

**Técnicas de Machine Learning aplicadas en el seguimiento de egresados universitarios:  
explorando el acceso al mercado laboral y la movilidad social**

Ferney Mauricio Calderón

Mauricio Esteban Ramírez Guzmán

Nidia Danigza Lugo López

Director de Trabajo de Grado

Dayana Alejandra Barrera Buitrago

Codirector de Trabajo de Grado

Universidad Nacional Abierta y a Distancia - UNAD

Escuela de Ciencias Básicas, Tecnología e Ingeniería - ECBTI

Especialización en Ciencia de Datos y Analítica

2024

### **Dedicatoria**

A familiares por creer en nosotros y su apoyo incondicional.

A amigos y compañeros que estuvieron en el proceso.

A todas las personas que creyeron en nosotros.

## **Agradecimientos**

Los autores ofrecen sus agradecimientos a:

La directora y codirectora del trabajo de grado,  
por la confianza depositada, los aportes realizados y  
la colaboración brindada en revisión y retroalimentación.

La Universidad Nacional Abierta y a Distancia y  
la Escuela de Ciencias Básicas, Tecnología e Ingeniería,  
por la formación recibida y los medios para alcanzar esta meta.

Por último y en forma especial a familiares, amigos y  
compañeros del programa por el apoyo y trabajo colaborativo.

## Resumen

Esta monografía tiene como objetivo proponer recomendaciones para el diseño de un modelo de Machine Learning para el seguimiento de egresados universitarios, su inserción en el mercado laboral y su movilidad social. Para lo cual, se lleva a cabo una revisión bibliográfica que permite identificar tendencias y patrones en la trayectoria profesional de los graduados universitarios, así como para evaluar las contribuciones, limitaciones y desafíos en este ámbito. Al aplicar un análisis de datos y de herramientas de Machine Learning, se busca no solo comprender los factores determinantes que influyen en la empleabilidad y la movilidad social de los egresados, sino también desarrollar soluciones predictivas que impulsen su éxito profesional y social en un entorno laboral dinámico.

Los hallazgos de este estudio informan sobre la situación actual de los egresados universitarios y destacan la relevancia de las técnicas de ciencia de datos para orientar el diseño e implementación de políticas universitarias centradas en el seguimiento y apoyo efectivo de los graduados en su trayectoria profesional, lo cual crea posibilidades para la implementación de políticas más efectivas y dirigidas, que puedan responder de manera más precisa a las necesidades y desafíos que enfrentan los egresados en su camino hacia el éxito laboral y social.

***Palabras clave:*** Egresados, Mercado Laboral, Movilidad Social, Machine Learning

### **Abstract**

This monograph proposes recommendations for designing a Machine Learning model for tracking university graduates, their integration into the job market, and their social mobility. To achieve this, a literature review is conducted to identify trends and patterns in the professional trajectory of university graduates, as well as to evaluate the contributions, limitations, and challenges in this field. By applying data analysis and Machine Learning tools, the goal is not only to understand the determining factors that influence graduates' employability and social mobility but also to develop predictive solutions that enhance their professional and social success in a dynamic work environment.

The findings of this study provide insights into the current situation of university graduates and underscore the relevance of data science techniques in guiding the design and implementation of university policies focused on tracking and effectively supporting graduates in their professional trajectories. This creates opportunities for implementing more effective and targeted policies that can better address the specific needs and challenges graduates encounter on their path to achieving professional and social success.

***Keywords:*** Graduates, Labor Market, Social Mobility, Machine Learning

**Tabla de contenido**

	<b>Pág.</b>
Introducción .....	10
Planteamiento del problema.....	11
Justificación .....	15
Objetivos .....	18
Objetivo general.....	18
Objetivos específicos .....	18
Marco conceptual.....	19
Tendencias de acceso al mercado laboral y movilidad social en egresados .....	26
Análisis Bibliométrico con VOSviewer.....	32
Aplicaciones existentes de Machine Learning.....	34
Recomendaciones para la implementación del modelo .....	42
Conclusiones .....	48
Referencias bibliográficas.....	50
Apéndices.....	56

**Lista de tablas**

	<b>Pág.</b>
<b>Tabla 1.</b> <i>Clasificación de las técnicas de Machine Learning</i> .....	22
<b>Tabla 2.</b> <i>Palabras claves para la búsqueda</i> .....	27
<b>Tabla 3.</b> <i>Criterios de selección de referencias</i> .....	28
<b>Tabla 4.</b> <i>Antecedentes por idioma, tipo de documento y país</i> .....	36
<b>Tabla 5.</b> <i>Técnicas de Machine Learning utilizadas</i> .....	37

**Lista de figuras**

	<b>Pág.</b>
<b>Figura 1.</b> <i>Graduados según tipo de cotizante y nivel de formación</i> .....	12
<b>Figura 2.</b> <i>Referencias por países latinoamericanos en Scopus</i> .....	29
<b>Figura 3.</b> <i>Referencias por todos los países en Scopus</i> .....	31
<b>Figura 4.</b> <i>Referencias por todos los países en Scopus</i> .....	31
<b>Figura 5.</b> <i>Mapa de coocurrencia de palabras claves con VOSviewer</i> .....	33
<b>Figura 6.</b> <i>Técnicas de Machine Learning utilizadas</i> .....	38
<b>Figura 7.</b> <i>Recomendaciones propuestas para el modelo</i> .....	42
<b>Figura 8.</b> <i>Metodología de Machine Learning - RAPEAD</i> .....	47

**Lista de apéndices**

	<b>Pág.</b>
<b>Apéndice A.</b> <i>Resultados de referencias bibliográficas</i> .....	56

## Introducción

La transición de los graduados universitarios al mercado laboral es un proceso de gran importancia que no solo influye en la posición social individual de estos, sino también en el panorama socioeconómico de una sociedad, a nivel nacional, regional o local. Es por esto que el seguimiento de los egresados universitarios, tanto en su acceso al mercado laboral como en su movilidad social o cambio de posición económica, es crucial para entender las barreras y desafíos que enfrentan, con el objetivo de mejorar este proceso y generar un mayor valor social.

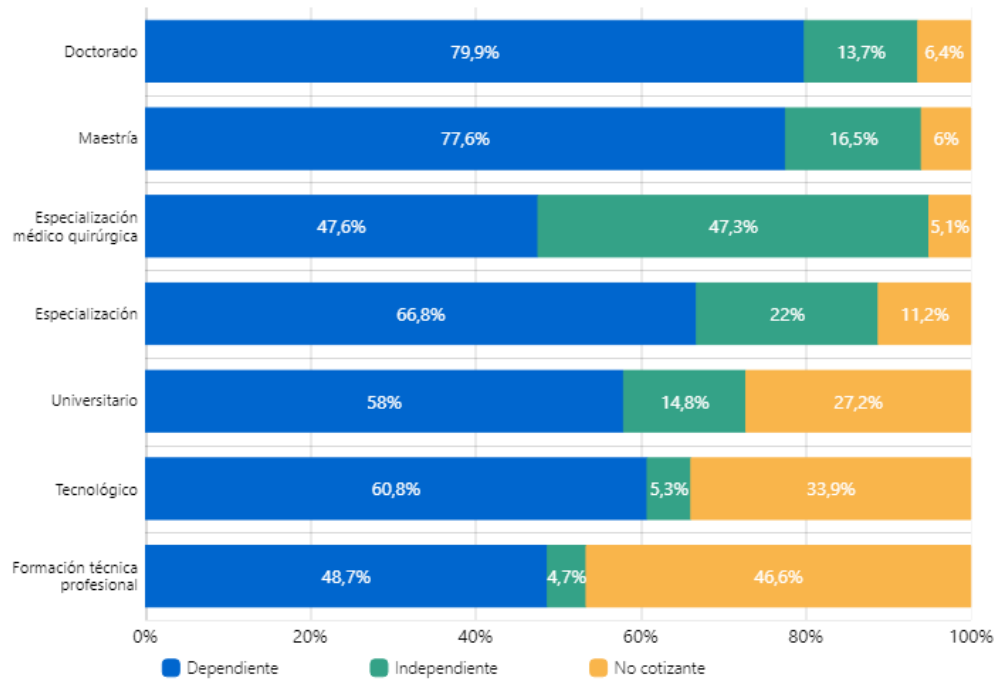
En esta investigación, se lleva a cabo una revisión bibliográfica sobre las tendencias y desafíos del mercado laboral y la movilidad social de los egresados en Latinoamérica y se revisan las aplicaciones de técnicas de Machine Learning en esta área, destacando los diversos enfoques y resultados obtenidos en diferentes países. Para garantizar la rigurosidad y la coherencia en el análisis se realiza una revisión de literatura mediante análisis bibliométrico, que proporciona una estructura que facilita la identificación, selección y síntesis de los estudios relevantes, permitiendo una comprensión profunda del panorama actual.

A partir de los resultados obtenidos, se proponen recomendaciones para la implementación de modelos de Machine Learning que permitan mejorar los procesos de inserción laboral y movilidad social de los egresados universitarios. Al aprovechar el potencial de estas técnicas, se busca no solo identificar los factores clave que influyen en la empleabilidad y la movilidad social, sino también desarrollar soluciones predictivas que impulsen el éxito profesional y social de los egresados en un entorno laboral en constante evolución.

### **Planteamiento del problema**

La educación superior es crucial en la formación de individuos y en la promoción de la movilidad social, ya que se espera que los egresados puedan acceder al mercado laboral y mejorar su posición socioeconómica, pero a pesar de los esfuerzos continuos en la expansión de las universidades persisten grandes desafíos. Actualmente, la desigualdad en el acceso al mercado laboral es uno de los principales problemas, relacionados con factores como el origen socioeconómico, el género, la raza y la ubicación geográfica. Estas diferencias pueden ocasionar que algunos egresados tengan problemas para encontrar empleo que vaya acorde con sus capacidades, su experiencia y en línea con su carrera profesional. En varias ocasiones la problemática inicia desde el desconocimiento de oportunidades disponibles, en el transcurso y al culminar su ciclo universitario, como la movilidad, posibilidad de becas para estudios de posgrado u ofertas laborales disponibles para su contratación.

En Colombia, el Ministerio de Educación Nacional (2023) presenta el Observatorio Laboral para la Educación (OLE) como un sistema de información especializado que analiza la relevancia de la educación superior, mediante el seguimiento de los graduados del país y su inserción en el mercado laboral colombiano. En el último reporte disponible de los graduados en el año 2020 - vinculados en el año 2021, se indica que de los 426.735 recién graduados en este periodo, el 75.1% se encuentran vinculados al mercado laboral formal de la economía, con 260.439 dependientes y 60.160 independientes. En la Figura 1 se observa la distribución de los graduados de acuerdo con el tipo de cotizante (dependiente, independiente o no cotizante) y nivel de formación (posgrados, pregrado, tecnológico o técnico) y en él se refleja que entre menor es el grado de formación, mayor es la informalidad o el desempleo.

**Figura 1.***Graduados según tipo de cotizante y nivel de formación*

Según los datos recopilados por El País (2023) acerca de los ingresos de los recién graduados en Colombia, se revela que el 39 % de los jóvenes que finalizan sus estudios universitarios, tras invertir en carreras costosas, perciben ingresos que oscilan entre 1 y 1,5 salarios mínimos. Considerando que en 2023 el salario mínimo legal vigente en Colombia es de aproximadamente \$1.160.000 (equivalente a unos US\$286), esto implicaría ganancias entre 1.160.000 y 1.740.000 pesos. Por otro lado, el 31,9 % de los graduados universitarios experimentan una situación financiera más favorable, con ingresos que van desde 1,5 hasta 2,5 salarios mínimos, mientras que solo el 28 % logra superar los 2,5 salarios mínimos en su primer año laboral. En cuanto a los técnicos profesionales, se observa que el 70 % percibe ingresos inferiores a 1,5 salarios mínimos.

Otros estudios, como el de Martínez (2022), mencionan que en Colombia el 45% de graduados de universidades públicas obtienen empleos no calificados, frente a solo 19% de universidades privadas, lo cual demuestra que existen brechas importantes según tipo de institución; los graduados de universidades de elite, reconocidas por su excelencia académica y contar con mejores recursos, tienen cinco veces más probabilidades de movilidad social que graduados de instituciones menos prestigiosas. Además, se observa que las mujeres graduadas tienen un 21% menos de probabilidades de experimentar movilidad social ascendente que los hombres y que solo el 12 % de graduados supera el estrato socioeconómico de sus padres, lo cual refleja una movilidad social ascendente limitada.

