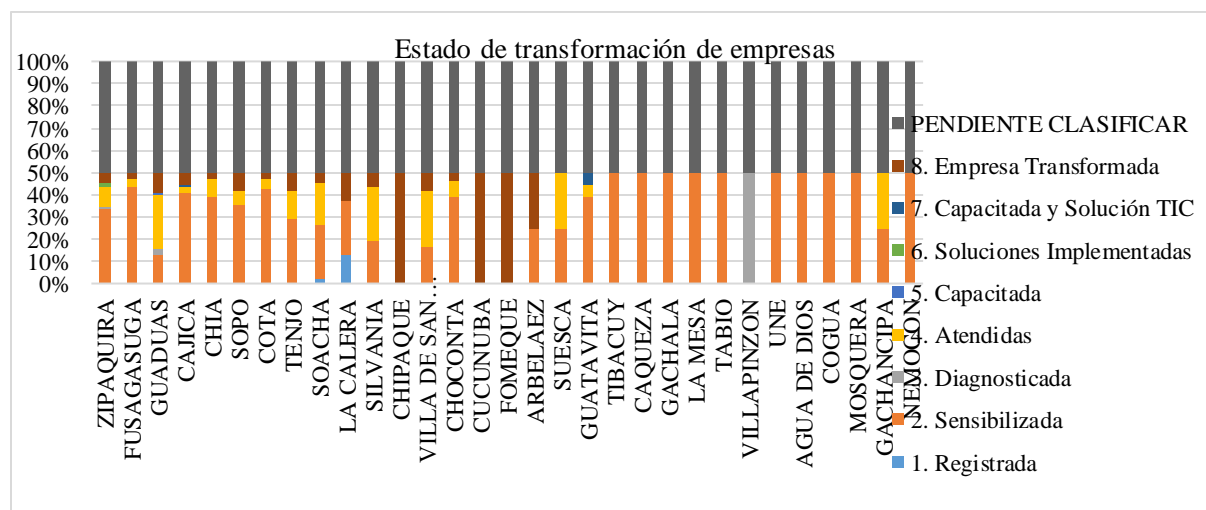


En donde departamento de interés, en este caso Cundinamarca, se encuentra en la posición 15 de los departamentos en donde se ha llevado a cabo transformación digital en las empresas, con el 8% de empresas transformadas del total de las registradas, sensibilizadas, diagnosticadas, atendidas y capacitadas.

A continuación se muestran los resultados del análisis por detalle de municipio dentro del departamento de Cundinamarca:

