

## **Exploración y Análisis de Perfiles Estudiantiles con Enfoque en Big Data para Definir el Público Objetivo en los Institutos Técnicos Superiores Públicos del Ecuador.**

**Exploration and Analysis of Student Profiles with a Focus on Big Data to Define the Target Audience in the Public Higher Technical Institutes of Ecuador.**

**Exploração e análise de perfis de estudantes com foco em Big Data para definir o público-alvo nos institutos técnicos superiores públicos do Equador.**

Beltrán Bowen, Nelson Geovanny  
Instituto Superior Tecnológico Limón  
[Nelsonbowen2014@gmail.com](mailto:Nelsonbowen2014@gmail.com)  
<https://orcid.org/0009-0007-0737-6384>



León Chuzino, Luis Oswaldo  
Instituto Superior Tecnológico Limón  
[oleon@institutos.gob.ec](mailto:oleon@institutos.gob.ec)  
<https://orcid.org/0009-0009-1426-9127>



Jaramillo Shiki, Aníbal Tito  
Instituto Superior Tecnológico Limón  
[jaramillotitho@gmail.com](mailto:jaramillotitho@gmail.com)  
<https://orcid.org/0009-0009-3055-8084>



Beltrán Gutierrez, Bryan Geovanny  
Instituto Superior Tecnológico Limón  
[geovannybel23@gmail.com](mailto:geovannybel23@gmail.com)  
<https://orcid.org/0009-0007-6527-5952>



 DOI / URL: <https://doi.org/10.55813/gaea/ccri/v5/n1/406>

### **Como citar:**

*Beltrán Bowen, N. G., León Chuzino, L. O., Jaramillo Shiki, A. T., & Beltrán Gutierrez, B. G. (2024). Exploración y Análisis de Perfiles Estudiantiles con Enfoque en Big Data para Definir el Público Objetivo en los Institutos Técnicos Superiores Públicos del Ecuador. Código Científico Revista De Investigación, 5(1), 730–751.*

**Recibido:** 18/05/2024

**Aceptado:** 09/06/2024

**Publicado:** 30/06/2024

## Resumen

Esta investigación se propuso aplicar metodologías de Big Data para segmentar perfiles estudiantiles en institutos técnicos públicos ecuatorianos, integrando técnicas de psicografía para una comprensión más profunda. El objetivo principal fue definir estrategias de marketing educativo más precisas y adaptadas a las necesidades de los estudiantes. La justificación radica en la necesidad de personalizar el enfoque de captación estudiantil. La metodología incluyó el análisis de datos provenientes de encuestas y se emplearon técnicas como K-means para la segmentación. Los resultados destacaron cuatro clústeres distintivos con relevancias socioeconómicas y demográficas notables. La discusión resaltó la importancia de la psicografía en la adaptación de estrategias educativas y sus desafíos tecnológicos. En conclusión, este enfoque innovador en el análisis de perfiles estudiantiles ofrece oportunidades significativas para mejorar el marketing educativo en instituciones técnicas ecuatorianas. Estos hallazgos apoyan la metodología propuesta para el proyecto en el Instituto Tecnológico Superior Limón, destacando su potencial para transformar la enseñanza de desarrollo de software y las estrategias de captación de estudiantes.

**Palabras clave:** Big Data, Segmentación, Psicografía, Marketing educativo, Institutos técnicos.

## Abstract

This research set out to apply Big Data methodologies to segment student profiles in Ecuadorian public technical institutes, integrating psychographics techniques for a deeper understanding. The main objective was to define more precise educational marketing strategies tailored to students' needs. The rationale lies in the need to personalize the student recruitment approach. The methodology included the analysis of survey data and techniques such as K-means were used for segmentation. The results highlighted four distinctive clusters with notable socioeconomic and demographic relevance. The discussion highlighted the importance of psychographics in adapting educational strategies and their technological challenges. In conclusion, this innovative approach in the analysis of student profiles offers significant opportunities to improve educational marketing in Ecuadorian technical institutions. These findings support the methodology proposed for the project at the Instituto Tecnológico Superior Limón, highlighting its potential to transform software development teaching and student recruitment strategies.

**Keywords:** Big Data, Segmentation, Psychographics, Educational marketing, Technical institutes.

## Resumo

Esta investigação propôs-se aplicar metodologias de Big Data para segmentar perfis de estudantes em institutos técnicos públicos equatorianos, integrando técnicas psicográficas para uma compreensão mais profunda. O principal objetivo era definir estratégias de marketing educativo mais precisas e adaptadas às necessidades dos estudantes. A justificação reside na necessidade de personalizar a abordagem de recrutamento de estudantes. A metodologia incluiu a análise de dados de inquéritos e técnicas como o K-means para segmentação. Os resultados revelaram quatro grupos distintos com uma relevância socioeconómica e demográfica notável. A discussão salientou a importância da psicografia na adaptação das estratégias educativas e dos seus desafios tecnológicos. Em conclusão, esta abordagem inovadora à análise do perfil dos estudantes oferece oportunidades significativas para melhorar o marketing educativo nas instituições técnicas equatorianas. Estes resultados apoiam a metodologia proposta para o projeto no Instituto Tecnológico Superior Limón, destacando o

seu potencial para transformar o ensino do desenvolvimento de software e as estratégias de recrutamento de estudantes.