A nivel de Latinoamérica la situación no varía mucho, el estudio de Filgueira et al. (2022) mencionan que en general, la movilidad social intergeneracional es baja, un hijo de familia pobre tiene solo 20 a 30% de probabilidades de alcanzar un estrato medio. Entre 2010 y 2019, la movilidad social regional paso de 41.1% a 40.3%, lo cual indica estancamiento y poca mejoría en la última década; solo 5 de 17 países latinoamericanos (Argentina, Chile, Ecuador, Nicaragua y República Dominicana) registran mejorías marginales en movilidad social intergeneracional dentro del período. Se requerirían entre 4 a 7 generaciones en promedio para que los descendientes de una familia del quintil más pobre alcancen el ingreso medio de cada sociedad y la educación parece tener un efecto ambiguo: expande expectativas, pero con limitada movilidad socioeconómica efectiva posterior.

Esta información coincide con lo mencionado en el Informe de Tendencias Sociales y Educativas del Sistema de Información de Tendencias Educativas de América Latina [SITEAL] (2021), donde menciona que solo el 34% de los jóvenes latinoamericanos logra acceder a la educación superior, con la cobertura más baja a nivel mundial. Además, el 63% de universitarios

de la región señala no haber recibido orientación efectiva sobre opciones laborales antes de su graduación y el desempleo juvenil de egresados universitarios alcanza niveles de 25% en promedio en la región. La tasa de informalidad laboral de recientes egresados universitarios es de 62% en Perú, 45% en Colombia y 38% en Chile. Esta brecha salarial entre egresados de diferente origen socioeconómico llega a ser de hasta 60% en países como México y Colombia.

Otro ejemplo por mencionar es el caso de México, donde según el estudio de Romero (2021) sobre inserción laboral de egresados universitarios, menciona que el 83% de los egresados universitarios de estratos socioeconómicos bajos acceder a trabajos vulnerables, en contraste con el 56% de estratos altos; además, que los egresados de universidades públicas obtienen una remuneración 17% menor respecto a egresados de instituciones privadas, controlando la carrera. Por otro lado, el salario inicial promedio de mujeres egresadas representa solo el 76% del ingreso de los hombres y el 60% de graduados de zonas rurales requiere migrar a ciudades para insertarse en empleos calificados.

Ante estos desafíos, surge la oportunidad de utilizar técnicas de Machine Learning para abordar el problema del acceso al mercado laboral y la movilidad social de los egresados universitarios. Estas técnicas pueden contribuir con el análisis de gran cantidad de datos, que permite clasificarlos para una mejor comprensión de los factores que influyen en la trayectoria de los egresados y la posibilidad de pronosticar oportunidades de mejora que puedan servir como estrategias efectivas para mejorar su perspectiva profesional. Para llevar a cabo el presente trabajo, se inicia con la pregunta de investigación: ¿Qué recomendaciones pueden ser utilizadas para diseñar un modelo de Machine Learning destinado a mejorar el acceso al mercado laboral y la movilidad social de egresados universitarios?

### **Justificación**

Las carencias con las que crecen las personas y la falta de oportunidades son las principales causas de la desigualdad para la movilidad social; se han encontrado variables como la educación de los padres, el género, el desarrollo de las regiones o la etnia, que deben considerarse determinantes al momento de revisar este fenómeno. Dado lo anterior, se considera que la educación superior junto a la posibilidad de empleo, son unos de los vehículos principales para que los individuos puedan tener movilidad social (Consuegra et al., 2020).

La educación superior es un motor clave para el progreso social y económico, se espera que esta brinde a los egresados las habilidades y oportunidades necesarias para tener éxito en el mercado laboral y mejorar su calidad de vida. La desigualdad socioeconómica, la brecha de oportunidades y las cambiantes dinámicas del mercado laboral, especialmente en un mundo en constante cambio, donde las tendencias laborales y las demandas evolucionan aceleradamente, los profesionales deben tener presente que la evolución de las dinámicas del mercado laboral ha transformado la naturaleza de los empleos y las habilidades requeridas.

El informe del Banco Interamericano de Desarrollo (BID) (2020) menciona datos relevantes sobre los cambios en la dinámica del mercado laboral, donde se puede observar que el 65% de los niños que empiezan la escuela primaria en América Latina, terminarán trabajando en puestos que aún no han sido inventados. Se debe tener en cuenta que más de 100 millones de empleos en América Latina podrían ser automatizados por máquinas en las próximas décadas; además, el 42% de las habilidades básicas requeridas para los trabajos de 2020 cambiarán completamente respecto a lo utilizado en 2015 y la penetración del trabajo independiente en países como Colombia y Argentina roza el 30% - 40% de la fuerza laboral.

Por su parte, el informe del World Economic Forum (2023) proyecta que cerca del 23% de los empleos tendrán cambios en los próximos cinco años y se resalta un mayor crecimiento en sectores como educación, agricultura y comercio digital. Además, se indica que la adopción de tecnología seguirá siendo crucial para la transformación empresarial, donde el 86% de los encuestados esperan que la ampliación del acceso digital transforme su organización y el 52% esperan que genere crecimiento laboral. Aunque no se proporciona información específica sobre movilidad social, se resalta que tecnologías como el análisis de datos, el cambio climático, la gestión ambiental y la ciberseguridad impulsan el crecimiento del empleo; también se destaca la importancia de capacitar los trabajadores en inteligencia artificial y otras tecnologías emergentes.

El informe del Desarrollo Mundial 2019 del Banco Mundial (Berryhill et al., 2019) destaca la importancia de la educación y la formación para mejorar la empleabilidad y la movilidad social de los jóvenes; además señala que la tecnología puede ser una herramienta útil para mejorar la calidad y la accesibilidad de la educación, y que la inteligencia artificial y el Machine Learning pueden ser utilizados para personalizar el aprendizaje y adaptarlo a las necesidades y habilidades de cada uno.

En la actualidad, es crucial considerar la Agenda 2030, como el compromiso firmado por más de 150 países durante la histórica Cumbre del Desarrollo Sostenible, donde se pacta mejorar las condiciones de vida y transformar el mundo en un lugar más próspero a través de los Objetivos de Desarrollo Sostenible (ODS) (United Nations, 2016). Entre los ODS, el más enfocado por las instituciones de educación superior es el número 4, que aborda la educación de calidad, seguido por otros como justicia y paz, igualdad de género, empleo y economía, salud y acceso al agua. Estos objetivos se han abordado con enfoques tanto disciplinarios como interdisciplinarios, según señala Consuegra et al. (2020).

Para afrontar estas metas y retos que se presentan, los avances tecnológicos y la ciencia de datos ofrecen nuevas posibilidades que permiten abordar problemas complejos y realizar análisis más profundos y personalizados. Sin embargo, el potencial de las técnicas de Machine Learning aún no se ha explorado completamente en el seguimiento de egresados y la mejora de la movilidad social. La presente monografía se centra en la revisión y análisis de investigaciones previas para comprender el seguimiento de egresados universitarios y su acceso al mercado laboral y la movilidad social; además, se explorará la aplicación de técnicas de Machine Learning como herramienta potencial para abordar esta problemática. Este trabajo busca proporcionar una base sólida para futuras investigaciones y políticas educativas destinadas a reducir la desigualdad y mejorar las oportunidades para los egresados universitarios en su transición al mundo laboral.

## **Objetivos**

### **Objetivo general**

Proporcionar recomendaciones para la implementación de técnicas de Machine Learning en un modelo de seguimiento a egresado que explore el acceso al mercado laboral y la movilidad social, partiendo de una revisión exhaustiva de literatura.

### **Objetivos específicos**

Identificar las tendencias y patrones actuales en el acceso al mercado laboral y la movilidad social de los egresados, basándose en referentes bibliográficos y datos históricos.

Analizar las aplicaciones existentes de Machine Learning con enfoque en el acceso al campo laboral y la movilidad social de los egresados.

Formular recomendaciones para la implementación de un modelo de Machine Learning de seguimiento de egresados, de acuerdo con la información recopilada.

### **Marco conceptual**

La movilidad social se refiere al cambio en la posición socioeconómica de un individuo a lo largo de su vida, lo que implica la capacidad de una persona para mejorar su posición económica y social en comparación con la de sus padres (Grajales et al., 2015). Este concepto es crucial para comprender la dinámica de una sociedad, además, es un indicador clave del éxito de la educación superior, ya que se espera que esta proporcione a los egresados las herramientas necesarias para acceder al mercado laboral y mejorar su posición socioeconómica. Sin embargo, la falta de movilidad social puede contribuir a la continuidad de desigualdades, donde se limita el potencial de desarrollo económico y social de una sociedad. En Latinoamérica, la movilidad social, especialmente las transformaciones estructurales, las disparidades socioeconómicas y las tendencias sociales, ha sido objeto de estudio y análisis en algunos informes y publicaciones de la Comisión Económica para América Latina y el Caribe (CEPAL, 2020).

De acuerdo con Aké-Uitz (2022) se menciona que en México la educación superior se percibe comúnmente como un factor determinante del éxito económico en el futuro de las personas, considerándose un posible catalizador de la movilidad social ascendente. Sin embargo, las opciones en el sistema educativo pueden dar lugar a patrones distintivos según el origen socioeconómico y los resultados financieros tras graduarse de una institución de educación superior específica. En muchas ocasiones se ve afectado el nivel educativo en la promoción de la movilidad social dependiendo de qué institución educativa provenga, es decir, que las personas de estratos socioeconómicos bajos que logran graduarse de Instituciones de Educación Superior (IES) podrían enfrentar obstáculos para progresar en términos socioeconómicos, por lo cual, es fundamental garantizar una educación de calidad y dar oportunidades equitativas a estudiantes de bajos recursos para acceder a instituciones prestigiosas y reconocidas en el campo educativo.

De acuerdo con el estudio de Ramirez (2022), donde se utilizan datos de una encuesta a egresados de la Universidad Nacional Autónoma de México (UNAM) del año 2012, con una muestra de 7.482 egresados, se menciona que la mayoría se enfrentan a condiciones laborales precarias, en promedio el 66.6% de los egresados se encuentran en condiciones bajas de ocupación, mientras que solo un 8.3% están en condiciones buenas y mejores. Además, el 25% de la población egresada se encuentra en una situación de precariedad laboral o empleabilidad espuria en los segmentos de licenciatura completa, incompleta y especialización. Estos resultados muestran un retorno de la inversión en los egresados en todos los segmentos, más acentuado en los niveles de maestría y doctorado, donde en promedio el 66.6% encuentra condiciones bajas de ocupación, contrastada con un promedio de 8.1% en buenas y solo el 0.2% en promedio en las mejores condiciones. Este trabajo resalta la importancia de vincular la educación superior con el sector productivo y mejorar las condiciones laborales.

El estudio de Maquera et al. (2021) presenta un modelo multivariado diseñado para analizar los factores socioeconómicos y culturales que afectan la inserción laboral de los graduados de la Universidad Nacional de Moquegua, en Perú. Basándose en datos recopilados de 121 egresados, el estudio revela varias percepciones entre los encuestados. Por ejemplo, el 44.6% de los graduados consideran que existen oportunidades laborales, pero afirman que la educación universitaria por sí sola no garantiza un empleo de calidad, además, el 30.6% opina que la oferta de trabajo en su campo es limitada, mientras que el 19% señala que, aunque hay oportunidades laborales, los contratos son inestables o precarios, y solo el 5.8% percibe una amplia oferta de trabajo. Es importante destacar que se observan resultados desfavorables entre los estudiantes de estratos socioeconómicos más bajos y aquellos de culturas diversas.

De igual forma Vieira (2023) en su estudio con modelos logísticos y multinomiales para analizar acceso al empleo formal y ocupaciones típicas en egresados de educación superior en Rio de Janeiro, Brasil, menciona que las desigualdades sociales en la entrada al empleo formal están relacionadas con la posición social de la familia de origen del graduado y con las diferencias cualitativas de la educación superior; aunque la educación superior puede ser una herramienta para igualar oportunidades de entrada al mercado laboral, hay factores adicionales que pueden influir en la posición social de un individuo. Además, se discuten algunas limitaciones, como el papel del Estado y las instituciones educativas para garantizar la equidad y la calidad, y la necesidad de principios de pertinencia y responsabilidad social en el diseño de políticas que permitan anular el efecto del origen socioeconómico sobre el destino social.