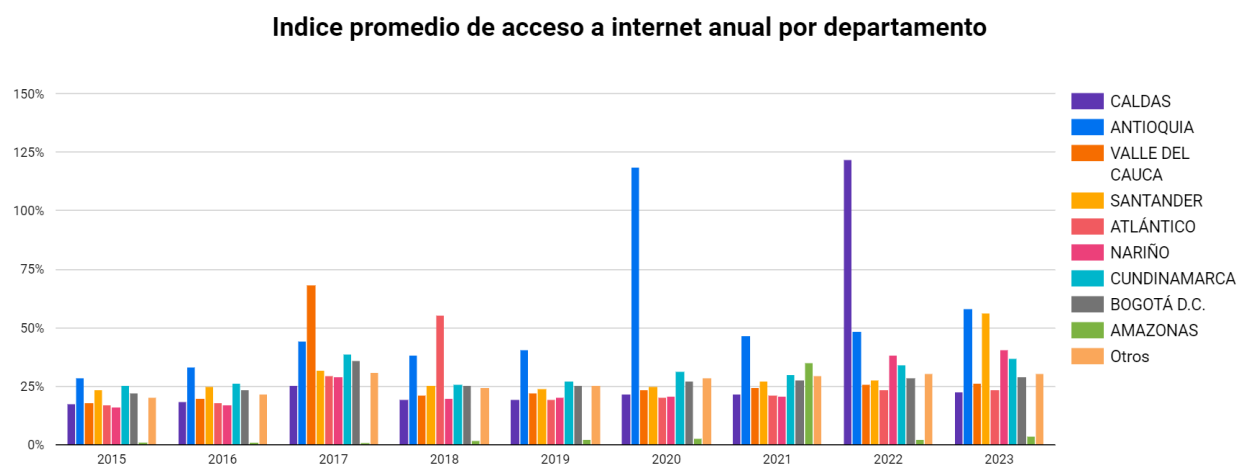
**Figura 5**

*Estado de Transformación de Empresas (Cundinamarca)*



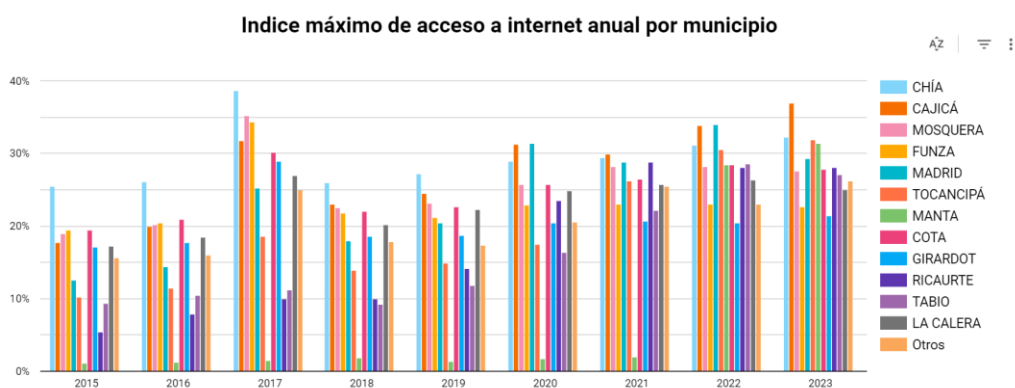
En donde los municipios de interés, que en este caso son La Calera, Tabio, Cajicá y Tocancipá, se encuentra en la posición 10, 24, 4 respectivamente, con excepción de Tocancipá ya que no registra ninguna empresa. Hasta este punto los municipios de interés serían Tabio y La Calera, por su ubicación en el ranking en cuanto a la cantidad de empresas transformadas.

Finalmente se analizó la causa sobre falta de cobertura tecnológica en la región y el efecto de difícil acceso a redes de conectividad, ya que para este caso se encontró un conjunto de datos suministrado por el ministerio de tecnologías de la información y las comunicaciones (MINTIC) (Comunicaciones, Datos, 2024), la cual contiene la penetración de internet fijo en los municipios del país de Colombia. El conjunto de datos contiene registros del año 2015 al año 2023 y tiene un total de 35.579 filas y 9 columnas. Luego de limpiar y transformar los datos usando el lenguaje de programación Python y visualizando los datos en LookerStudio de Google, las variables de interés fueron las siguientes: Año, Departamento, Municipio, No. Accesos, Índice y Población, obteniendo los siguientes resultados:

**Figura 6***Indice Promedio de Acceso a Internet (Departamentos)*

En donde departamento de interés, en este caso Cundinamarca, el índice promedio de acceso a internet en los últimos 8 años es del 30.3% muy cerca a Bogotá con un 28.4% y superado por Antioquia con un 42.7%.

A continuación se muestra los resultados del análisis por detalle de municipio dentro del departamento de Cundinamarca:

**Figura 7***Indice Promedio de Acceso a Internet (Municipios)*

Encontrando que los municipios de interés de acuerdo a los resultados del análisis de la primera base de datos, que para este punto son La Calera y Tabio, en donde el índice máximo anual de acceso a internet en los últimos 8 años en promedio es del 22.8% y un 16.2% respectivamente, lo cual ayudará a continuar con el siguiente objetivo específico.

### **Caracterización de una Mipyme en Zona Rural de Cundinamarca para Implementar Ciencia de Datos**

Luego de analizar datos sobre las posibles causas que nos permiten abordar la problemática sobre el déficit de datos, y las bases de datos con las que se realizó una segmentación de departamentos y municipios de interés, que en este caso los conjuntos de datos analizados fueron sobre los centros de transformación digital y la penetración de internet fijo en los municipios de Colombia, finalmente el municipio de interés es Tabio en el departamento de Cundinamarca, por su bajo alcance en transformación digital y baja penetración de internet fijo, según las estadísticas y las variables analizadas. en donde se llevará a cabo la selección de la MiPyme de interés, dando prioridad a la zona rural.

En este caso, se procedió a realizar un recorrido por la zona rural de Tabio, en donde se identificó una MiPyme correspondiente al sector primario de la economía, en este caso al sector agrícola, ya que se trata de un cultivo de lechugas hidropónicas ubicado más exactamente en la vereda de La Lucera, que para efectos de poder desarrollar el presente proyecto de grado aplicado, se procedió a realizar un acercamiento inicial con los encargados del cultivo, socializando el objetivo general del presente trabajo, para lo cual se percibió un interés y aprobación para iniciar con un diagnóstico inicial al interior de los procesos que se llevan a cabo allí.

En posteriores visitas, se realizó un recorrido por el cultivo de lechugas, realizando un levantamiento de procesos con el fin de entender el contexto del negocio y entendiendo los procesos, la capacidad instalada, inclusive las herramientas tecnológicas con las que cuentan y el acceso a internet. En este punto se encontró que cuentan con computadores en el área administrativa y en la parte de producción los operarios cuentan con dispositivos móviles, así como también un acceso a internet satelital, lo cual resulta siendo una ventaja al momento de comenzar a implementar una solución basada en ciencia de datos.

Luego del levantamiento de los procesos, lo siguiente que se hizo fue la identificación de problemas o las necesidades en donde sea viable implementar algunas de las soluciones de ciencia de datos, que permitan controlar los procesos, hacer seguimiento a los mismos y poder ayudar a la toma de decisiones por parte de los líderes. En este punto se identificaron variables que hacen parte de los procesos que se lleven a cabo, mas específicamente en la parte de producción, con el fin de poder realizar la construcción preliminar de una base de datos estructurada propuesta, en donde se puedan almacenar los datos.