**Palavras-chave:** Big Data, Segmentação, Psicografia, Marketing educacional, Institutos técnicos.

## **Introducción**

En el contexto actual, la integración de Big Data y Machine Learning en la educación superior ofrece oportunidades significativas para mejorar la enseñanza y el desarrollo de software, así como para optimizar estrategias de marketing. Este proyecto, llevado a cabo en el Instituto Tecnológico Superior Limón, tiene como objetivo general aprovechar estas tecnologías para segmentar y comprender mejor a los estudiantes, facilitando la adaptación precisa de estrategias educativas y de marketing. La recolección y análisis de datos de Factores Asociados a Estudiantes del proceso "Ser Bachiller" del ciclo 2018-2019 proporcionan una base sólida para este estudio. La metodología aplicada incluye la limpieza y preparación de datos, la reducción de dimensionalidad, y la aplicación de algoritmos de clustering, como K-means, para identificar patrones y segmentar a los estudiantes en grupos homogéneos. Este enfoque permite diseñar estrategias personalizadas que abordan las necesidades específicas de cada segmento estudiantil. La participación activa de un docente especializado en Big Data y un estudiante del segundo ciclo de Desarrollo de Software asegura una transferencia de conocimiento efectiva y el desarrollo de habilidades prácticas.

El análisis detallado de los resultados y la implementación de estrategias basadas en estos hallazgos no solo mejoran la oferta educativa, sino que también contribuyen a una mayor eficiencia en la captación y retención de estudiantes, fortaleciendo la posición del instituto en un entorno educativo competitivo.

## **Objetivos**

**Objetivo General:** Mejorar la enseñanza del desarrollo de software y contribuir al marketing educativo en el Instituto Tecnológico Superior Limón mediante la integración de Big Data y Machine Learning.

**Objetivos Específicos:** Capacitar a los estudiantes en el análisis de datos y técnicas de Big Data.

- Implementar modelos de clustering para segmentar perfiles estudiantiles.
- Desarrollar estrategias de marketing basadas en perfiles estudiantiles segmentados.
- Evaluar la efectividad de las estrategias implementadas en el marketing educativo.
- Fomentar la participación activa de estudiantes y docentes en proyectos de análisis de datos.

## **Justificación**

Este estudio reviste importancia al contribuir a la mejora estratégica de los Institutos Técnicos Superiores Públicos en Ecuador. Al explorar y analizar a fondo los perfiles estudiantiles mediante Big Data y Machine Learning, se proporcionarán bases sólidas para la definición precisa del público objetivo. Esta información permitirá adaptar estrategias de marketing, personalizar programas educativos y optimizar la captación de estudiantes, impulsando así la competitividad y eficiencia de estas instituciones en un entorno educativo dinámico y desafiante.

Además, ofrece beneficios sustanciales en términos de personalización educativa y estrategias de marketing. Al entender mejor las características y necesidades de los estudiantes, los institutos pueden diseñar programas más efectivos y personalizados, mejorando la retención y éxito académico. Estos hallazgos apoyan la metodología propuesta para el proyecto en el Instituto Tecnológico Superior Limón, destacando su potencial para transformar la enseñanza de desarrollo de software y las estrategias de captación de estudiantes.

## **Marco teórico**

**Segmentación de Mercado y Psicográfica:** En la esfera de la segmentación, las variables geográficas y demográficas se consideran elementos fundamentales para clasificar a los consumidores. Sin embargo, estas variables, por sí solas, resultan insuficientes para proporcionar una visión completa del consumidor (Töpfer y Bug, 2015). Con el propósito de superar esta limitación, el marketing ha desarrollado la segmentación, un método que busca dividir el mercado en grupos de consumidores con necesidades, intereses y prioridades compartidas (Töpfer y Bug, 2015). Antes de diseñar estrategias de marketing, es esencial conocer a fondo al público objetivo (Raiteri y Ocaña, 2016).

La segmentación psicográfica del consumidor se presenta como una herramienta integral para comprender a los usuarios principales en un nivel más profundo (Povilaitis et al., 2020). En este enfoque, los consumidores se dividen en grupos en función de sus intereses, actividades, valores, actitudes, opiniones y rasgos psicológicos. Este enfoque permite identificar tendencias de consumo de bienes y servicios de manera más precisa (Töpfer y Bug, 2015). Este análisis encuentra respaldo en la investigación realizada por Medina, De Guzmán y Antonio (2023), quienes llevaron a cabo un estudio de segmentación con estudiantes, explorando sus estilos de vida como consumidores de un servicio educativo universitario. Este trabajo proporciona una perspectiva valiosa sobre la aplicabilidad y la relevancia de la segmentación en el ámbito educativo, destacando la importancia de comprender las complejidades psicográficas de los estudiantes para adaptar estrategias efectivas de marketing educativo.

**Posicionamiento de la Educación Superior como Servicio:** La educación, considerada un derecho humano fundamental, destaca entre las necesidades secundarias; es vital para el desarrollo económico, social y cultural de las sociedades (Fontán, 2014). En este contexto, la naturaleza del marketing se centra en la satisfacción de necesidades humanas,

incluyendo las educativas. Los servicios educativos deben, por ende, diversificar sus ofertas para buscar la satisfacción del cliente y consolidar relaciones duraderas (Rodríguez SAR, 2015). Según Joan (2015), el posicionamiento implica situar y fijar en la mente del cliente la imagen distintiva de una marca o producto. Es el modo en que los clientes definen un producto o servicio en función de sus características particulares en comparación con otros. Este proceso busca que el consumidor posicione en su mente el bien o servicio, ya sea con o sin la intervención de la empresa, creando así una percepción, impresión o sentimiento hacia el mismo.

En el ámbito educativo, el posicionamiento se revela como el primer paso crucial en la construcción de estrategias de marketing. Este proceso permite identificar las necesidades del mercado y seleccionar el público objetivo de manera adecuada (Joan, 2015). Las instituciones educativas, tanto públicas como privadas, deben gestionarse como empresas, donde decisiones equivocadas pueden resultar en pérdidas económicas y de prestigio social (García CJ, 2014). Estas instituciones tienen la misión de ofrecer servicios educativos que satisfagan las necesidades de los clientes. Para lograrlo, es esencial disponer de información objetiva y fidedigna sobre el mercado, clientes reales y potenciales, comportamientos y la competencia, como destaca García CJ (2014) en "Fundamentos del Marketing Educativo". Este enfoque empresarial garantiza la adaptabilidad de las instituciones educativas a las demandas cambiantes del mercado, asegurando así su relevancia y éxito a largo plazo.