El Machine Learning, como parte integral de la inteligencia artificial, permite a las máquinas aprender a partir de datos y la aplicación de algoritmos específicos que le permiten ejecutar tareas sin una programación manual explícita para cada acción. Según lo señalado por Goodfellow et al. (2016), este campo se enfoca en el desarrollo de algoritmos y técnicas que permiten a los sistemas mejorar su desempeño mediante la experiencia y datos disponibles. Estos algoritmos les confieren la capacidad de identificar patrones en grandes volúmenes de datos, clasificar información y realizar predicciones (Raschka & Vahid, 2017).

Según un artículo de Forbes (2022), las habilidades más demandadas en Inteligencia Artificial (IA) incluyen programación en Python, R, C++ y Java, conocimientos matemáticos en campos como estadística, álgebra lineal y cálculo inferencial, así como habilidades en Machine Learning y Deep Learning. De igual forma, el Informe 2023 Artificial Intelligence Index (Maslej et al., 2023) destaca que el Machine Learning es una de las técnicas más utilizadas en la actualidad, y que se espera que su uso siga creciendo en el futuro. Por lo cual, se observa una

oportunidad de aplicación de estas técnicas en la movilidad social y el mercado laboral de egresados universitarios, con el fin de analizar datos y extraer patrones que permitan identificar oportunidades de empleo y mejorar la toma de decisiones, así como desarrollar sistemas de recomendación personalizados que ayuden a los egresados a encontrar empleos acordes a sus habilidades y preferencias. En el ámbito de la formación, el Machine Learning puede ser utilizado para personalizar el aprendizaje y adaptarlo a las necesidades y habilidades de cada estudiante, lo que puede mejorar su rendimiento académico y su empleabilidad.

En la Tabla 1 se menciona la clasificación de las técnicas de Machine Learning mayormente utilizadas, junto a su principal característica y aplicaciones:

**Tabla 1.**

*Clasificación de las técnicas de Machine Learning*

	<b>Característica</b>	<b>Aplicaciones</b>
Supervisado	Los algoritmos aprenden a partir de datos etiquetados, donde se proporcionan tanto las entradas como las salidas deseadas.	Modelos de clasificación y regresión
No supervisado	Los algoritmos obtienen conocimiento únicamente de los datos de entrada sin etiquetas, descubriendo patrones ocultos o agrupaciones.	Clustering, asociación, reducción de dimensionalidad
Por refuerzo	Se basa en el concepto de incentivos y consecuencias, donde los agentes aprenden mediante la interacción con su entorno.	Juegos, robótica, toma de decisiones secuenciales
Semi-Supervisado	Fusiona características del aprendizaje supervisado y no supervisado al emplear un conjunto limitado de datos etiquetados con un volumen considerable de datos sin etiquetar para el entrenamiento de modelos.	Reconocimiento de voz, imágenes o texto en procesamiento de lenguaje natural (NLP)

Fuente: Adaptado de Hastie et al. (2009) y Chapelle et al. (2006).

Al aplicar técnicas de Machine Learning es fundamental medir los modelos mediante métricas que permitan evaluar su rendimiento, como por ejemplo, la exactitud que mide la proporción de predicciones correctas, la precisión que evalúa la proporción de verdaderos positivos entre los casos predichos como positivos, la recuperación o sensibilidad que mide la proporción de verdaderos positivos sobre todos los casos que realmente son positivos o la puntuación F1 como la media armónica de la precisión y la recuperación. También existe la curva ROC (Receiver Operating Characteristic) que muestra la tasa de verdaderos positivos frente a la tasa de falsos positivos a diferentes umbrales de decisión y su área bajo la curva (AUC) que mide la capacidad del modelo para distinguir entre clases. Por otra parte, la matriz de confusión detalla las predicciones correctas e incorrectas por clase para obtener una visión detallada del rendimiento del modelo y finalmente, la validación cruzada que permite dividir el conjunto de datos en subconjuntos para realizar múltiples rondas de entrenamiento y evaluación en diferentes combinaciones, lo cual permite una estimación más precisa de su rendimiento. Estas métricas combinadas aseguran una evaluación exhaustiva y robusta (Hastie et al., 2009).

En la investigación de Mason et al. (2023) sobre decisiones profesionales y capacitación en egresados universitarios, se realiza un análisis de datos existentes en línea, junto con taxonomías de habilidades y combinación con técnicas de Machine Learning. El estudio muestra cómo estos conjuntos de datos integrados pueden ser utilizados para lograr una mayor alineación entre las necesidades y ofertas de educadores, empleadores y buscadores de empleo. Los datos integrados pueden ser utilizados para ayudar a los trabajadores que se ven afectados por la tecnología a identificar trayectorias profesionales alternativas, asistir a los educadores a identificar brechas en sus ofertas de cursos y apoyar a los estudiantes para comunicar el valor de su formación a los empleadores Venegas et al. (2021) realiza un estudio en el que examina los

factores que determinaron los resultados en el mercado laboral de los recién graduados de educación técnica en Chile, para lo cual usa cinco métodos de aprendizaje automático, SVM, Naive Bayes, CART, Random Forest y XGBoost, junto a la recolección de datos de encuesta a graduados sobre atributos sociodemográficos, capital humano, desempeño académico. Este estudio resalta que ciertos atributos personales, académicos y de capital humano son más determinantes de los resultados laborales que el origen socioeconómico entre los egresados.

En el estudio de Moreno et al. (2020) se explora el potencial de las técnicas de minería de datos para estimar la reputación universitaria a partir de datos de empleabilidad de graduados, donde analizó datos de empleo de graduados universitarios, incluyendo variables como salario, tipo de contrato, sector económico, entre otros. Mediante técnicas como reglas de asociación, árboles de decisión y algoritmos genéticos, se logra establecer asociaciones entre características de las universidades y resultados de empleabilidad. Este estudio menciona que estas técnicas permiten explicar alrededor del 65% de la variabilidad en los indicadores de reputación universitaria y destaca el potencial de utilizar datos de graduados para crear nuevos rankings universitarios o mejorar los existentes.

En el estudio de Azpíroz et al. (2023) se emplea una clasificación basada en redes neuronales con el fin de pronosticar el nivel de satisfacción de algunos graduados de distintos programas de posgrado en el campo de la salud de una institución educativa latinoamericana que emplea una metodología de aprendizaje en línea (e-learning). Como resultado, se obtuvo un instrumento válido y confiable para evaluar la satisfacción con el programa académico cursado. El estudio incluyó el desarrollo de un modelo de red neuronal destinado a clasificar y predecir el grado de satisfacción, lo cual contribuye a reducir la brecha entre la calidad educativa ofrecida y la satisfacción de los egresados mediante el uso de AI. El índice medio de satisfacción global de

los egresados es 2.66 sobre 4, con un 80% reportando nivel de satisfacción media-alta. Se observa una mejor valoración de los recursos disponibles, mientras que la comunicación y el contexto socioeconómico fueron áreas con calificaciones más bajas. Además, se logra una precisión del 96.8% en la clasificación mediante Red Neuronal Artificial entrenada.

Respecto a los salarios de egresados, en el estudio de Wang et al. (2022) se utiliza modelos de Regresión Logística para predecir probabilidad de salarios altos. Este trabajo realiza estudio empírico de los factores que influyen en el salario inicial de los estudiantes universitarios de economía y finanzas, donde se propone una prueba sobre cómo mejorar su competitividad y obtener un salario alto, pronosticando las posibilidades para obtenerlo. El estudio concluye que los factores principales son el capital humano (educación, habilidades), capital social (ser líder estudiantil) y segmentación del mercado laboral (regiones y sectores con mayores salarios), además que la profesión influye en obtener un salario inicial elevado, por ejemplo, los médicos tienen más posibilidades de un salario alto que los especialistas en comunicaciones inalámbricas e informática móvil y se sugiere que los estudiantes pueden aumentar su probabilidad al mejorar su capital humano o habilidades obtenidas mediante la educación, formación y desarrollo personal (estudiando carreras mejor remuneradas).

En Colombia, Buitrago y Romero (2022) emplearon datos del Observatorio Laboral de Educación Superior (OLE) correspondientes al periodo 2012-2016, con una muestra de 21.140 graduados recientes de 177 instituciones de educación superior, con el fin de evaluar técnicas de Machine Learning para encontrar el modelo predictivo más preciso que permita identificar los factores que influyen en sus salarios, utilizando como variables de interés el género, el programa académico, la institución universitaria y su ranking, para predecir la categoría salarial. Los resultados fueron: Árbol de Decisión (DT) con 68%, Perceptrón Multicapa con 19%, Máquinas

de Soporte Vectorial (SVM) con 56% y K-Means con 70,19%. Por consiguiente, se procedió a desarrollar el modelo de K-Means utilizando el software WEKA, el cual se determina que tanto el género como el programa académico ejercen una influencia significativa en los salarios, lo que pone de manifiesto la brecha salarial de género entre los graduados.

Otro ejemplo de aplicación de Machine Learning para predecir la empleabilidad de los egresados, es el estudio de ElSharkawy et al. (2022), donde se realiza una investigación explicativa y predictiva, mediante encuesta a graduados de tecnología de la información y empleadores, para obtener datos sobre habilidades técnicas, blandas, capacitaciones, entre otras; además de la aplicación de 5 algoritmos: Regresión Logística, Árbol de Decisión (DT), Bosques Aleatorios (RF), Naïve Bayes Gaussiano (Gaussian NB) y Máquinas de Soporte Vectorial (SVM). Como resultado, se observa que el DT tuvo la mayor precisión (100%), para predecir la empleabilidad de los graduados y las habilidades más relevantes fueron las técnicas en demanda por la industria como análisis de datos e inteligencia artificial. Mientras que la segunda precisión más alta la logra Regresión Logística y SVM con 98% y la precisión más baja con 92% la logra Gaussian NB. Se concluye que el modelo puede ayudar a que las instituciones educativas adapten sus planes de estudio para una mejor preparación de los estudiantes, mejorando los métodos de enseñanza y aprendizaje e incluyendo las competencias demandadas por la industria.

### **Tendencias de acceso al mercado laboral y movilidad social en egresados**

Para la revisión bibliográfica sobre el uso de técnicas de Machine Learning aplicada al mercado laboral y la movilidad social en egresados universitarios, inicialmente se realizan búsquedas generales en sitios web abiertos y la plataforma de Google Académico, con la finalidad de tener una comprensión inicial de la temática y poder obtener palabras claves, junto

con sus sinónimos o variantes ortográficas, para la construcción de una matriz de palabras que sirve de base inicial para la organización de la ecuación de búsqueda que es utilizada para encontrar referencias más formales de bases de datos bibliográficas:

**Tabla 2.**

*Palabras claves para la búsqueda*

Graduate (Egresados)	Graduate students (estudiantes graduados)	University alumni (Exalumnos universitarios)	follow-up graduate (Seguimiento egresados)	Alumni survey (encuesta a exalumnos)
Labor market (Mercado laboral)	Job placement (Colocación laboral)	Employment outcomes (Resultados de empleo)	Social mobility (Movilidad social)	Income inequality (Desigualdad de ingresos)
Machine Learning (Machine Learning)	Predictive modeling (Modelado predictivo)	Data Analytics (Análisis de datos)	Data Mining (Procesamiento de datos)	Multivariate Analysis (Análisis multivariable)
Regression or Classification (Regresión o Clasificación)	Supervised Learning (Aprendizaje supervisado)	Unsupervised Learning (Aprendizaje no supervisado)	Clustering algorithm (Algoritmos de clustering)	Deep learning (Aprendizaje profundo)

A partir de estas palabras, se observa que existen tres grupos de palabras importantes, las relacionadas con los graduados, las del mercado laboral o la movilidad social y todo lo referente a Machine Learning. Con ellas se procede a construir la ecuación de búsqueda inicial:

*(graduate OR “graduate student” OR “university alumni” OR “follow-up graduate” OR “alumni survey”) AND (“labor market” OR “job placement” OR “employment outcome” OR “social mobility” OR “income inequality”) AND (“machine learning” OR “predictive modeling” OR “data analytic” OR “data mining” OR “multivariate analysis OR “regression model” OR “classification model” OR “supervised learning” OR “unsupervised learning” OR “clustering algorithm” OR “deep learning”)*

Se seleccionan las fuentes de información más relevantes para el estudio:

- Plataformas de investigación académica de acceso por Biblioteca como Scopus, Web of Sciences, EBSCO, Proquest, entre otros recursos con contenido de editoriales reconocidas como la IEEE, Elsevier, Springer, Emerald, Wiley, Taylor & Francis.
- Plataformas de investigación académica de acceso abierto como Google Scholar, ResearchGate, Academia.edu y repositorios institucionales.
- Recursos y publicaciones relacionadas con el análisis de datos en educación: Asociación Internacional de Analítica de Datos Educativos (EDM Society) y en Colombia el Observatorio Laboral para la Educación (OLE).

