

AÑO 30 NO. ESPECIAL 13, 2025
ENERO-JUNIO



AÑO 30 NO. ESPECIAL 13, 2025

ENERO-JUNIO



Revista Venezolana de Gerencia



UNIVERSIDAD DEL ZULIA (LUZ)
Facultad de Ciencias Económicas y Sociales
Centro de Estudios de la Empresa

ISSN 1315-9984

Esta obra está bajo una licencia de Creative Commons
Reconocimiento-NoComercial-CompartirIgual 3.0 Unported.
http://creativecommons.org/licenses/by-nc-sa/3.0/deed.es_ES

Como citar: Lara-Salazar, J. R., Morales-Morales, J., y Yee-Rendón, A.(2025). Intención Emprendedora: Segmentación de Perfiles a través de Inteligencia Artificial. *Revista Venezolana De Gerencia*, 30(Especial 13), 660-677. <https://doi.org/10.52080/rvgluz.30.especial13.42>

Universidad del Zulia (LUZ)
Revista Venezolana de Gerencia (RVG)
Año 30 No. Especial 13, 2025, 660-677
Enero-Junio
ISSN 1315-9984 / e-ISSN 2477-9423



Intención emprendedora: segmentación de perfiles a través de inteligencia artificial

Lara-Salazar, José Rosario*
Morales-Morales, Jaime**
Yee-Rendón, Arturo***

Resumen

El objetivo de este estudio es analizar la intención emprendedora mediante un enfoque cuantitativo que combina técnicas estadísticas clásicas y algoritmos de Inteligencia Artificial. A partir de datos del Global Entrepreneurship Monitor y una encuesta estructurada de 89 ítems, se evaluaron variables clave como la motivación para emprender, los valores personales, el ecosistema emprendedor, la diversidad del conocimiento y los esquemas de inversión. La validez del instrumento se comprobó mediante Análisis Factorial Exploratorio, obteniendo altos índices de confiabilidad (Alfa de Cronbach > .86) y adecuación muestral (KMO > .89). Posteriormente, se aplicó el algoritmo de K-medias técnica de aprendizaje no supervisado ampliamente utilizada por su eficiencia y escalabilidad, con el propósito de segmentar los datos en clústeres homogéneos. Para evaluar la relevancia de cada variable, se empleó la varianza intra-clúster como métrica de dispersión interna. Variables con menor varianza presentaron mayor peso en la formación de perfiles, lo cual optimiza la interpretación y precisión del modelo. Los hallazgos demuestran que esta metodología permite identificar factores determinantes en la intención emprendedora favoreciendo la reducción de dimensionalidad y aportando información clave para la toma de decisiones en entornos complejos y de alto volumen de datos. Se concluye intención emprendedora responde a una estructura multidimensional en la que convergen motivaciones económicas, sociales y personales.

Palabras clave: intención emprendedora; inteligencia artificial; clusterización; emprendimiento juvenil.

Recibido: 25.03.25

Aceptado: 03.06.25

* Universidad Autónoma de Sinaloa, Facultad de Contaduría y Administración. Miembro del Sistema Nacional de Investigadoras e Investigadores (SNII) nivel C. Email: joselarasalazar@uas.edu.mx. ORCID: <https://orcid.org/0000-0002-7174-4854>

** Universidad Autónoma de Sinaloa, Facultad de Contaduría y Administración. Miembro del Sistema Nacional de Investigadoras e Investigadores (SNII) nivel C. Email: jmorales@uas.edu.mx. ORCID: <https://orcid.org/0000-0001-7723-4137>. Autor de correspondencia.

*** Universidad Autónoma de Sinaloa, Facultad de Informática Culiacán. Miembro del Sistema Nacional de Investigadoras e Investigadores (SNII) nivel 1. Email: arturo.yee@uas.edu.mx. ORCID: <https://orcid.org/0000-0002-9052-6588>

Entrepreneurial Intention Analysis: Profile Segmentation through Artificial Intelligence

Abstract

This study aims to analyze entrepreneurial intention (EI) using a quantitative approach that combines classical statistical techniques with artificial intelligence algorithms. Based on data from the Global Entrepreneurship Monitor and a structured 89-item survey, the research evaluates key variables such as motivation to undertake, personal values, the entrepreneurial ecosystem, knowledge diversity, and investment structures. Instrument validity was assessed through Exploratory Factor Analysis, yielding high levels of reliability (Cronbach's Alpha > .86) and sampling adequacy (KMO > .89). Subsequently, the K-means algorithm—an unsupervised machine learning technique known for its computational efficiency and scalability—was applied to segment the data into homogeneous clusters. To determine the relevance of each variable, intra-cluster variance was used as a metric of internal dispersion. Variables with lower variance were considered more influential in shaping the profiles, thereby enhancing the interpretability and precision of the model. The findings demonstrate that this methodology effectively identifies key factors influencing EI, supporting dimensionality reduction and offering valuable insights for decision-making in complex, high-volume data environments. We conclude that entrepreneurial intention responds to a multidimensional structure in which economic, social and personal motivations converge.

Keywords: entrepreneurial intention; artificial intelligence; clustering; youth entrepreneurship.

1. Introducción

La intención emprendedora (IE) se ha consolidado como uno de los principales indicadores para predecir el comportamiento empresarial, al representar el punto de partida hacia la creación de nuevas iniciativas económicas. En un entorno global caracterizado por la aceleración tecnológica, la inestabilidad laboral y el aumento del autoempleo como opción profesional, comprender cómo se forma esta intención y qué factores la determinan resulta fundamental para promover el desarrollo económico sostenible y la formulación de políticas públicas eficaces.

En este contexto, el interés por

el estudio de la IE ha evolucionado progresivamente desde enfoques psicológicos y socioculturales hacia modelos analíticos que integran herramientas de procesamiento avanzado de datos. Particularmente, la inteligencia artificial (IA) ha cobrado relevancia por su capacidad para analizar grandes volúmenes de información, detectar patrones complejos y ofrecer nuevas perspectivas sobre fenómenos sociales. Entre estas herramientas, los algoritmos de aprendizaje no supervisado, como el K-medias (K-means), destacan por su eficacia para segmentar poblaciones en clústeres homogéneos, permitiendo una clasificación más precisa y significativa de los perfiles emprendedores.