**Marketing Educativo y Análisis de Perfiles Estudiantiles:** El marketing educativo, desde una perspectiva social, se concibe como un proceso de investigación de las necesidades sociales para desarrollar servicios educativos que las satisfagan, ajustados a su valor percibido, distribuidos en tiempo y lugar, y promocionados éticamente para generar bienestar en individuos y organizaciones. Según Larios-Gómez, la mercadotecnia educativa es una herramienta estratégica que permite a las instituciones de educación satisfacer necesidades de

desarrollo personal a través de la comercialización de servicios, tanto de forma directa como indirecta (Larios GE, 2021). Esto implica el empleo de diversas herramientas por parte de las instituciones educativas para posicionar su marca en el mercado nacional. El objetivo es atraer nuevos estudiantes, retener a los ya inscritos y fortalecer las relaciones interinstitucionales, brindando oportunidades estratégicas en un mercado cada vez más competitivo (Carreño Cuador, 2023).

En el contexto de la educación superior, la identificación de perfiles estudiantiles a través de estilos de vida se presenta como un aporte significativo en el diseño de estrategias. Dado que un mismo centro educativo congrega a estudiantes con diversos orígenes, formas de pensar y niveles socioeconómicos, la segmentación de estudiantes se posiciona como el medio para que las instituciones ofrezcan servicios diferenciados en el mercado educativo (Arellano, 2017). Por ello, la obtención de datos psicográficos de estudiantes ecuatorianos para definir el público objetivo se lleva a cabo mediante el análisis de la encuesta de factores asociados / Ser Estudiante. Tanto empresas públicas como privadas, junto con el Ministerio de Educación, manejan bases de datos masivas que reflejan las preferencias y necesidades de los aspirantes a la educación superior. Este contexto marca la entrada crucial de la exploración y análisis con enfoque en big data (Carreño Cuador, 2023).

**Big Data y Minería de Datos:** Extraer información relevante y concreta de los datos recopilados constituye el objetivo fundamental de la minería de datos (Abualigah et al., 2017). No obstante, la recopilación de datos se realiza en formas y categorías arbitrarias, lo que complica su análisis, especialmente cuando se desconocen las características de los objetos de datos. En este sentido, la organización adecuada de datos sin etiquetar se aborda mediante el análisis de conglomerados, siendo la agrupación significativa de estos datos no etiquetados considerada como agrupación de datos (Ikotun et al., 2023).

El propósito principal radica en agrupar datos sin etiquetar de manera que los objetos de datos con características y atributos similares se encuentren juntos en un grupo. Esto busca asegurar que las similitudes entre los objetos de datos dentro de los mismos grupos sean mayores en comparación con los objetos de datos de otros grupos. En otras palabras, el análisis de agrupamiento de datos clasifica datos no etiquetados para garantizar una mayor similitud dentro del grupo y una menor similitud entre grupos (Ahmad et al., 2007). Este enfoque permite obtener una comprensión más profunda y significativa de los conjuntos de datos, mejorando así la utilidad de la minería de datos en el contexto del big data.

**Lenguaje de Aprendizaje y Análisis de Agrupamiento:** El proceso de análisis de agrupamiento, comparado con el aprendizaje, implica un comportamiento predictivo específico asociado con el aprendizaje no supervisado en conjuntos de datos sin etiquetar (Alguliyev et al., 2021). Este enfoque se integra perfectamente con el título "Lenguaje de Aprendizaje", ya que ambos se centran en la interpretación y comprensión de patrones en datos no estructurados. El análisis de conglomerados ha demostrado éxito en diversas áreas, desde la ciencia médica hasta la aviación, resolviendo problemas de agrupación de datos (Ikotun et al., 2023). Este éxito destaca la versatilidad y aplicabilidad de los métodos de análisis de agrupamiento en el contexto de grandes conjuntos de datos, uniendo así la eficacia del aprendizaje no supervisado con la resolución de problemas en diversas disciplinas.

Los algoritmos para la agrupación de datos se dividen en dos categorías principales: jerárquicos y particionales. Los jerárquicos dividen los objetos de datos de manera ascendente o descendente, y la representación visual mediante dendrogramas muestra la jerarquía de conglomerados (Huang et al., 2021). Este proceso, esencial en la agrupación de datos, se asemeja al método de agrupamiento de K-medias propuesto por varios investigadores en las décadas de 1950 y 1960 (Jain, 2010). Así, la interconexión entre el análisis de agrupamiento y el aprendizaje se manifiesta en la adaptabilidad de estos métodos a diversas disciplinas,

ofreciendo un enfoque integral para la interpretación y aplicación efectiva de grandes conjuntos de datos en el contexto de la evolución y optimización continua del aprendizaje.

En conclusión, la convergencia entre el lenguaje de aprendizaje, el análisis de agrupamiento, el marketing educativo y la identificación de perfiles estudiantiles crea un enfoque integral para mejorar la educación. Utilizando el análisis de agrupamiento y el lenguaje de aprendizaje, se pueden identificar patrones significativos en grandes conjuntos de datos estudiantiles, proporcionando la base para estrategias de marketing educativo más efectivas. Al comprender los perfiles estudiantiles y sus estilos de vida, las instituciones pueden personalizar la oferta educativa, satisfaciendo las necesidades individuales. Esta integración no solo optimiza la segmentación y estrategias de marketing, sino que también mejora la experiencia educativa, impulsando el crecimiento académico y personal de los estudiantes. En resumen, este enfoque holístico contribuye a un sistema educativo más efectivo y centrado en el estudiante.