De ellas, se define qué para poder llevar a cabo la búsqueda sistemática de literatura y la selección de contenido de mayor relevancia en la investigación, se inicia incluyendo esta ecuación de búsqueda en la base de datos Scopus, con información referencial de una variedad de revistas científicas, libros, conferencias y otras fuentes con el mayor factor de impacto. Al realizar esto, se observa un total de 3,741 resultados, en los cuales se procede a implementar los criterios de selección establecidos para la filtración de las referencias más adecuadas:

**Tabla 3.**  
*Criterios de selección de referencias*

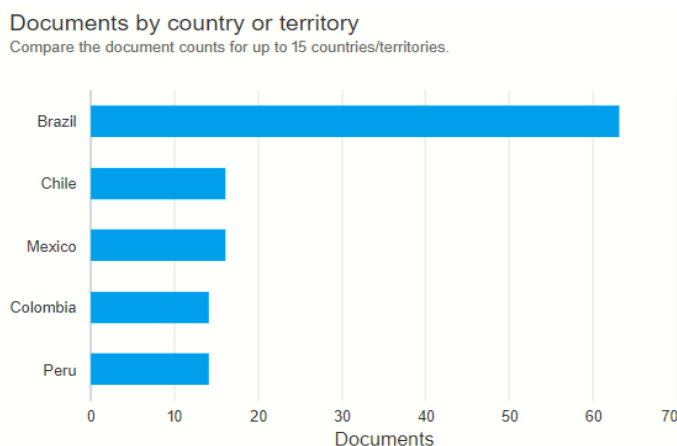
Criterios	Descripción
<b>De Inclusión</b>	<ul style="list-style-type: none"> <li>• Documentos en base de datos relevantes al tema de la revisión</li> <li>• Ventana de tiempo comprendida entre 2019- 2024</li> <li>• Área o categoría relacionada con la temática</li> <li>• Documentos en inglés y español</li> <li>• Tipos de documentos: Artículos, conferencias o libros</li> <li>• Inicialmente preferencia en Latinoamérica</li> </ul>

<b>De Exclusión</b>	<ul style="list-style-type: none"> <li>• Documento que son versión preliminar o resumida y no se ha reemplazado con otro documento completo</li> <li>• Boletines, comunicados de prensa o memorandos</li> </ul>
<b>De Calidad</b>	<ul style="list-style-type: none"> <li>• Artículos que, aunque cumplan con los términos de búsqueda, no tienen relación directa con el tema de interés.</li> </ul>

De acuerdo con estos criterios de filtración, se revisan los resultados obtenidos, encontrando que, al intentar seleccionar solo las referencias relacionadas con Latinoamérica, la cantidad de resultados se reduce a 118 documentos, de los cuales, la mayor cantidad está en Brasil con 63 resultados, seguida de Chile y México con 16, junto a Colombia y Perú con 14, como se observa en la Figura 3. Además, se encuentra una brecha u oportunidad, ya que estas referencias tienen gran relación con las instituciones de educación superior, los egresados, el mercado laboral y algo de movilidad social, pero de acuerdo con estos resultados, en estos países latinoamericanos se ha documentado muy poco sobre la implementación de técnicas de Machine Learning relacionadas con la temática de interés. De igual forma, esta información es base para la problemática y justificación de la investigación.

### Figura 2.

#### *Referencias por países latinoamericanos en Scopus*



Fuente: Tomado de Scopus

Teniendo en cuenta que no son tantas las referencias encontradas en estas regiones y para esta temática que se enfocan explícitamente en el uso de técnicas de Machine Learning, las ideas planteadas en estos estudios sirven para proporcionar consideraciones valiosas que pueden ser aplicadas localmente en futuras implementaciones de estas técnicas en el mercado laboral y la movilidad social de los graduados universitarios en Latinoamérica.

Por lo anterior se decide continuar con la revisión, ampliando la cobertura geográfica, más allá de Latinoamérica, pero asegurando que en los nuevos resultados obtenidos si se tenga relación directa con la implementación de técnicas de Machine Learning en la temática de investigación. Para lo cual, se realizan búsquedas que permitan obtener la información de las técnicas más usadas, para incluir en la ecuación de búsqueda algunos términos específicos, relacionados con los nombres más comunes. Esto permite completar la ecuación de búsqueda planteada inicialmente, al agregar los siguientes términos:

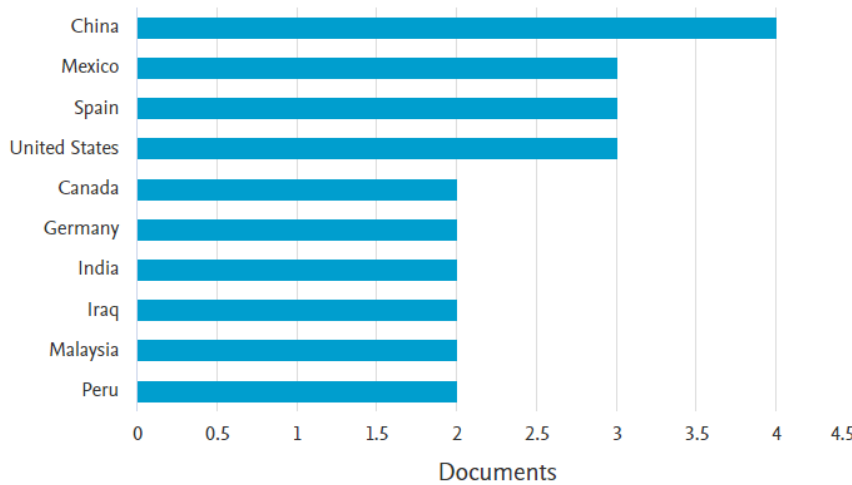
*(“linear regression” OR “logistic regression” OR “support vector” OR “SVM” OR “SVC” OR “SVR” OR “decision tree” OR “random forest” OR “k-nearest neighbor” OR “KNN” OR “neural networks” OR “naive bayes” OR “adaptive boosting” OR “gradient boosting” OR “light GBM” OR “k-mean” OR “hierarchical clustering” OR “component analysis” OR “PCA” OR “self-organizing map” OR “Apriori” OR “T-SNE” OR “bayesian network” OR “generative adversarial network” OR “autoencoder” OR “ARIMA” OR “Q-Learning” OR “Q-Networks”))*

Al realizar nuevas búsquedas, inicialmente en Scopus, se evidencia la distribución de las referencias bibliográficas, por país y por área de conocimiento (Figura 3 y Figura 4), donde la mayoría de las implementaciones de Machine Learning, en esta temática, provienen de países como China, México, España, Estados Unidos entre otros; además, las áreas de mayor interés son Ciencias Sociales, Ciencias de la Computación e Ingenierías. Cabe resaltar que el Machine

Learning está en tendencia y cada año se obtienen más publicaciones oficiales, donde se implementan estas técnicas para clasificar y predecir datos relevantes en diferentes sectores.

**Figura 3.**

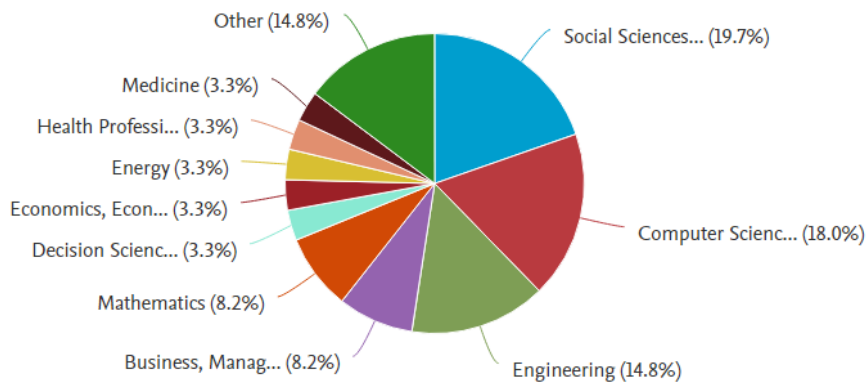
*Referencias por todos los países en Scopus*



Fuente: Tomado de Scopus

**Figura 4.**

*Referencias por todos los países en Scopus*



Fuente: Tomado de Scopus

Estos resultados de Scopus, con todos sus datos referenciales, son exportados en archivo con valores separados con comas (.csv) y se complementan con algunas otras publicaciones de interés que no necesariamente se encuentran publicados en las revistas de mayor factor de

impacto, pero también sirven como referencia claves para la investigación y se pueden recuperar mediante bases de datos más generales como Google Académico, EBSCO y ProQuest.

### **Análisis Bibliométrico con VOSviewer**

Con este archivo de las referencias bibliográficas finalmente seleccionadas, se procede a realizar un análisis de coocurrencia de términos y visualizar las redes bibliométricas en la aplicación VOSviewer, el cual permite observar gráficamente los nodos que representan las repeticiones entre las palabras clave y las conexiones en las relaciones de coocurrencia.

En la Figura 5 se observa que las tendencias se enfocan hacia cinco grupos principales:

- Color amarillo referente a la educación superior, universidades, currículo, enseñanza y entrenamiento o formación.
- Color morado referente al mercado laboral, trabajo, salario y algunas características como el sector público, los logros educativos, desigualdad o discriminación.
- Color azul referente al humano, su carrera, educación profesional, desempeño laboral, satisfacción, toma de decisiones e inteligencia emocional.
- Color verde referente a estudios realizados, junto a variables que se pueden tener en cuenta como el género, educación primaria, migración, estado civil y ocupación.
- Color rojo referente a estudiantes, graduados, desempeño académico, empleo, movilidad social y en este grupo se presentan algunas técnicas de procesamiento de datos para predicción y pronóstico mediante machine learning, como regresión logística, arboles de decisión y bosque aleatorio.



resultados de la investigación, agregando columnas que ayudan a la selección de información para la clasificación por país, año, técnicas de Machine Learning utilizadas, variables utilizadas y resultados finales. El análisis de este archivo (es cual se resume en el Apéndice A) es insumo para los resultados obtenidos.

### **Aplicaciones existentes de Machine Learning**

Después de realizar una revisión exhaustiva de las referencias seleccionadas para complementar los datos requeridos, se observa que algunas abordan la temática del mercado laboral y movilidad social de egresados o emplean herramientas estadísticas y tecnológicas, pero no implementan directamente técnicas de Machine Learning; mientras que otras fuentes que si las aplican en contextos relacionados con egresados, contienen antecedente como la evaluación de calidad universitaria, el desempeño estudiantes, la deserción universitaria, la decisión de posgrado, entre otras aspectos que no son directamente relacionados. Por lo anterior, se realiza una depuración final, donde se seleccionan 34 referencias que cumplen con todos los criterios para el análisis, aplicando técnicas de Machine Learning en modelos relacionados con el mercado laboral y la movilidad social de los graduados, en diversos países del mundo.

En la Tabla 4, se observa la distribución de estas investigaciones, donde el 83.3% están en idioma inglés y el 14.7% en español. Además, en tipo de documentos se obtiene el 70.6% de artículos y 29.4% conferencias. Al observar esta distribución por países, la mayoría de las aplicaciones provienen de Asia con un 41.2%, también se cuenta con trabajos realizados en Europa con un 23.5%, Latinoamérica con un 20.6% y África con un 14.7%.

Es importante mencionar que algunas de las investigaciones donde se aplican técnicas de Machine Learning en Latinoamérica, no fueron contempladas inicialmente, ya que no todas han

sido publicadas en revistas de alto impacto, pero se pueden recuperar por acceso abierto. Por ejemplo, en México, la aplicación de técnicas como la regresión logística y la regresión lineal han arrojado importantes hallazgos en diversas áreas. El estudio de Hirose et al. (2023) encuentra que el 75% de egresados de odontología tenía un trabajo relacionado con su formación profesional, pero presentan insatisfacción con los ingresos y la preparación para el trabajo clínico; además, en el estudio de Ramírez et al. (2021) se descubre que la satisfacción laboral y el cumplimiento de expectativas estaban estrechamente relacionados con el tipo de trabajo y los ingresos. Por otra parte, Gomez et al. (2022) realizan un estudio sobre la predicción de ingresos de exalumnos, donde muestra que variables como edad, género y posición laboral son predictores significativos del ingreso actual, revelando una brecha salarial de género.

