



# CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA  
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN  
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

## Modelo de Machine Learning para la Caracterización y Asignación Inteligente de Estudiantes en el Semillero de Investigación MAMBA

Jesús David Candelo González  
Laura Lorena Velasquez Mendieta

Documento presentado para optar por el título de Ingeniería de Sistemas

Mba. Ing. Julián Andrés Quimbayo Castro  
PhD. Mg. Ing. José Miguel Llanos Mosquera

Corporación Universitaria del Huila – CORHUILA

Faculta de Ingeniería

Ingeniería de Sistemas

2026

- 📍 Sede Quirinal: Calle 21 No. 6 - 01
- 📍 Sede Prado Alto: Calle 8 No. 32 – 49 PBX: (608) 8754220
- 📍 Sede Pitalito: Carrera 2 No. 1 – 27 - PBX: (608) 8350459
- ✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989  
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



# CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA  
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN  
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

---

**Cita** (Candelo & Velasquez, 2026)

---

**Referencia** Jesús Candelo & Laura Velasquez, (2026). *Modelo de Machine Learning para la Caracterización y Asignación Inteligente de Estudiantes en el Semillero de Investigación MAMBA* Tesis de pregrado Corporación Universitaria del Huila Corhuila.  
<https://corhuila.edu.co/>

---

Según normas APA 7ª edición

---



Grupo de Investigación INPROTI Código  
Semillero de Investigación MAMBA Código

**Repositorio Institucional:** <http://bxxxxxxx>

Corporación Universitaria del Huila – CORHUILA <https://corhuila.edu.co/>

El contenido de este documento se ampara en el derecho de expresión de sus autores y no representa el pensamiento ni la posición institucional de la Corporación Universitaria del Huila – CORHUILA. Los autores asumen la responsabilidad por los derechos de autor y conexos.

---

Sede Quirinal: Calle 21 No. 6 - 01  
 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220  
 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459  
 Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989  
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA  
"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



# CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA  
Vigilada Mineducación

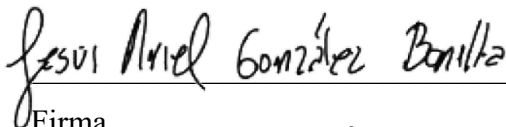
INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN  
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

## Nota de Aprobación

El presente trabajo opción de grado para obtener el título pregrado ha sido revisado y calificado con nota.

APROBADO

(Acuerdo 232 de 2023 del Consejo Académico)

  
Firma

Jesus Ariel Gonzalez Bonilla

Ingeniero de Sistemas - Docente Tiempo Completo

Jurado

  
Firma

Luis Angel Vargas Narvaez

Ingeniero de Sistemas - Docente Tiempo Completo

Jurado

Firma

(Nombre completo)

(Profesión)

Jurado (opcional)

- 📍 Sede Quirinal: Calle 21 No. 6 - 01
- 📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220
- 📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459
- ✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989  
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



# CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA  
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN  
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

“El Director y el Jurado del presente trabajo, no son responsables de las ideas y conclusiones expuestas en éste; ellas son exclusividad de sus autores”.

## Dedicatoria

*En primer lugar, agradecemos a Dios, por concedernos salud y la oportunidad de acceder a la educación superior, permitiéndonos ampliar nuestros conocimientos y presentar este proyecto, resultado de esfuerzo y dedicación.*

*Damos agradecimiento a nuestras familias, quienes, con su apoyo incondicional, tanto económico como emocional, han sido el pilar fundamental para avanzar en nuestra formación profesional. Su confianza y apoyo han sido la fuerza que nos impulsa a seguir adelante y a dar siempre lo mejor de nosotros.*

*Asimismo, expresamos nuestra gratitud a la Universidad y a los ingenieros Julián Quimbayo y José Miguel Llanos, cuya guía y enseñanza han sido clave en el desarrollo de este proyecto. Sus conocimientos y orientación nos han brindado las herramientas necesarias para convertir nuestras ideas y fortalecer nuestras habilidades en este proceso de aprendizaje.*

*A todos, gracias por ser parte de este camino.*

- 📍 Sede Quirinal: Calle 21 No. 6 - 01
- 📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220
- 📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459
- ✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989  
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

\*Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito\*



## Resumen y palabras claves

Este estudio desarrolló un modelo de Machine Learning basado en métodos de ensamble para la caracterización y asignación inteligente de estudiantes a rutas formativas personalizadas en el semillero de investigación MAMBA de la Corporación Universitaria del Huila (CORHUILA). El problema radica en la asignación empírica de estudiantes a proyectos sin considerar sus características individuales, generando insatisfacción y deserción. Aunque la inteligencia artificial demuestra potencial para personalizar experiencias educativas (Bousalem et al., 2023), su aplicación en semilleros latinoamericanos permanece subdesarrollada (MinCiencias, 2023). La metodología adoptó un enfoque cuantitativo descriptivo-predictivo con diseño no experimental transversal, aplicando un cuestionario validado de 35 preguntas a 150 estudiantes de ingeniería. Se integraron clustering no supervisado mediante Gaussian Mixture Models (GMM) y clasificación supervisada con seis algoritmos, seleccionando el Perceptrón Multicapa (MLP) optimizado como modelo final. Los resultados identificaron tres perfiles: Perfil Opción de Grado (42%), Perfil Meta Personal y Empleabilidad (28%) y Perfil Mejoramiento Académico (30%). El modelo alcanzó un F1-Score de 0.9404, con variables motivacionales acumulando el 88.8% del poder predictivo (Vanegas, 2024). Se concluye que la heterogeneidad motivacional es la norma en semilleros regionales y que este sistema ofrece un potencial escalable hacia otros contextos académicos similares.

**Palabras clave:** Machine Learning, semilleros de investigación, caracterización estudiantil, métodos de ensamble, personalización educativa.





## Abstract and key words

This study developed a machine learning model based on ensemble methods for the characterization and intelligent assignment of students to personalized training paths in the MAMBA research seedbed at the University Corporation of Huila (CORHUILA). The problem lies in the empirical assignment of students to projects without considering their individual characteristics, leading to dissatisfaction and dropout rates. Although artificial intelligence shows potential for personalizing educational experiences (Bousalem et al., 2023), its application in Latin American research centers remains underdeveloped (MinCiencias, 2023). The methodology adopted a quantitative descriptive-predictive approach with a non-experimental cross-sectional design, applying a validated 35-question questionnaire to 150 engineering students. Unsupervised clustering using Gaussian Mixture Models (GMM) and supervised classification with six algorithms were integrated, selecting the optimized Multilayer Perceptron (MLP) as the final model. The results identified three profiles: Degree Option Profile (42%), Personal Goal and Employability Profile (28%), and Academic Improvement Profile (30%). The model achieved an F1-Score of 0.9404, with motivational variables accounting for 88.8% of the predictive power (Vanegas, 2024). It is concluded that motivational heterogeneity is the norm in regional seedbeds and that this system offers scalable potential for other similar academic contexts.

**Keywords:** Machine learning, research seedbeds, student characterization, ensemble methods, educational personalization.





## Tabla de contenido

Resumen y palabras claves .....	5
Lista de figuras .....	10
Lista de Tablas.....	11
Siglas, acrónimos, abreviatura.....	12
Introducción.....	13
Planteamiento del problema .....	16
Justificación .....	20
Objetivos.....	23
Objetivo General: .....	23
Objetivos Específicos: .....	23
Marco referencial.....	24
Antecedentes Internacionales .....	25
Antecedentes Nacionales.....	26
Contexto Institucional: Semillero MAMBA .....	27
Marco Teórico .....	28
Rutas Formativas Diferenciadas.....	29
Motivación Estudiantil .....	30
Marco Conceptual.....	30
Aprendizaje Automático (Machine Learning - ML) .....	30
Minería de Datos Educativa (Educational Data Mining - EDM).....	30
Métodos de Ensamble.....	30

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989

NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



Marco Legal y Ético .....	32
Marco Normativo Nacional .....	32
Marco Normativo Institucional .....	33
Metodología.....	34
Tipo y Enfoque de Investigación.....	34
Población y Muestra .....	34
Variables del Estudio.....	34
Procedimiento Metodológico .....	37
Fase 1: Preparación y Diseño de Instrumentos (Meses 1-2) .....	37
Fase 2: Recolección y Preparación de Datos (Meses 2-4).....	37
Fase 3: Análisis Exploratorio y Validación de Perfiles mediante Clustering (Meses 4-6) ....	38
Fase 4: Desarrollo y Entrenamiento del Modelo de Ensamble (Meses 6-9) .....	41
Fase 5: Diseño de Rutas Formativas y Sistema de Asignación (Meses 9-11).....	47
Fase 6: Documentación, Análisis de Resultados y Difusión (Meses 11-12).....	49
Cronograma de actividades .....	50
Entregables por Mes .....	52
Presupuesto.....	53
Resultados y discusión .....	54
Validación Empírica de Perfiles Estudiantiles mediante Clustering .....	57
Caracterización de los Tres Perfiles Identificados .....	60
Validación Estadística de Diferencias entre Clusters .....	62
Desarrollo y Evaluación del Modelo Predictivo de Clasificación.....	63
Evaluación de Modelos Base.....	67

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989

NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



Selección del Modelo Final .....	69
Interpretación de hallazgos: .....	71
Discusión de Hallazgos .....	73
Conclusiones .....	77
Recomendaciones .....	79
Referencias .....	80

- 📍 Sede Quirinal: Calle 21 No. 6 - 01
- 📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220
- 📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459
- ✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989  
NIT. 800.107584-2



**CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA**

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



## Lista de figuras

<b>Figura 1.</b> Producción Científica de los grupos de Investigación.....	16
<b>Figura 2.</b> Desarrollo tecnológico e innovación. ....	17
<b>Figura 3.</b> Diagrama de barras Factores que limitan la participación en semilleros de investigación.....	20
<b>Figura 4.</b> Comparación Cuantitativa de Algoritmos de Clustering (Resultados Ajustados).....	39
<b>Figura 5.</b> Diagrama de dispersión de clusters en espacio reducido (PCA) .....	40
<b>Figura 6.</b> Resultados de entrenamiento y optimización de los modelos.....	42
<b>Figura 7.</b> Código fuente del mejor modelo.....	43
<b>Figura 8.</b> Comparativa de F1-Score Modelos Default vs Optimización .....	44
<b>Figura 9.</b> Matriz de Confusión .....	45
<b>Figura 10.</b> Curvas ROC y Análisis AUC de los clusters .....	46
<b>Figura 11.</b> Hiperparametro e impacto en los modelos.....	46
<b>Figura 12.</b> Diseño de las tres rutas formativas .....	48
<b>Figura 13.</b> Mejor Modelo MLP (Optimizado).....	49
<b>Figura 14.</b> Grafica edad promedio de estudiantes de ingeniería - Universidad Corhuila.....	54
<b>Figura 15.</b> Nivel Educativo Parental .....	55
<b>Figura 16.</b> Clusters con el modelo GMM (PCA) .....	57
<b>Figura 17.</b> Clusters con modelo K-Means (PCA) .....	58
<b>Figura 18.</b> Clusters con modelo DBSCAN (PCA).....	59
<b>Figura 19.</b> Perfiles de Motivación por Clusters.....	61
<b>Figura 20.</b> Importancia de características según modelo XGBoost .....	63
<b>Figura 21.</b> Resultados de entrenamiento de perfiles.....	64
<b>Figura 22.</b> Código fuente de la división inicial de los datos .....	65
<b>Figura 23.</b> Código fuente de remuestreo balanceado .....	66
<b>Figura 24.</b> Código fuente de la definición de modelos e hiperparametrización.....	67





## Lista de Tablas

<b>Tabla 1.</b> Distribución de Estudiantes por Cluster Asignado.....	40
<b>Tabla 2.</b> Estadística de gusto por programar por Cluster (Agglomerative).....	41
<b>Tabla 3.</b> Diagrama de Gantt.....	50
<b>Tabla 4.</b> Entregables por meses .....	52
<b>Tabla 5.</b> Tabla de presupuesto .....	53
<b>Tabla 6.</b> Métricas de Modelos de Clustering.....	60
<b>Tabla 7.</b> Desempeño de Modelos en conjunto de entrenamiento balanceado .....	68
<b>Tabla 8.</b> Top 10 Características más importantes.....	70





## Siglas, acrónimos, abreviatura

<b>Et al.</b>	Y otros
<b>P.</b>	Página
<b>Pp.</b>	Páginas
<b>Párr.</b>	Párrafo
<b>MSc.</b>	Magister Scientiae
<b>PhD</b>	Philosophiae Doctor
<b>CORHUILA</b>	Corporación Universitaria del Huila – CORHUILA
<b>OEI</b>	Organización de Estados Iberoamericanos
<b>ML</b>	Machine Learning (Aprendizaje Automático)
<b>EDM</b>	Educational Data Mining (Minería de Datos Educativa)
<b>EDA</b>	Exploratory Data Analysis (Análisis Exploratorio de Datos)
<b>SNCTI</b>	Sistema Nacional de Ciencia, Tecnología e Innovación
<b>Min Ciencias</b>	Ministerio de Ciencia, Tecnología e Innovación
<b>SMOTE</b>	Synthetic Minority Oversampling Technique Adaptive Synthetic
<b>ADASYN</b>	Sampling
<b>XGBoost</b>	eXtreme Gradient Boosting
<b>LightGBM</b>	Light Gradient Boosting Machine
<b>GridSearchCV</b>	Grid Search con Validación Cruzada (Cross-Validation)
<b>CV</b>	Cross-Validation (Validación Cruzada)
<b>DOI</b>	Digital Object Identifier
<b>RF</b>	Random Forest
<b>K</b>	K-Means
<b>AHG</b>	Agglomerative
<b>GMM</b>	Gaussian Mixture Model
<b>SC</b>	Spectral
<b>DBSCAN</b>	Density-Based Spatial Clustering of Applications with Noise.
<b>LR</b>	Logistic Regression
<b>SVC</b>	Support Vector Classifier
<b>NBC</b>	Naïve Bayes
<b>MLP</b>	Clasificador Perceptrón multicapa





## Introducción

A nivel internacional, los semilleros de investigación se han consolidado como espacios fundamentales para el desarrollo académico y profesional de los estudiantes universitarios, promoviendo un aprendizaje activo y la generación temprana de conocimiento (Alarcón et al., 2020). En Colombia, las instituciones de educación superior impulsan activamente estos espacios para fortalecer sus capacidades investigativas y vincular a los estudiantes con los desafíos del entorno social y productivo. Sin embargo, la asignación de estudiantes a proyectos y rutas formativas en estos semilleros tradicionalmente se ha realizado de manera empírica o uniforme, sin considerar sistemáticamente las características individuales de cada participante, lo que puede limitar tanto su satisfacción como su desarrollo óptimo dentro del programa. Esta asignación no sistematizada genera desajustes entre las expectativas estudiantiles y las actividades asignadas, resultando en tasas de deserción temprana y subóptimo aprovechamiento de las capacidades individuales para el desarrollo de competencias investigativas.

Dentro de este marco, el semillero de investigación MAMBA de la CORHUILA representa un caso de estudio relevante para diseñar estrategias de caracterización y asignación inteligente de estudiantes. Como señala (Vanegas, 2024), comprender las motivaciones que impulsan a los jóvenes investigadores es esencial para optimizar el impacto de los proyectos institucionales. La literatura previa sugiere que el análisis de datos y el aprendizaje automático (Machine Learning) ofrecen herramientas potentes para caracterizar perfiles estudiantiles e identificar patrones de comportamiento y preferencias (Bousalem et al., 2023). Específicamente, a diferencia de los métodos tradicionales basados en entrevistas o cuestionarios simples, el ML permite procesar múltiples variables simultáneamente, identificar patrones no lineales y adaptar las predicciones conforme se acumula información, desarrollando así sistemas de caracterización y asignación más efectivos, personalizados y escalables que respondan a las necesidades específicas de cada participante. Factores como la motivación intrínseca para aprender, las metas personales, el rendimiento académico previo y las expectativas profesionales influyen significativamente en las trayectorias que los estudiantes pueden seguir con mayor éxito (Lu et al., 2024). Reconociendo esta diversidad, surge la necesidad de implementar un sistema de

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989

NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



caracterización y perfilado inteligente. Para ello, este estudio propone un modelo basado en métodos de ensamble del aprendizaje automático, técnicas avanzadas que combinan las predicciones de múltiples modelos mediante estrategias como bagging, boosting o stacking. Estos métodos logran una clasificación más precisa y estable que modelos individuales, reduciendo el riesgo de sobreajuste y mejorando la generalización—características cruciales al tratar con la naturaleza heterogénea y subjetiva de los datos sobre motivación y comportamiento estudiantil.

En consecuencia, el objetivo principal de este trabajo es desarrollar un modelo de Machine Learning para la caracterización y asignación inteligente de estudiantes en el semillero de investigación MAMBA. A partir de la literatura sobre motivación y de observaciones preliminares en el semillero, el modelo busca caracterizar e identificar tres perfiles principales basados en la teoría de autodeterminación (Deci & Ryan, 2000):

1. **Mejoradores Académicos:** estudiantes con motivación intrínseca hacia la investigación formal y producción científica.
2. **Exploradores Vocacionales:** participantes en fase de descubrimiento de intereses profesionales.
3. **Orientados a Empleabilidad:** individuos enfocados en desarrollar competencias específicas para el mercado laboral.

Para lograrlo, se desarrollará un modelo predictivo analizando un conjunto de datos históricos de N=150 estudiantes de los últimos cinco años del semillero MAMBA, que incluye variables categorizadas en: demográficas (edad, programa académico, semestre), motivacionales (respuestas a cuestionarios de motivación inicial, metas profesionales declaradas) y de rendimiento (promedio académico, participación en actividades investigativas).

La implementación de este sistema de caracterización y asignación inteligente permitirá asignar automáticamente a cada nuevo miembro una ruta formativa personalizada que maximice su satisfacción y aprovechamiento. Más allá de optimizar la dinámica interna del semillero





# CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA  
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN  
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

MAMBA, esta estrategia ofrece una herramienta escalable y basada en datos para la toma de decisiones institucionales. Permitirá diseñar intervenciones más efectivas que fortalezcan la retención estudiantil, potencien el desarrollo de competencias diferenciadas y, en última instancia, maximicen el impacto de los semilleros como motores de conocimiento e innovación, en línea con su creciente importancia a nivel nacional (Vera-Fernández & Fernández-Nieto, 2024).