En este punto se encontró una necesidad de controlar el inventario de las lechugas, debido a que en ese momento el proceso que se hace es empírico, ya que una vez se siembra una plántula de lechuga, debe pasar cierto periodo de tiempo hasta que la lechuga sea apta para la venta y eso se hace únicamente con la experticia visual de quienes allí realizan el proceso. Lo cual representa un riesgo ya que no se lleva un control de periodos o ciclos de siembra y mucho menos un análisis detallado sobre las cantidad en cada una de las etapas del cultivo, así como las pérdidas y la disponibilidad de los bancos.

El cultivo tiene una capacidad instalada de 243 bancos, en donde cada banco tiene una capacidad para cultivar 408 lechugas, es decir que la capacidad total del cultivo es de 99.144 lechugas. Se cultivan las siguientes 17 variedades de lechugas:

**Tabla 1**

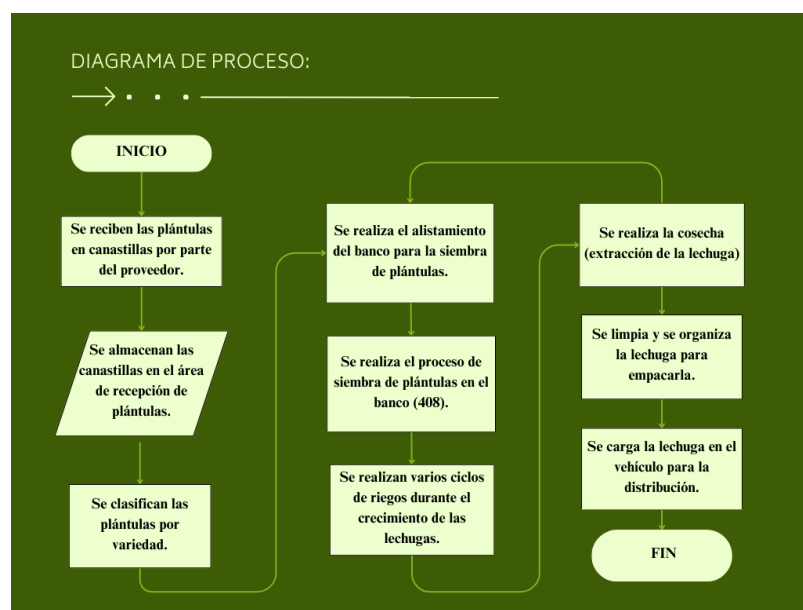
*Variedad de Lechugas*

Variedad	Variedad	Variedad
Verde Crespas	Carmin	Falbala
Cogollo derbi	Cogollo duende	Romana
Lisa verde	Salanova Triplex	Crespas Normal
Roble verde	Roble morada	Escarlet
Escarola	Mizuna verde	Mizuna Morada
Cogollo morado	Romana morada	

*Nota.* En esta tabla se muestran las variedades de lechugas que se cultivan en la MiPyme.

Hasta este punto es importante tener en cuenta el tamaño del cultivo, ya que de esto dependerá la cantidad de variables que intervienen en los procesos y la cantidad de registros que se generarían al momento de construir y alimentar una base de datos.

A continuación se muestra gráficamente como funciona la parte de producción del cultivo, en la cual comenzaría la implementación de herramientas de ciencia de datos:

**Figura 8***Diagrama de Flujo Proceso Cultivo de Lechugas Hidropónicas*

De los procesos que se identificaron en el anterior diagrama, es importante dar claridad sobre el concepto “banco”, ya que se hace referencia a una estructura elevada con madera y tubería agujerada en donde se siembran las plántulas, la cual tiene una capacidad para 408 lechugas. También es importante detallar la parte del proceso donde se realiza el alistamiento del banco para la siembra de plántulas, ya que en este punto lo que se hace es un drenado de agua y se limpian las tuberías del banco. Así como también especificar que en la parte de siembra de plántulas el tiempo de crecimiento varía entre 38 y 50 días dependiendo de la variedad.

En esta parte ya se identifica la manera en la que se deben capturar los datos y teniendo en cuenta las necesidades encontradas anteriormente, el primer registro se realizaría al momento del alistamiento del banco, ya que en este punto de partida comienza a correr el tiempo en que un banco termina la cosecha y comienza a estar disponible para una nueva siembra. Al momento de iniciar la siembra de las plántulas, se realizaría un segundo registro, el cual marcaría el inicio del

ciclo del proceso de crecimiento de las lechugas, siendo una parte fundamental por ejemplo para llevar la trazabilidad del tiempo en días para saber las lechugas próximas a estar disponibles para la cosecha. Y un último registro sería al momento de la cosecha, en donde se cerraría el ciclo y marcaría el inicio del siguiente, tal y como se visualiza en la figura 6.