Bajo este enfoque, el presente estudio tiene como objetivo general analizar la intención emprendedora mediante el uso combinado de técnicas estadísticas tradicionales y algoritmos de IA. Para ello, se adoptó una metodología cuantitativa basada en datos del Global Entrepreneurship Monitor (GEM) y en una encuesta estructurada de 89 ítems. Los objetivos específicos incluyen: (1) validar la estructura y consistencia del instrumento de medición mediante análisis factorial exploratorio (AFE), y (2) aplicar el algoritmo K-medias para identificar perfiles emprendedores y evaluar la importancia relativa de las variables a través de la varianza intra-clúster.

En consonancia con estos objetivos, se planteó la siguiente hipótesis de investigación: *la aplicación de técnicas de inteligencia artificial, específicamente el algoritmo K-medias, permite identificar perfiles diferenciados de intención emprendedora a partir de variables clave como la motivación económica, los valores sociales y las estructuras de inversión informal.*

Derivado de esta hipótesis, emergen dos preguntas de investigación que guían el presente trabajo de investigación:

¿Cuáles son las variables que presentan mayor capacidad discriminativa para la formación de perfiles de intención emprendedora mediante el algoritmo K-medias?

¿Qué características comparten los perfiles de emprendedores identificados a través de la clusterización, y cómo varían según el número de clústeres definidos?

En consecuencia, la estructura del artículo se organiza de la siguiente manera: introducción; después se presenta una revisión de la literatura sobre

la intención emprendedora y el uso de IA en estudios sociales; posteriormente se describe la metodología utilizada, incluyendo el diseño del instrumento y el proceso de análisis; además, se exponen los resultados obtenidos y, finalmente, se presentan las conclusiones.

2. Intención emprendedora: estado del arte

Entender el fenómeno del emprendimiento conlleva a un análisis a profundidad de diversos aspectos para tomar decisiones de manera más efectiva con información que brinde mejores herramientas a los emprendedores. En este tenor, se menciona que tener solamente intención que no vaya en sintonía con la acción no será suficiente para comenzar un emprendimiento, aunque tener la visión de comenzar un negocio es el precursor inevitable para hacerlo Zeynalov y Dođantan (2025).

Dicho lo anterior, es fundamental que bajo esta perspectiva se trabajen cuestiones de fomenten la intención emprendedora (IE) desde tempranas edades, para que con ello, se potencie este fenómeno como un proceso más natural en los individuos (Wang et al., 2023), por lo que, la educación emprendedora que se pueda tener desde niveles básicos de educación o inclusive desde casa, podría dar mayor crecimiento a este pensamiento en los individuos desde que son niños.

Aunando a lo anteriormente establecido, Rafiq et al. (2024) enfatizan que incluso brindar conocimientos de sustentabilidad en el área de negocios eleva la posibilidad que el individuo desde que es estudiante pueda tener buena intención emprendedora, así como estimular un gran nivel de IE sobre todo en todo lo relacionado con

los negocios verdes (hablando así de aquellas organizaciones que se enfocan fuertemente en apoyar al medio ambiente).

La intención emprendedora ha sido ampliamente estudiada en la literatura como un factor clave en la creación de nuevas empresas y el desarrollo económico. Su análisis, tradicionalmente abordado desde enfoques psicológicos y socioculturales, ha evolucionado con la incorporación de modelos cuantitativos que permiten una comprensión más precisa de sus determinantes. En este contexto, el uso de la Inteligencia Artificial (IA) representa una oportunidad innovadora para el tratamiento y análisis de datos a gran escala, facilitando la identificación de patrones y tendencias en el comportamiento emprendedor.

En el estudio de Aristizábal et al. (2024) se analizó la intención emprendedora de investigadores colombianos mediante aprendizaje automático. Se aplicaron técnicas de clasificación no supervisada, identificando tres grupos, dos con alta predisposición a emprender. La IE estuvo influenciada por la vinculación con el sector privado y la falta de apoyo institucional. Además, las normas sociales clave fueron la preferencia por el emprendimiento y su valoración social. Los hallazgos resaltaron el potencial del aprendizaje automático para comprender mejor la IE en el ámbito académico. Por otro lado, Taktak-Kallel (2025) exploró el uso de la Inteligencia Artificial (IA) en la investigación sobre intenciones emprendedoras en estudiantes y su impacto en el conocimiento del campo. Realizó una revisión de alcance de estudios recientes, identificando una tendencia hacia el uso de técnicas de clasificación en aprendizaje automático supervisado. Los hallazgos reafirmaron

la complejidad de las intenciones emprendedoras y destacaron enfoques pioneros basados en el análisis de grandes volúmenes de datos. Además, se observó una inclinación hacia la "augmentación estadística" para mejorar la precisión predictiva, evidenciando un resurgimiento del positivismo en la investigación sobre emprendimiento.

Otro estudio de Sólorzano et al. (2024) se analizó la aceptación de la IA y su impacto en las intenciones emprendedoras de estudiantes de comercio internacional en Perú y Ecuador, aplicando un enfoque cuantitativo y un diseño no experimental. A través de una encuesta en línea y el uso de los softwares Amos 24 y SmartPLS, se evaluaron los índices de ajuste del modelo y se probaron hipótesis mediante ecuaciones estructurales (SEM). Los resultados revelaron que la intención de utilizar herramientas de IA influye directamente en la intención emprendedora, mientras que factores como género y edad no mostraron un efecto moderador. Además, la motivación hedónica, el hábito y las expectativas de rendimiento fueron determinantes en la disposición de los estudiantes a adoptar estas tecnologías, destacando su papel en la formación de futuras iniciativas empresariales.