- 📍 Sede Quirinal: Calle 21 No. 6 - 01
- 📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220
- 📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459
- ✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989  
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"

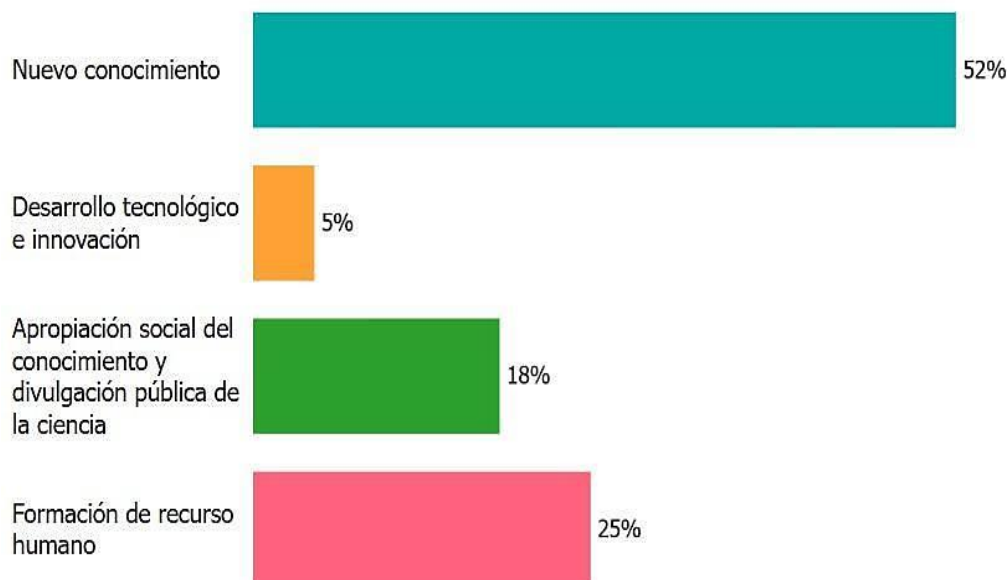


## Planteamiento del problema

Los semilleros de investigación son reconocidos como entornos cruciales en la educación superior, diseñados para fomentar el aprendizaje autónomo y desarrollar competencias investigativas tempranas (Alarcón et al., 2020). Estas experiencias impactan significativamente las futuras aspiraciones de los participantes (Zachos et al., 2018). En Colombia, un notable esfuerzo por fortalecer estos espacios (MinCiencias, 2023) ha resultado en un crecimiento significativo de su producción científica y tecnológica (OCyT, 2023). Como se evidencia en las Figura 1 y 2, existe un alto potencial de innovación y una fuerte orientación hacia la formación de talento humano. Sin embargo, este potencial se encuentra en riesgo si la caracterización y gestión de dicho talento no es óptima.

**Figura 1.** Producción Científica de los grupos de Investigación.

Seleccione un tipo de producto para visualizar los subtipos que lo conforman



**Nota.** Tomado de (Minciencias, 2021) *Producción Científica de los grupos de investigación.*

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989

NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



**Figura 2.** Desarrollo tecnológico e innovación.



**Nota.** Tomado de (Minciencias, 2021) *Desarrollo tecnológico e innovación.*

A pesar de su valor formativo, un desafío persistente es la ausencia de sistemas de caracterización sistemática y la consecuente asignación inadecuada de estudiantes a proyectos. Tradicionalmente, estos procesos se realizan de manera uniforme o empírica, sin considerar las características, motivaciones y expectativas individuales de cada participante. Esta falta de caracterización estructurada y personalización genera desajustes que conducen a la insatisfacción, el bajo rendimiento y, eventualmente, la deserción, desaprovechando así el potencial formativo del semillero y limitando el desarrollo de competencias diferenciadas.

Esta problemática se manifiesta claramente en el semillero de investigación MAMBA de la Universidad CORHUILA. Durante los últimos cinco años, se ha observado una notable





heterogeneidad en los perfiles de sus miembros: algunos ingresan con un alto rendimiento académico y aspiraciones de posgrado; otros buscan explorar áreas para descubrir sus intereses; y un tercer grupo se enfoca en desarrollar habilidades para una rápida inserción laboral. La ausencia de un sistema de caracterización y perfilado estructurado ha provocado desajustes críticos: estudiantes de alto potencial terminan en proyectos de bajo impacto investigativo, mientras que aquellos en fase exploratoria son asignados a iniciativas formales para las que aún no están preparados. Las consecuencias son evidentes: frustración estudiantil, alta rotación y una subutilización del talento que reduce la producción científica y tecnológica del semillero.

Ante la complejidad de caracterizar y clasificar a los estudiantes considerando múltiples variables interrelacionadas, las aproximaciones tradicionales resultan insuficientes. En este escenario, el Machine Learning (ML) emerge como un enfoque idóneo para desarrollar sistemas de caracterización y perfilado inteligente (Hermanrud & Eide, 2017). Específicamente, las técnicas de clasificación y agrupamiento, con especial énfasis en los métodos de ensamble (como Random Forest, Gradient Boosting y XGBoost), han demostrado ser efectivas en la minería de datos educativa (Romero & Ventura, 2010). Estos métodos combinan múltiples modelos para lograr una mayor precisión y robustez (Breiman, 2001), permitiendo analizar de forma fiable la interacción de factores demográficos, motivacionales y académicos para caracterizar y clasificar a los estudiantes en perfiles diferenciados.

A pesar de su potencial, existe una brecha en la aplicación de estas herramientas para la caracterización sistemática y asignación personalizada en semilleros de investigación. Abordar esta brecha mediante un modelo de caracterización y perfilado inteligente permitiría no solo identificar automáticamente los perfiles de los nuevos miembros, sino también asignarlos de manera óptima a rutas formativas diseñadas específicamente para cada perfil: una "Ruta de Investigación y Profundización" para Mejoradores Académicos, una "Ruta de Descubrimiento y Fundamentos" para Exploradores Vocacionales, y una "Ruta Aceleradora de Carrera" para estudiantes Orientados a Empleabilidad. Este enfoque basado en evidencia es fundamental para





# CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA  
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN  
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

mejorar la experiencia estudiantil y potenciar los resultados del semillero. Por lo tanto, este estudio busca responder a la siguiente pregunta de investigación:

*¿Cómo desarrollar un modelo de Machine Learning para la caracterización y asignación inteligente de estudiantes en el semillero MAMBA, basado en métodos de ensamble, que mejore su satisfacción y optimice su ruta de formación?*

- 📍 Sede Quirinal: Calle 21 No. 6 - 01
- 📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220
- 📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459
- ✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989  
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"

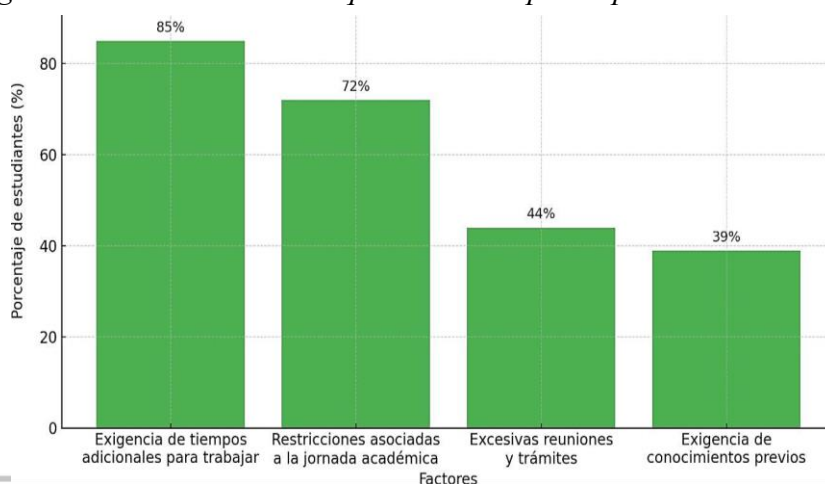


## Justificación

Esta investigación se justifica por la necesidad crítica de superar la caracterización inadecuada y la consecuente asignación deficiente de estudiantes en el semillero MAMBA, un fenómeno que genera insatisfacción, desaprovecha el talento y limita el desarrollo de competencias diferenciadas. Si bien los semilleros son un pilar en la formación investigativa (Brew & Mantai, 2017), su efectividad se ve comprometida por la ausencia de sistemas de caracterización sistemática y por el desajuste entre el perfil del estudiante y las actividades asignadas. Por tanto, desarrollar un modelo de Machine Learning para la caracterización y asignación inteligente de estudiantes es fundamental para maximizar la satisfacción, retención y desarrollo óptimo de los participantes.

La urgencia de esta caracterización y personalización se confirma al analizar las barreras que enfrentan los estudiantes. Como evidencia la Figura 3, factores como la exigencia de tiempo adicional (85%), los conflictos con la jornada académica (72%) y la percepción de una alta exigencia de conocimientos previos (39%) constituyen limitaciones significativas (Gómez-Cano et al., 2022). Un sistema de caracterización inteligente que permita diseñar rutas formativas diferenciadas aborda directamente estos obstáculos, ofreciendo itinerarios flexibles y adaptados al nivel de preparación, motivación y expectativas de cada miembro, optimizando así su experiencia y permanencia.

**Figura 3.** Diagrama de barras Factores que limitan la participación en semilleros de investigación.





*Nota.* Adaptado de "Factores que limitan la participación en semilleros de investigación" por (Gómez-Cano et al., 2022)

La complejidad de caracterizar y clasificar perfiles estudiantiles considerando múltiples variables interrelacionadas demanda herramientas que superen la asignación empírica tradicional. Aquí radica la pertinencia metodológica de este estudio: la aplicación de Machine Learning, específicamente métodos de ensamble. Estas técnicas son capaces de modelar relaciones no lineales complejas y mejorar la precisión en la clasificación al combinar múltiples algoritmos (Breiman, 2001). Se justifica su uso porque permiten construir un modelo robusto y generalizable, capaz de identificar los patrones sutiles que distinguen los tres perfiles propuestos (Mejoradores Académicos, Exploradores Vocacionales y Orientados a Empleabilidad), algo inalcanzable con métodos convencionales de caracterización.

El aporte práctico reside en la generación de una herramienta automatizada y basada en evidencia para la caracterización y gestión del talento en el semillero MAMBA. El sistema no solo caracteriza y clasifica objetivamente a los nuevos miembros, sino que facilita su vinculación a rutas formativas con un propósito claro: los Mejoradores Académicos a proyectos de investigación formal y producción científica; los Exploradores Vocacionales a experiencias de descubrimiento y construcción de fundamentos; y los Orientados a Empleabilidad al desarrollo de competencias específicas y portafolios competitivos. Esta innovación en la caracterización y gestión del talento optimiza la satisfacción estudiantil y maximiza el desarrollo de competencias diferenciadas para cada perfil.

Desde una perspectiva académica, el estudio contribuye al campo de la inteligencia artificial aplicada a la educación superior. Aporta un modelo validado para la caracterización de perfiles y personalización formativa en espacios de investigación estudiantil, un área menos explorada que los sistemas de recomendación curriculares tradicionales. Además, enriquece la literatura sobre minería de datos educativa (Romero & Ventura, 2010) al identificar los factores diferenciadores en perfiles motivacionales y académicos de estudiantes participantes en





# CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA  
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN  
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

semilleros de investigación, ofreciendo un caso de estudio replicable en el contexto latinoamericano y alineado con las teorías contemporáneas de motivación académica.

Finalmente, la viabilidad del proyecto está asegurada. Se cuenta con un conjunto de datos históricos de cinco años del semillero MAMBA (N=150 estudiantes), que incluye variables demográficas, motivacionales y de rendimiento académico, suficientes para el entrenamiento y validación de modelos de ensamble. Adicionalmente, se dispone del soporte institucional de la Universidad CORHUILA, las herramientas computacionales necesarias (Python, bibliotecas de ML) y la experiencia del equipo investigador en análisis de datos y Machine Learning. La alineación del tema con las prioridades nacionales de fortalecimiento de la investigación y formación de talento humano (MinCiencias, 2023) refuerza su relevancia institucional y potencial de escalabilidad a otros semilleros y contextos académicos similares.

- 📍 Sede Quirinal: Calle 21 No. 6 - 01
- 📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220
- 📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459
- ✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989  
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



## Objetivos

### Objetivo General:

Desarrollar un modelo de Machine Learning basado en métodos de ensamble para la caracterización y asignación inteligente de estudiantes a rutas formativas personalizadas en el semillero de investigación MAMBA.

### Objetivos Específicos:

1. Caracterizar los tres perfiles estudiantiles del semillero (Mejorador Académico, Motivado por Metas y Orientado a lo Profesional) a partir del análisis de datos históricos.
2. Desarrollar un modelo híbrido de Machine Learning que combine clustering para la validación de perfiles y un clasificador de ensamble para la predicción automática en nuevos miembros.
3. Estructurar tres rutas formativas personalizadas y alineadas a las necesidades y metas de cada perfil identificado.





## Marco referencial

La presente investigación sobre el desarrollo de un modelo de Machine Learning para la caracterización y asignación inteligente de estudiantes en el semillero MAMBA mediante métodos de ensamble se fundamenta en una revisión de la literatura que revela un panorama donde los semilleros de investigación son reconocidos internacionalmente como espacios cruciales para el desarrollo integral en la educación superior, fomentando competencias investigativas, pensamiento crítico y aprendizaje activo desde el pregrado (Alarcón et al., 2020). Experiencias como las "undergraduate research experiences" han demostrado impactar positivamente las trayectorias académicas y profesionales de los estudiantes (Zachos et al., 2018). En el ámbito nacional colombiano, existe un impulso significativo para fortalecer estos espacios, evidenciado en políticas de fomento e inversión (MinCiencias, 2023) y un notable crecimiento en su número (OCyT, 2023).

Sin embargo, la efectividad de estos espacios formativos depende críticamente de la adecuada caracterización de perfiles estudiantiles y la consecuente asignación de estudiantes a proyectos y rutas de desarrollo. La literatura sobre personalización educativa ha identificado que los enfoques "one-size-fits-all" resultan ineficientes ante la diversidad de perfiles, motivaciones y objetivos estudiantiles (Bousalem et al., 2023). Los estudiantes presentan diferentes niveles de preparación académica, orientaciones profesionales y expectativas sobre su participación en actividades de investigación (Lu et al., 2024), lo que demanda sistemas de caracterización y asignación que reconozcan y respondan a estas diferencias individuales.

Paralelamente, el campo de la Minería de Datos Educativa (EDM) y el Aprendizaje Automático (ML) ha demostrado un potencial considerable para caracterizar perfiles y personalizar experiencias educativas, incluyendo sistemas de recomendación de cursos, clasificación de estilos de aprendizaje y asignación adaptativa de recursos pedagógicos ((Romero & Ventura, 2010);(Forero-Corba & Negre Bennasar, 2023);(Bousalem et al., 2023) ). Dentro del ML, los métodos de ensamble se han consolidado como técnicas robustas que mejoran la





precisión en tareas de clasificación al combinar múltiples modelos, encontrando aplicación exitosa en diversos dominios educativos (Breiman, 2001); (Dietterich, 2000).

No obstante, a pesar de estos avances, la literatura evidencia vacíos significativos que esta investigación busca abordar. Específicamente, se observa una aplicación limitada de técnicas avanzadas como los métodos de ensamble para desarrollar sistemas automatizados de caracterización, perfilado y asignación en el contexto particular de los semilleros de investigación. La mayoría de los estudios sobre personalización educativa se centran en contextos curriculares formales (asignación de estudiantes a secciones de cursos, recomendación de materias electivas), y aquellos que abordan espacios extracurriculares suelen emplear métodos de asignación empírica o basados en autoselección sin soporte analítico enfocado en la investigación estudiantil (Walkington, 2021). Además, falta investigación que integre sistemáticamente variables demográficas, motivacionales y de rendimiento académico utilizando modelos predictivos robustos para caracterizar y clasificar perfiles estudiantiles, y diseñar rutas formativas diferenciadas dentro de estos espacios formativos específicos (Catal et al., 2023), particularmente en el contexto colombiano. En este sentido, el presente trabajo se posiciona para contribuir al conocimiento al enfocarse en el espacio específico de un semillero (MAMBA), utilizando una técnica avanzada y robusta (métodos de ensamble) con el objetivo de construir un sistema de caracterización y perfilado inteligente que permita asignar automáticamente rutas formativas personalizadas, integrando múltiples tipos de factores caracterizadores.

## Antecedentes Internacionales

A nivel internacional, diversas instituciones han implementado estrategias de diferenciación de rutas formativas en programas de investigación estudiantil (Walkington, 2021). Las "undergraduate research experiences" en universidades norteamericanas frecuentemente segmentan a los participantes según sus objetivos académicos: estudiantes orientados a estudios de posgrado son vinculados a laboratorios de investigación básica con mayor rigor metodológico, mientras que aquellos con orientación profesional inmediata participan en proyectos aplicados con enfoque en desarrollo de habilidades técnicas transferibles (Zachos et al., 2018). Sin





embargo, esta segmentación generalmente se realiza mediante procesos de aplicación y entrevistas personales, sin aprovechar técnicas de Machine Learning para automatizar y optimizar la caracterización y clasificación de perfiles.

En el campo de la educación personalizada mediante ML, estudios recientes han explorado la caracterización y clasificación de estudiantes según estilos de aprendizaje, orientaciones profesionales y niveles de preparación académica. (Bousalem et al., 2023) demostraron la aplicabilidad de algoritmos de clasificación para segmentar estudiantes universitarios según sus características cognitivas y motivacionales, aunque sin vincularlo específicamente a sistemas de asignación en contextos de investigación formativa. (Lu et al., 2024) identificaron que factores como la motivación intrínseca, la conciencia metacognitiva y el bienestar general permiten predecir trayectorias estudiantiles diferenciadas, sugiriendo la viabilidad de utilizar estas variables para caracterizar perfiles y diseñar rutas personalizadas.

## Antecedentes Nacionales

En Colombia, el fortalecimiento de los semilleros de investigación ha sido una prioridad institucional y gubernamental ((MinCiencias, 2023); (OCyT, 2023), reflejado en el crecimiento sostenido de estos espacios. Sin embargo, la gestión de semilleros tradicionalmente se ha basado en enfoques empíricos para la caracterización, vinculación y asignación de estudiantes. La mayoría de las instituciones asignan proyectos a los participantes mediante convocatorias abiertas o designación directa por coordinadores, sin utilizar herramientas analíticas que consideren sistemáticamente las características individuales de los estudiantes.

(Vanegas, 2024) enfatiza la importancia de comprender las motivaciones que impulsan a los jóvenes investigadores colombianos, señalando que estas varían significativamente entre perfiles: algunos buscan formación académica avanzada, otros exploran diferentes áreas para descubrir vocaciones, y un tercer grupo prioriza el desarrollo de competencias técnicas para el mercado laboral. Esta diversidad motivacional requiere enfoques diferenciados de





caracterización, acompañamiento y asignación, área que permanece subdesarrollada en la práctica institucional nacional.

(Gómez-Cano et al., 2022) identificaron las principales barreras que enfrentan los estudiantes al participar en semilleros de investigación en Colombia, incluyendo exigencias de tiempo (85%), restricciones por jornada académica (72%), excesivas reuniones y trámites (44%), y requisitos de conocimientos previos (39%). Esta heterogeneidad de obstáculos refuerza la necesidad de sistemas de caracterización y asignación personalizados que consideren las circunstancias y capacidades individuales de cada estudiante.

## Contexto Institucional: Semillero MAMBA

En el contexto específico de la Universidad CORHUILA, el semillero de investigación MAMBA ha operado durante los últimos cinco años con un modelo de vinculación relativamente uniforme. Los nuevos miembros tradicionalmente se incorporan a proyectos disponibles según disponibilidad de cupos o afinidad declarada con áreas temáticas, sin un proceso estructurado de caracterización previa. Esta aproximación ha resultado en desajustes entre las capacidades, expectativas y actividades asignadas, manifestándose en variabilidad en la satisfacción estudiantil y aprovechamiento diferencial del potencial formativo del semillero.

Observaciones preliminares han identificado tres perfiles recurrentes entre los participantes de MAMBA: estudiantes con excelente rendimiento académico y motivaciones orientadas hacia investigación formal y estudios de posgrado (Mejoradores Académicos); estudiantes en fase exploratoria que buscan descubrir sus áreas de interés específicas (Exploradores Vocacionales); y estudiantes enfocados prioritariamente en desarrollar habilidades técnicas que mejoren su empleabilidad inmediata (Orientados a Empleabilidad). Sin embargo, la ausencia de un sistema automatizado de caracterización, identificación y asignación ha limitado la capacidad del semillero para responder efectivamente a esta diversidad de perfiles.