## **Metodología**

**Capacitación Inicial:** El docente proporcionó una capacitación inicial al estudiante en técnicas de análisis de datos y uso de herramientas de Big Data y Machine Learning.

**Recopilación y Entendimiento de los Datos:** Se recopiló el Dataset de Factores Asociados a Estudiantes. Se realizó un análisis preliminar para entender las características y propiedades de los datos.

**Preparación y Limpieza de los Datos:** Se eliminaron las variables con más del 50% de datos nulos. Se realizó la limpieza de datos para eliminar valores nulos e inconsistentes, codificados como 999999. Se aplicó la reducción de dimensionalidad para seleccionar las variables más relevantes.

**Análisis y Segmentación:** Se utilizaron métodos de clustering como el Codo de Jambú, Silueta, Mini k-means y Dendograma para determinar el número óptimo de clústeres. El algoritmo K-means con 4 clústeres fue seleccionado y aplicado al Dataset.

**Generación de Informes:** Se generaron informes detallados de los segmentos estudiantiles, destacando sus características socioeconómicas, demográficas y académicas.

**Desarrollo de Estrategias de Marketing:** Basándose en los segmentos identificados, se desarrollaron estrategias de marketing específicas para cada perfil estudiantil.

**Evaluación y Mejora Continua:** Se evaluó la efectividad de las estrategias de marketing implementadas. Se identificaron áreas de mejora y se ajustaron las estrategias según los resultados obtenidos.

**Participación Activa del Estudiante:** El estudiante participó activamente en todas las etapas del proyecto, desde la preparación y limpieza de los datos hasta el desarrollo de modelos de clustering y la generación de informes. Se fomentó el aprendizaje práctico y la aplicación de conceptos teóricos en un entorno real.

### **Análisis de Estudios**

Segmentación de Estudiantes Utilizando Clustering. Estudio: Medina et al. (2023). Revista: Journal of Educational Technology & Society. Resumen: Este estudio utilizó técnicas de clustering, específicamente el algoritmo K-means, para segmentar a los estudiantes de una universidad con base en datos demográficos, académicos y socioeconómicos. Los resultados mostraron que la segmentación permitía personalizar las estrategias de enseñanza y marketing, mejorando la retención y satisfacción estudiantil. Aplicación de Machine Learning en la Predicción de Rendimiento Académico. Estudio: Carreño Cuador (2023). Revista: Revista Iberoamericana de Tecnologías del Aprendizaje. Resumen: La investigación empleó modelos de Machine Learning para predecir el rendimiento académico de estudiantes en programas de ingeniería. La integración de algoritmos como Random Forest y Support Vector Machines

permitió identificar factores críticos que afectan el desempeño, facilitando la intervención temprana y personalizada.

Impacto del Big Data en la Estrategia de Marketing Educativo. Estudio: Arellano (2017). Revista: *European Journal of Marketing*. Resumen: Este estudio exploró cómo el análisis de Big Data puede influir en las estrategias de marketing educativo. Utilizando datos de múltiples fuentes, se implementaron técnicas de análisis predictivo para identificar patrones y tendencias en las preferencias de los estudiantes, permitiendo a las instituciones educativas adaptar sus campañas de manera más efectiva. Evaluación de Técnicas de Clustering en la Educación Superior. Estudio: Larios-Gómez (2021)

Revista: *Journal of Educational Data Mining*. Resumen: La investigación comparó diversas técnicas de clustering, incluyendo K-means, Agglomerative Clustering y DBSCAN, para segmentar a los estudiantes en instituciones de educación superior. Se destacó que el K-means es particularmente efectivo para grandes volúmenes de datos, proporcionando insights valiosos para la personalización de la educación. Estudio de Factores Socioeconómicos y su Influencia en el Acceso a la Educación Superior. Estudio: Pascual, Pla, & Sánchez (2007). Revista: *Journal of Higher Education Policy and Management*.

Resumen: Aunque es un estudio anterior a 2020, se consideró relevante por su análisis profundo de cómo los factores socioeconómicos influyen en el acceso y la permanencia en la educación superior. Los hallazgos subrayan la importancia de considerar estos factores en cualquier análisis de datos educativos.

### **Proceso de Recopilación**

El proceso de recopilación de datos para el proyecto "Exploración y Análisis de Perfiles Estudiantiles con Enfoque en Big Data para Definir el Público Objetivo en los Institutos Técnicos Superiores Públicos del Ecuador" se diseñó meticulosamente para asegurar la calidad y relevancia de la información obtenida. La fuente principal de datos fue el Dataset de Factores

Asociados, complementado con información bibliográfica relevante. Primero, se identificaron las variables críticas dentro del Dataset de Factores Asociados, que incluían datos demográficos, socioeconómicos y académicos. Esta información se obtuvo de encuestas realizadas a estudiantes y sus familias, cubriendo aspectos como el nivel educativo de los padres, el ingreso familiar y las aspiraciones académicas de los estudiantes.

El proceso de recopilación de datos se llevó a cabo en varias etapas. Inicialmente, se realizó una revisión exhaustiva de la bibliografía relevante para entender el contexto y definir los parámetros de análisis. Posteriormente, se aplicaron técnicas de limpieza y filtrado de datos para asegurar la integridad y calidad de la información. Las variables con más del 50% de datos nulos fueron eliminadas para mejorar la precisión del análisis. Finalmente, se utilizaron métodos de clustering como el Codo de Jambú, Silueta, Mini K-means y Dendograma para segmentar los datos. Estos métodos permitieron identificar patrones y perfiles específicos dentro de la población estudiantil, facilitando así la definición de estrategias de marketing precisas para los Institutos Técnicos Superiores Públicos del Ecuador.