Por otro lado, Obschonka y Audretsch (2020) analizaron el impacto de la Inteligencia Artificial (IA) y el big data en el emprendimiento, un campo en el que su influencia aún es limitada. Con un enfoque exploratorio, se examinaron investigaciones previas y casos concretos para comprender cómo estas tecnologías pueden transformar tanto la teoría como la práctica emprendedora. Los resultados indicaron que la IA y el big data pueden potenciar el desarrollo de un "emprendimiento inteligente",

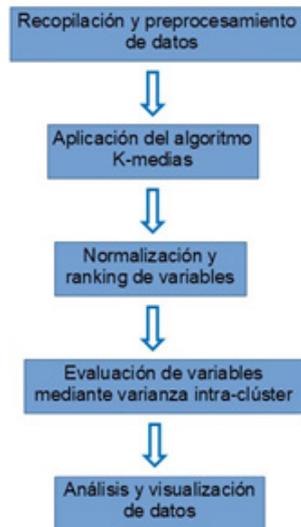
aunque también plantean desafíos éticos y metodológicos, especialmente en la relación entre la incertidumbre empresarial y la racionalidad algorítmica.

En los últimos años, diversas investigaciones han analizado la evolución del emprendimiento social. Un ejemplo es el estudio de Guatemala et al. (2023), cuyo propósito fue identificar tendencias en marketing digital aplicables a este tipo de emprendimientos en plataformas digitales, destacando el impacto de tecnologías como la Inteligencia Artificial, la realidad virtual y la realidad aumentada. Esto sugiere que, al iniciar un proyecto, los emprendedores pueden enfrentar la necesidad de integrar herramientas tecnológicas avanzadas, un desafío clave en la actualidad desde la perspectiva de la intención de emprender.

3. Consideraciones metodológicas de la investigación

La metodología empleada en esta investigación sigue un enfoque sistemático para la agrupación de datos y evaluación de variables mediante el algoritmo K-medias. Este algoritmo es ampliamente utilizado debido a su eficiencia computacional y escalabilidad, lo que lo convierte en una opción ideal para la segmentación de datos en diversas aplicaciones. La calidad del agrupamiento se garantiza mediante la minimización de la varianza intra-clúster, permitiendo una mejor interpretación y optimización de los resultados (diagrama 1).

Diagrama 1
Esquema metodológico



El procedimiento se compone de las siguientes fases: a) recopilación y

preprocesamiento de datos: se realiza la limpieza y normalización de los datos

para asegurar una correcta ejecución del algoritmo; b) aplicación del Algoritmo K-medias: se define el número de clústeres K y se ejecuta el algoritmo para agrupar los datos en subconjuntos homogéneos; c) evaluación de la importancia de las variables: se calcula la varianza intra-clúster para determinar el impacto de cada variable en la segmentación; d) normalización y ranking de variables: se invierte la relación entre la varianza y la importancia de la variable, normalizando los valores obtenidos para facilitar la interpretación de resultados; y, e) análisis y visualización de datos: se generan representaciones gráficas que destacan las variables más influyentes en la clusterización de los datos.

En este trabajo de investigación, se implementa un procedimiento que evalúa la importancia de cada variable utilizando la varianza intra-clúster, una métrica que mide la dispersión de los datos dentro de cada grupo identificado. Cuanto menor sea la varianza intra-clúster, mayor será la similitud entre los elementos del mismo clúster, lo que implica una agrupación más efectiva (Xu y Wunsch, 2005). Además, se adopta un enfoque de selección de variables basado en la relación entre la dispersión dentro de los clústeres y la contribución de cada variable al proceso de agrupamiento, lo que permite mejorar

la calidad de los resultados y optimizar la interpretación de los datos (Arthur y Vassilvitskii, 2007). Este enfoque no solo facilita la interpretación de los datos agrupados, sino que también proporciona una representación visual de la importancia relativa de cada variable dentro del conjunto de datos.

• Descripción de los datos

Para el trabajo de investigación se determinó el objeto y sujetos de estudio a través de revisión de literatura, así como la revisión de una base importante mundial sobre temas de emprendimiento como lo es el Global Entrepreneurship Monitor. Asimismo, el estudio deviene de uno previo en el cual se utilizó una metodología cuantitativa a través de datos del GEM para poder medir la intención emprendedora y cómo esta puede verse influida por otras variables.

Se realizó una encuesta de 89 ítems en la cual se abordaron las variables de estudio independientes valores personales y comportamiento, ecosistema emprendedor, diversidad del conocimiento e inversión; así como de variable dependiente se analizó la anteriormente mencionada intención emprendedora. Para una mayor comprensión de las dimensiones que se han trabajado, se muestra el siguiente cuadro 1.

Cuadro 1
Dimensiones del estudio de investigación

Variable	Dimensiones
Intención emprendedora	Tradicón familiar; generar mayores ingresos; desempleo/condiciones laborales; hacer del mundo un lugar mejor
Valores personales y comportamiento	Valores individuales; valores colectivos
Ecosistema emprendedor	Educación; servicios de incubación; programas de apoyo
Diversidad de conocimiento	Diversidad de educación; diversidad de experiencias; diversidad de información
Inversión	Inversión formal; inversión informal; estructuras de inversión informal

El tratamiento de los datos se llevó a cabo el SPSS versión 29, con un Análisis Factorial Exploratorio (AFE) a través de un análisis de fiabilidad mediante la herramienta de Alfa de Cronbach así como una reducción de dimensiones para obtener el Kaiser-Meyer Olkin (KMO) lo que permitió medir el grado de predicción de una variable sobre las demás dentro del estudio. Posteriormente, se realizó una matriz de componentes rotados, para ver la alineación que podrían tener entre todos los indicadores de cada una de las variables; acción que permite observar que matemáticamente los datos recabados a través del instrumento son congruentes con la literatura que se ha trabajado.

• **Análisis factorial exploratorio**

La técnica estadística Análisis Factorial Exploratorio tiene como propósito explorar el conjunto de variables latentes o factores comunes que explican las respuestas a los ítems de un test (Lloret-Segura et al., 2017), en este contexto, el propósito es identificar la estructura subyacente de un conjunto de datos mediante el análisis de las interrelaciones o correlaciones entre sus

variables (Pizarro y Martínez, 2020); de manera que exista una fuerte asociación entre las variables (López-Aguado y Gutiérrez-Provecho, 2019).