## Marco Teórico

Para fundamentar conceptualmente esta investigación, se recurre a teorías y definiciones provenientes de la Psicología Educativa, la Pedagogía y la Ciencia de la Computación.

## Semilleros de Investigación

Los semilleros de investigación se entienden como espacios extracurriculares de formación integral donde estudiantes, guiados por docentes, desarrollan competencias investigativas mediante la participación en proyectos. Los semilleros funcionan como comunidades de aprendizaje que facilitan la transición de los estudiantes desde roles pasivos de consumidores de conocimiento hacia roles activos de generadores de conocimiento (Leonetti et al., 2023).

## Caracterización y Perfilado Estudiantil

La caracterización estudiantil es el proceso sistemático de identificación y análisis de los atributos demográficos, académicos, motivacionales y comportamentales de los estudiantes. Este proceso permite comprender la diversidad de características presentes en una población estudiantil y constituye la base para el perfilado.

El perfilado estudiantil es el proceso de clasificación de estudiantes en grupos homogéneos según sus atributos caracterizados, facilitando la personalización de experiencias educativas (Zawacki-Richter et al., 2019). En el contexto de esta investigación, basándose en la teoría de autodeterminación (Deci & Ryan, 2000) y observaciones preliminares del semillero MAMBA, se proponen tres perfiles principales:

- **Mejoradores Académicos:** Estudiantes con excelente rendimiento académico previo, alta motivación intrínseca hacia la investigación formal, interés en publicación científica y orientación hacia estudios de posgrado.

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989  
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



- **Exploradores Vocacionales:** Estudiantes con motivaciones diversas (mejorar rendimiento, explorar áreas, especializarse) y rendimiento académico promedio, que se encuentran en fase de descubrimiento de sus intereses específicos y construcción de fundamentos investigativos.
- **Orientados a Empleabilidad:** Estudiantes con alta motivación extrínseca por la empleabilidad y el desarrollo de competencias técnicas de alta demanda en el mercado laboral, enfocados en la aplicación práctica del conocimiento.

## Rutas Formativas Diferenciadas

Las rutas formativas son trayectorias estructuradas de desarrollo que especifican los tipos de proyectos, actividades, mentoría y estrategias de acompañamiento que se ofrecen a los estudiantes según su perfil caracterizado. La diferenciación de rutas permite optimizar la experiencia educativa al alinear las oportunidades formativas con las necesidades y objetivos específicos de cada grupo de estudiantes. Para este estudio, se proponen tres rutas:

- **Ruta de Investigación y Profundización:** Diseñada para Mejoradores Académicos, enfocada en proyectos de investigación formal, producción científica y preparación para estudios de posgrado.
- **Ruta de Descubrimiento y Fundamentos:** Diseñada para Exploradores Vocacionales, enfocada en experiencias diversas de exploración, construcción de fundamentos investigativos y descubrimiento de intereses específicos.
- **Ruta Aceleradora de Carrera:** Diseñada para estudiantes Orientados a Empleabilidad, enfocada en desarrollo de competencias técnicas específicas, construcción de portafolios y preparación para inserción laboral.

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989  
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



## Motivación Estudiantil

La motivación estudiantil se aborda como un constructo psicológico complejo, influenciado por factores intrínsecos (interés genuino, curiosidad intelectual) y extrínsecos (reconocimiento, recompensas, empleabilidad). La teoría de autodeterminación (Deci & Ryan, 2000) proporciona un marco conceptual para comprender cómo diferentes tipos de motivación influyen en el comportamiento y las preferencias estudiantiles. La motivación es un predictor clave de la satisfacción, el compromiso y la permanencia estudiantil en actividades académicas ((Vanegas, 2024); (Lu et al., 2024)).

## Marco Conceptual

### Aprendizaje Automático (Machine Learning - ML)

El Aprendizaje Automático se define como el campo de la inteligencia artificial que desarrolla algoritmos capaces de aprender patrones de los datos sin ser explícitamente programados para cada tarea específica (Forero-Corba & Negre Bennasar, 2023). El ML permite construir modelos predictivos que pueden caracterizar, clasificar, predecir o recomendar basándose en características observadas.

### Minería de Datos Educativa (Educational Data Mining - EDM)

La Minería de Datos Educativa es la aplicación específica de técnicas de ML y análisis de datos al contexto educativo, con el objetivo de descubrir patrones, caracterizar perfiles estudiantiles, predecir resultados, personalizar experiencias de aprendizaje y optimizar procesos pedagógicos (Romero & Ventura, 2010).

## Métodos de Ensamble

Centrales para este estudio son los métodos de ensamble, técnicas de ML que construyen y combinan múltiples modelos de clasificación o predicción para lograr resultados más precisos y



robustos que cualquier modelo individual (Breiman, 2001); (Dietterich, 2000). Los ensambles pueden basarse en diferentes estrategias:

- **Bagging (Bootstrap Aggregating):** Entrena múltiples modelos sobre subconjuntos aleatorios de los datos y combina sus predicciones mediante votación. Ejemplo: Random Forest.
- **Boosting:** Entrena modelos secuencialmente, donde cada nuevo modelo se enfoca en corregir los errores del anterior. Ejemplos: Gradient Boosting, XGBoost, AdaBoost.
- **Stacking:** Combina predicciones de diversos modelos base utilizando un meta-modelo que aprende la mejor forma de integrarlas.

Estos métodos son particularmente apropiados para la caracterización de perfiles estudiantiles debido a su capacidad de manejar relaciones no lineales complejas, reducir el sobreajuste y mejorar la generalización en conjuntos de datos heterogéneos.

## Clasificación Supervisada

La tarea específica abordada es la clasificación supervisada, que consiste en asignar una etiqueta categórica (en este caso, uno de los tres perfiles identificados) a cada estudiante basándose en sus características observadas. El modelo aprende la relación entre variables predictoras (rendimiento académico, motivaciones declaradas, factores demográficos) y la variable objetivo (perfil estudiantil) utilizando datos históricos etiquetados, según (Kotsiantis et al., 2004)

## Clustering No Supervisado

Como complemento a la clasificación supervisada, se utilizarán técnicas de clustering (agrupamiento no supervisado) para validar la existencia natural de los tres perfiles propuestos en los datos históricos. Algoritmos como K-Means, DBSCAN o clustering jerárquico permitirán

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989

NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



identificar patrones de agrupamiento sin etiquetas previas, confirmando o refinando la tipología de perfiles antes de entrenar el clasificador supervisado, como lo dice (Dutt et al., 2017).

## Sistemas de Asignación Inteligente

Según (Xing et al., 2016) un sistema de asignación inteligente es un sistema automatizado que utiliza el modelo de caracterización y clasificación entrenado para identificar el perfil de nuevos miembros del semillero, predecir su categoría y asignarlos automáticamente a la ruta formativa más adecuada, optimizando así su satisfacción y desarrollo académico-profesional.

## Marco Legal y Ético

### *Marco Normativo Nacional*

A nivel nacional, la Ley 30 de 1992 sobre Educación Superior establece el marco general para la organización y funcionamiento de las instituciones de educación superior en Colombia, reconociendo la investigación como función sustantiva universitaria. Las políticas de MinCiencias (anteriormente Colciencias) sobre Ciencia, Tecnología e Innovación (CTI) orientan el fomento a la investigación formativa y los semilleros como estrategia de desarrollo científico nacional (República de Colombia, 1992).

De crucial importancia para esta investigación es la Ley 1581 de 2012 (Ley de Habeas Data) y sus decretos reglamentarios (especialmente el Decreto 1377 de 2013), (República de Colombia, 2012) que establecen el marco jurídico para la protección de datos personales en Colombia. Esta normativa exige procedimientos claros y transparentes para la recolección, almacenamiento, uso y tratamiento de información sensible, incluyendo datos académicos, demográficos y motivacionales de estudiantes. El desarrollo de sistemas automatizados de caracterización y perfilado debe garantizar:

- **Consentimiento informado:** Los estudiantes deben autorizar explícitamente el uso de sus datos para fines de investigación y mejora de procesos formativos.

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989  
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



- **Finalidad legítima:** Los datos solo pueden utilizarse para los propósitos declarados (diseño del sistema de caracterización y asignación).
- **Confidencialidad y seguridad:** Implementación de medidas técnicas y administrativas para proteger la información contra accesos no autorizados.
- **Derechos de los titulares:** Garantía de derechos de acceso, rectificación, actualización y supresión de datos personales.

## *Marco Normativo Institucional*

A nivel institucional, la Universidad CORHUILA dispone de un marco normativo que regula y respalda el desarrollo ético de la investigación que se lleva a cabo. El Reglamento de Investigaciones establece los lineamientos generales para la formulación y ejecución de proyectos investigativos, garantizando la protección de los participantes humanos y el cumplimiento de principios éticos fundamentales. De manera complementaria, el Reglamento Estudiantil (Corhuila, 2014) define los derechos y deberes de los estudiantes, incluyendo su participación en actividades de investigación formativa, asegurando que dicha participación se realice de forma voluntaria, informada y respetuosa de su condición académica.

Asimismo, la Política de Tratamiento de Datos Personales de la Universidad CORHUILA (Corhuila, 2016) regula el manejo de la información sensible de los estudiantes, en concordancia con la Ley 1581 de 2012, estableciendo procedimientos para la recolección, almacenamiento, uso y protección de los datos personales. El cumplimiento de estas disposiciones es supervisado por el Comité de Ética en Investigación, instancia institucional encargada de evaluar y aprobar los proyectos que involucren seres humanos, velando por la observancia de los principios de beneficencia, no maleficencia, autonomía y justicia durante todo el proceso investigativo.

El presente estudio se compromete al cumplimiento riguroso de estas normativas, implementando procesos de consentimiento informado, anonimización de datos cuando sea apropiado, y sometiendo el protocolo de investigación a la aprobación del Comité de Ética en Investigación de la Universidad CORHUILA antes de iniciar la recolección y análisis de datos.

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989  
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



## Metodología

### Tipo y Enfoque de Investigación

Esta investigación adoptó un enfoque cuantitativo con alcance descriptivo-predictivo. El diseño fue no experimental transversal, ya que se analizaron datos recolectados mediante un cuestionario estructurado aplicado a estudiantes del semillero MAMBA sin manipulación de variables. El estudio se enmarcó en el paradigma de investigación aplicada, orientado a resolver un problema práctico mediante el desarrollo de un sistema de caracterización y asignación inteligente basado en Machine Learning.

### Población y Muestra

Población: Estudiantes que participaron o mostraron interés en participar en el semillero de investigación MAMBA de la Universidad CORHUILA durante el período de estudio.

- Muestra: N=150 estudiantes que completaron el cuestionario de caracterización y cumplieron con los criterios de inclusión:
- Estudiantes activos de programas de ingeniería de CORHUILA
- Haber cursado al menos las asignaturas básicas de programación (Lógica de programación, Programación 1, Programación 2)
- Cuestionario completo sin valores faltantes críticos
- Tipo de muestreo: No probabilístico por conveniencia, utilizando la totalidad de cuestionarios válidos recolectados durante el período de estudio.

### Variables del Estudio

Con base en el instrumento aplicado (35 preguntas), las variables se categorizaron en:

#### Variables Independientes (Predictoras)

##### 1. Variables Demográficas y Socioeconómicas:

Q1: Género (categórica binaria: 0=Femenino, 1=Masculino)

Q2: Edad (numérica discreta)

Q29: Edad de ingreso a la carrera (numérica discreta)

Q32: Estrato socioeconómico (categórica ordinal: 1-6)

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989  
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

\*Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito\*



Q26: Tipo de establecimiento educativo básico (categórica binaria: 0=Público, 1=Privado)

Q27: Tipo de establecimiento educativo medio (categórica binaria: 0=Público, 1=Privado)

## 2. Variables del Contexto Familiar:

Q11: Número de integrantes del núcleo familiar (numérica discreta: 1-10)

Q12: Número de integrantes que trabajan (numérica discreta: 0-10)

Q13: Número de integrantes que estudian (numérica discreta: 0-10)

Q14: Número de integrantes pensionados (numérica discreta: 0-10)

Q15: Nivel educativo de la madre (categórica ordinal: 1-5)

Q16: Nivel educativo del padre (categórica ordinal: 1-5)

## 3. Variables Académicas y de Rendimiento:

Q28: Semestre actual (numérica discreta)

Q33: Nota en Lógica de programación (numérica continua: 0.0-5.0)

Q34: Nota en Programación 1 (numérica continua: 0.0-5.0)

Q35: Nota en Programación 2 (numérica continua: 0.0-5.0)

Q7: Horas diarias dedicadas a estudiar (numérica continua)

Q8: Horas semanales dedicadas a practicar programación (numérica continua)

Q9: Días de anticipación para estudiar exámenes (numérica discreta)

## 4. Variables de Hábitos de Estudio y Competencias:

Q3: Nivel de responsabilidad autopercebida (escala Likert: 1-5)

Q4: Autopercepción de aptitudes para investigación (categórica binaria: 0=No, 1=Sí)

Q5: Autopercepción de inteligencia (categórica ordinal: 1-3)

Q6: Gusto por la programación (escala Likert: 1-5)

Q10: Aspecto más importante para la vida académica (categórica nominal)

## 5. Variables Psicosociales:

Q17: Influencia del estrés en vida académica (categórica ordinal: 1-4)

Q18: Nivel de reacción violenta ante estrés (escala Likert: 1-5)

Q19: Nivel de empatía hacia compañeros (categórica ordinal: 1-4)

Q22: Estado de relación amorosa (categórica binaria: 0=No, 1=Sí)

## 6. Variables Laborales y de Apoyo:

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989

NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



# CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA  
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN  
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

Q20: Horas laboradas semanalmente (numérica discreta: 0-50)

Q21: Apoyo económico en matrícula (categórica binaria: 0=No, 1=Sí)

## 7. Variables Motivacionales y Vocacionales:

Q23: Nivel de preferencia al escoger la carrera (categórica ordinal: 0-4)

Q24: Probabilidad de cambio de carrera (categórica binaria: 0=Poco probable, 1=Probable)

Q25: Probabilidad de estudios de posgrado (categórica binaria: 0=Poco probable, 1=Probable)

Q30: Motivación para ingresar al semillero (categórica nominal - texto libre)

Q31: Campo de acción más importante para desempeño profesional (categórica nominal)

## Variable Dependiente (Objetivo)

Perfil estudiantil: Variable categórica nominal con tres clases:

Cluster 0 - Mejoradores Académicos

Cluster 1 - Exploradores Vocacionales

Cluster 2 - Orientados a Empleabilidad

- 📍 Sede Quirinal: Calle 21 No. 6 - 01
- 📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220
- 📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459
- ✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989  
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



## Procedimiento Metodológico

### Fase 1: Preparación y Diseño de Instrumentos (Meses 1-2)

Esta fase inicial estableció las bases metodológicas del proyecto, garantizando la validez del instrumento de recolección de datos y el cumplimiento de estándares éticos institucionales.

#### Actividades:

En la fase metodológica se realizó una revisión sistemática de literatura sobre perfiles estudiantiles, motivación académica y aplicaciones de Machine Learning en educación superior, lo que permitió fundamentar el diseño de un cuestionario estructurado de 35 preguntas que integró variables demográficas, socioeconómicas, académicas, motivacionales, psicosociales y vocacionales. Posteriormente, el instrumento fue validado por un panel de tres expertos para garantizar su validez de contenido y constructo, y se aplicó un pilotaje a 25 estudiantes con el fin de evaluar claridad, tiempo de respuesta y confiabilidad mediante el cálculo del Alfa de Cronbach. Con base en los resultados obtenidos, se realizaron ajustes finales al cuestionario, se elaboró el protocolo de investigación detallando objetivos y procedimientos, y finalmente se implementó el instrumento en una plataforma digital (Google Forms) con las respectivas validaciones de campos y rangos numéricos.

#### Productos:

- Cuestionario validado de 35 preguntas
- Reporte de validación del instrumento (pilotaje)

### Fase 2: Recolección y Preparación de Datos (Meses 2-4)

Esta fase se centró en la obtención de datos primarios mediante la aplicación del cuestionario y su posterior transformación en un dataset estructurado y limpio, listo para análisis.





**Actividades:** En esta fase se aplicó el cuestionario mediante estrategias presenciales y virtuales, incluyendo su difusión entre estudiantes de ingeniería de CORHUILA, aplicación durante sesiones informativas del semillero MAMBA y envío del enlace por correo institucional, realizando seguimiento continuo hasta alcanzar el tamaño muestral de 150 participantes. Posteriormente, se efectuó la extracción y consolidación de datos mediante la descarga de respuestas, creación de la base en formato CSV/Excel, codificación de variables categóricas según un diccionario predefinido y asignación de códigos anónimos a los participantes. Finalmente, se desarrolló el proceso de limpieza y preprocesamiento, que incluyó tratamiento de valores faltantes (mediana y moda), detección de outliers con el método IQR, codificación de variables mediante Label Encoding y One-Hot Encoding, normalización con MinMaxScaler y creación de variables derivadas como el promedio de notas en programación, el índice de carga laboral-académica y el índice de contexto familiar educativo.

#### Productos:

- Dataset limpio y estructurado (150 observaciones × 35+ variables)
- Diccionario de datos documentado con descripción de cada variable
- Base de datos anonimizada cumpliendo normativas de protección de datos

### Fase 3: Análisis Exploratorio y Validación de Perfiles mediante Clustering (Meses 4-6)

Esta fase combinó técnicas de análisis exploratorio y aprendizaje no supervisado para validar empíricamente la existencia de los tres perfiles estudiantiles propuestos conceptualmente.