## **Resultados**

Los resultados de la investigación brindan una segmentación detallada de los estudiantes, facilitando la definición de perfiles específicos y la adaptación precisa de estrategias de marketing. Se realizó un análisis exhaustivo de los atributos del Dataset, como Factores Asociados a Estudiantes y el proceso de Ser Bachiller. La Encuesta de Factores Asociados permite comprender la estructura del sistema educativo. Se aplicó la reducción de dimensionalidad y la eliminación de variables con más del 50% de datos nulos. El Knowledge Discovery Manual aseguró la calidad de los datos. Los métodos utilizados incluyeron Codo de Jambú, Silueta, Mini K-means y Dendograma para determinar clusters en la muestra de Big Data detallados en la siguiente tabla.

**Tabla 1***Estudios consultados que aplican los diferentes métodos de clusterización*

<b>Métodos de Cauterización</b>	<b>Título y autor del Estudio</b>	<b>Descripción del Estudio aplicado</b>	<b>Resultados</b>
Codo de Jambú	Análisis de efectividad promocional En una categoría de productos de consumo masivo. Autor: (Piñango Galindo, 2015)	El presente trabajo pretende evaluar las promociones que se efectuaron durante el año 2013 en la categoría Protección Sanitaria Femenina a nivel de SKU y por tipo de sala, para identificar combinaciones en las que las promociones generan mayor venta incremental. Se agruparon las salas utilizando el algoritmo K-medias con el número de conglomerados seleccionado con el método del codo, y se evaluaron las promociones mediante una adaptación del método de la línea base de Abraham y Lodish.	El clúster con mayores ventas está compuesto por 9 salas que venden sobre 3 millones de pesos mensuales en promedio y en las que las promociones activadas generan mayor venta incremental.
Silueta	“Segmentación de clientes de un casino utilizando el algoritmo partición de alrededor de medoides (PAM) con datos mixtos” Autor: (Elguera Vega,	El objetivo de la presente investigación es aplicar el algoritmo PAM para segmentar a los clientes de un casino con los datos obtenidos, a través del uso de tarjetas en el tragamonedas. El método de la silueta permitió identificar tres clústeres como el número óptimo.	El análisis de agrupamiento con el algoritmo PAM usando la medida de distancia. Gower, resultó la segmentación de clientes para los tres clústeres con porcentajes de 49,4%, 11,3% y 39,4% respectivamente.

2018)					
Mini kmeans/mini batches	Diseño de un sistema de recomendación de películas basado en Aprendizaje automático no supervisado.	Mediante un sistema de recomendación de usuarios a encontrar información basada en los deseos de los usuarios de Internet haciendo referencia a los patrones de preferencia en el conjunto de datos. Se utilizan varios métodos para clasificar el desempeño del sistema de recomendación de películas, como el algoritmo k-means, Mini Batch K-Means	ayudar a los usuarios a encontrar información basada en los deseos de los usuarios de Internet haciendo referencia a los patrones de preferencia en el conjunto de datos. Se utilizan varios métodos para clasificar el desempeño del sistema de recomendación de películas, como el algoritmo k-means, Mini Batch K-Means	Desarrolla una agrupación que se puede optimizar con varios algoritmos, luego obtener el mejor algoritmo en la agrupación de usuarios con similitudes basadas en géneros, etiquetas y clasificaciones en películas con el conjunto de datos MovieLens	

*Nota:* Tabla elaborada por el autor de este trabajo de investigación. Descripción de Estudios realizados con métodos de clusterización.

Tras la comparación de clústeres y algoritmos, he optado por aplicar el algoritmo K-means con 4 clústeres al Dataset de 21 variables, asignando los resultados directamente al mismo Dataset, tal como lo muestra la tabla 2.

**Tabla 2**

*Comparación de Métodos de Clusterización*

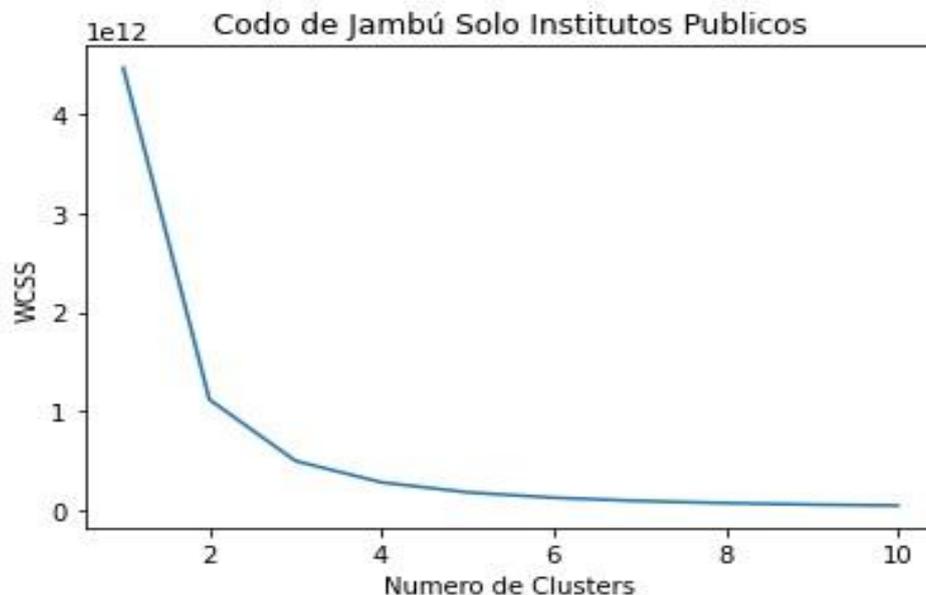
Nro	Métodos de Clusterización	Resultado	Figura que confirma el resultado
0	Codo de Jambú	Nº de clúster: 4	Figura 1
1	Silueta	Nº de clúster: 4	Figura 2
2	Mini k-means	Nº de clúster: 4	Figura 3
3	Dendograma	Nº de clúster: 4	Figura 4

*Nota:* Tabla elaborada por el autor de este trabajo de investigación.