Para realizar un AFE, es fundamental tomar decisiones metodológicas basadas en evidencia, ya que existen múltiples opciones de análisis, pero algunas son más adecuadas que otras, según lo han señalado diversos estudios (Lloret-Segura et al., 2017 y Watkins, 2018). A continuación, se describe la ecuación (1) matemática del AFE.

$$X_1 = V_{1(1)}F_{(1)} + V_{1(2)}F_{(2)} + \dots + V_{1(m)}F_{(m)} + e_1$$

$$X_2 = V_{2(1)}F_{(1)} + V_{2(2)}F_{(2)} + \dots + V_{2(m)}F_{(m)} + e_2$$

(Ecuación 1)

$$X_p = V_{p(1)}F_{(1)} + V_{p(2)}F_{(2)} + \dots + V_{p(m)}F_{(m)} + e_p$$

(Ecuación 2)

Donde X_i , F_i y e_j , contiene la puntuación de una persona en el ítem X_i , el factor común F_i , y el vector específico e_j , m : número de factores comunes, p : número de ítems, F : factor común de $V_j(i)$ peso del factor común i ésimo asociado a la j ésima variable observada o ítem, $i = 1, 2, \dots, m$; $j = 1, 2, \dots, p$; e_j : factor común único, $j = 1, 2, \dots, p$ (cuadro 2).

Cuadro 2
Ficha técnica de la investigación

Propuesta experimental con técnicas IA	
Instrumento	Encuesta
No. ítems	89
Tipo de ítems	Numéricos
Medición	Ordinal (o intervalos)
Técnica de IA	K-medias
Software de tratamiento de información	R, <i>flexclust</i> y <i>FeatureImpCluster</i>
Sistema operativo	Ubuntu 16.04

Así pues, como fue mencionado anteriormente, se hizo un estudio cuantitativo con el AFE para medir

la fiabilidad del estudio arrojando los siguientes datos (tabla 1):

Tabla 1
Alfa de Cronbach variables del estudio

Variable independiente motivación a emprender	
Tradición familiar	.929
Generar mayores ingresos	.971
Desempleo/condiciones laborales	.963
Hacer del mundo un lugar mejor	.974
Variable independiente valores personales y comportamiento	
Valores individuales	.900
Valores colectivos	.880
Variable independiente ecosistema emprendedor	
Educación	.971
Servicios de incubación	.961
Programas de apoyo	.949
Variable independiente diversidad de conocimiento	
Diversidad de educación	.965
Diversidad de experiencias	.910
Diversidad de información	.925
Variable independiente inversión	
Inversión formal	.866
Inversión informal	.890
Estructuras inversión informal	.930
Variable dependiente intención emprendedora	
Intención emprendedora	.951

Como se puede observar en la tabla 2, se pudo obtener la fiabilidad y consistencia de cada una de las dimensiones (o ítems), ya que superaron .800, lo que permite avanzar con la certeza que la información que se obtendrá será de un nivel de validez alto.

Por otro lado, también se realizó el KMO, véase cuadro 4 entre las dimensiones y sus ítems para sobre ello observar el nivel de predicción que puedan existir entre ellas para la formación de la variable

Tabla 2
KMO del instrumento

Variable independiente: Motivación a emprender	.972
Variable independiente: Valores personales y comportamiento	.928
Variable independiente: Ecosistema emprendedor	.965
Variable independiente: Diversidad de conocimiento	.947
Variable independiente: Inversión	.911
Variable dependiente: Intención emprendedora	.897

Así como fue hecho el análisis anterior, se puede observar todas las variables independientes obtuvieron un KMO superior al .900 que al igual que en la medición del Alfa de Cronbach, se califica como excelente; mientras que en la variable dependiente este fue de .897 lo cual es muy bueno casi llegando al nivel excelente, por lo cual junto el análisis anterior, se considera así que el AFE de esta investigación ha sido satisfactorio

- **Descripción de los métodos basado en inteligencia artificial**

El análisis de datos en grandes volúmenes requiere de técnicas avanzadas que permitan la identificación de patrones y estructuras subyacentes. En este contexto, la clusterización se ha convertido en una herramienta fundamental dentro del aprendizaje automático y la estadística, permitiendo la agrupación de datos en conjuntos homogéneos llamados clústeres (Jain et al., 1999; Xu y Wunsch, 2005). Esta técnica se ha aplicado ampliamente en minería de datos y *big data*, facilitando la organización de la información y el descubrimiento de conocimientos ocultos en conjuntos de datos masivos (Aggarwal y Reddy, 2013).

La presente metodología describe un enfoque basado en la técnica de k-medias para la agrupación de datos y la selección de variables más relevantes en la clasificación de los mismos. El método de k-medias es ampliamente utilizado debido a su eficiencia computacional y su capacidad de escalabilidad, lo que lo convierte en una opción ideal para la agrupación de datos en diversas aplicaciones, desde el

análisis de clientes hasta la biomedicina (MacQueen, 1967; Han et al., 2022).

Además, como parte clave de esta metodología, se propone determinar la importancia de cada variable a partir de la varianza intra-clúster, la cual cuantifica el grado de dispersión de los datos dentro de cada grupo generado. Una menor varianza intraclúster indica una mayor similitud entre los elementos del mismo clúster, lo que favorece una agrupación más precisa y homogénea (Xu y Wunsch, 2005). Asimismo, se emplea un enfoque de selección de variables que considera la relación entre la dispersión intra-clúster y el impacto de cada variable en el proceso de agrupamiento, permitiendo así mejorar la calidad de la agrupación y optimizar la interpretación de los resultados (Arthur y Vassilvitskii, 2007). Además de mejorar la organización de los datos, esta estrategia facilita la visualización de la importancia relativa de cada variable dentro del conjunto analizado, favoreciendo una mejor toma de decisiones basada en la estructura subyacente de los datos.

La combinación de clusterización y análisis de relevancia de variables permite mejorar la precisión de los modelos predictivos, reducir la dimensionalidad de los datos y proporcionar información valiosa para la toma de decisiones en entornos de alto volumen de información (Han et al., 2022).

- **Descripción del algoritmo de K-medias**

La clusterización es una técnica de aprendizaje no supervisado que permite agrupar datos sin la necesidad de etiquetas previas. Su principal objetivo es organizar los datos en grupos o clústeres, de manera que los elementos dentro de un mismo grupo sean más

similares entre sí en comparación con aquellos pertenecientes a otros grupos (Ikotun et al., 2023).