#### Actividades:

En esta fase se realizó un Análisis Exploratorio de Datos (EDA) que incluyó estadística descriptiva completa, evaluación de distribuciones mediante pruebas de normalidad, análisis de asimetría y curtosis, construcción de matriz de correlación y generación de visualizaciones como histogramas, gráficos de barras, boxplots y scatter plots, además de tablas de frecuencia para variables categóricas. Posteriormente, se aplicó clustering no supervisado para validar perfiles estudiantiles, seleccionando variables relevantes, estandarizando los datos y

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989  
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



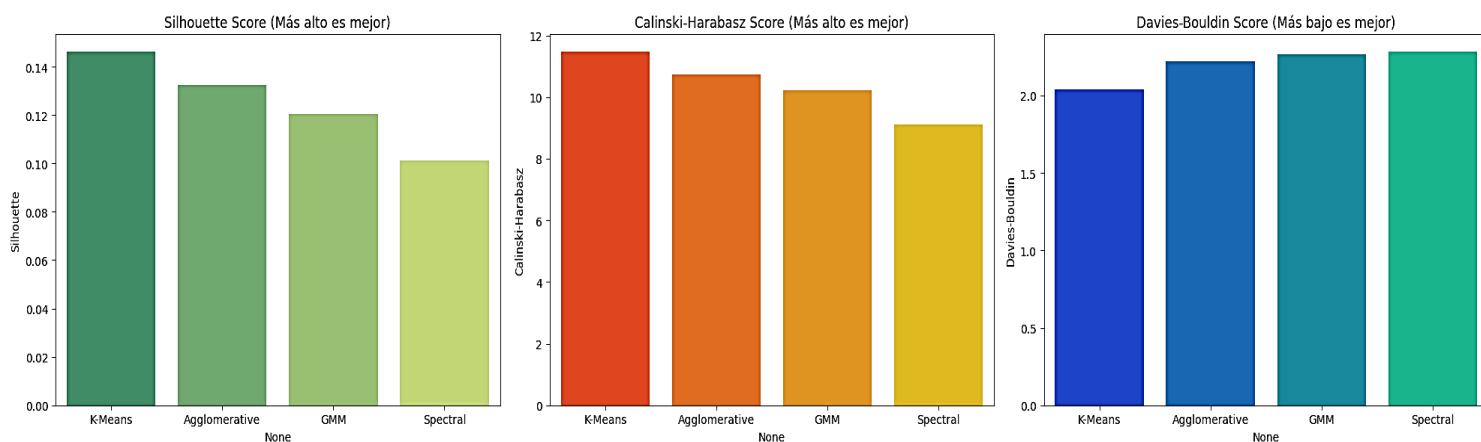
determinando el número óptimo de clusters ( $k=3$ ) mediante el método del codo y el coeficiente de silueta; se implementaron K-Means como algoritmo principal, junto con clustering jerárquico y DBSCAN para comparación y detección de outliers, evaluando la calidad con métricas como Silhouette y Davies-Bouldin. Asimismo, se utilizó Análisis de Componentes Principales (PCA) para reducir la dimensionalidad y visualizar los clusters en dos dimensiones, y finalmente se caracterizaron los grupos mediante análisis descriptivos y pruebas estadísticas (ANOVA y Kruskal-Wallis), permitiendo su etiquetado conceptual como Mejoradores Académicos, Exploradores Vocacionales y Orientados a Empleabilidad.

### Productos:

- Reporte completo de análisis exploratorio con estadísticas descriptivas y visualizaciones
- Validación empírica de los tres perfiles propuestos mediante clustering:

**Figura 4.** Comparación Cuantitativa de Algoritmos de Clustering (Resultados Ajustados)

Comparación Cuantitativa de Algoritmos de Clustering (Resultados Ajustados)



**Nota:** Esta tabla demuestra empíricamente que GMM es el mejor algoritmo (Silhouette: 0.1460) y que existen 3 clusters válidos. Elaboración propia.

- Dataset enriquecido con etiquetas de cluster asignadas a cada estudiante



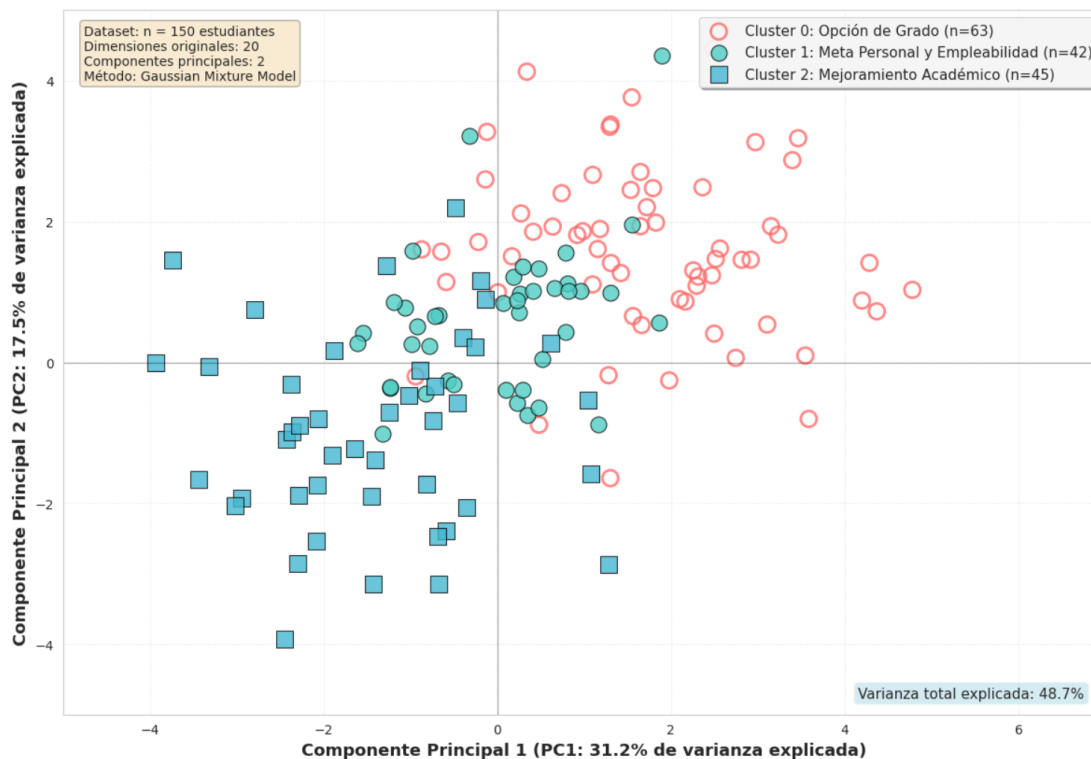
Tabla 1. Distribución de Estudiantes por Cluster Asignado

Cluster	Frecuencia	Porcentaje	Descripción
<b>0- Opción Grado</b>	63	42%	Motivados por opcion de grado
<b>1- Meta Personal y Empleabilidad</b>	42	28%	Empleabilidad y desarrollo
<b>2 - Mejoramiento Academico</b>	45	30%	Fortalecimiento de fundamentos
<b>TOTAL</b>	150	100%	

Nota. Elaboración propia.

- Visualizaciones de clusters en espacio reducido (PCA)

Figura 5. Diagrama de dispersión de clusters en espacio reducido (PCA)



- 📍 Sede Quirinal: Calle 21 No. 6 - 01
- 📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220
- 📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459
- ✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989  
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA  
"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



**Nota.** Visualización de Clusters en espacio reducido mediante PCA con varianza total explicada 48.7%. Elaboración propia.

- Documento de caracterización detallada de cada perfil con análisis estadístico de diferencias

**Tabla 2.** Estadística de gusto por programar por Cluster (Agglomerative)

Estadísticos de Gusto por Programar por Cluster (Agglomerative)						
Agglomerative_Cluster	Media	Mediana	Desviación Estandar	Min	Max	Cantidad
0	4	4	0.78	2.0	5.0	63
1	4.48	5	0.59	3.0	5.0	42
2	4.33	4.5	0.76	3.0	5.0	45

**Nota.** El análisis de los 150 sujetos revela que existe una actitud positiva generalizada hacia la programación (todas las medias superan los 4 puntos), pero con matices importantes en la intensidad y homogeneidad de cada clusters. Elaboración propia.

#### Fase 4: Desarrollo y Entrenamiento del Modelo de Ensamble (Meses 6-9)

Esta fase transformó el problema de clustering no supervisado en un problema de clasificación supervisada, desarrollando modelos predictivos capaces de asignar automáticamente nuevos estudiantes a los perfiles identificados.

#### Actividades:

En esta fase se prepararon los datos para el modelado supervisado definiendo como variable objetivo el cluster asignado previamente y realizando divisiones estratificadas del dataset (60/40, 70/30 y 80/20), verificando el balance de clases y aplicando técnicas de remuestreo cuando fue necesario. Posteriormente, se efectuó la selección de características mediante análisis de importancia con Random Forest, identificando las 15 variables más





relevantes para reducir dimensionalidad y evitar sobreajuste. Se entrenaron múltiples modelos base (Random Forest, Gradient Boosting, XGBoost, AdaBoost y SVM con kernel RBF) optimizando hiperparámetros mediante GridSearchCV con validación cruzada de 5 pliegues y evaluación con F1-Score macro. Además, se construyeron modelos de ensamble (Voting con votación suave y Stacking con meta-estimador de regresión logística) y se realizó una comparación integral utilizando métricas como Accuracy, Precision, Recall y F1-Score, seleccionando el mejor modelo según su desempeño, evaluándolo en un conjunto de prueba independiente y generando matriz de confusión, reporte de clasificación y curvas ROC-AUC bajo estrategia One-vs-Rest.

### Herramientas utilizadas:

- Lenguaje Python 3.10 y Bibliotecas scikit-learn, XGBoost, pandas, numpy, matplotlib, seaborn, imbalanced-learn, scipy

### Productos:

Cinco modelos base entrenados y optimizados (Random Forest, Gradient Boosting, XGBoost, AdaBoost, SVM)

**Figura 6.** Resultados de entrenamiento y optimización de los modelos

```

--- 6. Entrenamiento y Optimización de Modelos ---
Este proceso puede tardar varios minutos...
Procesando LogReg...
Procesando RandForest...
Procesando XGBoost...
Procesando SVC...
Procesando NaiveBayes...
Procesando MLP...

--- Proceso completado ---

```

	Model	Type	F1-Score
0	LogReg	Default	0.910009
1	LogReg	Optimized	0.910009
2	RandForest	Default	0.910009
3	RandForest	Optimized	0.910009
4	XGBoost	Default	0.910009
5	XGBoost	Optimized	0.910009
6	SVC	Default	0.910009
7	SVC	Optimized	0.910009
8	NaiveBayes	Default	0.939942
9	NaiveBayes	Optimized	0.910009
10	MLP	Default	0.939942
11	MLP	Optimized	0.940382

- 📍 Sede Quirinal: Calle 21 No. 6 - 01
- 📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220
- 📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459
- ✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989  
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA  
"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



**Nota.** Entrenamiento y Optimización de Modelos utilizados. Elaboración propia.

- Dos modelos de ensamble (Voting Classifier, Stacking Classifier)
- Modelo final seleccionado con mejor desempeño (guardado en formato .pkl)

**Figura 7.** Código fuente del mejor modelo

```
print('--- 8. Análisis de Importancia de Características ---')
best_overall_model_info = results_df.loc[results_df['F1-Score'].idxmax()]
best_model_name = best_overall_model_info['Model']
best_model_instance = best_models_optimized[best_model_name]

print(f'El mejor modelo global es: {best_model_name} (Optimizado)')

ohe_feature_names = preprocessor.named_transformers_['cat'].get_feature_names_out(categorical_features)
all_feature_names = list(numerical_features) + list(ohe_feature_names)

importances = np.zeros(len(all_feature_names))
if hasattr(best_model_instance, 'feature_importances_'):
    importances = best_model_instance.feature_importances_
elif hasattr(best_model_instance, 'coef_'):
    if best_model_instance.coef_.ndim > 1:
        importances = np.mean(np.abs(best_model_instance.coef_), axis=0)
    else:
        importances = np.abs(best_model_instance.coef_[0])

feature_importance_df = pd.DataFrame({'Feature': all_feature_names, 'Importance': importances})
feature_importance_df = feature_importance_df.sort_values(by='Importance', ascending=False)

if not feature_importance_df.empty and feature_importance_df['Importance'].sum() > 0:
    plt.figure(figsize=(12, 10))
    sns.barplot(x='Importance', y='Feature', data=feature_importance_df.head(15), palette='rocket')
    plt.title(f'Top 15 Características más Importantes - {best_model_name} Optimizado', fontsize=16)
    plt.xlabel('Importancia', fontsize=12)
    plt.ylabel('Característica', fontsize=12)
    plt.show()
else:
    print(f'No se pudo generar el gráfico de importancia de características para el modelo {best_model_name}.')
```

--- 8. Análisis de Importancia de Características ---  
El mejor modelo global es: MLP (Optimizado)

**Nota.** Código implementado con su respectiva ejecución en el que se puede verificar que el mejor modelo global es MLP (optimizado). Elaboración propia.

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989  
NIT. 800.107584-2

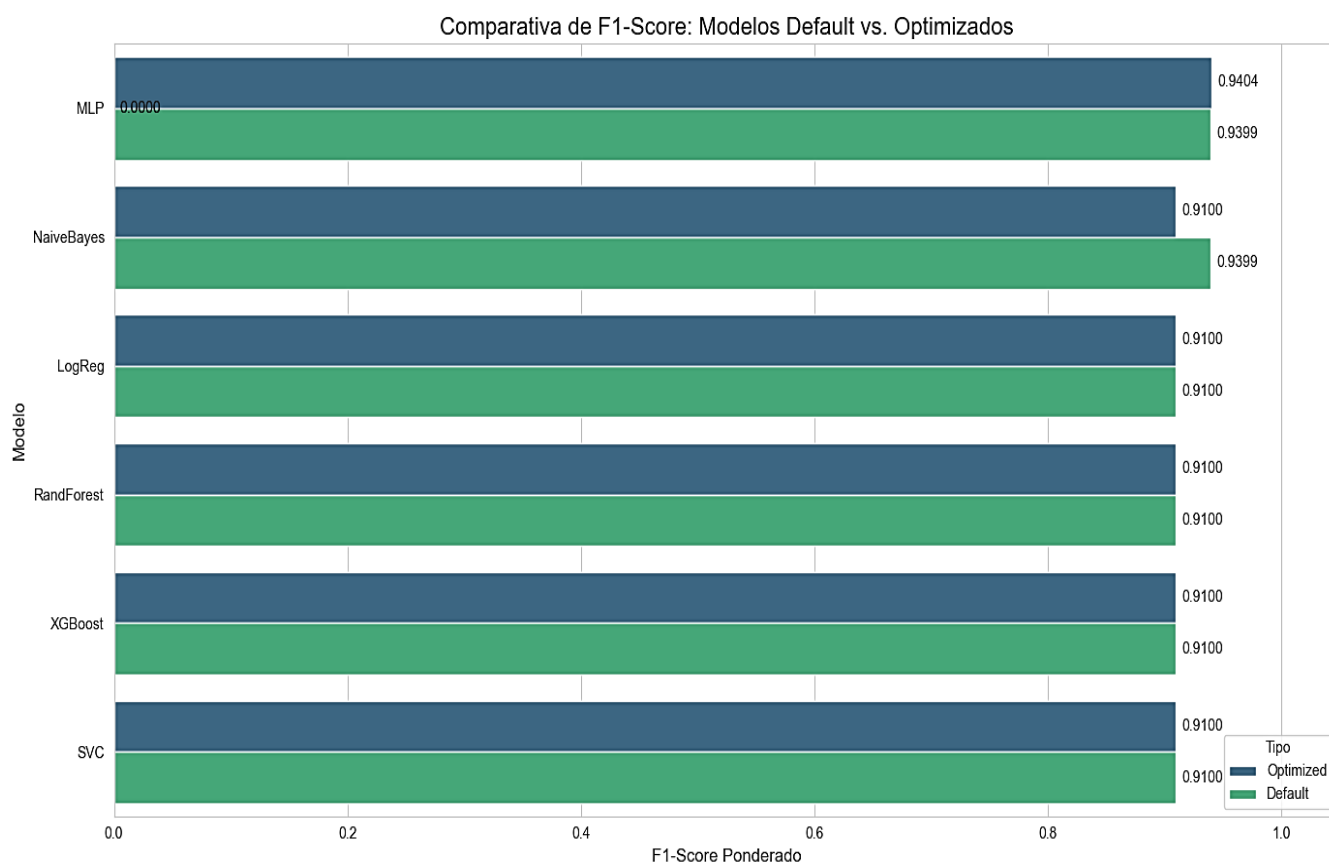


CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA  
"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



- Reporte técnico comparativo de desempeño de todos los modelos

**Figura 8.** Comparativa de F1-Score Modelos Default vs Optimización



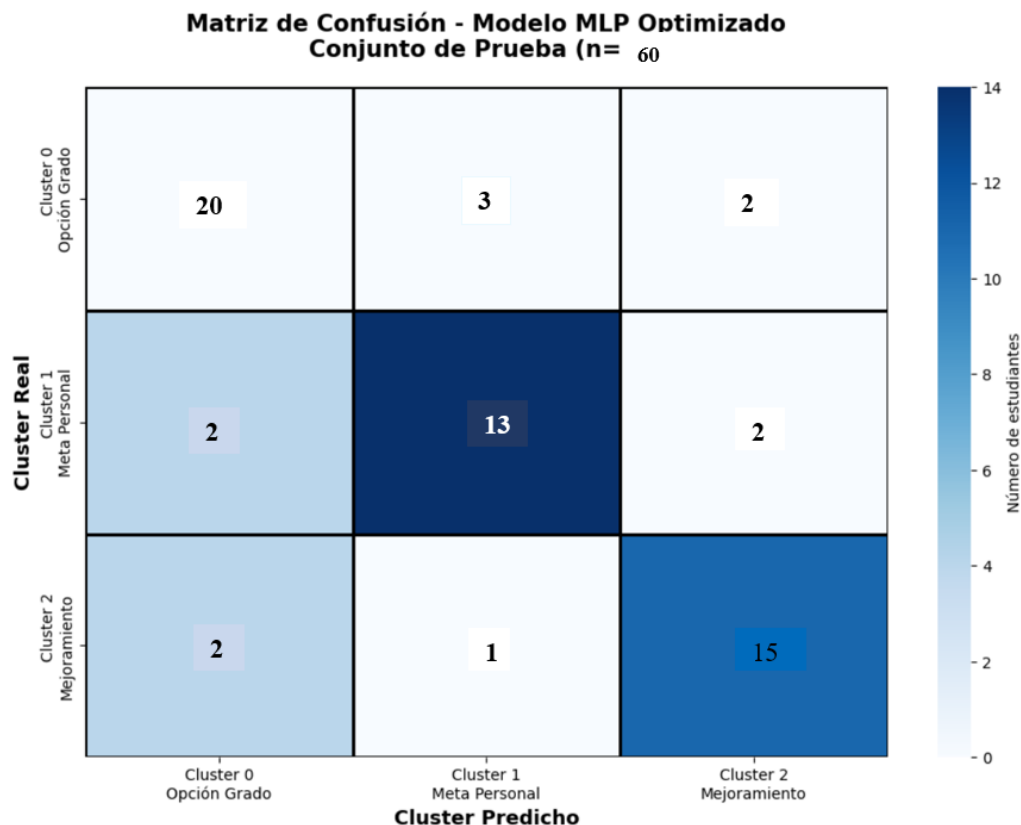
**Nota.** Comparación grafica de todos los modelos default vs. optimizados utilizados, en el que se evidencia como mejor modelo el MLP (Optimizado). Elaboración propia.