Teniendo en cuenta lo anterior, se procede a construir una base de datos propuesta en una hoja de cálculo de Google (Google Sheet) con el fin de manejar un almacenamiento en tiempo real en la nube y de costos nulos, ya que la cantidad de registros prevista no requeriría un almacenamiento robusto. En la base de datos se incluyeron las siguientes variables:

**Tabla 2**

*Variables de la Base de Datos*

Variables 1	Variables 2	Variables 3
Banco (número de banco del 1 a 243).	Días de ciclo.	
Inventario inicial.	Tipo de Banco (normal o mixto).	
Fecha terminado.	Ciclo (el número del ciclo que se encuentra el	No cumple calidad.
Dobletada.	banco).	Inventario final.
Días en proceso.	Enferma.	Las variedades en las siguientes columnas.
Estado (alistamiento, en proceso y terminado).	% Pérdida.	Fecha en proceso.
Inventario inicial mixto.	Usuario (correo de quien hace el registro).	Pérdida accidental.
Vacio.	Variedad.	Días en almacenamiento.
Total pérdida.	Fecha alistamiento.	

*Nota.* En esta tabla se muestran los nombres de las columnas registradas en la base de datos construida.

Una vez estructurada la base de datos incluyendo todas las variables mencionadas anteriormente en columnas, se procede a desarrollar una herramienta que permita diligenciar la

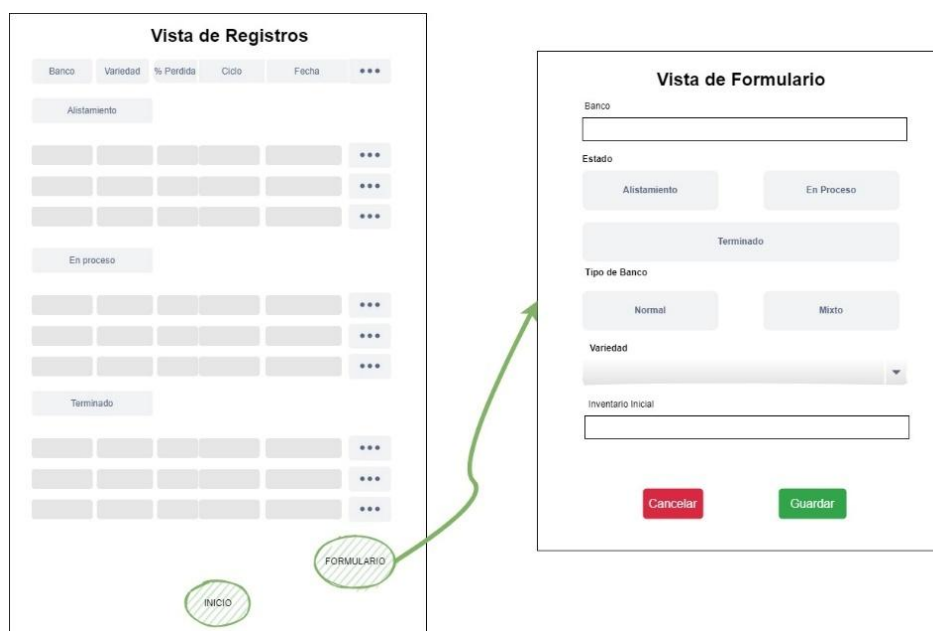
base de datos inicial, la cual sea flexible, muy intuitiva al usuario, de bajo costo y fácil conectividad con las demás herramientas, además que se pueda utilizar en los dispositivos con los que se cuentan en el cultivo (computador y móvil).

La herramienta o aplicación para capturar datos, se desarrolló utilizando Google AppSheet, ya que permite una fácil conectividad a la base de datos creada en Google Sheet y permite ser utilizada en dispositivos móviles y computadores, además no toma mucho tiempo en ser desarrollada ya que se tienen claras las variables y la cantidad de usuarios, que en este caso serían tres personas las encargadas de realizar el registro y control de los procesos productivos del cultivo, para lo cual hasta este punto no es necesario contar con una cuenta de pago por la cantidad de usuarios que la utilizarían.

Una vez se capacitó al personal sobre la herramienta de captura de datos y varias pruebas pilotos en campo, en la medida en que se comiencen a capturar datos, se proceden a realizar los respectivos ajustes de acuerdo a la retroalimentación por parte de los usuarios que la utilizan, dejando una versión final y funcional.