Uno de los métodos más utilizados en este campo es K-medias, el cual busca dividir un conjunto de datos en K grupos predefinidos. La agrupación se basa en minimizar la distancia entre los puntos de datos y los centroides de los clústeres a los que son asignados (Gribel y Vidal, 2019; Fränti y Sieranoja, 2019). A diferencia de otros algoritmos de agrupamiento, K-medias se caracteriza por su eficiencia computacional y capacidad de escalabilidad, lo que lo hace adecuado para grandes volúmenes de datos (Bagirov et al., 2016).

El proceso de agrupamiento en K-medias se basa en la repetición de dos pasos fundamentales: la asignación de cada observación al centroide más cercano y la actualización de los centroides mediante el cálculo de la media de los puntos asignados a cada clúster. Este procedimiento continúa hasta que los centroides dejan de cambiar significativamente o se alcanza un número máximo de iteraciones, garantizando así la convergencia del algoritmo (Bagirov et al., 2016).

Una de las ventajas clave de K-medias es su eficiencia computacional, lo que lo hace adecuado para trabajar con grandes volúmenes de datos. Sin embargo, su desempeño depende en gran medida de la elección del número de clústeres (K), que debe determinarse utilizando criterios como el método del codo o el índice de silueta (Schubert, 2023; Rousseeuw, 1987). La correcta determinación del número de clústeres es fundamental para evitar una segmentación deficiente que pueda afectar la interpretabilidad y utilidad de los resultados.

El procedimiento detallado del

algoritmo 1 K-medias se describe a continuación, donde se presentan los pasos necesarios para su implementación y convergencia. Paso 1: selección de K (número de clústeres); paso 2: inicialización de los Centroides; paso 3: asignación de cada punto al clúster más cercano; paso 4: cálculo de nuevos centroides; y, paso 5: repetición del proceso (se repiten los pasos 3 y 4 hasta que los centroides dejan de cambiar significativamente o se alcanza un número máximo de iteraciones.)

En el paso 1, Se define el número de clústeres en los que se desea dividir el conjunto de datos (hiperparámetros). Una mala elección de K puede generar agrupamientos de baja calidad. En el paso 2, Se seleccionan K puntos iniciales al azar dentro del conjunto de datos. Estos puntos servirán como centros iniciales de cada clúster. En el paso 3, para cada punto del conjunto de datos, se calcula la distancia euclidiana a cada centroide. Se asigna el punto al clúster cuyo centroide esté más cerca. En el paso 4, para cada clúster, se calcula el nuevo centroide como la media aritmética de todos los puntos asignados a ese clúster. Véase fórmula (2) para el nuevo centroide está definida como:

$$C_j = \frac{1}{N_j} \sum_{i=1}^{N_j} x_i$$

(Ecuación 3)

Donde C_j es el nuevo centroide del clúster j , N_j es el número de puntos en ese clúster y x_i son los puntos asignados a dicho clúster. Paso 5, Se repiten los pasos 3 y 4 hasta que los centroides dejan de cambiar significativamente o se alcanza un número máximo de iteraciones.

• Varianza Intra-clúster en la selección de variables

La varianza intra-clúster es una métrica fundamental en la evaluación de la calidad de los grupos generados por el algoritmo K-medias (Jain, 2010; Oti et al., 2020), ya que mide la dispersión de los datos dentro de cada clúster en relación con su centroide. Un valor bajo de varianza intra-clúster sugiere que los datos están bien agrupados y que la segmentación realizada es efectiva, dado que los elementos dentro de un mismo clúster presentan un alto grado de similitud. Por el contrario, una varianza intra-clúster elevada indica una distribución dispersa de los datos dentro del clúster, lo que puede sugerir la presencia de grupos mal definidos o la necesidad de ajustar el número de clústeres (K).

En este estudio, la varianza intra-clúster se utiliza como criterio para evaluar la importancia relativa de cada variable en la segmentación de los datos. La minimización de la varianza intra-clúster es una estrategia clave para mejorar la compactación de los datos dentro de los clústeres y, por lo tanto, optimizar la segmentación (Ding y He, 2004). Además, la varianza intra-clúster juega un papel esencial en la interpretación de modelos de agrupación, ya que su reducción permite mejorar la precisión de los resultados y evitar agrupamientos poco representativos. En estudios recientes, se ha demostrado que la selección de variables basada en la contribución a la varianza intra-clúster es una estrategia efectiva para mejorar la calidad del agrupamiento, especialmente en datos de alta dimensión.

Este enfoque permite la selección de características clave para futuros análisis exploratorios y la optimización

de modelos predictivos, ya que la identificación de las variables más influyentes en la agrupación de los datos facilita la reducción de dimensionalidad y mejora la interpretabilidad de los resultados. Además, la metodología propuesta puede aplicarse en diversos contextos, como la segmentación de clientes, la identificación de patrones en datos biomédicos y la detección de factores críticos en procesos industriales, donde la caracterización precisa de los grupos es fundamental para la toma de decisiones basada en datos.

Matemáticamente, la varianza intra-clúster se expresa en la fórmula (3):

$$V_{intra_clase} = \sum_{j=1}^K \sum_{i \in C_j} d(x_i, C_j)^2$$

(Ecuación 4)

Donde:

- V_{intra_clase} es la suma de los cuadrados de las distancias intra-clúster.
- K es el número de clústeres.
- C_j es el centroide del clúster j .
- x_i representa cada punto dentro del clúster j .
- $d(x_i, C_j)$ es la distancia entre el punto x_i y su centroide.

Para obtener la importancia relativa de cada variable, se invierte esta métrica y se normaliza, garantizando que la variable con menor varianza intra-clúster tenga el mayor peso dentro del análisis.

4. Intención emprendedora mediante un enfoque cuantitativo: análisis de los resultados

En esta sección se detalla la fase experimental de la investigación y se exponen los resultados obtenidos a partir de la aplicación de modelos

computacionales bajo la metodología propuesta. El procesamiento de datos mediante técnicas de Inteligencia Artificial se llevó a cabo a partir de una encuesta estructurada de 89 ítems, diseñada con base en dimensiones teóricas previamente validadas. Los resultados fueron analizados empleando algoritmos de clusterización no supervisada, específicamente el algoritmo K-medias, que ha demostrado ser eficaz para segmentar poblaciones en conjuntos homogéneos a partir de patrones latentes en los datos.