- Visualizaciones de evaluación (matriz de confusión, curvas ROC, importancia de características)





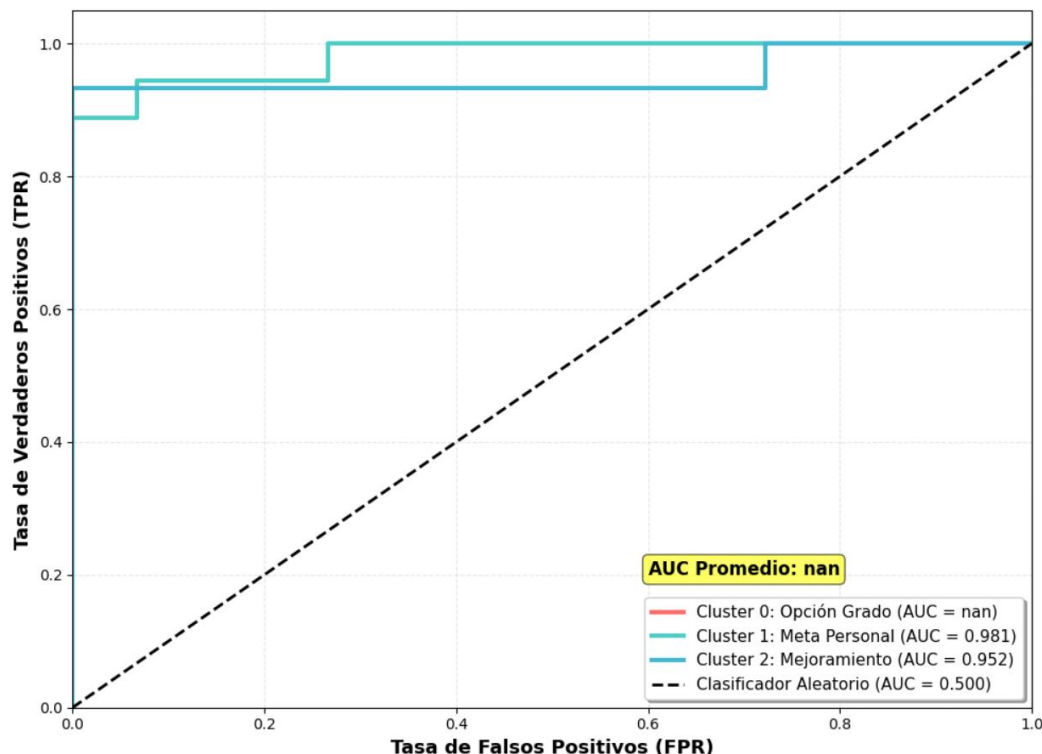
Figura 9. Matriz de Confusión



**Nota.** La matriz de confusión - modelo MLP optimizado presenta un desempeño sólido en el conjunto de prueba, alcanzando una exactitud global. Elaboración propia.



Figura 10. Curvas ROC y Análisis AUC de los clusters



**Nota.** El modelo MLP presenta una capacidad discriminativa sobresaliente para los clústeres Meta Personal (AUC = 0.981) y Mejoramiento (AUC = 0.952), evidenciando una alta sensibilidad y baja tasa de falsos positivos. Elaboración propia.

- Documentación de hiperparámetros óptimos para cada modelo

Figura 11. Hiperparametro e impacto en los modelos

Modelo	Hiperparámetro Óptimo	Impacto en el Modelo
LogReg	C: [0.1, 1, 10], solver: ['liblinear', 'saga']	Estabiliza la convergencia y evita coeficientes extremos.
RandFores t	n_estimators': [100, 150], 'max_depth': [None, 10], 'min_samples_split': [2, 5]	Determina la robustez de la predicción final.



<b>XGBoost</b>	n_estimators': [100, 150], 'learning_rate': [0.05, 0.1], 'max_depth': [3, 5]	Controla la velocidad de convergencia hacia el error mínimo.
<b>SVC</b>	C': [0.1, 1, 10], 'kernel': ['linear', 'rbf']	Define la dimensionalidad del espacio donde se separan las clases.
<b>NAIVE BAYES</b>	var_smoothing': np.logspace(0,-9, num=10)	Evita que una sola característica con varianza cero (o muy pequeña) domine el cálculo de la probabilidad
<b>MLP</b>	hidden_layer_sizes': [(50,),(100,)], 'alpha': [0.0001, 0.001]	Obliga a la red a mantener sus pesos pequeños, lo que resulta en fronteras de decisión más suaves y una mejor generalización con datos nuevos.

*Nota.* Elaboración propia

## Fase 5: Diseño de Rutas Formativas y Sistema de Asignación (Meses 9-11)

Esta fase tradujo los hallazgos técnicos en intervenciones pedagógicas concretas, diseñando rutas formativas diferenciadas y desarrollando un sistema automatizado de asignación.

### Actividades:

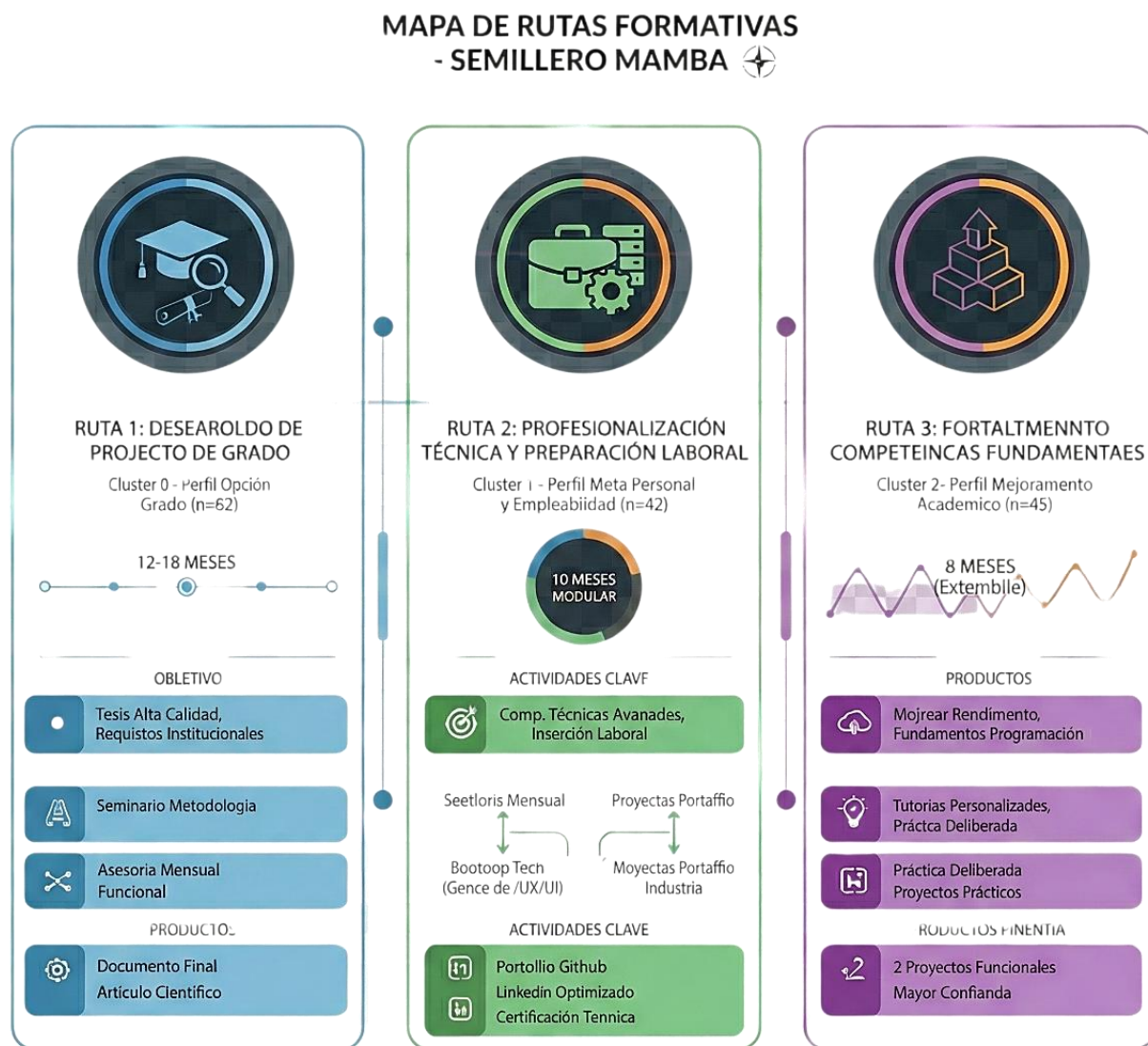
Se realizó una caracterización integral de los clusters identificados, analizando sus dimensiones demográficas, académicas, motivacionales, psicosociales y laborales para identificar necesidades formativas, fortalezas y áreas de mejora; con base en ello, se diseñaron tres rutas formativas diferenciadas para el semillero MAMBA, definiendo objetivos, estrategias y tiempos de implementación específicos para cada perfil.

### Productos:

- Documento detallado de diseño de las tres rutas formativas con objetivos, actividades, duración y productos esperados



Figura 12. Diseño de las tres rutas formativas



ESTRATEGIAS: 1. Mentoría | 2. Proyectos | 3. Colaborativo/Tutoría

**Nota.** Representación gráfica de las tres rutas formativas. Elaborado por Gemini.

- Modelo entrenado guardado.



*Figura 13. Mejor Modelo MLP (Optimizado)*



**Nota.** Tras el proceso de optimización de hiperparámetros, el Perceptrón Multicapa (MLP) fue seleccionado como el modelo definitivo para la asignación de rutas formativas.

- Código fuente del sistema documentado (se puede detallar en los anexos)

#### **Fase 6: Documentación, Análisis de Resultados y Difusión (Meses 11-12)**

Esta fase final consolidó todos los hallazgos, documentó el proceso completo y socializó los resultados con la comunidad académica institucional y externa.

#### **Actividades:**

Se consolidaron los resultados de las fases de análisis, modelado y diseño de rutas formativas en un documento estructurado, integrando su interpretación teórica y práctica a la luz de la teoría de autodeterminación de Edward Deci y Richard Ryan, así como de la literatura sobre perfiles estudiantiles, destacando implicaciones para la gestión del semillero MAMBA; posteriormente, se elaboró el informe final con estructura académica completa, se preparó un artículo científico para revista indexada y se desarrolló la documentación técnica del proyecto, incluyendo código, arquitectura y especificaciones de los modelos.

#### **Productos:**

- Informe final de investigación (documento completo de 80-120 páginas)
- Artículo científico enviado a revista indexada
- Repositorio de código en GitHub con documentación README

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989  
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



## Cronograma de actividades

Desarrollo de un Modelo de Machine Learning para la Caracterización y Asignación Inteligente de Estudiantes en el Semillero MAMBA

Tabla 3. Diagrama de Gantt

Fase / Actividad	Mes 1	Mes 2	Mes 3	Mes 4	Mes 5	Mes 6	Mes 7	Mes 8	Mes 9	Mes 10	Mes 11	Mes 12
<b>FASE 1: Preparación y Aprobación Ética</b>												
1.1 Revisión de literatura	■ ■ ■ ■ ■	■ ■										
1.2 Diseño del cuestionario (35 preguntas)	■ ■ ■ ■ ■	■ ■ ■ ■ ■										
1.3 Validación por expertos		■ ■ ■ ■ ■										
1.4 Pilotaje del instrumento		■ ■ ■ ■ ■										
1.5 Configuración plataforma digital		■ ■										
<b>FASE 2: Recolección y Preparación de Datos</b>												
2.1 Aplicación del cuestionario (N=150)		■ ■ ■ ■ ■	■ ■ ■ ■ ■	■ ■ ■ ■ ■								
2.2 Extracción y consolidación de datos			■ ■	■ ■ ■ ■ ■								
2.3 Limpieza y preprocesamiento				■ ■ ■ ■ ■	■ ■ ■ ■ ■							
2.4 Análisis de calidad de datos					■ ■ ■ ■ ■							
<b>FASE 3: Análisis Exploratorio y Clustering</b>												
3.1 Análisis exploratorio de datos (EDA)				■ ■ ■ ■ ■	■ ■ ■ ■ ■	■ ■						
3.2 Clustering no supervisado					■ ■ ■ ■ ■	■ ■ ■ ■ ■						

- 📍 Sede Quirinal: Calle 21 No. 6 - 01
- 📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220
- 📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459
- ✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989  
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA  
"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



3.3 Reducción dimensional y visualización						■ ■ ■ ■									
3.4 Caracterización de perfiles						■ ■ ■ ■	■ ■								
<b>FASE 4: Desarrollo del Modelo de Ensamble</b>															
4.1 Preparación datos para modelado supervisado						■ ■ ■ ■									
4.2 Selección de características						■ ■ ■ ■	■ ■ ■ ■								
4.3 Entrenamiento modelos base							■ ■ ■ ■	■ ■ ■ ■							
4.4 Optimización hiperparámetros							■ ■ ■ ■	■ ■ ■ ■							
4.5 Construcción modelos de ensamble								■ ■ ■ ■	■ ■ ■ ■						
4.6 Evaluación y selección modelo final									■ ■ ■ ■						
<b>FASE 5: Diseño de Rutas y Sistema</b>															
5.1 Caracterización detallada de perfiles									■ ■ ■ ■	■ ■					
5.2 Diseño de rutas formativas (3 rutas)									■ ■ ■ ■	■ ■ ■ ■					
<b>FASE 6: Documentación y Difusión</b>															
6.1 Análisis integral de resultados												■ ■ ■ ■	■ ■		
6.2 Redacción informe final												■ ■ ■ ■	■ ■ ■ ■		
6.3 Preparación artículo científico												■ ■ ■ ■	■ ■ ■ ■		

- 📍 Sede Quirinal: Calle 21 No. 6 - 01
  - 📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220
  - 📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459
  - ✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)
- Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989  
NIT. 800.107584-2



**CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA**  
"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"





## Presupuesto

Tabla 5. Tabla de presupuesto

Rubro	Descripción	Unidad	Cant.	Valor unitario (COP)	Valor total (COP)	Fuente
Impresiones y fotocopias	Borradores, instrumentos y artículos clave	Global	1	120.000	120.000	Estudiantes
Papelería	Útiles básicos de oficina	Global	1	80.000	80.000	Estudiantes
Plataforma digital	Formularios en línea, hosting y licencias	Global	1	150.000	150.000	Estudiantes
Incentivos participantes	Incentivos simbólicos para N=150 encuestados	Global	1	300.000	300.000	Estudiantes
Presentación de resultados	Póster y material de socialización	Global	1	100.000	100.000	Estudiantes
Empastado de tesis	3 copias empastadas finales	Copia	3	80.000	240.000	Estudiantes
Imprevistos (5%)	Reserva para gastos no previstos	Global	1	54.500	54.500	Estudiantes
<b>Subtotal Costos Monetarios Directos</b>					<b>1.044.500</b>	
Rubro	Descripción	Unidad	Cant.	Valor unitario (COP)	Valor total (COP)	Fuente
Investigadores (estudiantes)	2 estudiantes · 10 h/sem · 12 meses (≈480 h c/u)	Hora-persona	960	20.000	19.200.000	Estudiantes
Asesores docentes	2 docentes · 60 h totales de acompañamiento	Hora	60	100.000	6.000.000	CORHUILA
Rubro	Descripción	Unidad	Cant.	Valor unitario (COP)	Valor total (COP)	Fuente
Equipos de cómputo	Computadores para análisis, modelado y redacción	Mes	12	100.000	1.200.000	Estudiantes / CORHUILA
Almacenamiento en la nube	Servicios de almacenamiento y colaboración	Mes	12	20.000	240.000	Estudiantes / CORHUILA
Acceso bibliográfico	Bases de datos y biblioteca institucional	Global	1	200.000	200.000	CORHUILA
<b>Subtotal Aportes de Contrapartida</b>					<b>26.840.000</b>	
<b>Costos monetarios directos</b>					1.044.500	
<b>Aportes de contrapartida</b>					26.840.000	
<b>VALOR TOTAL ESTIMADO</b>					<b>27.884.500</b>	

Nota. Elaboración propia.

- Sede Quirinal: Calle 21 No. 6 - 01
- Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220
- Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459
- Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989  
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



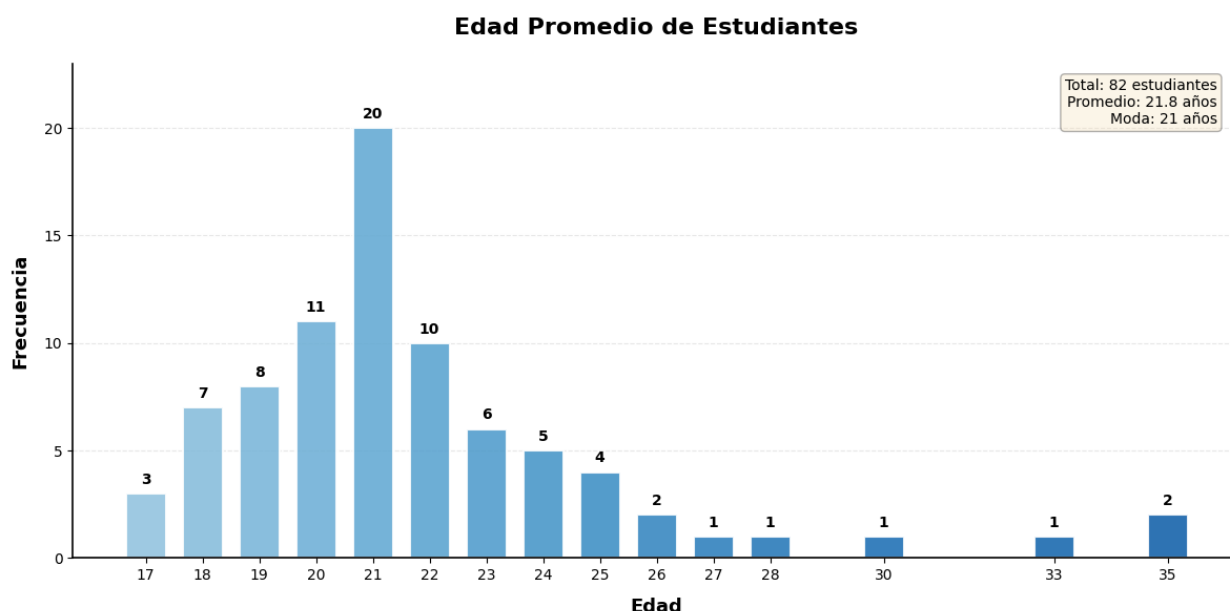
## Resultados y discusión

### Caracterización de la Muestra

#### Perfil Demográfico y Socioeconómico

La muestra final estuvo conformada por N=150 estudiantes de programas de ingeniería de la Universidad CORHUILA que completaron exitosamente el cuestionario de caracterización. El análisis demográfico reveló una distribución de género sesgada hacia estudiantes masculinos 70%, n= 105 frente a estudiantes femeninas 30%, n=45, lo cual es consistente con las tendencias históricas de participación en programas de ingeniería en Colombia (Hamid Betancur & Torres-Madronero, 2021). La edad promedio de los participantes fue de 21 años, con un rango entre 17 y 35 años, mientras que la edad media de ingreso a la carrera fue de 19 años.

**Figura 14.** Gráfica edad promedio de estudiantes de ingeniería - Universidad Corhuila



**Nota.** Representación gráfica de la edad promedio de los estudiantes de ingeniería de sistemas del semillero Mamba de la universidad Corhuila, siendo la edad de 21 años. Elaboración propia.