En otros países como Perú, se utilizan modelos de regresión logística multivariada para algunos hallazgos importantes como que el género, la ubicación de la universidad y el tipo de trabajo están relacionados con la discriminación salarial y la dificultad para encontrar empleo en enfermeras (Quispe, 2019) y que el estado laboral del egresado, los servicios básicos de electricidad, la preferencia por actividades como el cine y los medios de comunicación, como la televisión y la radio, tenían un impacto significativo en la inserción laboral (Maquera, 2021). Por su parte, en Chile, el estudio hecho por De la Fuente et al. (2020) identifica que las recomendaciones de terceros y los estudios de postgrado influyen significativamente en la inserción laboral de graduados de contabilidad. Finalmente, en Colombia, Buitrago y Romero (2022) desarrollan un modelo de machine learning para predecir los salarios de recién egresados, encontrando que el género y el programa académico eran variables influyentes.

**Tabla 4.***Antecedentes por idioma, tipo de documento y país*

<b>Idioma</b>	<b>Antecedentes investigativos</b>			<b>Total</b>	<b>%</b>
	<b>Mercado laboral</b>	<b>Movilidad social</b>	<b>Ambos Temas</b>		
Español	3	0	2	5	14,7%
Inglés	24	3	2	29	85,3%
<b>Tipo de Documento</b>					
Artículo	18	3	3	24	70,6%
Conferencias	9	0	1	10	29,4%
<b>País</b>					
India	5	0	0	5	14,7%
China	3	0	1	4	11,8%
Malasia	1	0	0	1	2,9%
Bangladesh	1	0	0	1	2,9%
Kazajistán	1	0	0	1	2,9%
Arabia Saudita	0	0	1	1	2,9%
Iraq	0	1	0	1	2,9%
España	1	1	0	2	5,9%
Polonia	2	0	0	2	5,9%
Portugal	1	0	0	1	2,9%
Alemania	1	0	0	1	2,9%
Rumania	1	0	0	1	2,9%
Austria	1	0	0	1	2,9%
México	1	1	1	3	8,8%
Perú	2	0	0	2	5,9%
Colombia	0	0	1	1	2,9%
Chile	1	0	0	1	2,9%
Nigeria	3	0	0	3	8,8%
Egipto	1	0	0	1	2,9%
Marruecos	1	0	0	1	2,9%

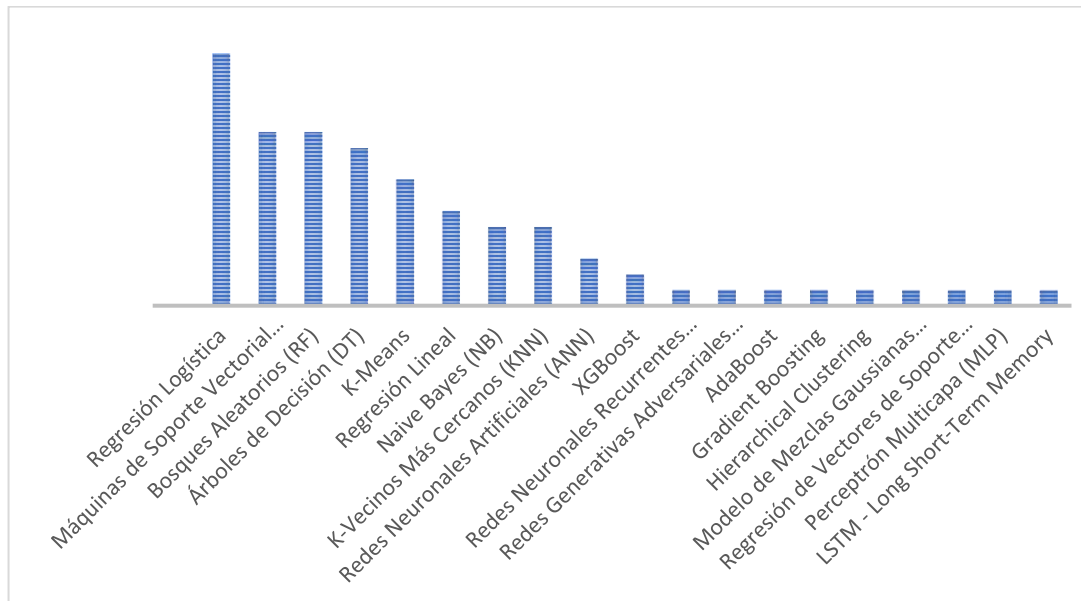
El análisis realizado permite identificar que se han aplicado técnicas o algoritmos de Machine Learning, con el fin de clasificar o predecir el acceso al mercado laboral y la movilidad social en egresados, las cuales se observan en la Tabla 5 y Figura 7, donde las diez de mayor uso, en orden descendente son: Regresión Logística, Maquinas de Soporte Vectorial (SVM), Bosques Aleatorios (RF), Árboles de Decisión, K-Means, Regresión Lineal, Naives Bayes (NB), K-Vecinos Más Cercanos (KNN), Redes Neuronales Artificiales (ANN) y XGBoost.

**Tabla 5.**  
*Técnicas de Machine Learning utilizadas*

Técnica de Machine Learning	Antecedentes investigativos			Total	%
	Mercado laboral	Movilidad social	Ambos Temas		
Regresión Logística	14	1	1	16	47,1%
Máquinas de Soporte Vectorial (SVM)	9	1	1	11	32,4%
Bosques Aleatorios (RF)	8	1	2	11	32,4%
Árboles de Decisión (DT)	9	0	1	10	29,4%
K-Means	7	0	1	8	23,5%
Regresión Lineal	3	3	0	6	17,6%
Naive Bayes (NB)	5	0	0	5	14,7%
K-Vecinos Más Cercanos (KNN)	5	0	0	5	14,7%
Redes Neuronales Artificiales (ANN)	2	0	1	3	8,8%
XGBoost	2	0	0	2	5,9%
Redes Neuronales Recurrentes (RNN)	1	0	0	1	2,9%
Redes Generativas Adversariales (GAN)	1	0	0	1	2,9%
AdaBoost	1	0	0	1	2,9%
Gradient Boosting	1	0	0	1	2,9%
Hierarchical Clustering	1	0	0	1	2,9%
Modelo de Mezclas Gaussianas (GMM)	0	0	1	1	2,9%
Regresión de Vectores de Soporte (SVR)	0	0	1	1	2,9%
Perceptrón Multicapa (MLP)	0	0	1	1	2,9%
LSTM - Long Short-Term Memory	1	0	0	1	2,9%

**Figura 6.**

*Técnicas de Machine Learning utilizadas*



El rendimiento de estas aplicaciones se ha evaluado en términos de exactitud, precisión, recuperación, puntuación F1 y el Área Bajo la Curva ROC(AUC), donde algunos han logrado índices de precisión elevados. Por ejemplo, en el estudio de Zayed et al. (2023) el algoritmo de Bosque Aleatorio (RF) ajustado demuestra una precisión del 98.43%, superando a otros, como Árboles de Decisión (DT) y SVM, para predecir la empleabilidad de los graduados e identificar características importantes como la escuela secundaria y el título o programa que se estudia.

En el estudio de ElSharkawy et al., (2022) se han aplicado algoritmos como Regresión Logística, Árbol de Decisión (DT), SVM, Bosques Aleatorios (RF) y Naïve Bayes Gaussiano (Gaussian NB), para predecir la capacidad de empleo de los graduados en TI, con el objetivo de satisfacer las demandas del mercado laboral, encontrando resultados que evalúan el rendimiento de estos algoritmos para la predicción de la capacidad de empleo, donde la precisión más alta la logra DT con 100%, seguido de Regresión Logística y SVM con 98%, mientras que la precisión más baja con 92% la logra Gaussian NB.

De igual forma en el estudio de Muraina et al. (2022) se implementan técnicas como Árboles de Decisión (DT), SVM y K-Means, donde los tres clasificadores con respecto a exactitud, precisión, recuperación y medida F; muestran como resultado que el DT tenía la mejor corrección de clasificación, así como la medida precisión (90%), recuperación (89%) y medida F (89%), además, se observan errores mínimos en el clasificador del DT que en otros dos.

En el estudio de Rao et al. (2023) donde se evalúa el rendimiento del modelo Extreme Gradient Boost (XGBoost) en comparación con las técnicas de clasificación estándar, se encuentra como resultado que esta es la más relevante para predecir trayectorias profesionales de estudiantes universitarios, con una precisión de 86%, ofrece resultados significativos en la influencia de competencias profesionales y coherencia conductual. Por otra parte, Haque et al. (2024) aplica algoritmos de clasificación como Regresión Logística, Bosque Aleatorio (RF), Naïve Bayes (NB), SVM, XGBoost y Redes Neuronales Artificiales (ANN), donde los resultados indican que ANN alcanza la mayor precisión, aproximadamente 80%, además, que características demográficas y académicas de los estudiantes, junto con el nivel de satisfacción respecto a instalaciones universitarias (como la biblioteca y el servicio de asesoramiento), son aspectos importantes para predicciones de empleabilidad de estudiantes de posgrado.

Baffa, et al. (2023) utiliza modelos como Regresión Logística, Árboles de Decisión (DT), Bosque Aleatorio (RF) y K-Means para anticipar la empleabilidad de los estudiantes antes de su graduación, considerando atributos académicos y de experiencia laboral, como el promedio acumulativo de calificaciones, el esquema de experiencia laboral, actividades curriculares, género y afiliaciones sindicales previas a la graduación. Los resultados evidencian que el RF predice con mayor precisión la empleabilidad de los estudiantes, con una precisión del 98% y una puntuación F1 de 0.99.

En algunas universidades se han utilizado conjuntos de datos de empleo para construir Redes Neuronales Artificiales (ANN) y modelos de Mezcla Gaussiana (GMM) para predecir las perspectivas de empleo de los estudiantes, como es el caso de China, donde se menciona en la investigación de Minghui (2023) que el GMM funciona mal y no puede resolver la aleatoriedad en la predicción del empleo de los graduados, mientras que el modelo de ANN se puede aplicar bien a la predicción de los tipos de empleo de los graduados universitarios, con un valor AUC (área bajo la curva ROC) de 0,771 como métrica de evaluación.

Respecto a las variables utilizadas en la aplicación de técnicas de Machine Learning en las referencias que han sido de apoyo para la investigación, se encuentra que Rao et al. (2023), Zayed et al. (2023), ElSharkawy (2022) y Wang et al. (2022) se basan en datos académicos, demográficos, evaluaciones internas y atributos familiares o personales del mercado objetivo para el estudio, como la edad, género, nacionalidad, nivel de educación y competencias profesiones, entre otras. Por su parte, Wei et al. (2023) tiene en cuenta variables como su experiencia como líder estudiantil o situación familiar para predicción del fenómeno de "slow employment" en el que los graduados universitarios experimentan dificultades para encontrar empleo de manera inmediata después de graduarse.

En el trabajo realizado por Gadegaonkar et al. (2023) donde se crea una aplicación para recomendar trabajos a los usuarios en función de sus intereses se utilizan variables como estimación salarial, descripción del trabajo, calificación de la empresa, ubicación e industria. Por su parte, Gomez (2022) en su trabajo de análisis predictivo para los ingresos de los graduados, toman en cuenta variables como la edad, género, horas de trabajo por semana, primer ingreso y variables relacionadas con la posición laboral y empresa del exalumno.

En general, los factores o variables a utilizar en la implementación de modelos de Machine Learning, dependen de lo que se quiera clasificar o predecir y los objetivos que se desean lograr, de acuerdo con la revisión realizada, se pueden considerar algunos datos sobre los graduados y estudiantes próximos a graduarse de las instituciones de interés, como:

- Atributos demográficos como la edad, género, lugar de residencia, estado civil.
- Atributos académicos, como carrera universitaria, rendimiento académico, promedio, nivel de estudio y la satisfacción de los estudiantes.
- Atributos institucionales, como la calidad y renombre de la institución.
- Atributos socioeconómicos, como el nivel de ingresos, estrato, sector y estabilidad laboral, tanto del egresado como de sus padres.
- Experiencia laboral, incluyendo pasantías, prácticas profesionales o trabajos previos.
- Habilidades relevantes, áreas de interés y capacidades adquiridas durante su formación.
- Necesidades de la industria, como criterios exigidos para la empleabilidad.