Los experimentos se ejecutaron en un entorno computacional configurado con las siguientes especificaciones técnicas: procesador: Intel Xeon W-2133 (6 núcleos, 12 hilos); memoria RAM: 16GB DDR4; sistema operativo: Ubuntu 16.04 LTS; lenguaje computacional: R versión 3.6.3; y, biblioteca principal: stats, función kmeans.

Se realizaron múltiples corridas del algoritmo K-medias con distintos valores de K (3, 5, 7 y 9) para identificar la segmentación más adecuada de los perfiles de intención emprendedora:

- K = 3: Punto de partida que ofrece una segmentación general (necesidad, oportunidad, impacto social), pero simplifica en exceso la diversidad del fenómeno.
- K = 5: Aporta mayor detalle, revelando perfiles intermedios como emprendedores institucionalizados y académicos.
- K = 7: Considerado óptimo según el método del codo, equilibra precisión y claridad, diferenciando

perfiles por motivaciones y factores estructurales, útil para diseñar intervenciones específicas.

- K = 9: Aunque permitió observar más clústeres, redujo la claridad e incrementó la superposición, afectando la utilidad práctica para políticas públicas.

En cada ejecución del algoritmo se calculó la varianza intra-clúster por variable, generando una métrica de homogeneidad interna. Estos valores fueron invertidos (1/varianza), normalizados y ajustados a una escala unitaria (suma igual a 1), permitiendo estimar la importancia relativa de cada variable en la formación de los clústeres. Esta estrategia facilitó la identificación de dimensiones clave —como motivación social, diversidad de conocimiento y acceso a inversión formal— y su visualización mediante gráficos y rankings.

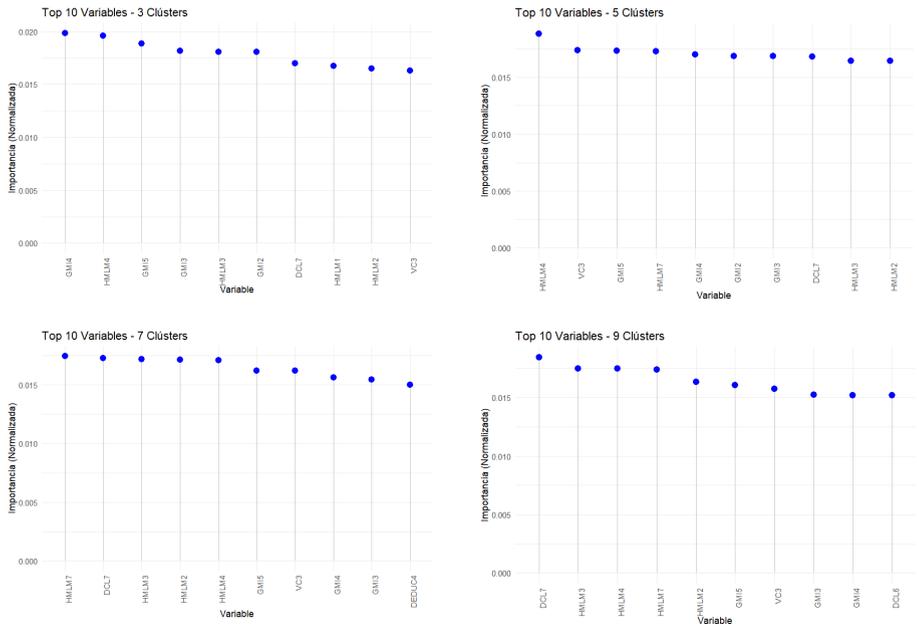
Para facilitar una interpretación clara y focalizada, se visualizaron únicamente las 10 variables con mayor capacidad discriminativa en cada configuración de clústeres. La selección se basó en los valores normalizados de la varianza intra-clúster invertida, priorizando las variables que más contribuyeron a la segmentación. Esta estrategia optimiza la claridad del análisis, evita la sobrecarga informativa y permite identificar con mayor precisión los factores clave de cada grupo (tabla 1 e ilustración 2, que muestran la relevancia de las dimensiones más influyentes).

Tabla 1
Resultados del clúster 3

Código	Variable	Importancia
Resultado clúster 3		
GMI4	Generar mayores ingresos (ítem 4)	0.02
HMLM4	Hacer del mundo un lugar mejor (ítem 4)	0.0198
GMI5	Generar mayores ingresos (ítem 5)	0.0193
GMI3	Generar mayores ingresos (ítem 3)	0.0183
HMLM3	Hacer del mundo un lugar mejor (ítem 3)	0.0181
GMI2	Generar mayores ingresos (ítem 2)	0.0181
DCL7	Desempleo/condiciones laborales (ítem 7)	0.0175
HMLM1	Hacer del mundo un lugar mejor (ítem 1)	0.0173
HMLM2	Hacer del mundo un lugar mejor (ítem 2)	0.0172
VC3	Valores colectivos (ítem 3)	0.017
Resultado clúster 5		
HMLM4	Hacer del mundo un lugar mejor (ítem 4)	0.0182
CV3	Valores colectivos (ítem 3)	0.0175
GMI5	Generar mayores ingresos (ítem 5)	0.0173
HMLM7	Hacer del mundo un lugar mejor (ítem 7)	0.0172
GMI4	Generar mayores ingresos (ítem 4)	0.017
GMI2	Generar mayores ingresos (ítem 2)	0.017
GMI3	Generar mayores ingresos (ítem 3)	0.0169
DCL7	Desempleo/condiciones laborales (ítem 7)	0.0166
HMLM3	Hacer del mundo un lugar mejor (ítem 3)	0.0165
HMLM2	Hacer del mundo un lugar mejor (ítem 2)	0.0164
Resultado clúster 7		
HMLM7	Hacer del mundo un lugar mejor (ítem 7)	0.018
DCL7	Desempleo/condiciones laborales (ítem 7)	0.018
HMLM3	Hacer del mundo un lugar mejor (ítem 3)	0.0179
HMLM2	Hacer del mundo un lugar mejor (ítem 2)	0.0179
HMLM4	Hacer del mundo un lugar mejor (ítem 4)	0.0178
GMI5	Generar mayores ingresos (ítem 5)	0.0172
VC3	Valores colectivos (ítem 3)	0.0165
GMI4	Generar mayores ingresos (ítem 4)	0.0161
GMI3	Generar mayores ingresos (ítem 3)	0.016
DEDUC4	Diversidad de educación (ítem 4)	0.0151
Resultado clúster 9		
DCL7	Desempleo/condiciones laborales (ítem 7)	0.0177
HMLM3	Hacer del mundo un lugar mejor (ítem 3)	0.0174
HMLM4	Hacer del mundo un lugar mejor (ítem 4)	0.0174
HMLM7	Hacer del mundo un lugar mejor (ítem 7)	0.0173
HMLM2	Hacer del mundo un lugar mejor (ítem 2)	0.0166
GMI5	Generar mayores ingresos (ítem 5)	0.0163
VC3	Valores colectivos (ítem 3)	0.016
GMI3	Generar mayores ingresos (ítem 3)	0.0154
GMI4	Generar mayores ingresos (ítem 4)	0.0153
DCL5	Desempleo/condiciones laborales (ítem 5)	0.0153