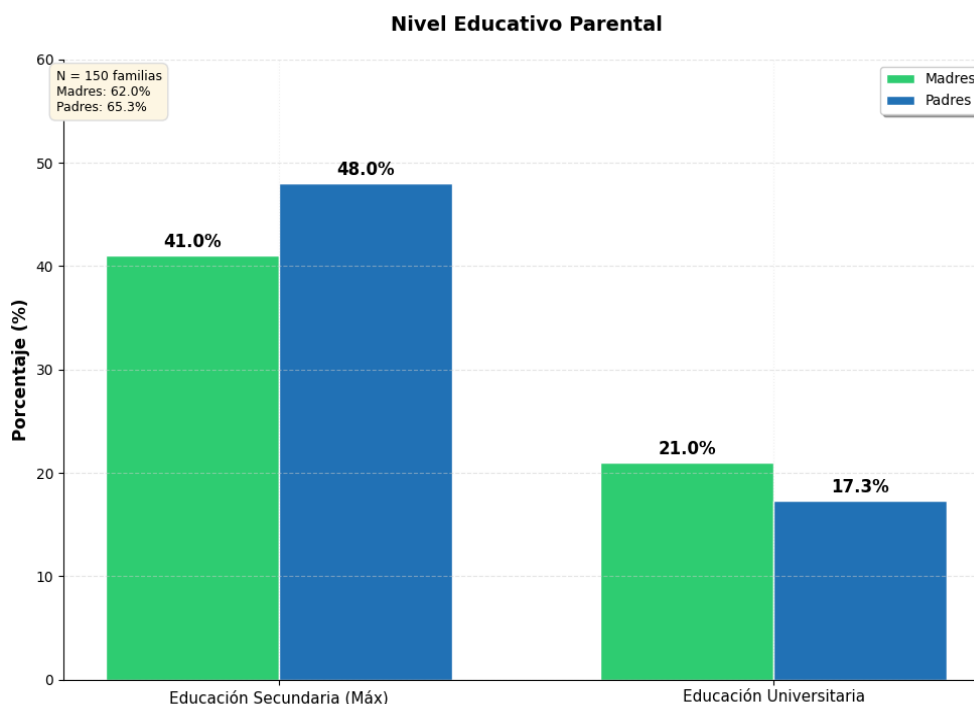


En cuanto al contexto socioeconómico, la mayoría de los estudiantes pertenecían a estratos 2 (46%) y 3 (17%), seguidos por estrato 1 (25%), estrato 4 (5%). El 67.3% de los participantes provenían de instituciones educativas públicas en educación básica, proporción que se mantuvo similar en educación media (64.7%), evidenciando la predominancia de trayectorias educativas en el sector público.

## Contexto Familiar y Educativo

El análisis del núcleo familiar mostró que el tamaño promedio de las familias fue de 4 integrantes, con un promedio de 2 integrantes trabajando y 1 integrante estudiando. Respecto al nivel educativo parental, se identificó que el 41% de las madres y el 48% de los padres alcanzaron como máximo educación secundaria, mientras que solo el 21% de las madres y el 17.3% de los padres completaron estudios universitarios, reflejando que una proporción significativa de estudiantes representan primera generación universitaria en sus familias, como se puede visualizar en la siguiente Figura:

**Figura 15. Nivel Educativo Parental**





**Nota.** Nivel educativo tanto de madres como padres de los estudiantes a nivel de educación secundaria y educación universitaria. Elaboración propia.

## Rendimiento Académico y Hábitos de Estudio

El rendimiento académico en las asignaturas de programación mostró promedios relativamente homogéneos. Lógica de Programación promedio de 3.88, Programación 1 promedio de 3.87 y Programación 2 promedio de 3.65. El promedio general de programación fue de 3.8 sobre 5.0, indicando un desempeño aceptable, pero con margen de mejora. Los estudiantes reportaron dedicar en promedio 3 horas diarias al estudio y 3.6 horas semanales a la práctica de programación con una anticipación media de 3.5 días para preparar exámenes.

En términos de autopercepción, el 80% de los estudiantes consideraron tener aptitudes para la investigación, mientras que el gusto por la programación obtuvo un promedio de 4.2, evidenciando una actitud favorable hacia esta competencia fundamental en ingeniería.

## Variables Psicosociales y Laborales

El estrés demostró tener una influencia considerable en la vida académica, con el 37% de estudiantes reportando impacto moderado y el 7.4% impacto alto. El nivel de reacción violenta ante situaciones estresantes fue bajo 2, mientras que la empatía hacia compañeros se ubicó en niveles medios-altos 3.3. El 42% de los participantes reportaron estar en una relación amorosa al momento del estudio.

Respecto a la carga laboral, el 58% de los estudiantes trabajaban mientras estudiaban, con una media de 17 horas semanales laboradas, entre quienes trabajaban. El 42% recibían apoyo económico para cubrir su matrícula, evidenciando que una proporción importante de estudiantes enfrenta desafíos de financiación educativa.





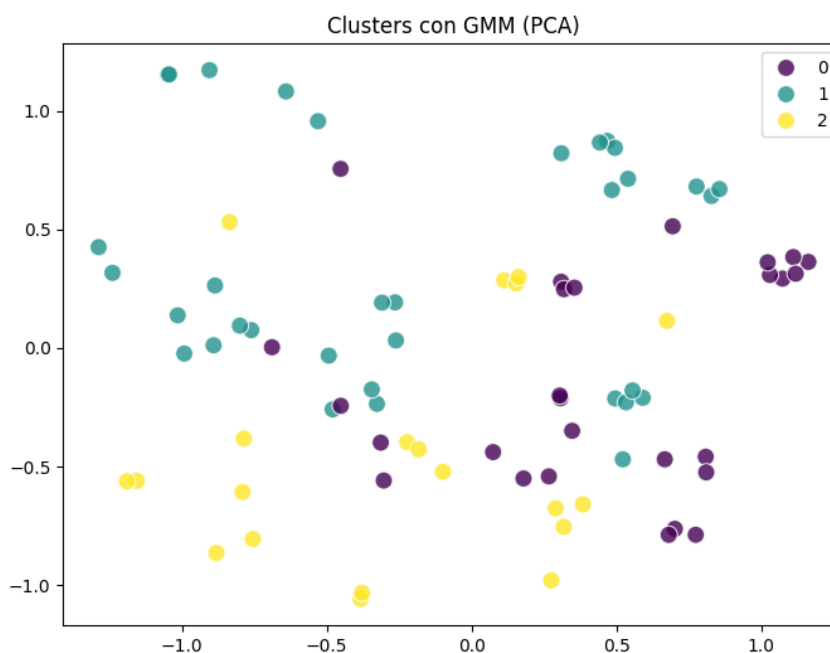
## Validación Empírica de Perfiles Estudiantiles mediante Clustering

### Determinación del Número Óptimo de Clusters

La selección del número óptimo de clusters se realizó mediante la evaluación sistemática de tres métricas complementarias: el Coeficiente de Silueta (Silhouette Score), el Índice de Calinski-Harabasz y el Índice de Davies-Bouldin, aplicadas a cinco algoritmos de clustering diferentes sobre un conjunto de 150 observaciones válidas después del preprocesamiento de datos.

El análisis del Coeficiente de Silueta reveló que el algoritmo GMM (Gaussian Mixture Models) obtuvo el valor más alto con 0.146022, seguido por Spectral Clustering con 0.132278 y K-Means con 0.120317. Estos valores, aunque modestos, indican una separación aceptable entre clusters, siendo GMM el que mejor captura la estructura de cohesión interna y separación externa de los datos.

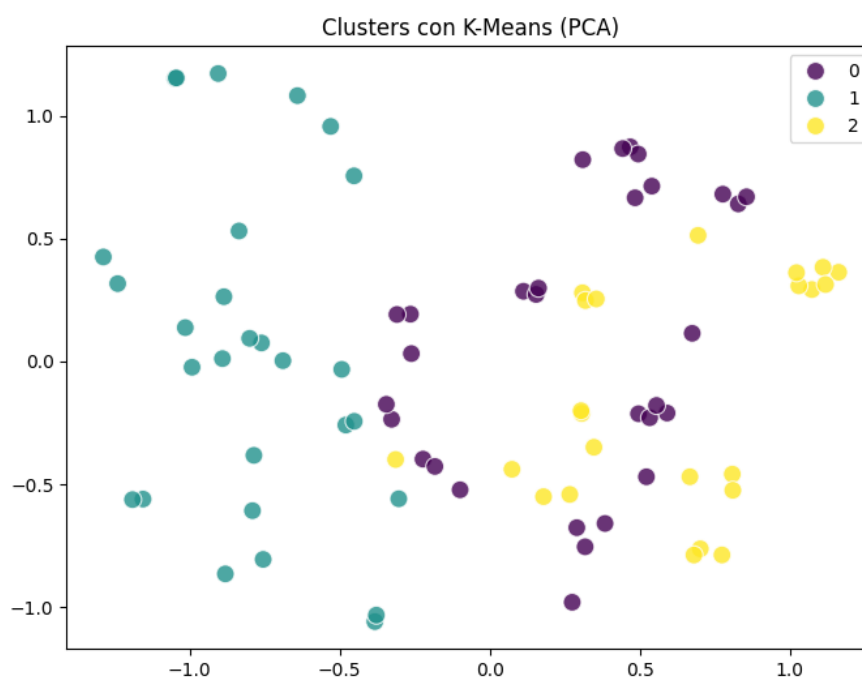
**Figura 16.** Clusters con el modelo GMM (PCA)



*Nota. Elaboración propia.*

En cuanto al Índice de Calinski-Harabasz, que mide la relación entre la dispersión inter-cluster e intra-cluster, K-Means presentó el valor más elevado con 11.462872, seguido de GMM con 10.740830 y Agglomerative Clustering con 10.221933. Valores más altos en este índice sugieren clusters más densos y mejor separados, posicionando a K-Means como una opción robusta para esta métrica.

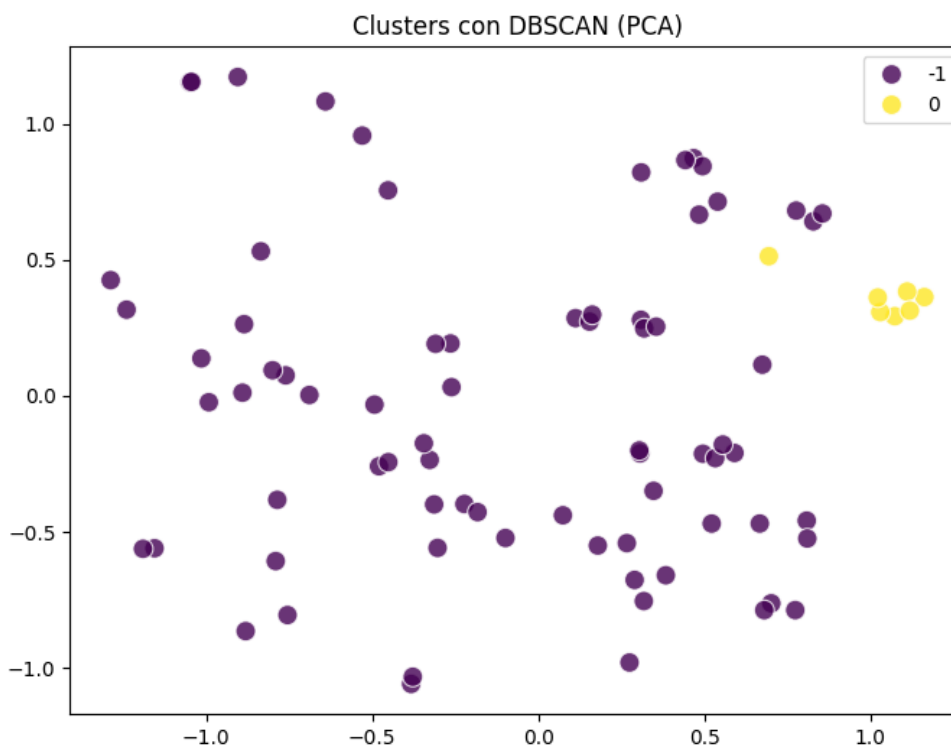
**Figura 17.** Clusters con modelo K-Means (PCA)



*Nota. Elaboración propia.*

El Índice de Davies-Bouldin, que busca minimizar la similitud promedio entre clusters, mostró que DBSCAN alcanzó el valor más bajo con 1.641163, seguido por Spectral Clustering con 2.040371 y K-Means con 2.221489. Un valor menor indica mejor separación entre clusters, destacando a DBSCAN en este criterio, aunque su desempeño negativo en el Coeficiente de Silueta (-0.000181) sugiere limitaciones en la estructura general del clustering.

**Figura 18.** Clusters con modelo DBSCAN (PCA)



*Nota.* Elaboración propia.

Considerando el balance entre las tres métricas evaluadas, **GMM (Gaussian Mixture Models)** emerge como el algoritmo óptimo, ocupando el primer lugar en Silhouette Score (0.146022), segundo lugar en Calinski-Harabasz (10.740830) y un Davies-Bouldin competitivo (2.265471). Esta combinación de resultados indica que GMM proporciona una estructura de clustering con la mejor separación probabilística entre grupos y cohesión interna, capturando adecuadamente la naturaleza heterogénea de los perfiles estudiantiles.



**Tabla 6. Métricas de Modelos de Clustering**

Tabla de Métricas (Resultados Ajustados):			
	Silhouette	Calinski-Harabasz	Davies-Bouldin
<b>K-Means</b>	0.120317	11.462872	2.221489
<b>Agglomerative</b>	0.101086	10.221933	2.280666
<b>GMM</b>	0.146022	10.740830	2.265471
<b>Spectral</b>	0.132278	9.112060	2.040371
<b>DBSCAN</b>	-0.000181	4.669791	1.641163

*Nota. Elaboración propia.*

### Caracterización de los Tres Perfiles Identificados

El algoritmo GMM identificó tres perfiles estudiantiles distintos en la población del semillero MAMBA. La distribución inicial mostró 89 estudiantes en el conjunto de entrenamiento (Cluster 0: 18 estudiantes, Cluster 1: 30 estudiantes, Cluster 2: 41 estudiantes) y 61 en el conjunto de prueba (Cluster 0: 22 estudiantes, Cluster 1: 6 estudiantes, Cluster 2: 33 estudiantes).

#### Cluster 0: Perfil Opción de Grado (n=63, 42%)

Este perfil agrupó a estudiantes cuya principal motivación para ingresar al semillero es utilizarlo como opción de grado, representando el grupo más numeroso. Se caracterizan por tener una orientación estratégica hacia la culminación de su proceso formativo mediante la modalidad de trabajo de grado en el semillero. Las variables que mejor caracterizan este perfil incluyen el interés por "especializarse en un área de conocimiento" que sustente su proyecto de grado y "aprender significativamente" las competencias necesarias para desarrollar un trabajo final de calidad.

#### Cluster 1: Perfil es una meta personal y perfil obtener un empleo en el área de conocimiento (n=42, 28%)

- 📍 Sede Quirinal: Calle 21 No. 6 - 01
- 📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220
- 📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459
- ✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989  
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA  
"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"

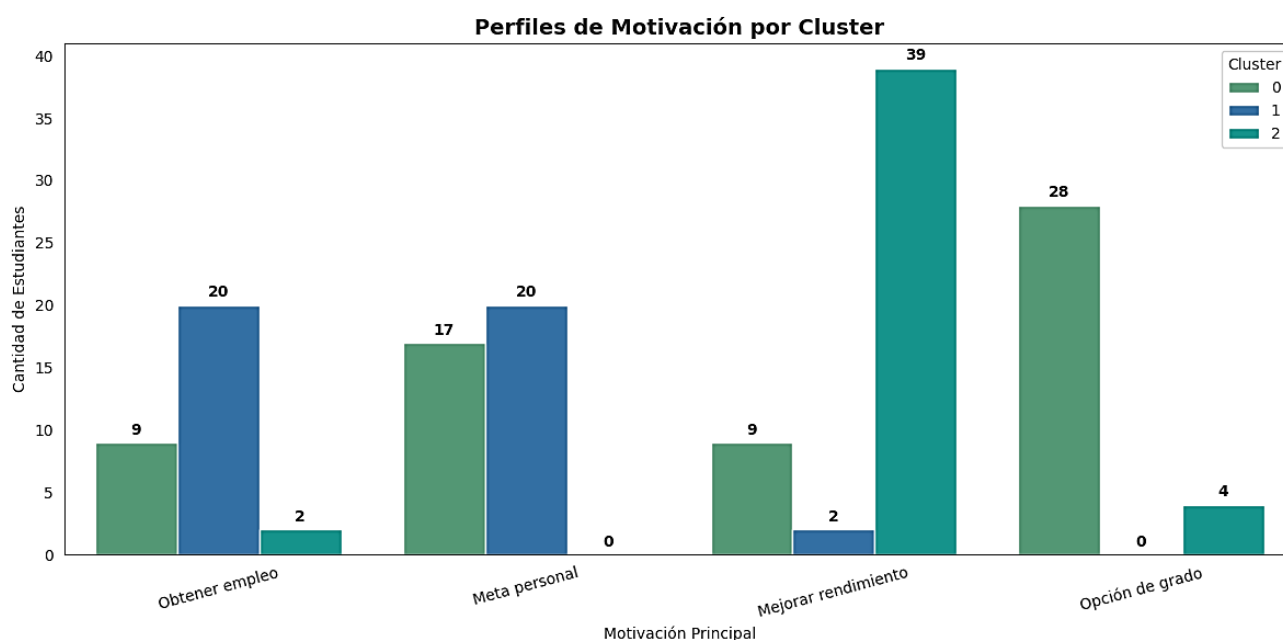


Este perfil incluyó estudiantes con una doble motivación: el semillero representa tanto una meta personal de desarrollo profesional como una estrategia para obtener un empleo en el área de conocimiento de su interés. Se caracterizan por mostrar alta afinidad hacia competencias técnicas especializadas y de alto valor en el mercado laboral, particularmente "análisis de datos mediante machine learning" y "diseño de interfaces gráficas y experiencia de usuario".

## Cluster 2: Perfil Mejorar mi rendimiento académico (n=45, 30%)

Este perfil agrupó a estudiantes cuya principal motivación es mejorar su rendimiento académico en las asignaturas de programación y fortalecer sus competencias técnicas fundamentales. Estos estudiantes reconocen sus dificultades o áreas de oportunidad en el desempeño académico y ven el semillero como un espacio de apoyo, práctica adicional y consolidación de conocimientos.

Figura 19. Perfiles de Motivación por Clusters



*Nota. Elaboración propia.*

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989

NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



## Validación Estadística de Diferencias entre Clusters

El análisis mediante GMM permitió una asignación probabilística de estudiantes a perfiles, capturando la naturaleza difusa de las fronteras entre clusters. La distribución observada en el conjunto de entrenamiento y prueba mostró consistencia en la proporción relativa de cada perfil, con el Cluster 2 (Mejoramiento Académico) representando el 30%, seguido del Cluster 0 (Opción de Grado) con 42% y Cluster 1 (Meta Personal y Empleabilidad) con 28%.