## Figura 9

### *Ezquema de la Aplicación*

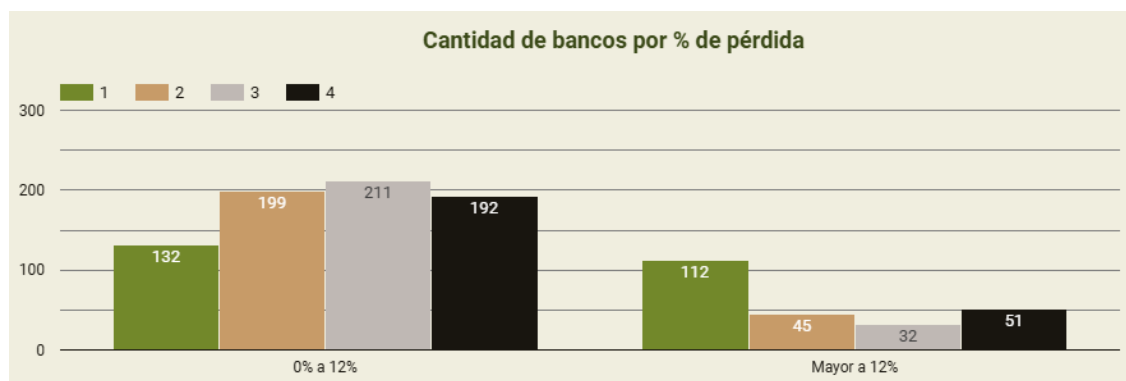


Posteriormente, después de generar varios registros, o por lo menos completar el primer ciclo, se realizaría un monitoreo para validar el funcionamiento de la base de datos y la aplicación de captura de datos, con el fin de comenzar a construir un dashboard utilizando otra herramienta de Google como lo es LookerStudio, con indicadores que permitan a los encargados de los procesos visualizar los datos y generar controles para cumplir con el propósito de la ciencia de datos, que en este caso es aportar a la toma de decisiones basadas en datos y de esta manera cumplir con el último objetivo específico y terminar de completar el objetivo general.

Después de dos meses desde la captura de datos, se procede a construir el respectivo dashboard en LookerStudio, atendiendo las necesidades previstas inicialmente, obteniendo las siguientes métricas:

**Figura 10**

*Métricas Generadas a Partir de los Datos (Visualización Novedades)*



El informe contiene métricas de interés para la MiPyme, el cual contiene tres módulos, en donde en el primero que se denomina “Novedades” se encuentran todas las gráficas que permiten visualizar las novedades presentadas en el cultivo, ya que es de vital importancia para los líderes del proceso poder controlar la cantidad de bancos por porcentaje de pérdida, específicamente por dos grupos: del 0% al 12% y mayor al 12% de pérdida por ciclo como se observa en la anterior Figura 10. Esto les permite evaluar en cada ciclo la cantidad de bancos que registraron pérdida, para de esta manera abordarlos en el siguiente ciclo.

También como se visualiza en la figura 9, se puede observar que el dashboard inicia con un apartado de filtros, los cuales les permiten especificar el estado del proceso, el tipo de banco, el ciclo y graduar el porcentaje de pérdida deseado a consultar.

**Figura 11**

*Métricas Generadas a Partir de los Datos (Visualización Seguimiento)*

		Ciclo / Dias en Proceso			Ciclo / Inventario Final			
Banco	Variedad	1	3	6	Variedad	1	3	Total
241	Cogollo derbi	-	-	-	Verde crespa	401	408	41.363
239	Cogollo derbi	-	-	-	Cogollo derbi	-	-	16.569
235	Cogollo derbi	-	-	-	Verde Caipira	-	-	5.712
240	Cogollo derbi	-	-	-	Carmin	-	-	4.452
76	Carmin	-	-	-	Romana	-	-	2.037
119	Verde crespa	48	-	-				
118	Verde crespa	-	-	-				

En la anterior figura se muestra parte de la segunda página del dashboard denominada “Seguimiento”, que contiene la parte de seguimiento del cultivo en función del tiempo y cantidades, ya que como se observa en la anterior Figura 11, se puede hacer seguimiento y control a los días en proceso que se encuentra cada banco por ciclo, lo cual permite saber y tomar la decisión de los bancos que han cumplido con su ciclo y de esta manera proceder a la fase de cosecha. También se puede ver las cantidades del inventario por variedad y ciclo, ayudando a obtener cantidades en tiempo real a las personas encargadas de los procesos, para de esta manera tener la disponibilidad de lechugas en la parte comercial, quien finalmente las vende a sus clientes.

También es importante en este punto mencionar que cada gráfico o tabla del dashboard se puede descargar en Excel para realizar un análisis adicional en caso de ser necesario.

**Figura 12**

*Métricas Generadas a Partir de los Datos (Visualización Calidad de Datos)*

Banco	Estado	Cantidad ▾	Estado / Dias en Alistamiento		
55	En proceso	1	Ciclo	Banco	Alistamiento
239	En proceso	1	7	154	25
140	En proceso	1		153	25
235	En proceso	1		152	25
35	En proceso	1		156	25
		1 - 100 / 189 < >		150	22

En la última parte del dashboard denominada “Calidad de datos”, en la cual se muestra mediante gráficos de tablas, la cantidad de bancos que se encuentran por estado y por ciclo, ya que de esta manera se puede controlar que los registros en cada etapa del proceso se hayan realizado de manera correcta. En caso de existir alguna inconsistencia, se podrán tomar acciones para corregir dicho registro, ya que se puede obtener la trazabilidad del mismo fácilmente, así como en la aplicación donde se registra, la cual permitirá realizar este tipo de ajustes de novedades.

## Conclusiones

El presente proyecto aplicado ha demostrado la viabilidad de implementar ciencia de datos en una MiPyme rural, específicamente en un cultivo de lechugas hidropónicas en el municipio de Tabio, Cundinamarca. A través de la identificación de una problemática real y la aplicación de herramientas de análisis de datos, se logró desarrollar una solución que aporta valor a la empresa y a su proceso productivo por medio de métricas que permiten generar controles y tomar decisiones basados en datos, desde la recolección y limpieza de datos hasta la construcción de un dashboard interactivo para la visualización y análisis de información de interés para el cultivo de lechugas.