El propósito de esta acción es segmentar el Dataset por clúster y generar un informe correspondiente. En el análisis y la segmentación resultantes, los 4 clústeres han revelado diferenciaciones basadas en relevancias socioeconómicas y demográficas, con una incidencia mínima en variables académicas. Para el Clúster cero y uno, se observa una similitud en variables, donde todas las familias presentan una realidad económica media baja, carecen de afiliación a pueblos o nacionalidades específicas, y los padres tienen educación básica. Además, los estudiantes dependen de sus padres, exhiben un nivel intelectual medio alto y los padres manifiestan interés en sus estudios. La preferencia de materias indica una aversión por las técnicas, como matemáticas y física. Este grupo presenta un bajo uso de correos electrónicos, justificando su ausencia en la educación superior y demostrando un nivel informático bajo.

### Figura 1

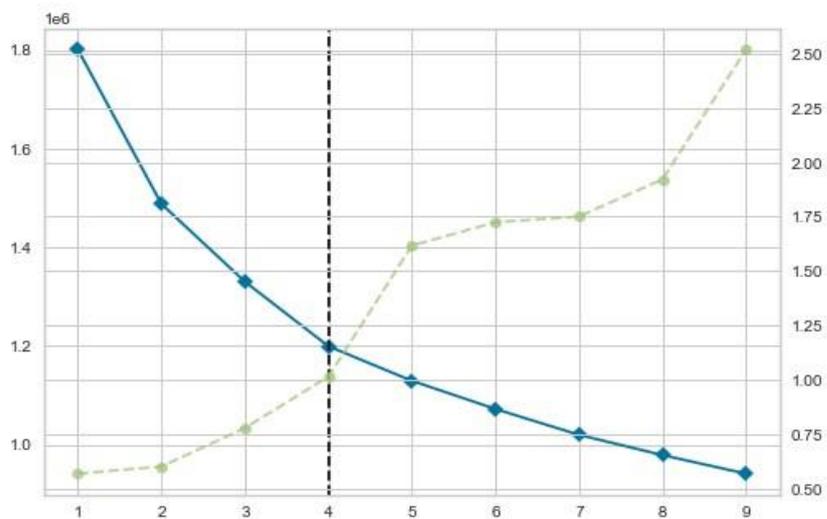
*Captura de pantalla resultado del método de Codo de Jambú.*



*Nota:* Autores (2024)

**Figura 2**

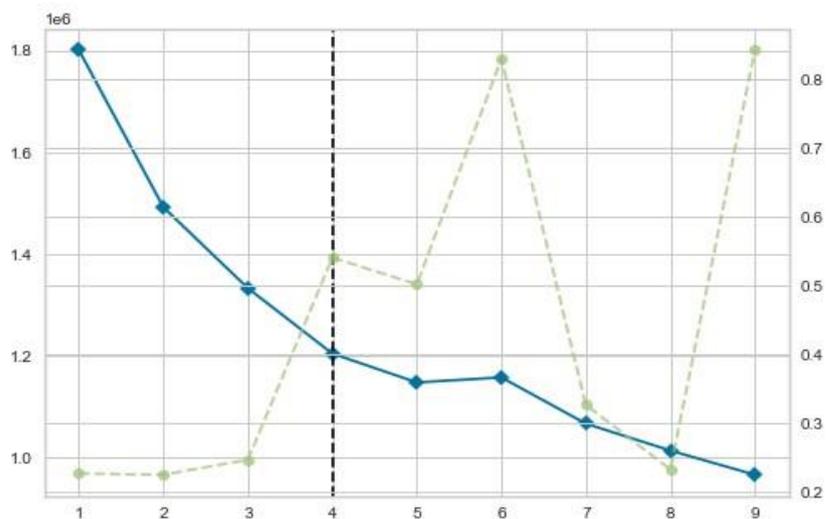
*Captura de pantalla resultado del Método de Silueta*



*Nota: Autores (2024)*

**Figura 3**

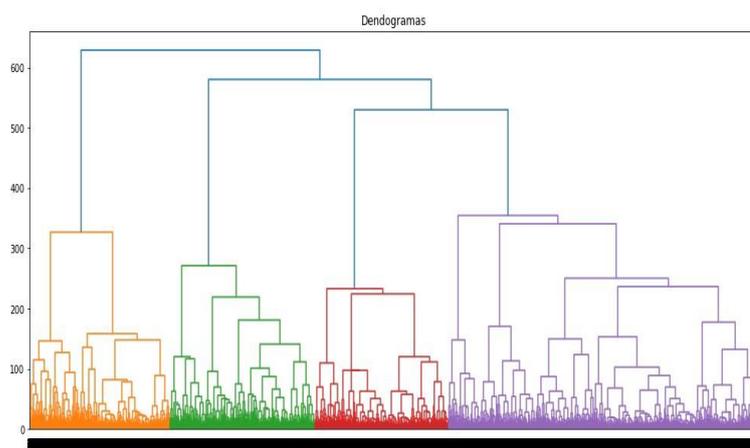
*Captura de pantalla resultado del Método de Mini k-means*



*Nota: Autores (2024)*

## Figura 4

Captura de pantalla resultado del Método de Dendograma.



Nota: Autores (2024)

En cuanto a los Clústeres dos y tres, se identifica una semejanza en las variables socioeconómicas y demográficas, así como en las preferencias académicas. Al igual que en los Clústeres cero y uno, las familias tienen una realidad económica media baja, no tienen afiliación a pueblos o nacionalidades y los padres cuentan con educación básica. Los estudiantes son dependientes de sus padres, poseen un nivel intelectual medio alto y los padres están involucrados en sus estudios. Sin embargo, en estos clústeres, se observa una mayor preocupación por parte de los padres, un cambio hacia materias más intelectuales y un mayor interés en cursar una carrera específica. Aunque mantienen un bajo uso de correos electrónicos, explican su falta de ingreso a la educación superior y muestran un nivel informático bajo.