Ilustración 2

Dimensiones más importantes extraído del entorno computacional configurado



El análisis de la frecuencia de aparición de variables con mayor importancia relativa en las configuraciones de clústeres ($K = 3, 5, 7$ y 9) reveló cinco variables constantes: GM14, GM13, HMLM4, HMLM3 y DCL7. Su recurrencia sugiere un alto poder discriminativo en la segmentación de perfiles de intención emprendedora, independientemente del número de clústeres.

- GM14 y GM13 reflejan motivaciones económicas y laborales, alineadas con teorías del emprendimiento por oportunidad y por necesidad.
- HMLM4 y HMLM3 representan dimensiones axiológicas y sociales, indicando una orientación hacia

el impacto positivo más allá del beneficio económico.

- DCL7, asociada a inversión informal, señala la importancia del acceso a financiamiento no tradicional, posiblemente vinculado al capital social y redes de apoyo.

La consistencia de estas variables valida su relevancia transversal y las posiciona como indicadores clave para investigaciones futuras, diseño de políticas públicas y programas de apoyo diferenciados.

En relación con las variables analizadas, aunque la IE es un precursor esencial para iniciar un negocio (Zeynalov y Doğantan, 2025), debe ir acompañada de acciones concretas

y de un entorno propicio desde etapas tempranas (Wang et al., 2023). Esto coincide con la literatura reciente que resalta el papel de la educación —formal y no formal— en el desarrollo de una mentalidad emprendedora sostenible.

De acuerdo con Rafiq et al. (2024), integrar la sostenibilidad en la educación empresarial aumenta la IE. Nuestros resultados apoyan esta afirmación, al mostrar que la conciencia hacia la sostenibilidad no solo incrementa la intención de emprender, sino que también promueve proyectos con impacto social y ambiental.

Mientras Aristizábal et al. (2024) destacan el papel del sector privado y del apoyo institucional, nuestros hallazgos muestran una mayor relevancia de la motivación intrínseca, como la sostenibilidad, por encima de factores externos. Esto sugiere diferencias contextuales, como el tipo de formación o la zona geográfica, que deben explorarse en futuras investigaciones.

Asimismo, aunque Sólorzano et al. (2024) sostienen que variables como edad y género no influyen en la intención emprendedora, en este estudio se observó un efecto indirecto: los jóvenes, familiarizados con tecnologías digitales, mostraron mayor disposición a emprender con herramientas de inteligencia artificial.

Asimismo, Obschonka y Audretsch (2020) y Guatemala et al. (2023) confirman que el emprendimiento actual exige habilidades tecnológicas avanzadas. Esto representa una oportunidad para quienes cuentan con dichas competencias, pero también un reto para quienes no las poseen, lo que subraya la necesidad de fortalecer la formación tecnológica de los futuros emprendedores.

Finalmente, el uso de inteligencia artificial en el análisis de datos, como

lo demuestran estudios recientes (Aristizábal et al., 2024; Taktak-Kallel, 2025; Sólorzano et al., 2024), está creciendo en las ciencias sociales. Las técnicas de aprendizaje automático, como K-medias, permiten una segmentación más precisa que los métodos convencionales, captando matices relevantes. Por ejemplo, Morales-Morales et al. (2023) aplicaron K-medias para identificar los principales determinantes de la calidad del servicio en el sector hotelero, lo que demuestra la aplicabilidad de estas técnicas en diversos contextos sociales para una comprensión más profunda y una mejor toma de decisiones.

5. Conclusiones

La intención emprendedora responde a una estructura multidimensional en la que convergen motivaciones económicas, sociales y personales, lo cual refuerza la necesidad de enfoques diferenciados en la formación y el acompañamiento emprendedor.

La segmentación de perfiles mediante técnicas de análisis factorial y algoritmos de agrupamiento no supervisado permitió identificar patrones estables y significativos que amplían la comprensión del fenómeno más allá de las aproximaciones tradicionales.

Se destaca la creciente relevancia de factores intrínsecos como la sostenibilidad y el impacto social, por encima de estímulos externos, así como la influencia indirecta de competencias tecnológicas en la disposición a emprender. Estos hallazgos ofrecen insumos valiosos para el diseño de políticas y programas educativos orientados a contextos diversos, y demuestran el potencial de la inteligencia

artificial como herramienta analítica para abordar problemáticas complejas en el ámbito social, aportando así a una toma de decisiones más informada, integral y contextualizada.