También se lograron identificar los principales factores que impiden la implementación de ciencia de datos en MiPymes colombianas, como la falta de cobertura tecnológica y la ausencia de una cultura de datos, mediante el uso de herramientas utilizadas en la ciencia de datos, como lo es el procesamiento de datos con Python y el uso de la nube para la visualización de los mismos.

En los resultados y aportes del presente proyecto de grado, se logró el desarrollo de una herramienta de captura de datos a la medida, la cual se realizó mediante el diseño e implementación de una aplicación móvil utilizando Google AppSheet, la cual permite a los empleados registrar los datos de manera sencilla y eficiente. También se logró la construcción de una base de datos estructurada en Google Sheets para almacenar la información recopilada en la aplicación, lo que facilita el análisis y la generación de reportes. Y la implementación de un dashboard interactivo en la nube, en este caso en Looker Studio, la cual proporciona una visualización clara y concisa de los datos, permitiendo a los tomadores de decisiones identificar tendencias, patrones y oportunidades de mejora.

Con la implementación de las anteriores herramientas de bussines intelligence mencionadas, se abre la oportunidad a las partes interesadas a mejorar la toma de decisiones, ya que lo anterior a permitido a la MiPyme contar con información precisa y actualizada para tomar decisiones más informadas y estratégicas, optimizando así sus procesos productivos, inclusive ajustando la forma de hacer los procesos operativos, como lo fue el mejoramiento en la calificación de las lechugas con novedades (perdidas), para poder separarlas y registrarlas en la herramienta implementada.

Así como también se aportó al fortalecimiento de la cultura de datos, ya que por medio del presente proyecto se ha contribuido a fomentar una cultura de datos en la empresa seleccionada, incentivando el uso de la información para la mejora continua.

La principal enseñanza de este proyecto ha sido demostrar que las MiPymes rurales, aunque enfrentan limitaciones de infraestructura y cultura tecnológica, pueden adoptar soluciones de ciencia de datos siempre que estas sean simples, escalables y personalizadas. El éxito del proyecto radicó en diseñar una herramienta accesible a la medida que abordara específicamente los puntos críticos de la recolección y análisis de datos.

A partir de este trabajo, se logró identificar y empatizar sobre los desafíos específicos que enfrentan las MiPymes en entornos rurales para integrar soluciones tecnológicas. Se logra evidenciar en los resultados obtenidos en la Tabla 2, que se logró abordar en su totalidad las variables que se involucran en las actividades operativas de la empresa, las cuales carecían de registro estructurado. Este hallazgo dirigió el diseño de una solución que no solo automatizó estos procesos, sino que también fomentó el aprendizaje organizacional en torno al uso de datos.

Al abordar la problemática inicial — déficit de información — mediante herramientas específicas, como lo fue la aplicación móvil desarrollada, basada en Google AppSheet, se

permitió registrar un promedio de 30 entradas diarias de datos, alimentando de esta manera una base de datos estructurada y seguimiento constante por parte de las personas involucradas en los procesos. Además, el análisis realizado en el dashboard interactivo permitió identificar y cuantificar la situación actual del cultivo de lechugas, permitiendo llevar así un inventario real del mismo e impactando en la productividad al ajustar procesos en tiempo real.

Esta solución puede ser replicada en otras empresas del sector agrícola y en otros sectores productivos, adaptando las herramientas y metodologías a las características específicas de cada caso.

Finalmente se logra demostrar que la ciencia de datos puede ser una herramienta poderosa para impulsar el crecimiento y la competitividad de las MiPymes en Colombia. Al facilitar la toma de decisiones basadas en datos, abriendo de esta manera un camino para que otras empresas adopten prácticas similares y aprovechen los beneficios de la transformación digital.

## Recomendaciones

Se recomienda ampliar la solución a otros procesos de la empresa, como la comercialización y la gestión de inventarios, mediante la exploración de la integración de la solución con otras herramientas de gestión empresarial, como sistemas de planificación de recursos empresariales (ERP). Así como también se recomienda implementar programas de capacitación para los empleados de la empresa, con el objetivo de garantizar la sostenibilidad de la solución en el tiempo y poder generar una mejora continua a las herramientas desarrolladas. También es importante realizar estudios más profundos sobre el impacto de la implementación de la solución en la rentabilidad y la sostenibilidad de la empresa, en la medida en que se obtengan mas datos o al lograr completar por lo menos uno o dos años para poder construir métricas en función de tiempo y de esta manera analizar un panorama completo de datos históricos.

Es importante resaltar la importancia del uso de otras herramientas que brinda la ciencia de datos, como por ejemplo se recomienda explorar la implementación de modelos de predicción basados en machine learning, ya que para momento en el que se terminó de desarrollar el presente proyecto aplicado, no se contaba con una base de datos considerable para entrenar un modelo. Una vez se obtenga una base de datos con registros de por lo menos un año, por medio del uso de los mismos, se podrán entrenar modelos predictivos que permitan anticipar variables críticas, como el rendimiento del cultivo, la detección temprana de plagas o enfermedades, y la optimización del consumo de recursos (agua, nutrientes, energía). Lo cual no solo ayudaría a minimizar pérdidas, sino que también facilitaría una planificación más estratégica de las actividades agrícolas, como la siembra y cosecha, maximizando la productividad del cultivo.