## Discusión

La segmentación de estudiantes a través del algoritmo K-means ha revelado patrones significativos en sus perfiles, respaldando la idea de que la segmentación psicográfica es esencial para el diseño de estrategias de marketing educativo (Medina et al., 2023). Los resultados obtenidos proporcionan una visión detallada de los perfiles estudiantiles,

permitiendo una orientación más precisa de las estrategias de marketing para los Institutos Técnicos Superiores Públicos en Ecuador. La relevancia socioeconómica en los Clústeres cero y uno es notoria, donde prevalece una realidad económica media baja. La falta de afiliación a pueblos o nacionalidades específicas y la presencia de padres con educación básica se alinean con las características identificadas en el marketing educativo (Carreño Cuador, 2023). Este grupo de estudiantes se muestra totalmente dependiente de sus padres y presenta un nivel intelectual medio-alto. En cuanto a la selección de materias, se observa una aversión hacia aquellas de índole técnica, como matemáticas, física y cálculo. La baja utilización de correos electrónicos y el bajo nivel informático explican, en parte, la reticencia a ingresar a la educación superior, aspecto que concuerda con las barreras identificadas en la toma de decisiones de los futuros estudiantes (Pascual, Pla, & Sánchez, 2007; Palmeira, 2022).

En los Clústeres dos y tres, aunque persisten similitudes socioeconómicas, se destaca una mayor preocupación por parte de los padres y un cambio hacia materias más intelectuales. Aquí, los estudiantes muestran un interés específico en carreras particulares, lo que resalta la necesidad de adaptar la oferta educativa a las aspiraciones individuales de los estudiantes, como sugiere el análisis de perfiles estudiantiles en el marco teórico (Arellano, 2017; Larios-Gómez, 2021). Este cambio en la dinámica de interés y la mayor participación de los padres en la elección de carreras indica una oportunidad para las instituciones educativas de ajustar sus estrategias de marketing y oferta académica. Es importante señalar que estos hallazgos confirman la importancia de la segmentación en el marketing educativo, donde adaptar las estrategias a las características específicas de cada grupo puede mejorar la efectividad de las campañas de captación y retención de estudiantes (Medina et al., 2023; Carreño Cuador, 2023).

La persistencia en ambos grupos de un bajo uso de correos electrónicos y un nivel informático reducido destaca un área de mejora en las estrategias de marketing adoptadas por las instituciones de educación superior. La adaptación de estas estrategias para abordar las

limitaciones tecnológicas podría facilitar una mayor conexión con los estudiantes y facilitar el proceso de toma de decisiones. La discusión de los resultados también arroja luz sobre la necesidad de una diferenciación más clara en las estrategias de marketing para los Institutos Técnicos Superiores Públicos en Ecuador. Los Clústeres cero y uno demandan un enfoque que aborde sus preocupaciones socioeconómicas y académicas, mientras que los Clústeres dos y tres requieren una estrategia más especializada que capitalice su mayor interés intelectual y la participación activa de los padres en la elección de carreras.

Adicionalmente, la aplicación de técnicas de agrupamiento como el K-means en este estudio se alinea con el enfoque contemporáneo de utilizar Big Data para analizar y comprender mejor a la audiencia (Algulyev et al., 2021). La combinación de técnicas de análisis de datos masivos y marketing educativo permite una toma de decisiones más fundamentada y centrada en el estudiante. Sin embargo, es crucial reconocer las limitaciones de este estudio. La efectividad de las estrategias de marketing propuestas debe evaluarse en la práctica, considerando la dinámica cambiante del mercado y las preferencias de los estudiantes. Además, la naturaleza dinámica de las tendencias en educación y tecnología requiere un monitoreo continuo para ajustar las estrategias de manera efectiva. En esta perspectiva, la segmentación detallada de perfiles estudiantiles a través de técnicas de Big Data y análisis de clústeres ofrece una base sólida para la optimización de estrategias de marketing en Institutos Técnicos Superiores Públicos en Ecuador. La adaptación de estas estrategias para abordar las características específicas de cada grupo y la incorporación de un enfoque tecnológico más fuerte pueden mejorar la efectividad de las instituciones en la captación y retención de estudiantes en un entorno educativo cada vez más competitivo y dinámico.

## Conclusión

En conclusión, el proyecto "Exploración y Análisis de Perfiles Estudiantiles con Enfoque en Big Data para Definir el Público Objetivo en los Institutos Técnicos Superiores Públicos del Ecuador" ha logrado avances significativos.

**Segmentación Efectiva:** La aplicación de técnicas de Big Data, como el algoritmo K-means, ha permitido una segmentación efectiva de los estudiantes en cuatro clústeres, destacando similitudes socioeconómicas y demográficas. **Relevancia de la Psicografía:** Los resultados refuerzan la importancia de la psicografía en la segmentación, revelando patrones de comportamiento y preferencias que van más allá de las variables tradicionales. **Adaptación de Estrategias:** La comprensión detallada de cada clúster facilita la adaptación de estrategias de marketing, abordando necesidades específicas y maximizando la efectividad de las campañas de captación. **Desafíos Tecnológicos:** La identificación de limitaciones tecnológicas resalta la necesidad de inversiones y estrategias específicas para mejorar la infraestructura y abordar barreras de acceso a la información.

En síntesis, la integración de Big Data en el análisis de perfiles estudiantiles no solo redefine la segmentación, sino que también ofrece una visión estratégica para el marketing educativo. Estas conclusiones respaldan la relevancia y el impacto positivo de la implementación de enfoques innovadores en el ámbito educativo.