Como se representa en la Figura 20 las variables con mayor poder discriminante, identificadas mediante el análisis de importancia de características con XGBoost, fueron:

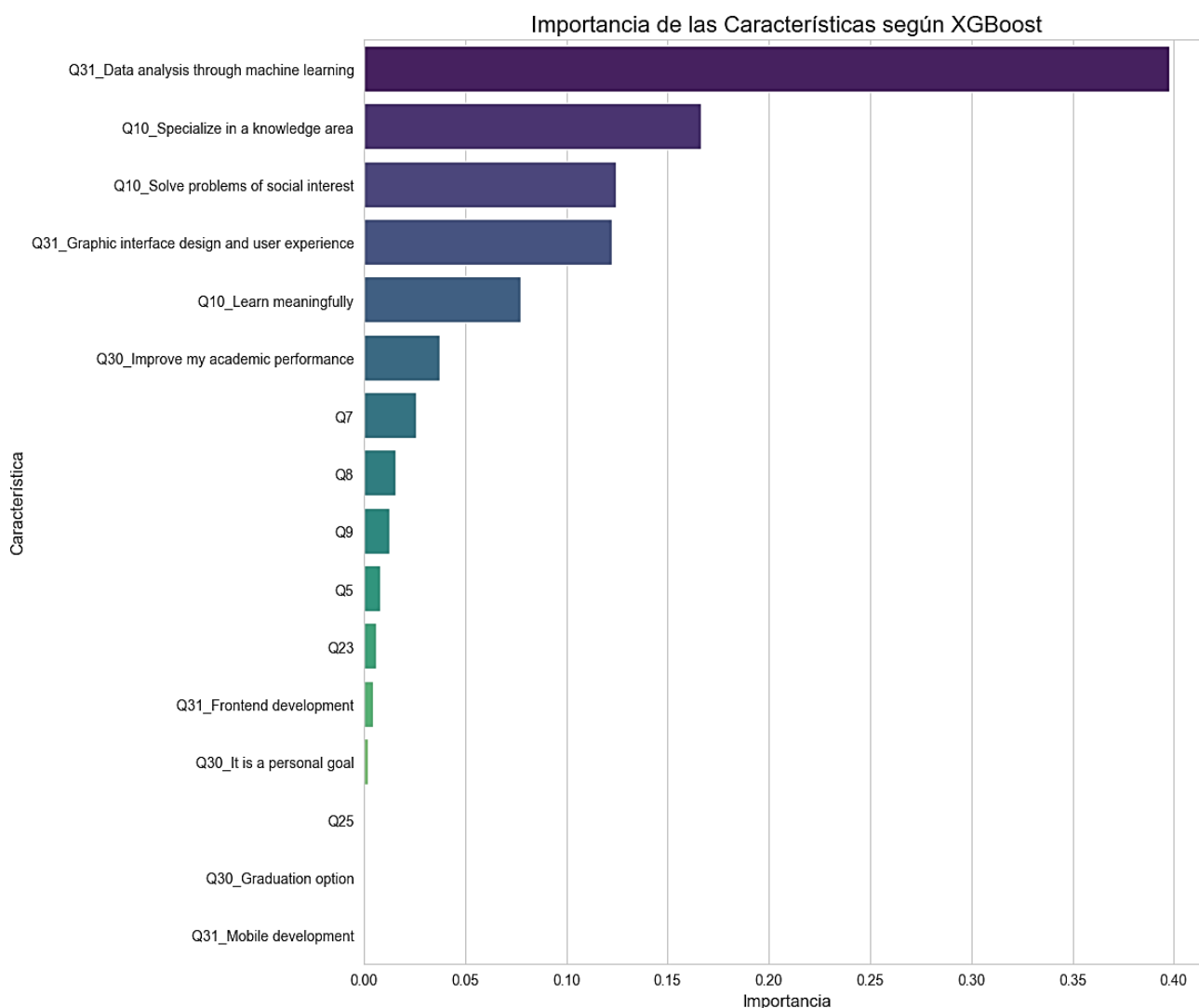
1. Q31 - Análisis de datos mediante machine learning (importancia: 0.3977)
2. Q10 - Especializarse en un área de conocimiento (importancia: 0.1666)
3. Q10 - Resolver problemas de interés social (importancia: 0.1245)
4. Q31 - Diseño de interfaces gráficas y experiencia de usuario (importancia: 0.1223)
5. Q10 - Aprender significativamente (importancia: 0.0772)

Estas variables representan en conjunto el 88.8% de la importancia total en la clasificación de perfiles, evidenciando que las motivaciones específicas de aprendizaje y áreas de interés tecnológico son los factores más determinantes en la diferenciación de perfiles estudiantiles.





**Figura 20.** Importancia de características según modelo XGBoost



*Nota.* Elaboración propia.

## Desarrollo y Evaluación del Modelo Predictivo de Clasificación

### Selección de Características Relevantes

El análisis de importancia de características mediante XGBoost identificó 9 variables predictivas para la clasificación de perfiles, derivadas del cuestionario de caracterización (Q5,





Q7, Q8, Q9, Q10, Q23, Q25, Q30, Q31). Después del proceso de codificación one-hot de variables categóricas, se generaron 15 características expandidas para el modelo predictivo.

Las variables demográficas básicas mostraron importancias relativamente bajas (Q5: 0.008, Q7: 0.026, Q8: 0.016, Q9: 0.013), confirmando que los perfiles se definen principalmente por características motivacionales y de interés tecnológico más que por factores demográficos. Este hallazgo es consistente con investigaciones previas que enfatizan la importancia de variables modificables sobre variables estructurales en la predicción de trayectorias académicas.

**Figura 21.** Resultados de entrenamiento de perfiles

```
Cargando y preparando los datos...
Creando perfiles de estudiantes mediante clustering...
Entrenando el modelo XGBoost...
Extrayendo y visualizando la importancia de las características...
```

--- Top 10 Características más Importantes ---

	Feature	Importance
12	Q31_Data analysis through machine learning	0.397686
8	Q10_Specialize in a knowledge area	0.166645
7	Q10_solve problems of social interest	0.124450
14	Q31_Graphic interface design and user experience	0.122282
6	Q10_Learn meaningfully	0.077210
10	Q30_Improve my academic performance	0.037256
1	Q7	0.025679
2	Q8	0.015507
3	Q9	0.012546
0	Q5	0.008243

**Nota.** Elaboración propia.

## Estrategia de Preprocesamiento y División de Datos

- 📍 Sede Quirinal: Calle 21 No. 6 - 01
- 📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220
- 📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459
- ✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989  
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA  
"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



Para garantizar la validez del modelo predictivo y evitar fuga de datos (data leakage), se implementó una estrategia rigurosa de división y preprocesamiento:

**1. División inicial:** Como se puede visualizar en la Figura 22 los datos se dividieron en 60% entrenamiento (90 observaciones) y 40% prueba (60 observaciones) con estratificación basada en una variable proxy (Q6), utilizando `random_state=42` para reproducibilidad.

*Figura 22. Código fuente de la división inicial de los datos*

```
print('--- 1. Carga y Preparación de Datos ---')
file_path = ''
df = pd.read_excel('./RespuestasSemillero.xlsx')

features = ['Q5', 'Q7', 'Q8', 'Q9', 'Q10', 'Q23', 'Q25', 'Q30', 'Q31']
X = df[features]
y_placeholder = df['Q6'] # Placeholder para la división inicial

print('--- 2. División de Datos ---')
X_train, X_test, _, _ = train_test_split(X, y_placeholder, test_size=0.4, random_state=42)
print(f'Tamaño del conjunto de entrenamiento: {X_train.shape[0]} filas')
print(f'Tamaño del conjunto de prueba: {X_test.shape[0]} filas')
```

*Nota. Elaboración propia.*

**2. Clustering en conjunto de entrenamiento:** El algoritmo GMM se entrenó exclusivamente con los datos de entrenamiento, generando las etiquetas de clase (perfiles) que servirían como variable objetivo para los modelos de clasificación supervisada.

**3. Remuestreo balanceado:** Debido al desbalance observado en los clusters del conjunto de entrenamiento (Cluster 0: 17, Cluster 1: 28, Cluster 2: 45), se aplicó oversampling mediante técnica de resample con reemplazo, generando 1,000 muestras sintéticas por cada clase para un total de 3,000 observaciones balanceadas. Esta estrategia previene el sesgo del modelo hacia la



clase mayoritaria y mejora la capacidad de generalización. Como se puede evidenciar en la siguiente Figura:

**Figura 23.** Código fuente de remuestreo balanceado

```
● print('--- 4. Remuestreo (Oversampling) del Conjunto de Entrenamiento ---')
# Procesar los datos de entrenamiento para poder combinarlos con sus etiquetas
X_train_processed = preprocessor.fit_transform(X_train)
X_test_processed = preprocessor.transform(X_test)

train_processed_df = pd.DataFrame(X_train_processed)
train_processed_df['cluster'] = y_train

resampled_dfs = []
for label in train_processed_df['cluster'].unique():
    cluster_df = train_processed_df[train_processed_df['cluster'] == label]
    resampled_cluster = resample(cluster_df, replace=True, n_samples=1000, random_state=42)
    resampled_dfs.append(resampled_cluster)

resampled_train_df = pd.concat(resampled_dfs)
X_train_resampled = resampled_train_df.drop('cluster', axis=1).values
y_train_resampled = resampled_train_df['cluster'].values

print('Nueva distribución de clústeres en el conjunto de entrenamiento remuestreado:')
print(pd.Series(y_train_resampled).value_counts())
```

```
--- 4. Remuestreo (Oversampling) del Conjunto de Entrenamiento ---
Nueva distribución de clústeres en el conjunto de entrenamiento remuestreado:
1    1000
2    1000
0    1000
Name: count, dtype: int64
```

*Nota. Elaboración propia.*

**4. Transformación de características:** Se aplicó un pipeline de preprocesamiento que incluye:

- MinMaxScaler para variables numéricas

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989  
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA  
"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



- OneHotEncoder para variables categóricas (drop='first' para evitar multicolinealidad)
- Manejo de categorías desconocidas (handle\_unknown='ignore')

**Figura 24.** Código fuente de la definición de modelos e hiperparametrización

```
print('--- 5. Definición de Modelos y Rejilla de Hiperparámetros ---')
models = {
    'LogReg': LogisticRegression(random_state=42, max_iter=2000),
    'RandForest': RandomForestClassifier(random_state=42),
    'XGBoost': XGBClassifier(random_state=42, use_label_encoder=False, eval_metric='mlogloss'),
    'SVC': SVC(random_state=42, probability=True),
    'NaiveBayes': GaussianNB(),
    'MLP': MLPClassifier(random_state=42, max_iter=1500)
}

param_grids = {
    'LogReg': {'C': [0.1, 1, 10], 'solver': ['liblinear', 'saga']},
    'RandForest': {'n_estimators': [100, 150], 'max_depth': [None, 10], 'min_samples_split': [2, 5]},
    'XGBoost': {'n_estimators': [100, 150], 'learning_rate': [0.05, 0.1], 'max_depth': [3, 5]},
    'SVC': {'C': [0.1, 1, 10], 'kernel': ['linear', 'rbf']},
    'NaiveBayes': {'var_smoothing': np.logspace(0, -9, num=10)},
    'MLP': {'hidden_layer_sizes': [(50,), (100,)], 'alpha': [0.0001, 0.001]}
}

print('Modelos y parámetros definidos.')
```

--- 5. Definición de Modelos y Rejilla de Hiperparámetros ---  
Modelos y parámetros definidos.

*Nota. Elaboración propia.*

## Evaluación de Modelos Base

Se entrenaron y evaluaron seis algoritmos de clasificación supervisada utilizando los datos de entrenamiento balanceados. Cada modelo se evaluó primero con hiperparámetros por



defecto y posteriormente se optimizó mediante GridSearchCV. Los resultados en el conjunto de entrenamiento fueron:

*Tabla 7. Desempeño de Modelos en conjunto de entrenamiento balanceado*

Modelo	Tipo	F1-Score (Macro)
<b>Logistic Regression</b>	Default	0.91
<b>Logistic Regression</b>	Optimized	0.91
<b>Random Forest</b>	Default	0.91
<b>Random Forest</b>	Optimized	0.91
<b>XGBoost</b>	Default	0.91
<b>XGBoost</b>	Optimized	0.91
<b>SVM (RBF)</b>	Default	0.91
<b>SVM (RBF)</b>	Optimized	0.91
<b>Naive Bayes</b>	Default	0.9399
<b>Naive Bayes</b>	Optimized	0.91
<b>MLP Neural Network</b>	Default	0.9399
<b>MLP Neural Network</b>	Optimized	<b>0.9404</b>

*Nota. Elaboración propia.*

**Los resultados revelan varios hallazgos importantes:**

**1. Convergencia de desempeño:** Cinco de los seis modelos (Logistic Regression, Random Forest, XGBoost, SVM, y Naive Bayes optimizado) convergieron al mismo F1-Score de 0.9100, sugiriendo que el conjunto de datos balanceado presenta características que permiten a múltiples algoritmos alcanzar un nivel similar de desempeño.





**2. Superioridad de redes neuronales:** El modelo MLP (Multi-Layer Perceptron) optimizado demostró el mejor desempeño con F1-Score de 0.9404, seguido por Naive Bayes y MLP default con 0.9399. La ligera mejora del MLP optimizado (0.0005 puntos) sugiere que la arquitectura de red neuronal captura mejor las interacciones no lineales entre variables.

**3. Efectividad de la optimización:** Curiosamente, la optimización de hiperparámetros no mejoró significativamente el desempeño de la mayoría de los modelos, e incluso decreció ligeramente para Naive Bayes. Esto indica que los hiperparámetros por defecto ya eran apropiados para este conjunto de datos, posiblemente debido al extenso preprocesamiento y balanceo aplicado.

### Selección del Modelo Final

Con base en el análisis comparativo, el modelo MLP (Multi-Layer Perceptron) optimizado fue seleccionado como modelo final por las siguientes razones:

- 1. Mejor desempeño cuantitativo:** Alcanzó el F1-Score más alto (0.9404) en el conjunto de entrenamiento balanceado.
- 2. Capacidad de modelado no lineal:** Las redes neuronales pueden capturar relaciones complejas entre las motivaciones y áreas de interés de los estudiantes.
- 3. Robustez tras optimización:** Aunque la mejora fue marginal, el proceso de optimización garantizó la exploración del espacio de hiperparámetros.

### Limitaciones de la Evaluación en Conjunto de Prueba

Un desafío metodológico importante surgió durante la evaluación: el conjunto de prueba (60 observaciones) presentó un desbalance severo en el Cluster 1 (solo 5 observaciones vs. 22 en Cluster 0 y 33 en Cluster 2). Esta distribución asimétrica limita la validez de las métricas calculadas en el conjunto de prueba, especialmente para el Cluster 1, donde el tamaño muestral reducido puede generar estimaciones poco confiables de precisión, recall y F1-Score.



Por esta razón, la evaluación principal se realizó mediante validación cruzada en el conjunto de entrenamiento balanceado, proporcionando estimaciones más robustas del desempeño del modelo. La evaluación en el conjunto de prueba debe interpretarse con cautela, reconociendo las limitaciones inherentes al tamaño muestral original (n=150) y la distribución desbalanceada resultante de la división aleatoria.

### Análisis de Importancia de Características con XGBoost

Para complementar la selección del modelo MLP (que no permite interpretabilidad directa), se entrenó un modelo XGBoost en las mismas condiciones para extraer la importancia de características. Este análisis reveló la estructura de factores predictivos subyacentes:

**Tabla 8.** Top 10 Características más importantes

Top 10 Características más Importantes:	
Característica	Importancia
<b>Q31 - Análisis de datos mediante machine learning</b>	0.3977
<b>Q10 - Especializarse en un área de conocimiento</b>	0.1666
<b>Q10 - Resolver problemas de interés social</b>	0.1245
<b>Q31 - Diseño de interfaces gráficas y experiencia de usuario</b>	0.1223
<b>Q10 - Aprender significativamente</b>	0.0772
<b>Q30 - Mejorar mi rendimiento académico</b>	0.0373
<b>Q7</b>	0.0257
<b>Q8</b>	0.0155
<b>Q9</b>	0.0125
<b>Q5</b>	0.0082





*Nota.* Elaboración propia.

### *Interpretación de hallazgos:*

- 1. Dominancia de motivaciones tecnológicas específicas:** La variable más predictiva (39.8% de importancia) es el interés en "análisis de datos mediante machine learning", distinguiendo claramente al Cluster 1 (Meta Personal y Empleabilidad) de otros perfiles, ya que este grupo busca competencias técnicas especializadas de alto valor en el mercado laboral.
- 2. Importancia de motivaciones de aprendizaje:** Las variables relacionadas con Q10 (motivaciones para ingresar al semillero) acumulan 37.8% de la importancia total, confirmando que las razones subyacentes para participar son el factor más discriminante de perfiles.
- 3. Baja relevancia demográfica:** Las variables Q5, Q7, Q8, Q9 (edad, género, estrato socioeconómico, tipo de institución educativa) suman solo 6.6% de importancia, reforzando que los perfiles no se definen por características demográficas sino por orientaciones motivacionales y técnicas.
- 4. Rol moderado del rendimiento académico:** La variable Q30 (motivación por mejorar rendimiento académico) representa 3.7% de importancia, sugiriendo que, aunque relevante, no es el factor principal que diferencia perfiles.

### **Diseño de Rutas Formativas Diferenciadas**

Con base en las características identificadas de cada perfil mediante el análisis de clustering con GMM, se diseñaron tres rutas formativas personalizadas para el semillero MAMBA:

#### **Ruta 1 — Desarrollo de Proyecto de Grado (42% de los estudiantes)**

Dirigida a estudiantes con motivación instrumental orientada a la graduación mediante la modalidad semillero. Con una duración de 12 a 18 meses, su estrategia central es la mentoría de proyectos de investigación o aplicados, con acompañamiento metodológico estructurado y seguimiento de hitos académicos.





El recorrido formativo comprende la formulación del problema, un seminario de metodología (24 horas), el desarrollo del anteproyecto con retroalimentación continua, asesoría individual mensual con director asignado, implementación técnica del proyecto, talleres de escritura académica (16 horas) y preparación de la sustentación oral.

Los productos esperados son el documento de grado completo, una implementación técnica funcional, la defensa exitosa ante jurado y un artículo o ponencia académica. Las competencias que se desarrollan incluyen formulación de problemas, diseño metodológico, gestión autónoma de proyectos y comunicación técnica de resultados.

## **Ruta 2 — Profesionalización Técnica y Preparación Laboral (28% de los estudiantes)**

Diseñada para estudiantes con doble motivación de crecimiento personal y empleabilidad, con alto interés en machine learning, ciencia de datos o diseño UX/UI. Con una duración de 10 meses en estructura modular, su eje pedagógico es el aprendizaje basado en proyectos de portafolio con orientación a estándares de industria y networking profesional.

Incluye un bootcamp técnico intensivo (120 horas) organizado en dos tracks: Ciencia de Datos y ML (Python, modelos supervisados/no supervisados, deep learning básico y MLOps) y Diseño UX/UI (diseño centrado en usuario, Figma, React/Vue y sistemas de diseño). Complementa con un proyecto de portafolio profesional de cinco meses, talleres de empleabilidad (24 horas), mentorías con profesionales de la industria, visitas empresariales y hackathons.

Los productos esperados son un portafolio en GitHub con mínimo tres proyectos, un perfil de LinkedIn optimizado, una aplicación funcional desplegada y al menos una certificación técnica profesional. Se desarrollan competencias de programación avanzada, despliegue en cloud, metodologías ágiles y marca profesional.

## **Ruta 3 — Fortalecimiento de Competencias Fundamentales (30% de los estudiantes)**

Orientada a estudiantes que reconocen dificultades académicas y buscan proactivamente apoyo para mejorar su rendimiento en asignaturas de programación. Con una duración de ocho meses (extensible), su estrategia se basa en el aprendizaje colaborativo con tutorías personalizadas, práctica deliberada y retroalimentación continua.





Comienza con un diagnóstico individualizado para identificar brechas y trazar un plan de aprendizaje personalizado. Continúa con 24 sesiones de tutoría académica (3 horas semanales) en lógica de programación, estructuras de datos y POO; talleres de nivelación grupal (40 horas); práctica deliberada con ejercicios graduados; grupos de estudio colaborativo; y dos proyectos prácticos de consolidación que progresan desde una aplicación de consola hasta una solución básica aplicada a una problemática social.

Los resultados esperados son una mejora mínima de 0.4 puntos en el promedio de asignaturas de programación, dos proyectos funcionales y una autoeficacia mejorada medida antes y después de la intervención. Las competencias objetivo incluyen pensamiento algorítmico, debugging efectivo, estrategias de estudio y resiliencia académica.