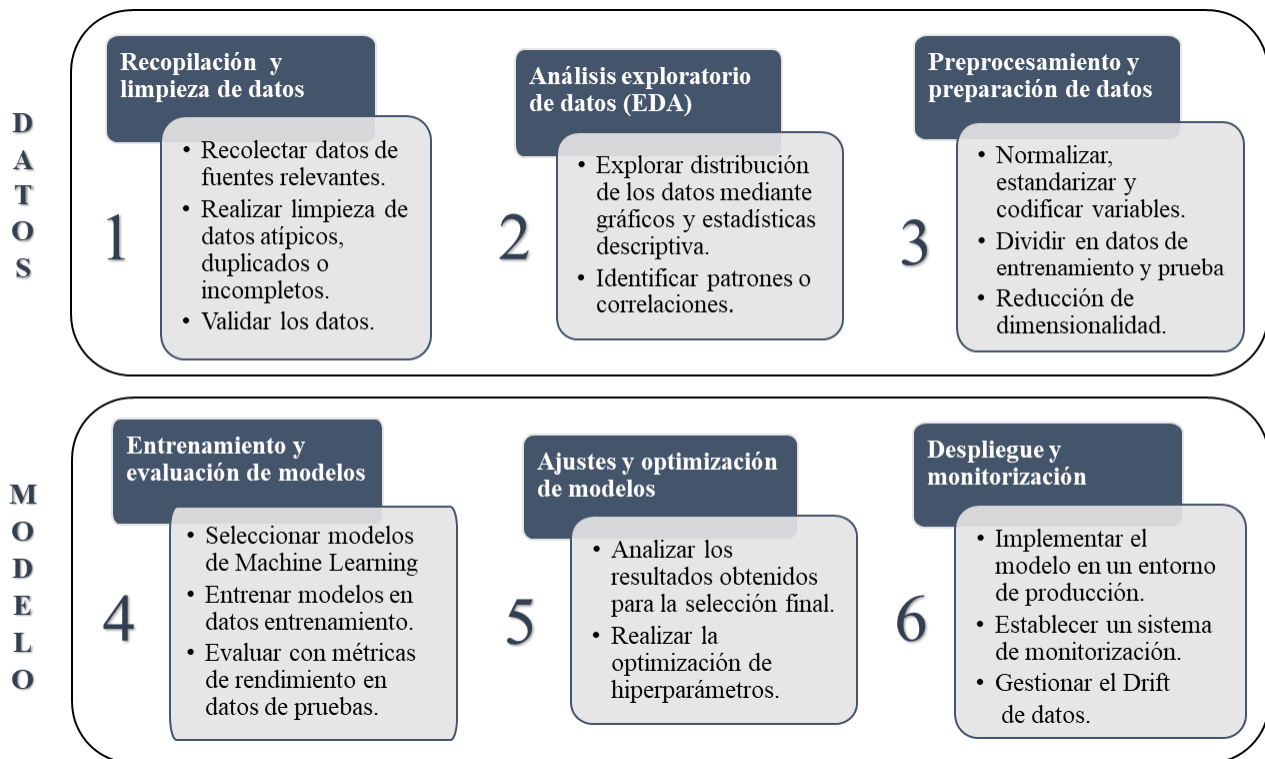
En Colombia, la obtención de algunos datos de interés para el estudio puede realizarse mediante encuestas dirigidas al público objetivo o a través de información proveniente de instituciones de educación superior, el Ministerio de Educación Nacional (MEN), el Sistema Nacional de Información de la Educación Superior (SNIES) y el Instituto Colombiano para la Evaluación de la Educación (ICFES), entre otras fuentes disponibles. Según el estudio realizado por Buitrago y Romero (2022) se deberían tener en cuenta variables como el género, la institución universitaria, el ranking de esta y el programa académico, como factores influyentes en la predicción del salario de los egresados. Estos datos pueden ser obtenidos del portal del Observatorio Laboral para la Educación (OLE) durante el período de interés, exceptuando el ranking, que puede estar disponible en el portal Webometrics.

### Recomendaciones para la implementación del modelo

Para la implementación del modelo de Machine Learning que permita recopilar, procesar y analizar información relevante sobre el mercado laboral y la movilidad social en el seguimiento a egresados, con el objetivo de generar hallazgos significativos y predictivos que maximicen la efectividad y utilidad en la toma de decisiones en este ámbito; se proponen un conjunto de recomendaciones que abarcan desde la recopilación y limpieza de datos hasta la implementación y despliegue del modelo, como se muestra a continuación:

**Figura 7.**

*Recomendaciones propuestas para el modelo*



### 1. Recopilación y limpieza de datos.

- Recolectar datos de las principales fuentes de información como bases de datos institucionales, datos de redes profesionales o gubernamentales, encuestas a egresados o empleadores, así como información socioeconómica, tendencias del mercado laboral y datos demográficos.
- Realizar la limpieza de datos que permite eliminar valores atípicos y procesar los datos duplicados o incompletos, mediante técnicas como la detección y manejo de valores atípicos, junto a la imputación de valores faltantes.
- Validación los datos limpiados para asegurar que no haya datos faltantes o inconsistencias graves (integridad), que estén en los formatos adecuados y consistentes para cada variable (consistencia) y la precisión de los datos.

### 2. Análisis exploratorio de los datos (EDA).

- Explorar la distribución de los datos mediante gráficos como histogramas o diagramas de caja y estadísticas descriptivas como la media, la mediana, la desviación estándar y los percentiles, que permita comprender su estructura.
- Identificar posibles patrones o correlaciones entre las variables, mediante análisis de correlación y gráficos de dispersión. Esto puede ayudar a identificar las características que tienen relación significativa con la variable objetivo, sugerir nuevas o requerir transformarlas para mejorar el rendimiento del modelo.

### 3. Preprocesamiento y preparación de datos.

- Preparar los datos mediante la normalización y estandarización de variables numéricas para asegurar que tengan la misma escala, así como la codificación de variables categóricas utilizando técnicas como one-hot encoding o label encoding.

- Dividir los datos de forma aleatoria para asegurar que tanto el conjunto de entrenamiento como el de prueba sean representativos de la distribución y evitar sesgos al evaluar su rendimiento con datos no usados previamente.
  - Aplicar técnicas de reducción de dimensionalidad es recomendable cuando el número de características es muy elevado y puede dificultar el entrenamiento del modelo, o si se requiere un espacio menor para visualizar los datos e identificar patrones y relaciones. Técnicas como PCA (Análisis de Componentes Principales) o LDA (Análisis de Discriminante Lineal) ayudan a simplificar el modelo y mejorar el rendimiento computacional.
4. Entrenamiento y evaluación de modelos.
- Seleccionar los modelos de Machine Learning apropiados para la clasificación o predicción de acuerdo con los datos seleccionados.
  - Entrenar los modelos utilizando el conjunto de entrenamiento. Para esto se recomienda el uso de lenguaje de programación Python (con librerías como Pandas, NumPy, Matplotlib, Seaborn y Scikit Learn) o R (con librerías como caret, tidyverse, ggplot2, rpart).
  - Evaluar el rendimiento del modelo utilizando los datos de prueba, mediante métricas como la exactitud, precisión, recuperación o sensibilidad, puntuación F1 o área bajo la curva ROC, junto con la matriz de confusión. Es importante tener en cuenta que la elección adecuada de métricas de rendimiento y la interpretación de los resultados dependen del contexto del problema que se está abordando.

- Utilizar técnicas de validación cruzada que permitan evaluar el rendimiento del modelo usando diferentes subconjuntos de datos de entrenamiento y prueba, con el fin de obtener estimaciones más precisas.
5. Ajustes y optimización de modelos.
- Analizar los resultados obtenidos, junto a las métricas, la matriz de confusión o la validación cruzada, para la selección final del modelo. Si ninguno alcanza los niveles de rendimiento deseados se deben realizar ajustes adicionales.
  - Realizar la optimización de hiperparámetros para encontrar la mejor combinación de parámetros ajustables del modelo para mejorar su rendimiento, mediante técnicas como la búsqueda en cuadrícula, donde se evalúan sistemáticamente todas las combinaciones posibles de hiperparámetros dentro de un rango, búsqueda aleatoria donde se selecciona combinaciones de manera aleatoria u optimización bayesiana, donde se utiliza métodos probabilísticos para encontrar la configuración óptima basada en resultados anteriores de evaluación del modelo.
6. Despliegue y monitorización del modelo.
- Implementar el modelo en un entorno de producción utilizando herramientas como contenedores Docker o servicios en la nube. Para este proceso, es importante considerar la inclusión de un sistema de control de versiones para la gestión y seguimiento de los cambios realizados.
  - Establecer un sistema de monitorización para llevar un seguimiento constante del rendimiento del modelo y detectar cualquier inconveniente presentado, teniendo en cuenta la definición de métricas claves y la frecuencia de monitorización, para garantizar un rendimiento óptimo.

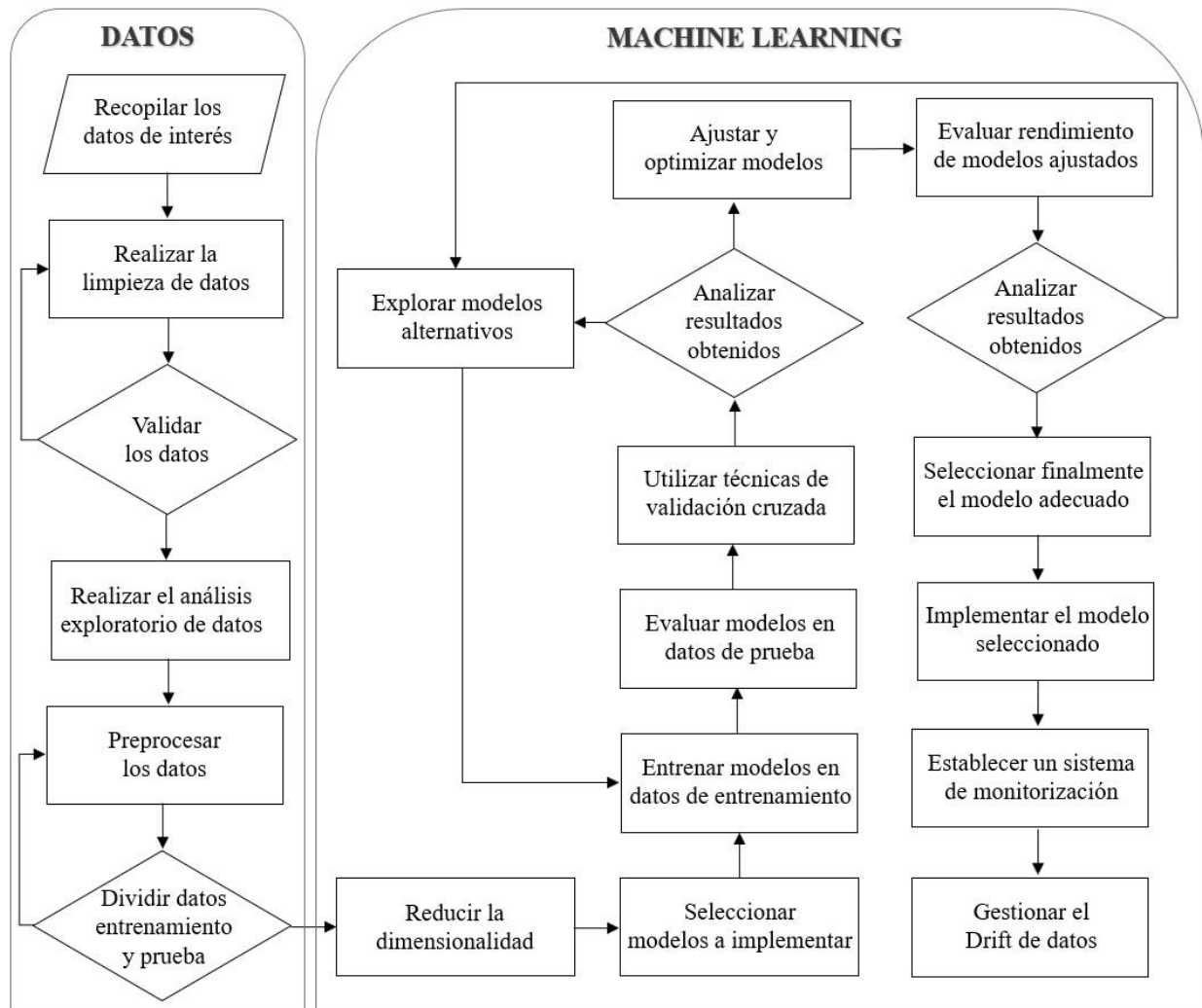
- Gestionar el Drift de datos que permita manejar posibles cambios, como el comportamiento de los usuarios, cambios en el entorno, las condiciones del mercado o errores en la recolección de datos, con el fin de corregir los posibles problemas y ajustar el modelo para mantener o mejorar su rendimiento.