## Referencias bibliográficas

- Abualigah, L.M., et al. (2017). A novel hybridization strategy for krill herd algorithm applied to clustering techniques. *Applied Soft Computing*, 55, 536-545. <https://doi.org/10.1016/j.asoc.2017.02.015>
- Alguliyev, R.M., et al. (2020). A comprehensive survey of the harmony search algorithm in clustering applications. *Applied Sciences*, 10(4), 1237. <https://doi.org/10.3390/app10041237>
- Alguliyev, R.M., et al. (2021). Parallel batch K-means for big data clustering. *Computers & Industrial Engineering*, 160, 107292. <https://doi.org/10.1016/j.cie.2021.107292>
- Bonilla Garzón, A. (2023). Segmentación del estudiantado universitario: el caso de la Udima.

- Cañaveral Vergara, A. (2023). Propuesta de una metodología de implementación de Big data para la medición de la percepción de los estudiantes sobre la satisfacción en los procedimientos administrativos en universidades.
- Carreño Cuador, J. (2023). Marketing para la comercialización de servicios educativos. Un reto para la competitividad internacional. *Revista Cubana de Salud Pública*, 49(1).
- Ceballos Holguín, D. (2023). La segmentación no es una tendencia.
- Centeno Martín-Romero, A. (2020). Big Data. Técnicas de machine learning para la creación de modelos predictivos para empresas.
- Chavez, R. (2020, mayo 22). *ricovictor.com*. Retrieved from <https://ricovictor.com/index.php/2020/05/22/clustering-metodo-k-means-en-python/>
- Coronel, Y., & Avila, J. (2020). Aplicabilidad de la inteligencia artificial y la tecnología blockchain en el derecho contractual privado. *ALMONACID SIERRA, Derecho Privado*, 119.
- Flores López, J. C. (2021). Aplicación de Machine Learning sin supervisión. *Revista Varianza*, 20.
- Ganesha, Leu, & Seda. (2020). Diseño de un sistema de recomendación de películas basado en aprendizaje automático no supervisado. *Symmetry*, 12(185), 185. doi:sym12020185
- García, R. E. A. (2021). Sistema analítico basado en un modelo predictivo de procesamiento de datos en la big data en la educación superior.
- González, L. (2020, febrero 11). *Lidgi Gonzales Aprende todo sobre inteligencia artificial*. Retrieved from <https://lidgigonzaez.com/aprendizaje-no-supervisado-machinelearning/>
- Ikotun, A.M., et al. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622, 178-210. <https://doi.org/10.1016/j.ins.2021.10.007>
- Jain, A.K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651-666. <https://doi.org/10.1016/j.patrec.2009.09.011>
- Mamani Rodriguez, Z. (2022). Proceso de machine learning para determinar la demanda social de puestos de empleo de profesionales de TI. *Industrial Data*, 25(2), 275-300.
- MathWorks ©. (2020). *Cluster usando el modelo de mezcla gaussiana*. Retrieved from <https://la.mathworks.com/help/stats/clustering-using-gaussian-mixture-models.html>
- Mulero Merino, V. (2023). Estudio de técnicas de clustering aplicadas a una competición profesional de fútbol.
- Naranjo-Villota, D., Guaña-Moya, J., Acosta-Vargas, P., & Muirragui-Irrazábal, V. (2020). Evaluación de la accesibilidad web en institutos acreditados de educación superior del Ecuador. *Revista Espacios*, 41(04).
- Osorio-Sanabria, M. A., Amaya Fernández, F. O., & González-Zabala, M. (2020). Análisis de datos abiertos de instituciones de educación superior colombianas como apoyo a la relación Universidad-Entorno. *Entramado*, 16(1), 272-284.
- Palmeira, A. C. M. (2022). Marketing das instituições de ensino superior na captação de novos alunos.
- Pérez Suasnavas, A., Cela, K., & Hasperúe, W. (2020). Beneficios del uso de técnicas de minería de datos para extraer y analizar datos de twitter aplicados en la educación superior: una revisión sistemática de la literatura. *Teoría de la Educación. Revista Interuniversitaria Vol.32 num 2*.
- Ruiz Cruz, G. M. (2023). Antecedentes, situación actual y futuro del big-data en el sector turístico de España.
- Sánchez Guamán, J. L. (2021). *Propuesta de un modelo de inteligencia de marketing SIM para la Universidad Técnica Particular de Loja* (Bachelor's thesis).

- SENESCYT. (2020). *Educación Superior, Ciencia, Tecnología, Innovación y Saberes Ancestrales en cifras*. Quito - Ecuador: Secretaría de Educación Superior, Ciencia, Tecnología e Innovación.
- Töpfer, A., Bug, M. (2015). Segmentación de Mercado, Psicográfica y Posicionamiento de la Educación Superior.
- Universidad de Alcalá. (2020). *master-data-scientist*. Retrieved 2020, from master-datascientist: <https://www.master-data-scientist.com/diferencias-entre-master-big-data-ydata-science/>
- Valls, J., & Badiella, L. (s/f). *sct.uab.cat*. Retrieved 2020, from sct.uab.cat: <https://sct.uab.cat/estadistica/sites/sct.uab.cat.estadistica/files/ManualSAS.PDF>
- Vargas Toscanelli, H. S. C. (2023). Sistema web basado en técnicas de machine learning para predecir el desempeño de los docentes en el nivel de educación superior universitaria.
- Villarreal, J. O. G., Gómez, L. R. F., & Pineda-Ríos, W. (2021). Estimación de las principales causas de la deserción universitaria mediante el uso de técnicas de machine learning. *Aglala*, 12(2), 293-311.
- Zúñiga, F., Poveda, D. A. M., & Llerena, W. V. L. (2023). El big data y su implicación en el marketing. *Revista de Comunicación de la SEECI*, 56, 302-321.