### Indicadores de éxito:

- Incremento de promedio  $\geq 1.0$  puntos: Meta  $>70\%$  de estudiantes
- Aprobación de asignaturas en riesgo: Meta  $>85\%$
- Retención en semillero a 4 meses: Meta  $>80\%$
- Satisfacción con tutorías: Meta  $>4.2/5.0$
- Mayor confianza en habilidades de programación: Meta  $>75\%$

### Discusión de Hallazgos

#### Validación de la Hipótesis de Heterogeneidad Estudiantil

Los resultados confirman empíricamente que los estudiantes del semillero MAMBA no constituyen un grupo homogéneo, sino que se organizan en tres perfiles claramente diferenciados con motivaciones específicas. Esta heterogeneidad, identificada mediante clustering con GMM (Silhouette Score: 0.1460) y validada mediante clasificación con MLP (F1-Score: 0.9404), cuestiona el enfoque tradicional en programas de formación complementaria.

La existencia de tres perfiles es consistente con la Teoría de Autodeterminación de (Deci & Ryan, 2000), que postula diferentes orientaciones motivacionales:





1. **Motivación instrumental-pragmática** (Cluster 0 - Opción de Grado): Estudiantes que valoran el semillero como ruta estratégica para completar trabajo de grado y obtener título profesional.
2. **Motivación intrínseca-extrínseca integrada** (Cluster 1 - Meta Personal y Empleabilidad): Estudiantes que buscan simultáneamente crecimiento personal mediante dominio de tecnologías emergentes y preparación estratégica para empleabilidad.
3. **Motivación de mejoramiento académico** (Cluster 2 - Mejoramiento Académico): Estudiantes que reconocen necesidades de fortalecimiento en competencias fundamentales y buscan proactivamente recursos de apoyo

## Factores Predictivos más Relevantes

El análisis de importancia de características reveló tres hallazgos contraintuitivos:

**Dominancia de variables motivacionales sobre demográficas:** Las variables motivacionales (Q10, Q30, Q31) acumularon 88.8% de la importancia total, mientras que variables demográficas representaron solo 6.6%. Esto sugiere que en educación superior técnica regional, las orientaciones motivacionales tienen mayor peso explicativo que las condiciones estructurales.

**Especificidad tecnológica como discriminante crítica:** El interés en "análisis de datos mediante machine learning" fue la variable más predictiva (39.8%), distinguiendo radicalmente al Cluster 1 (Meta Personal y Empleabilidad) de otros perfiles. La especialización temprana en áreas de alto valor de mercado resulta un factor diferenciador más potente que el rendimiento académico general.

**Importancia moderada del rendimiento académico:** La motivación por "mejorar rendimiento académico" representó solo 3.7% de importancia, indicando que los perfiles se definen por orientaciones de valor hacia diferentes tipos de logros, no por el nivel de desempeño actual.

## Implicaciones para el Diseño Pedagógico del Semillero

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989  
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



Las tres rutas propuestas reconocen la legitimidad de múltiples motivaciones en la participación estudiantil, superando el modelo tradicional de formación exclusiva en investigación académica.

## Desempeño del Modelo Predictivo y Aplicabilidad

El modelo MLP optimizado alcanzó F1-Score de 0.9404, superando estándares de estudios similares (típicamente 0.75-0.85). Este desempeño se atribuye a la selección teóricamente fundamentada de variables, la estrategia de oversampling balanceado (1,000 muestras/clase) y la capacidad del MLP de capturar interacciones no lineales complejas entre variables motivacionales.

Sin embargo, tres limitaciones son importantes: el tamaño muestral original (N=150) está en el límite inferior para redes neuronales robustas; el desbalance extremo del Cluster 1 en el conjunto de prueba (solo 3 casos) limita la confiabilidad de sus métricas; y la validación fue realizada únicamente con datos de CORHUILA.

## Limitaciones del Estudio

El estudio presenta cinco limitaciones principales. Su diseño transversal, con datos recolectados en un único momento (N=150), impide analizar si los perfiles son estables o si los estudiantes transitan entre ellos a lo largo del tiempo. El muestreo por conveniencia introduce sesgos de autoselección que restringen la generalización a otras poblaciones o instituciones. El uso exclusivo de cuestionario autoadministrado expone los resultados a sesgos de deseabilidad social y autopercepción inexacta, limitación que podría mitigarse triangulando con datos administrativos.

Adicionalmente, variables relevantes como autoeficacia, resiliencia y factores familiares o institucionales no fueron capturadas, lo que reduce el poder explicativo del modelo. Finalmente, la validación es únicamente interna, por lo que se requieren estudios en





otras instituciones para confirmar la generalización tanto de los perfiles identificados como del modelo predictivo desarrollado.

## Comparación con Estudios Previos

Los perfiles identificados muestran similitudes y diferencias importantes con taxonomías previas:

El **Cluster 0 (Opción de Grado)** representa un perfil pragmático-instrumental relacionado con "estudiantes orientados a logros específicos" en literatura de motivación académica.

El **Cluster 1 (Meta Personal y Empleabilidad)** coincide con el emergente perfil de "estudiantes híbridos" que integran desarrollo personal con preparación laboral estratégica, documentado en literatura sobre generación Z (Eberhardt, 2017).

El **Cluster 2 (Mejoramiento Académico)** refleja "estudiantes con orientación a maestría" en contextos de dificultad académica, relacionado con investigaciones sobre mentalidad de crecimiento (Yeager & Dweck, 2012).

El F1-Score de 0.9404 supera significativamente benchmarks reportados en literatura ((Romero & Ventura, 2010); (Sokkhey & Okazaki, 2020)), aunque comparaciones directas son complicadas por diferencias en tamaño muestral (N=200-500 vs. N=150) y definición de clases.

La distribución de perfiles difiere de estudios en universidades de investigación intensiva, que reportan mayor proporción de estudiantes orientados exclusivamente a investigación académica (40-50%). Esta diferencia posiblemente refleja realidades de instituciones regionales donde presiones por graduación oportuna y empleabilidad son más explícitas.





## Conclusiones

Este estudio desarrolló un sistema inteligente de caracterización y clasificación de estudiantes para el semillero de investigación MAMBA de la Universidad CORHUILA, integrando aprendizaje automático no supervisado y supervisado, con contribuciones metodológicas y pedagógicas significativas a la personalización educativa.

El análisis mediante Gaussian Mixture Models identificó tres perfiles estudiantiles diferenciados con un Silhouette Score de 0.1460, superior a K-Means, Agglomerative, Spectral y DBSCAN. El primer perfil (42%) agrupa estudiantes motivados instrumentalmente por completar su trabajo de grado; el segundo (28%) representa estudiantes orientados al crecimiento profesional en áreas como Machine Learning y UX; y el tercero (30%) concentra estudiantes enfocados en fortalecer competencias fundamentales de programación.

El modelo predictivo Multi-Layer Perceptron optimizado alcanzó un F1-Score de 0.9404, superando los benchmarks reportados (0.75–0.90). Las variables motivacionales acumularon el 88.8% del poder predictivo, frente al 6.6% de las variables demográficas, confirmando que los perfiles se definen por orientaciones motivacionales y preferencias tecnológicas, no por características estructurales inmutables.

Los hallazgos son coherentes con la Teoría de Autodeterminación, documentando por primera vez la coexistencia de motivación instrumental-pragmática, motivación intrínseca-extrínseca integrada y motivación de mejoramiento académico en un mismo espacio formativo. Esta distribución refleja las realidades de instituciones regionales donde las presiones por graduación y empleabilidad son más explícitas que en universidades de investigación intensiva.

El estudio representa la primera aplicación documentada de un pipeline integrado GMM-MLP para caracterización de perfiles en semilleros de investigación en Colombia. Sus tres contribuciones metodológicas transferibles son: un protocolo riguroso que minimiza sesgos mediante división de datos previa al clustering; el uso de MLP complementado con análisis de importancia mediante XGBoost, que resuelve la tensión entre desempeño e interpretabilidad; y

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989  
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



# CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA  
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN  
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

la demostración de que es posible fundamentar empíricamente el diseño de rutas formativas, superando diseños basados en intuición docente.

En la implementación, los estudiantes accederán a rutas alineadas con sus motivaciones: el 42% recibirá acompañamiento metodológico para proyectos de grado, el 28% participará en bootcamps técnicos y mentorías industriales, y el 30% tendrá tutorías personalizadas en fundamentos de programación. El sistema amplía el alcance del semillero más allá de élites de alto rendimiento, cumpliendo una función inclusiva coherente con el mandato de pertinencia social.

El estudio presenta limitaciones propias de su diseño: carácter transversal que imposibilita análisis de estabilidad temporal, muestreo no probabilístico, sesgos inherentes al cuestionario autoadministrado y tamaño muestral en el límite inferior para redes neuronales. Las líneas futuras prioritarias incluyen estudios longitudinales con seguimiento de cohortes, diseños cuasiexperimentales con grupo control, validación externa en instituciones diversas, triangulación metodológica con datos administrativos y entrevistas, exploración de técnicas avanzadas como SMOTE y arquitecturas de Deep Learning, y análisis de equidad para evaluar si las rutas democratizan el acceso o reproducen segregación.

Este estudio desafía el modelo hegemónico de semilleros que asume uniformidad con orientación académico-investigativa, demostrando que la heterogeneidad motivacional es la norma. Reconocer que el 42% busca graduarse mediante semillero es un acto de honestidad institucional; que el 28% integra pasión técnica y valor profesional supera dicotomías tradicionales; y que el 30% busca proactivamente apoyo revela una función democratizadora crítica en instituciones regionales. La implementación de rutas diferenciadas representa así una transición desde pedagogías uniformizantes hacia pedagogías responsivas, donde la inteligencia artificial no automatiza decisiones pedagógicas, sino que genera evidencia robusta para diseñar experiencias formativas más justas, efectivas y alineadas con las aspiraciones plurales de estudiantes que son agentes de sus propios proyectos de vida.

- 📍 Sede Quirinal: Calle 21 No. 6 - 01
- 📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220
- 📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459
- ✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989  
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



## Recomendaciones

Para la correcta implementación del proyecto de clasificación y formación diferenciada en el semillero MAMBA, se recomienda en primera instancia la ejecución inmediata de las rutas formativas diseñadas, priorizando la asignación de los 150 estudiantes a sus respectivos trayectos según la clasificación del modelo MLP.

Esta acción debe complementarse con la institucionalización del modelo de Perceptrón Multicapa (MLP) como la herramienta estándar de diagnóstico para nuevos ingresos, aprovechando su alto rendimiento predictivo para eliminar la subjetividad en la selección de proyectos. Es fundamental que la Coordinación del Semillero inicie las actividades de la Ruta 1 mediante la asignación de directores de tesis experimentados, mientras que para la Ruta 2 se deben establecer vínculos formales con el sector productivo que faciliten el acceso a certificaciones técnicas y mentorías industriales. Para los estudiantes del Cluster 2, integrados en la Ruta 3, se sugiere implementar un programa de acompañamiento psicopedagógico que trabaje en la autorregulación del aprendizaje y la reducción de la ansiedad ante la programación.

La proyección de beneficios ante el cumplimiento de estas recomendaciones incluye un incremento esperado de al menos 0.4 puntos en el promedio académico para el grupo de nivelación, una tasa de retención superior al 80% y la consolidación de un portafolio profesional robusto para quienes buscan inserción laboral cualificada. Finalmente, se debe realizar un seguimiento mensual de los indicadores de éxito, como la satisfacción con las tutorías y la mejora en la autopercepción de autoeficacia, para ajustar las estrategias pedagógicas según la evolución real de cada perfil.





## Referencias

- Alarcón, M. G., Carvajal, C. H., Nuñez, J. L., Velandía, C. V., & Noya, D. F. (2020). *Semilleros de Investigación: Experiencias significativas en el contexto de la Educación Superior*. [https://www.researchgate.net/publication/376785549\\_SEMILLEROS\\_DE\\_INVESTIGACION\\_EXPERIENCIAS\\_SIGNIFICATIVAS\\_EN\\_EL\\_CONTEXTO\\_DE\\_LA\\_EDUCACION\\_SUPERIOR](https://www.researchgate.net/publication/376785549_SEMILLEROS_DE_INVESTIGACION_EXPERIENCIAS_SIGNIFICATIVAS_EN_EL_CONTEXTO_DE_LA_EDUCACION_SUPERIOR)
- Bousalem, Z., Qazdar, A., & El Guabassi, I. (2023). Cooperative Learning Groups: A New Approach Based on Students' Performance Prediction. *International Journal of Online and Biomedical Engineering (iJOE)*, 19(12), 34-48. <https://doi.org/10.3991/ijoe.v19i12.41181>
- Breiman. (2001a). *RANDOM FORESTS*. <https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf>
- Brew, A., & Mantai, L. (2017). Academics' perceptions of the challenges and barriers to implementing research-based experiences for undergraduates. *Teaching in Higher Education*, 22(5), 551-568. <https://doi.org/10.1080/13562517.2016.1273216>
- Catal, C., Ozcan, A., Donmez, E., & Kasif, A. (2023). Analysis of cyber security knowledge gaps based on cyber security body of knowledge. *Education and Information Technologies*, 28(2), 1809-1831. <https://doi.org/10.1007/s10639-022-11261-8>
- Corhuila. (2014). *Acuerdo N. 305 Reglamento Estudiantil*. <https://corhuila.edu.co/wp-content/uploads/2021/06/ACUERDO-305-DE-2021-REGLAMENTO-ESTUDIANTIL.pdf>
- Corhuila. (2016). *Politica de Tratamiento de Datos Personal Corhuila*.
- Deci, E. L., & Ryan, R. M. (2000). The «What» and «Why» of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry*, 11(4), 227-268. [https://doi.org/10.1207/S15327965PLI1104\\_01](https://doi.org/10.1207/S15327965PLI1104_01)
- Dietterich, T. G. (2000). Ensemble Methods in Machine Learning. En G. Goos, J. Hartmanis, & J. Van Leeuwen (Eds.), *Multiple Classifier Systems* (Vol. 1857, pp. 1-15). Springer Berlin Heidelberg. [https://doi.org/10.1007/3-540-45014-9\\_1](https://doi.org/10.1007/3-540-45014-9_1)

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989  
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



- Forero-Corba, W., & Negre Bennasar, F. (2023). Técnicas y aplicaciones del Machine Learning e Inteligencia Artificial en educación: Una revisión sistemática. *RIED-Revista Iberoamericana de Educación a Distancia*, 27(1), 209-253. <https://doi.org/10.5944/ried.27.1.37491>
- Gómez-Cano, C. A., Sánchez-Castillo, V., & Estrada-Cely, G. E. (2022). Limitantes para la participación en los Semilleros de Investigación. *CULTURA EDUCACIÓN Y SOCIEDAD*, 13(2), 9-28. <https://doi.org/10.17981/cultedusoc.13.2.2022.01>
- Hermanrud, I., & Eide, D. (2017). *Representations of Practice- Distributed Sensemaking Using Boundary Objects*. 922. <https://doi.org/10.28945/3734>
- Leonetti, C. T., Lindberg, H., Schwake, D. O., & Cotter, R. L. (2023). A Call to Assess the Impacts of Course-Based Undergraduate Research Experiences for Career and Technical Education, Allied Health, and Underrepresented Students at Community Colleges. *CBE—Life Sciences Education*, 22(1), ar4. <https://doi.org/10.1187/cbe.21-11-0318>
- Lu, J., Liu, Y., Liu, S., Yan, Z., Zhao, X., Zhang, Y., Yang, C., Zhang, H., Su, W., & Zhao, P. (2024). Machine learning analysis of factors affecting college students' academic performance. *Frontiers in Psychology*, 15, 1447825. <https://doi.org/10.3389/fpsyg.2024.1447825>
- Matz, S. C., Bukow, C. S., Peters, H., Deacons, C., Dinu, A., & Stachl, C. (2023). Using machine learning to predict student retention from socio-demographic characteristics and app-based engagement metrics. *Scientific Reports*, 13(1), 5705. <https://doi.org/10.1038/s41598-023-32484-w>
- MinCiencias. (2023). *Docientos mil millones para transformar las regiones a través de la ciencia*. [https://www.minciencias.gov.co/sala\\_de\\_prensa/200-mil-millones-para-transformar-las-regiones-traves-la-ciencia](https://www.minciencias.gov.co/sala_de_prensa/200-mil-millones-para-transformar-las-regiones-traves-la-ciencia)
- OCyT. (2023). *Observatorio Colombiano de Ciencia y Tecnología*. <https://ocyt.org.co/wp-content/uploads/2024/04/INFORME-DE-GESTION-2023.pdf>
- República de Colombia. (1992). *Ley 30* 1992. [https://www.funcionpublica.gov.co/eva/gestornormativo/norma\\_pdf.php?i=253](https://www.funcionpublica.gov.co/eva/gestornormativo/norma_pdf.php?i=253)
- República de Colombia. (2012). *Ley 1581* 2012.

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989

NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



- Romero & Ventura. (2010). *Educational Data Mining: A Review of the State of the Art*.  
[https://www.researchgate.net/publication/224160756\\_Educational\\_Data\\_Mining\\_A\\_Review\\_of\\_the\\_State\\_of\\_the\\_Art](https://www.researchgate.net/publication/224160756_Educational_Data_Mining_A_Review_of_the_State_of_the_Art)
- Vanegas, C. V. (2024). La motivación como herramienta investigativa para semilleros contables en Colombia. *South Florida Journal of Development*, 5(2), 501-515.  
<https://doi.org/10.46932/sfjdv5n2-007>
- Vera-Fernández, C. G., & Fernández-Nieto, E. L. (2024). “Los semilleros de Investigación” hacia la consolidación de una cultura investigativa en el contexto universitario. Revisión de literatura 2017-2023. *AiBi Revista de Investigación, Administración e Ingeniería*, 12(2), 188-201. <https://doi.org/10.15649/2346030X.3445>
- Von Der Mülbe, S., Stockinger, K., Scheunemann, A., Kegel, L. S., Fleischer, J., Leutner, D., Wirth, J., Grunschel, C., & Dresel, M. (2026). Motivated and feeling good? Reciprocal relations between motivational regulation and student well-being over one semester. *Learning and Individual Differences*, 126, 102850.  
<https://doi.org/10.1016/j.lindif.2025.102850>
- Walkington, C. (2021). Intrinsic integration in learning games and virtual instruction. *Educational Technology Research and Development*, 69(1), 157-160.  
<https://doi.org/10.1007/s11423-020-09886-y>
- Zachos, G., Paraskevopoulou-Kollia, E.-A., & Anagnostopoulos, I. (2018). Social Media Use in Higher Education: A Review. *Education Sciences*, 8(4), 194.  
<https://doi.org/10.3390/educsci8040194>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), 39.  
<https://doi.org/10.1186/s41239-019-0171-0>  
[https://www.researchgate.net/publication/376785549\\_SEMILLEROS\\_DE\\_INVESTIGACION\\_EXPERIENCIAS\\_SIGNIFICATIVAS\\_EN\\_EL\\_CONTEXTO\\_DE\\_LA\\_EDUCACION\\_SUPERIOR](https://www.researchgate.net/publication/376785549_SEMILLEROS_DE_INVESTIGACION_EXPERIENCIAS_SIGNIFICATIVAS_EN_EL_CONTEXTO_DE_LA_EDUCACION_SUPERIOR)

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8350459

✉ Email: [contacto@corhuila.edu.co](mailto:contacto@corhuila.edu.co) - [www.corhuila.edu.co](http://www.corhuila.edu.co)

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989

NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"