Se espera que al tener en cuenta estas recomendaciones propuestas, se puede diseñar un modelo de Machine Learning efectivo y aplicable para estudiar el mercado laboral y la movilidad social de los egresados, con el potencial de generar percepciones significativas y predictivas que ayuden a mejorar la toma de decisiones y las políticas relacionadas con el empleo y la educación.

Como resultado adicional de la investigación realizada y teniendo en cuenta que este proyecto servirá de insumo para encontrar la mejor técnica de Machine Learning que permita implementar un modelo de seguimiento a egresados universitarios, se propone organizar estas recomendaciones en una metodología que aplique no solo para esta situación, sino probablemente pueda servir de apoyo para futuros trabajos que requieran una estructura definida con el paso a paso desde el manejo de los datos hasta la puesta en producción del modelo seleccionado, como se observa en la Figura 8.

Es importante mencionar que la aplicación de la metodología propuesta dependerá de la problemática a abordar y en algunas ocasiones, puede no ser requerido aplicar en orden todos los pasos detallados, por lo cual se tiene la flexibilidad de omitir pasos innecesarios y avanzar al siguiente, para garantizar la eficiencia de la implementación en diferentes contextos.

**Figura 8.**  
*Metodología de Machine Learning – RAPEAD*



## Conclusiones

La revisión bibliográfica realizada en esta investigación evidencia que actualmente va en aumento la aplicación de tecnologías emergentes como el Machine Learning en el ámbito del mercado laboral y la movilidad social de los egresados en Latinoamérica. Teniendo en cuenta que la literatura actual en la región muestra un enfoque predominante en áreas sociales y humanas, se abren nuevas oportunidades para el desarrollo de modelos innovadores que puedan aprovechar algoritmos y técnicas avanzadas para mejorar la comprensión y predicción de esta y otras temáticas relacionadas.

En Colombia, se identificaron fuentes relevantes que pueden servir de insumo para la implementación de modelos, donde se puede sacar provecho de la información disponible en el Observatorio Laboral para la Educación (OLE) del Ministerio de Educación Nacional (MEN), con variables como la tasa de empleo de los egresados, el nivel salarial, las áreas o sector de empleo y las necesidades de formación y competencias demandadas por el mercado. Esta información puede complementarse con datos académicos, socioeconómicos y demográficos disponibles en las instituciones de educación superior. Además, se cuenta con el seguimiento de la evolución después de graduarse, junto a información primaria que pueda recolectarse u otras fuentes de datos externas, como bases de datos laborales o estudios de mercado, las cuales pueden servir de apoyo para enriquecer el análisis y mejorar la precisión de los modelos.

Los modelos de clasificación y predicción que utilizan el Machine Learning han surgido como herramientas esenciales que permiten evaluar el potencial de los estudiantes para conseguir empleo después de su graduación, así como su movilidad social; sin embargo, se debe tener en cuenta que la eficiencia y el rendimiento práctico de los resultados de su aplicación dependen del conocimiento profundo de la temática a intervenir, la claridad en los objetivos establecidos y la

disponibilidad de datos completos y relevantes para el proceso de modelado. De acuerdo con la investigación realizada, se observa que, aunque la Regresión Logística y SVM son las técnicas más utilizadas en la evaluación de modelos, no siempre son las más precisas, por lo cual, pueden ser usados como alternativas de comparación al momento de evaluar modelos para la implementación final, pero se recomienda considerar opciones que han demostrado ofrecer mejores resultados en aplicaciones similares, como Árbol de Decisión (DT), Bosques Aleatorios (RF), Redes Neuronales Artificiales (ANN) y el XGBoost,.

Las recomendaciones propuestas para encontrar e implementar el modelo de Machine Learning en el seguimiento a egresados, que explore el mercado laboral y la movilidad social, representan un resultado fundamental de esta investigación y como resultado adicional, se presenta la posibilidad de convertirse en una metodología sólida que puede ser usada en futuras implementaciones en otras áreas de estudio. Es esencial reconocer la interconexión y la importancia de cada etapa del proceso, desde la recopilación y limpieza inicial de datos hasta la monitorización continua del modelo en producción, cada paso contribuye de manera significativa a la eficacia y la precisión del modelo final. La combinación de herramientas, técnicas y prácticas, como la normalización de datos, la selección de características y la gestión del Drift de datos, garantiza la creación de modelos robustos y capaces de adaptarse en entornos dinámicos.

Como recomendación final, sería beneficioso para las universidades llevar a cabo un estudio longitudinal de seguimiento de egresados, que permita recopilar datos a lo largo del tiempo sobre la trayectoria laboral y social de los egresados, con el fin de proporcionar información completa y detallada sobre la evolución de las experiencias laborales y sus logros profesionales, así como de los factores que influyen en su movilidad social a lo largo del tiempo.

### Referencias bibliográficas

- Aké-Uitz, D. S. (2022). Higher education institutions as mediators of the intergenerational transmission of wealth in Mexico. *El trimestre económico*, 89(355), 865-900.
- Azpíroz, I. D., Villar, M. G., Ballester, J. B., Velasco, C. L. R., & Flores, E. S. (2023). Clasificación y pronóstico del nivel de satisfacción de egresados de programas de salud en el contexto de una metodología de aprendizaje automático: un análisis de caso orientado a posgrados online de una institución educativa iberoamericana. *MLS Educational Research (MLSER)*, 7(2).
- Baffa, M. H., Miyim, M. A., & Dauda, A. S. (2023). Machine Learning for Predicting Students' Employability. *UMYU Scientifica*, 2(1), 001-009.
- Banco Interamericano de Desarrollo (BID). (2020). ¿Está preparada América Latina para el futuro del trabajo? [https://publications.iadb.org/publications/spanish/document/El\\_futuro\\_del\\_trabajo\\_en\\_America\\_Latina\\_esta\\_preparada\\_la\\_region.pdf](https://publications.iadb.org/publications/spanish/document/El_futuro_del_trabajo_en_America_Latina_esta_preparada_la_region.pdf)
- Berryhill, J., Heang, K. K., Clogher, R., & McBride, K. (2019). Hello, World: Artificial intelligence and its use in the public sector. *OECD Working Papers on Public Governance (Vol. 36)*. <https://doi.org/10.1787/19934351>
- Buitrago, D., & Romero, D. C. (2022). Un modelo de Machine Learning para detectar factores que influyen en los salarios de los recién egresados de universidades colombianas. *In Revolución en la formación y la capacitación para el siglo XXI. Volúmenes I y II* (pp. 399-410). Instituto Antioqueño de Investigación (IAI).
- Chapelle, O., Scholkopf, B., & Zien, A. (2006). Semi-supervised learning. 2006. *Cambridge, Massachusetts: The MIT Press*.

Comisión Económica para América Latina y el Caribe (CEPAL). (2020). La movilidad social en América Latina: realidades y percepciones.

<https://www.cepal.org/es/publicaciones/46688-la-movilidad-social-america-latina-realidades-percepciones>

Consuegra Bolívar, J., Rodado Noriega, C., Berry, A., Corredor Pereira, C. F., Gaviria Uribe, A., Henríquez Guajardo, P., ... & Giha Tovar, Y. (2020). Desafíos de la educación y la movilidad social en Latinoamérica.

Daniel, B. (2015). Big Data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology*, 46(5), 904-920.

De la Fuente-Mella, H., Muñoz, C. G., Quezada, E. R., Ocares, C. I., & Pérez, M. L. (2020). Probabilistic Models of Job Placement and Positioning for Students with a Career in Chile. In *Advances in Human Factors, Business Management and Leadership: Proceedings of the AHFE 2020 Virtual Conferences on Human Factors, Business Management and Society, and Human Factors in Management and Leadership, July 16-20, 2020, USA* (pp. 516-522). Springer International Publishing.

El País. (2023). Preocupante cifra: esto es lo que gana un recién graduado en Colombia. *El País*.

ElSharkawy, G., Helmy, Y., & Yehia, E. (2022). Employability Prediction of Information Technology Graduates using Machine Learning Algorithms. *International Journal of Advanced Computer Science and Applications*, 13(10).

Filgueira, C. H., González-Rozada, M., Mallimaci, F., & Sánchez Aragón, R. (2022). Social mobility in Latin America: A new exploration. *World Development*, 152, 105673.

<https://doi.org/10.1016/j.worlddev.2021.105673>

Forbes, F. (2023). What Are The Most In-Demand AI Skills?. *Forbes*.

- Gadegaonkar, S., Lakhwani, D., Marwaha, S., & Salunke, A. (2023, February). Job Recommendation System using Machine Learning. *In 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS)* (pp. 596-603). IEEE.
- Gomez-Cravioto, D. A., Diaz-Ramos, R. E., Hernandez-Gress, N., Preciado, J. L., & Ceballos, H. G. (2022). Supervised Machine Learning Predictive Analytics for Alumni income. *Journal of Big Data*, 9(1), 11.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Grajales, R. V., Vázquez, R. M. C., & Fonseca, C. E. (2015). El concepto de movilidad social: dimensiones, medidas y estudios en México (No. 2015\_01).
- Haque, R., Quek, A., Ting, C. Y., Goh, H. N., & Hasan, M. R. (2024). Classification Techniques Using Machine Learning for Graduate Student Employability Predictions. *International Journal on Advanced Science, Engineering & Information Technology*, 14(1).
- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2, pp. 1-758). New York: Springer.
- Hirose-López, M., Sánchez-Mendiola, M., Borges-Yáñez, S. A., García-Minjares, M., & Martínez-González, A. (2024). Dental graduates: perceived academic training, job satisfaction, and fulfillment of performance expectations. *Investigación en Educación Médica*, 13(49), 40-52.
- La República. (2018). ¿Cuáles son las universidades de donde salen los egresados mejor cotizados del país?. *La República*.
- Maquera-Luque, P. J., Morales-Rocha, J. L., & Apaza-Panca, C. M. (2021). Socio-economic and cultural factors that influence the labor insertion of University Graduates, Peru. *Heliyon*, 7(7).

- Martínez-Franco, D. (2022). Movilidad socio-ocupacional de graduados universitarios en Colombia, una mirada a la educación como vector de movilidad social. *Desacatos*, 70, 110-127. <https://doi.org/10.29340/70.2553>
- Meraliyev, B., Alibekova, B., & Bekturganova, I. (2023, June). Machine Learning as an effective tool for Human Resource management in recruiting process in the higher educational field. *In 2023 17th International Conference on Electronics Computer and Computation (ICECCO)* (pp. 1-5). IEEE.
- Maslej, N., Fattorini, L., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., ... & Perrault, R. (2023). Artificial intelligence index report 2023. *Stanford University*.
- Mason, C. M., Chen, H., Evans, D., & Walker, G. (2023). Illustrating the application of a skills taxonomy, Machine Learning and online data to inform career and training decisions. *The International Journal of Information and Learning Technology*.
- Minghui, Z. (2023, September). Predicting Graduates' Career Paths Using Neural Networks and Gaussian Mixture Models. *In 2023 IEEE 6th International Conference on Information Systems and Computer Aided Education (ICISCAE)* (pp. 174-178). IEEE.
- Ministerio de Educación Nacional de Colombia. (2023). El Observatorio Laboral para la Educación (OLE). *MinEducación*.
- Quispe-Ilanzo, M. P., Pacovilca-Alejo, O. V., Zea-Montesinos, C. C., García-Delgado, R., & Oyola-García, A. E. (2019). Discriminatory barriers for professional practice in Peruvian nurses. *Revista del Cuerpo Médico Hospital Nacional Almanzor Aguinaga Asenjo*.
- Ramírez-Romero, J. L., Erizeira Ruiz-Chávez, A., & Peña-Medina, S. (2021). Relación entre Desempeño Laboral y Especialidades: El Caso de los Egresados de una Licenciatura en Enseñanza del Inglés. *Education Policy Analysis Archives*, 29.