Referencias

- Aggarwal, C. C., & Reddy, C. K. (2014). *Data Clustering Algorithms and Applications*. CRC.
- Aristizábal, J. M., Tarapuez, E., & Astudillo, C. A. (2024). Predicting entrepreneurial intention in Colombian academics: a machine learning approach. *Journal of Entrepreneurship in Emerging Economies*, 17(2), 260-288. <https://doi.org/10.1108/JEEE-04-2023-0141>
- Arthur, D., & Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. En S. SIAM Activity Group on Discrete Mathematics, *Proceeding* (págs. 1027 - 1035). Society for Industrial and Applied Mathematics. <https://doi.org/10.5555/1283383>
- Bagirov, A. M., Taheri, S., & Ugon, J. (2016). Nonsmooth DC programming approach to the minimum sum-of-squares clustering problems. *Pattern Recognition*, 53, 12-24. <https://doi.org/10.1016/j.patcog.2015.11.011>
- Ding, C., & He, X. (2004). K-means clustering via principal component analysis. *ICML '04: Proceedings of the twenty-first international conference on Machine learning*, 29. <https://doi.org/10.1145/1015330.1015408>
- Fränti, P., & Sieranoja, S. (2019). How much can k-means be improved by using better initialization and repeats? *Pattern Recognition*, 93, 95-112. <https://doi.org/10.1016/j.patcog.2019.04.014>
- Gribel, D., & Vidal, T. (2019). HG-means: A scalable hybrid genetic algorithm for minimum sum-of-squares clustering. *Pattern Recognition*, 88, 569-583. <https://doi.org/10.1016/j.patcog.2018.12.022>
- Guatemala, M. A., Martínez, P. G., y Martínez de Escobar, F. A. (2023). Tendencias de Marketing Digital para Emprendedores Sociales Basados en Plataformas Digitales. *Revista de Educación y Comunicación en la Sociedad del Conocimiento*, 23(1), 154-178. <https://doi.org/10.30827/eticanet.v23i1.27709>
- Han, J., Kamber, M., & Pei, J. (2022). *Data mining: concepts and techniques*. Morgan Kaufmann.
- Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (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.2022.11.139>
- Jain, A. K. (2010). Data clustering: 50 years beyond k-means. *Pattern Recognition*, 31(8), 651-666. <https://doi.org/10.1016/j.patrec.2009.09.011>
- Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM Computing Surveys (CSUR)*, 31(3), 264 - 323. <https://doi.org/10.1145/331499.331504>
- Lloret, S., Ferreres, A., Hernández, A., & Tomás, I. (2017). The exploratory factor analysis of items: guided analysis based on empirical data and software. *Anales de Psicología*, 33(2), 417-432. <https://doi.org/10.6018/analesps.33.2.270211>
- López-Aguado, M., & Gutiérrez-Provecho, L. (2019). Cómo realizar e interpretar un análisis factorial exploratorio utilizando SPSS. *REIRE Revista d'Innovació y Recerca*

en *Educació*, 1-14. <https://doi.org/10.1344/reire2019.12.227057>

MacQueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, 1(14), 281-297.

Morales-Morales, J., Valdez-Acosta, N., Morales-Morales, J. R., y Arturo, Y.-R. (2023). Calidad en el servicio hotelero: Propuesta de metodología experimental. *Revista Venezolana De Gerencia*, 28(104), 1717-1734. <https://doi.org/10.52080/rvgluz.28.104.21>

Obschonka, M., & Audretsch, D. B. (2020). Artificial intelligence and big data in entrepreneurship: a new era has begun. *Small Business Economics*, 55(3), 529-539. <https://doi.org/10.1007/s11187-019-00202-4>

Oti, E. U., Unyeagu, S., Nwankwo, C. H., Alvan, W. K., & Osuji, G. A. (2020). New K-means clustering methods that minimize the total intra-cluster variance. *African Journal of Mathematics and Statistics Studies*, 3(5), 42-54.

Pizarro, R. K., y Martínez, M. O. (2020). Análisis factorial exploratorio mediante el uso de las medidas de adecuación muestral kmo y esfericidad de bartlett para determinar factores principales. *Journal of Science and Research*, V(1), 903-924. <https://doi.org/10.5281/zenodo.4453224>

Rafiq, M., Yang, J., & Bashar, S. (2024). Impact of personality traits and sustainability education on green entrepreneurship behavior of university students: Mediating role of green entrepreneurial intention. *Journal of Global Entrepreneurship Research*, 14(14). <https://doi.org/10.1007/s40497-024-00384-6>

<https://doi.org/10.1007/s40497-024-00384-6>

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and applied mathematics*, 20, 53-65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)

Schubert, E. (2023). Stop using the elbow criterion for k-means and how to choose the number of clusters instead. *ACM SIGKDD Exploration Newsletter*, 25(1), 36-42. <https://doi.org/10.1145/3606274.3606278>

Sólorzano, S. S., Pizarro, R. J., Díaz, C. J., Arias, M. J., Zamora, C. M., Lozzelli, V. M., & Montes, N. (2024). Acceptance of artificial intelligence and its effect on entrepreneurial intention in foreign trade students: a mirror analysis. *Journal of Innovation and Entrepreneurship*, 13(59). <https://doi.org/10.1186/s13731-024-00412-5>

Taktak-Kallel, I. (2025). Advancing Entrepreneurship Knowledge with Artificial Intelligence (AI) Methods: Insights from Scoping Review of Emerging AI-Powered Research on Student Entrepreneurial Intentions. *Journal of Innovation Economics & Management*, 175-XXXIX. <https://doi.org/10.3917/jie.pr1.0175>

Wang, X.-H., You, X., Wang, H.-P., & Wang, B. (2023). The Effect of Entrepreneurship Education on Entrepreneurial Intention: Mediation of Entrepreneurial Self-Efficacy and Moderating Model of Psychological Capital. *Sustainability*, 15(3), 2562. <https://doi.org/10.3390/su15032562>

Watkins, M. W. (2018). Exploratory Factor Analysis: A Guide to Best Practice. *Journal of Black Psychology*, 44(3), 219-246. <https://doi.org/10.1177/0095798418771807>

Xu, R., & Wunsch, D. (2005). Survey

of clustering algorithms. *IEEE Transactions on neural networks*, 16(3), 645-678. <https://doi.org/10.1109/TNN.2005.845141>

Zeynalov, S., & Dođantan, E. (2025). The Effect of Digital Literacy and

Entrepreneurship Education on Digital Entrepreneurship Intention: The Mediating Role of Personal Innovativeness. *Technology, Knowledge and Learning*, 1-18. <https://doi.org/10.1007/s10758-025-09821-1>