Para fortalecer las capacidades de la solución desarrollada, se sugiere implementar el siguiente plan de acción, en el cual se detalla las fases, las actividades y los roles que podrían intervenir en cada etapa:

**Tabla 3**

*Plan de Acción Modelos de Machine Learning*

Fase	Actividad	Responsable
Preparación	Análisis de los datos recolectados en la base de datos.	
	Identificación de variables clave para modelos.	Ingeniero de Datos
Desarrollo	Limpieza y estructuración de los datos.	Analista de Datos
	Selección del modelo de <i>machine learning</i> adecuado (ej. regresión, clasificación).	Ingeniero de Datos
	Entrenamiento y validación del modelo.	Científico de Datos
	Integración del modelo a la base de datos.	Científico de Datos
	Prueba piloto del modelo con datos históricos.	Ingeniero de Datos
	Ajustes y optimización del modelo según resultados.	Científico de Datos
Pruebas	Implementación del modelo en el flujo de datos.	Ingeniero de Datos
	Capacitación al personal para el uso de nuevas funciones.	Analista de Datos
	Documentación y monitoreo del desempeño del modelo.	Equipo del Proyecto
Despliegue		

*Nota.* En esta tabla se muestra el detalle de las actividades para llevar a cabo del desarrollo e implementación de un modelo de machine learning al proyecto aplicado.

También es importante desarrollar un sistema de alertas automatizadas por medio de la implementación de algoritmos para identificar patrones anómalos en los datos, que puedan activar alertas en tiempo real. Por ejemplo, detectar variaciones inusuales en los parámetros de cultivo (como % de pérdida) que puedan requerir intervención inmediata. Lo cual incrementará

la capacidad de respuesta ante problemas operativos, reduciendo riesgos y asegurando la calidad del producto final.

En cuanto al dashboard interactivo, se recomienda complementarlo con funcionalidades predictivas y prescriptivas que sugieran acciones concretas basadas en los análisis, como por ejemplo, recomendaciones sobre fechas óptimas de siembra o cosecha en función de datos históricos, lo cual permitirá la toma de decisiones más estratégicas, basadas no solo en lo que ocurre, sino en lo que podría ocurrir y cómo actuar.

El presente proyecto aplicado implementado se podría replicar en otras MiPymes del sector agrícola o en diferentes cultivos, adaptando las herramientas y metodologías según las necesidades específicas ya que además de generar impacto positivo en otras empresas, la replicación del modelo puede abrir nuevas oportunidades de negocio, posicionando a la MiPyme como un caso de éxito en transformación digital. Para replicar la solución en otras MiPymes, se propone un plan detallado de implementación ajustado a las necesidades específicas de cada empresa:

**Tabla 4**

*Metodología para Replicar en Otras MiPymes*

Etapa	Actividad	Detalles
Diagnóstico Inicial	Recolección de información sobre la MiPyme y su operación.	- Análisis de la MiPyme, tamaño de la empresa y procesos.
Adaptación de Solución	Identificación de necesidades específicas.	- Evaluar los indicadores o KPIs son relevantes.
	Personalización de la base de datos en Google Sheets o BigQuery.	- Ajustar estructura según tipo de MiPyme.
	Desarrollo de una aplicación para recolección de datos en AppSheets.	- Adaptar formularios al tipo de producción.
	Configuración de dashboards en Looker Studio o Power BI.	

Capacitación	Capacitación al equipo para el uso de herramientas.	- Personalización de métricas (rendimientos, costos).
Implementación	Despliegue de la solución en la operación diaria.	- Talleres presenciales/virtuales de uso de la aplicación.
Monitoreo y Mejora	Evaluación periódica del impacto económico y optimización de los procesos.	- Monitoreo inicial de datos y ajustes necesarios. - Implementación de mejoras continuas.

*Nota.* En esta tabla se muestran las actividades para llevar la metodología en diferentes MiPymes.

### **Comentarios de Mejoramiento del Trabajo**

Es importante analizar y documentar el impacto económico generado por la implementación de la solución de inteligencia de negocios, destacando el ahorro en costos y el aumento en los ingresos. En cuanto a los costos, la digitalización de procesos y la automatización del registro de producción que posiblemente hayan reducido posibles errores humanos en la producción. Esto debe ser cuantificado mediante un análisis comparativo que incluya métricas específicas como el tiempo operativo, para demostrar cómo estas mejoras impactaron positivamente en la eficiencia operativa. Paralelamente, es esencial vincular este ahorro con un retorno de inversión (ROI) estimado, considerando los costos iniciales de implementación frente a los beneficios económicos obtenidos.