- Ramírez Escorcía, A. (2022). Ingreso al mercado laboral en condiciones de heterogeneidad de las y los egresados de la Universidad Nacional Autónoma de México en el 2012 (Movilidad Social y trayectoria ocupacional). *Revista de la educación superior*, 51(202), 27-51.
- Rao, K. E., Pydi, B. M., Vital, T. P., Naidu, P. A., Prasann, U. D., & Ravikumar, T. (2023). An Advanced Machine Learning Approach for Student Placement Prediction and Analysis. *International Journal of Performability Engineering*, 19(8), 536.
- Raschka, S., & Mirjalili, V. (2017). Python Machine Learning. Packt Publishing.
- Romero, M. (2021). Desigualdad e inserción laboral de egresados universitarios en México. *Revista Iberoamericana de Educación Superior*, 12(35), 3-23.  
<https://doi.org/10.22201/iisue.20072872e.2021.35.1584>
- SITEAL. (2021). Informe sobre Tendencias Sociales y Educativas en América Latina 2021. *IIEP UNESCO*. Recuperado de:  
[https://www.siteal.iiep.unesco.org/respuestas/informes/informe\\_sobre\\_tendencias\\_sociales\\_y\\_educativas\\_en\\_america\\_latina\\_2021](https://www.siteal.iiep.unesco.org/respuestas/informes/informe_sobre_tendencias_sociales_y_educativas_en_america_latina_2021)
- Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction. *MIT press*.
- Thilak, K. D., Devi, K. L., Kalaiselvi, K., & Teja, J. (2023, November). Revolutionizing University Graduate Employability: Leveraging Advanced Machine Learning Models to Optimize Campus Recruitment and Placement Strategies. *In 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE)* (pp. 1-6). IEEE.
- United Nations. (2016). Transforming our world: The 2030 agenda for sustainable development.

- Venegas-Muggli, J. I., Cifuentes-Donald, C., Rozas-Retamal, M., & González-Clares, M. J. (2021). Determining factors of labour market outcomes for recently graduated, underrepresented college students. *Australian Journal of Career Development*, 30(2), 150-162.
- Vieira, A. D. H. P. (2023). Desigualdades sociais na entrada no emprego formal entre graduados da educação superior. *Tempo Social*, 35, 131-161.
- Wang, P., Liao, W., Zhao, Z., & Miu, F. (2022). Prediction of factors influencing the starting salary of college graduates based on machine learning. *Wireless Communications and Mobile Computing*, 2022(1), 7845545.
- World Economic Forum. (2023). Informe sobre el futuro del empleo 2023. Recuperado de [https://www3.weforum.org/docs/WEF\\_Future\\_of\\_Jobs\\_2023\\_News\\_Release\\_ES.pdf](https://www3.weforum.org/docs/WEF_Future_of_Jobs_2023_News_Release_ES.pdf)
- Zayed, Y., Salman, Y., Awad, M., & Hasasneh, A. (2023). Employment Recommendation System for Graduates Using Machine Learning. *International Journal on Engineering Applications*, 11(5).

Apéndices

Apéndice A. Resultados de referencias bibliográficas

	<i>Cita</i>	<i>Idioma</i>	<i>Tipo Doc.</i>	<i>País</i>	<i>Antecedentes</i>	<i>Técnicas ML</i>	<i>Resultado</i>
1	(Rao et al., 2023)	Inglés	Artículo	India	Mercado Laboral	XGB / DT / RF / NB / SVM / KNN	XG Boost precisión 86%
2	(Rodríguez et al., 2019)	Español	Artículo	España	Mercado Laboral	LogR	LogR precisión 78,8%
3	(Chopra & Saini, 2023)	Inglés	Conferencia	India	Mercado Laboral	RNN / LSTM	LSTM precisión 72 %
4	(Moumen et al., 2020)	Inglés	Conferencia	Marruecos	Mercado Laboral	LogR / DT / NB / RF / ANN	ANN precisión 98%
5	(Hirose et al., 2024)	Español	Artículo	México	Mercado Laboral	LogR	LogR
6	(Quispe et al., 2019)	Español	Artículo	Perú	Mercado Laboral	LogR	LogR
7	(ElSharkawy et al., 2022)	Inglés	Artículo	Egipto	Mercado Laboral	LogR / DT / SVM / RF / NB	DT precisión 100%
8	(Zayed et al., 2023)	Inglés	Artículo	Arabia Saudita	Ambos Temas	DT / SVM / RF	RF precisión 98.43%
9	(Wei et al., 2023)	Inglés	Artículo	China	Mercado Laboral	SVM	SVM precisión 93.86%
10	(Wotaifi & Al-Shamery, 2020)	Inglés	Artículo	Iraq	Movilidad Social	LinR / RF / SVM	RF Modified
11	(Kannan et al., 2023)	Inglés	Artículo	India	Mercado Laboral	LinR / PNN	LinR
12	(Wang et al., 2022)	Inglés	Artículo	China	Mercado Laboral	NB / RF / SVM / XGB	SVM precisión 85,4%
13	(De la Fuente et al., 2020)	Inglés	Conferencia	Chile	Mercado Laboral	LogR	LogR p-value 0.990
14	(Maquera et al., 2021)	Inglés	Artículo	Perú	Mercado Laboral	LogR	LogR
15	(Ramírez et al., 2021)	Español	Artículo	México	Ambos Temas	LogR	LogR
16	(Valente et al., 2023)	Inglés	Artículo	España	Movilidad Social	LinR	LinR
17	(Hojda et al., 2022)	Inglés	Artículo	Polonia	Mercado Laboral	LogR	LogR
18	(Bittmann & Zorn, 2019)	Inglés	Artículo	Austria	Mercado Laboral	LinR / LogR	LinR / LogR
19	(Heinisch et al., 2020)	Inglés	Artículo	Alemania	Mercado Laboral	LogR / RF / AB	RF precisión 95,84%
20	(Stanca et al., 2022)	Inglés	Artículo	Rumania	Mercado Laboral	K-Means	K-Means
21	(Muraina & Agoi, 2022)	Inglés	Conferencia	Nigeria	Mercado Laboral	DT / SVM / KNN	DT precisión 89.6%
22	(Suleman & Suleman, 2022)	Inglés	Conferencia	Portugal	Mercado Laboral	K-Means	K-Means
23	(Muraina et al., 2022)	Inglés	Artículo	Nigeria	Mercado Laboral	DT / SVM / K-Means	DT precisión de 90%
24	(Guo et al., 2020)	Inglés	Conferencia	China	Mercado Laboral	LSTM / GAN	LSTM+GAN precisión 88%
25	(Gadegaonkar et al., 2023)	Inglés	Conferencia	India	Mercado Laboral	K-Means	K-Means precisión 87,6%.
26	(Baffa et al., 2023)	Inglés	Artículo	Nigeria	Mercado Laboral	LogR / DT / RF	RF precisión 98%
27	(Casuat et al., 2020)	Inglés	Artículo	Bangladesh	Mercado Laboral	LogR / DT / RF / SVM / KNN / K-Means	SVM logra mayor precisión
28	(Gomez et al., 2022)	Inglés	Artículo	México	Movilidad Social	LinR / LogR	LinR / LogR

29	(Piróg & Hibszer, 2024)	Inglés	Artículo	Polonia	Mercado Laboral	K-Means	K-Means
30	(Haque et al., 2024)	Inglés	Artículo	Malasia	Mercado Laboral	LogR / RF / NB / SVM / XGB / ANN	ANN precisión 80%
31	(Minghui, 2023)	Inglés	Conferencia	China	Ambos Temas	ANN / GMM	ANN valor AUC 0,771
32	(Thilak et al., 2023)	Inglés	Conferencia	India	Mercado Laboral	LogR / SVM / KNN / GB	GB precisión del 86%
33	(Merallyev et al., 2023)	Inglés	Conferencia	Kazajistán	Mercado Laboral	LogR / RF / KNN	RF precisión 72%
34	(Buitrago & Romero, 2022)	Español	Artículo	Colombia	Ambos Temas	RF / MLP / SVR / K-Means	K-Means precisión 70,19%
35	(Das et al., 2020)	Inglés	Conferencia	Bangladesh	Desempeño Estudiante	NB / DT / ANN / K-Means	N.A.
36	(Papík et al., 2022)	Inglés	Artículo	Eslovaquia	Desempleo Juvenil	LogR	N.A.
37	(Ünal & Gönülaçan, 2019)	Inglés	Artículo	Turquía	Pobreza del País	LogR	N.A.
38	(Mishra et al., 2023)	Inglés	Artículo	Alemania	Deserción Estudiantil	LinR	N.A.
39	(Litmeyer & Hennemann, 2024)	Inglés	Artículo	Alemania	Movilidad Estudiantes	LogR / DT / SVM / RF / NB	N.A.
40	(Gunn et al., 2019)	Inglés	Artículo	Canadá	Igualdad de Género	LinR	N.A.
41	(Wotaifi, 2019)	Inglés	Artículo	Estados Unidos	Deserción Estudiantil	LinR / RF / SVM	N.A.
42	(Da Silva et al., 2019)	Inglés	Conferencia	Brasil	Desempeño Estudiante	DT	N.A.
43	(Dobova et al., 2020)	Inglés	Artículo	México	Seguro de Salud	LogR	N.A.
44	(Jelks & Crain, 2020)	Inglés	Artículo	Estados Unidos	Profesional STEM	LogR	N.A.
45	(Jehn et al., 2022)	Inglés	Artículo	Canadá	Mercado Laboral	LinR	N.A.
46	(Zayed et al., 2022)	Inglés	Artículo	Palestina	Elección Carrera	DT / RF / SVM	N.A.
47	(Hao & Wang, 2022)	Inglés	Artículo	China	Mercado Laboral	HClustering	N.A.
48	(Zhu et al., 2021)	Inglés	Artículo	China	Alcance Profesional	N.A.	N.A.
49	(Ramírez, 2022)	Español	Artículo	México	Mercado Laboral	N.A.	N.A.
50	(Barrera et al., 2023)	Español	Artículo	México	Mercado Laboral	N.A.	N.A.
51	(Mason et al., 2023)	Inglés	Artículo	Australia	Decisión de Carrera	N.A.	N.A.
52	(Mejía et al., 2023)	Español	Artículo	México	Movilidad Social	N.A.	N.A.
53	(Valencia et al., 2023)	Inglés	Artículo	Colombia	Motivación para Posgrado	N.A.	N.A.
54	(Vieira, 2023)	Español	Artículo	Brasil	Movilidad Social	N.A.	N.A.
55	(Duryea et al., 2023)	Inglés	Artículo	Estados Unidos	Calidad de Universidad	N.A.	N.A.
56	(Moraes et al., 2022)	Inglés	Artículo	Brasil	Mercado Laboral	N.A.	N.A.
57	(Escabini, 2022)	Inglés	Conferencia	Perú	Mercado Laboral	N.A.	N.A.
58	(Aké-Uitz, 2022)	Español	Artículo	México	Movilidad Social	N.A.	N.A.
59	(Venegas et al., 2021)	Inglés	Artículo	Chile	Mercado Laboral	N.A.	N.A.
60	(Garza & Rivera, 2021)	Español	Conferencia	México	Mercado Laboral	N.A.	N.A.
61	(Concha et al., 2021)	Español	Artículo	Chile	Decisión de Carrera	N.A.	N.A.

62	(Schedlbauer et al., 2021)	Inglés	Artículo	Alemania	Mercado Laboral	N.A.	N.A.
63	(Zelada, 2021)	Inglés	Conferencia	Perú	Desarrollo Académico	N.A.	N.A.
64	(Romero & Becerra 2021)	Español	Artículo	México	Mercado Laboral	N.A.	N.A.
65	(Muñiz et al., 2021)	Español	Artículo	Argentina	Movilidad Social	N.A.	N.A.
66	(Flores & Rodríguez, 2021)	Español	Artículo	México	Movilidad Social	N.A.	N.A.
67	(Fernández et al., 2020)	Español	Artículo	México	Ambos Temas	N.A.	N.A.
68	(Pacovilca et al., 2020)	Español	Artículo	Perú	Ambos Temas	N.A.	N.A.
69	(Stanca et al., 2020)	Inglés	Artículo	Romania	Mercado Laboral	N.A.	N.A.
70	(Oyarce et al., 2020)	Español	Artículo	Perú	Mercado Laboral	N.A.	N.A.
71	(Bordón & Braga, 2020)	Inglés	Artículo	Chile	Mercado Laboral	N.A.	N.A.
72	(Flores & Vigier, 2020)	Inglés	Artículo	Ecuador	Modelo Educativo Dual	N.A.	N.A.
73	(Jiménez, 2019)	Español	Artículo	México	Mercado Laboral	N.A.	N.A.
74	(García & Cárdenas, 2018)	Español	Artículo	España	Mercado Laboral	N.A.	N.A.
75	(Azpíroz et al., 2023)	Español	Artículo	España	Mercado Laboral	N.A.	N.A.

Nota: De las 75 referencias bibliográficas encontradas, finalmente se utilizan para la evaluación de las técnicas de Machine Learning usadas en el mercado laboral y la movilidad social de egresados, las primeras 34; el resto sirven de apoyo para revisar las tendencias y desafíos generales de esta temática.