Por otro lado, el posible aumento en ingresos debe ser evaluado mediante indicadores claros, como el incremento en la producción de lechugas (debido al control de pérdidas) y decisiones comerciales más efectivas basadas en los insights generados. Documentar el impacto financiero incluye comparar los ingresos mensuales antes y después de la implementación, mostrando el crecimiento porcentual en ventas y el margen de beneficio por unidad. Asimismo, se recomienda incorporar tanto evidencia cualitativa, a través de entrevistas con trabajadores, como evidencia cuantitativa, mediante gráficos que ilustren la mejora operativa y el incremento en ingresos. Este análisis no solo fortalecerá las conclusiones del trabajo, sino que también servirá como modelo para replicar la solución en otras MiPymes del sector agrícola.

## Referencias Bibliográficas

- Beltran, C. Y. (2016). *"Kioscos Vive Digital": una estrategia digital y de conectividad para promover la inclusión social en comunidades rurales en Colombia*. 5.
- Comunicaciones, M. d. (30 de Mayo de 2024). *Datos*. Obtenido de [https://www.datos.gov.co/Ciencia-Tecnolog-a-e-Innovaci-n/Centros-de-Transformaci-n-digital/b84s-hkiv/about\\_data](https://www.datos.gov.co/Ciencia-Tecnolog-a-e-Innovaci-n/Centros-de-Transformaci-n-digital/b84s-hkiv/about_data)
- Comunicaciones, M. d. (30 de Mayo de 2024). *Datos*. Obtenido de [https://www.datos.gov.co/Ciencia-Tecnolog-a-e-Innovaci-n/Internet-Fijo-Penetraci-n-Municipio/fut2-keu8/about\\_data](https://www.datos.gov.co/Ciencia-Tecnolog-a-e-Innovaci-n/Internet-Fijo-Penetraci-n-Municipio/fut2-keu8/about_data)
- Daniela Rivero-Riqueme, Luis Felipe Ortiz-Clavijo. (2021). *Esquema de Flujo de Datos para la Toma de Decisiones en el Sector Público*. Revista Lasallista de Investigación, 11.
- Diego Armando Vanegas, Giovanni Mauricio Tarazona-Bermudez, Luz Andrea Rodríguez-Rojas. (2020). *Mejora de la toma de decisiones en ciclo de ventas del subsistema*. Revista Científica, 10.
- Efraín Alberto Oviedo Carrascal, Ana Isabel Oviedo Carrascal, Gloria Liliana Velez Saldarriaga. (2016). *Minería multimedia: hacia la construcción de una metodología y una herramienta de analítica de datos no estructurados*. 18.
- Eliana Esperanza Ordóñez Gómez, Nicolás Penagos Muñetón. (2015). *¿Cómo se entienden las tecnologías de la información y la comunicación (TIC) en los contextos educativos rural y urbano? Estudio comparativo*. 18.
- Iris Agustina Jiménez-Pitre, Raúl José Martelo Gómez, Juannys Chiquillo Rodelo, Denny de Jesús Lloreda Gracia, Maribel Sofía Morales Camacho. (2017). *Estrategias para el*

*empoderamiento digital y la integración de universidades con sectores clave para la sociedad colombiana.* 14.

Jiménez, S. A. (2021). *Retos del sistema financiero colombiano en la Cuarta Revolución Industrial.* 18.

Mónica Vélez O., Hermann Stuart Fuquen Gonzalez, Claudia Marcela Sánchez. (2020). *Desarrollo de nuevos procesos de operación y administración, soportados en innovación organizacional y herramientas de tecnologías de información para una empresa del sector automotor.* 6.

Pineda, R. (17 de 06 de 2022). *Aplicación de la Ciencia de Datos.* Obtenido de Unad:  
<https://repository.unad.edu.co/handle/10596/50443>

Pública, A. N. (07 de 03 de 2024). *Datos Abiertos.* Obtenido de  
[https://www.datos.gov.co/Gastos-Gubernamentales/SECOP-II-Contratos-Electronicos/jbjy-vk9h/about\\_data](https://www.datos.gov.co/Gastos-Gubernamentales/SECOP-II-Contratos-Electronicos/jbjy-vk9h/about_data)

Sarasa, A. (2016). *Introducción a las bases de datos NoSQL usando MongoDB.* Barcelona: Editorial UOC.

Toro, A. (22 de Marzo de 2024). *ubidots.* Obtenido de <https://es.ubidots.com/blog/open-implementing-precision-farming-in-kenya/>

UNAD, R. E. (13 de 10 de 2013). *Unad.* Obtenido de  
[https://sgeneral.unad.edu.co/images/documentos/consejoAcademico/acuerdos/2013/COA\\_C\\_ACUE\\_029\\_20131213.pdf](https://sgeneral.unad.edu.co/images/documentos/consejoAcademico/acuerdos/2013/COA_C_ACUE_029_20131213.pdf)

UNAD, R. E. (26 de 05 de 2020). *Unad.* Obtenido de  
[https://sgeneral.unad.edu.co/images/documentos/consejoAcademico/acuerdos/2020/COA\\_C\\_ACUE\\_034\\_20200526.pdf](https://sgeneral.unad.edu.co/images/documentos/consejoAcademico/acuerdos/2020/COA_C_ACUE_034_20200526.pdf)

Yurley T. Tovar, Andrés F. Calvo, Arley Bejarano. (2022). *Desarrollo de un sistema de clasificación de imágenes digitales para medir la humedad en granos de café*. 12.