



# Modelo de agregación para medir la influencia del machine learning y modelos predictivos de empleabilidad pertinente con la carrera universitaria

## Aggregation model to measure the influence of machine learning and predictive models of employability relevant to the college career

Silvio Amable Machuca Vivar<sup>1</sup>, Carlos Roberto Sampedro Guamán<sup>2</sup>, María Fernanda Pacheco Carrera<sup>3</sup> and Diego Paúl Palma Rivera<sup>4</sup>

<sup>1</sup> Universidad Regional autónoma de Los Andes, Santo Domingo. Ecuador. E-mail: [us.silviomachuca@uniandes.edu.ec](mailto:us.silviomachuca@uniandes.edu.ec)

<sup>2</sup> Instituto Superior Tecnológico Tsáchila. Ecuador. E-mail: [carlossampedro@tsachila.edu.ec](mailto:carlossampedro@tsachila.edu.ec)

<sup>3</sup> Instituto Superior Tecnológico Tsáchila. Ecuador. E-mail: [mariapacheco@tsachila.edu.ec](mailto:mariapacheco@tsachila.edu.ec)

<sup>4</sup> Universidad Regional autónoma de Los Andes, Santo Domingo. Ecuador. E-mail: [us.diegopalma@uniandes.edu.ec](mailto:us.diegopalma@uniandes.edu.ec)

**Resumen.** Las Instituciones de Educación Superior (IES) forman a los jóvenes profesionales en los aspectos técnico, teórico, competencias y habilidades en áreas específicas del conocimiento, para un mercado de trabajo con demandas muy cambiantes a las cuales a veces sus graduados no logran satisfacer las expectativas, es por ello la preocupación constante de la IES por identificar las causas asociadas con la tasa de empleabilidad de sus graduados. Por ello las IES deben prestar constante atención a la información que obtienen en su proceso de seguimiento a graduados, construir un modelo de optimización del proceso formativo para calcular y estimar los escenarios de producción de las IES en función de la demanda del mercado laboral. Los sistemas de seguimiento a graduados deben contar con técnicas de Minería de Datos que permitan identificar patrones de empleabilidad de los graduados para la toma de decisiones en procura de una mayor tasa de inserción laboral y una formación de sus profesionales que responda a la demanda laboral, es por ellos que se planteó como objetivo la creación y análisis de tres modelos de aprendizaje predictivo basado en algoritmos de machine learning, que permita predecir el resultado de una variable objetivo en función de varias variables predictoras, el caso estudiado es la Universidad Regional Autónoma de los Andes, en Ecuador. La presente investigación tiene como objetivo desarrollar un modelo de agregación para medir la influencia del machine learning y modelos predictivos de empleabilidad pertinente con la carrera universitaria. El modelo predictivo usó algoritmos de aprendizaje no supervisado, como: regresión lineal, árbol de decisiones y SVM generando los modelos con una precisión alta y con los ajustes necesarios demostrando su aplicabilidad.

**Palabras Claves:** Machine learning, seguimiento a graduados, educación superior, empleabilidad de los graduados, modelos predictivos, modelo neutrosófico.

**Abstract.** Higher Education Institutions (HEIs) train young professionals in technical and theoretical aspects, as well as competencies and skills in specific areas of knowledge, for a job market with very changing demands, which sometimes their graduates fail to meet expectations. This is why HEIs are constantly concerned about identifying the causes associated with the employability rate of their graduates. For this reason, HEIs must pay constant attention to the information they obtain in their graduate monitoring process, building a model to optimize the training process to calculate and estimate the HEI production scenarios based on the demand of the labor market. Graduate tracking systems must have Data Mining techniques that allow identifying graduate employability patterns for decision making in pursuit of a higher job insertion rate and a training of their professionals that responds to the labor demand, that is why the objective was raised to create and analyze three predictive learning models based on machine learning algorithms, which allow predicting the result of a target variable based on several predictor variables, the case studied is the Regional Autonomous University of the Andes, in Ecuador. The present research aims to develop an aggregation model to measure the influence of machine learning and predictive models of employability relevant to the university career. The predictive model used unsupervised learning algorithms, such as: linear regression, decision tree and SVM generating models with high precision and with the necessary adjustments demonstrating their applicability.

**Keywords:** Machine learning, graduate tracking, higher education, graduate employability, predictive models, neutrosophic model.

## 1 Introducción

Para las Instituciones de Educación Superior es clave el calcular los indicadores en cada uno de sus procesos sustantivos a partir de la recopilación, extracción, tratamiento y almacenamiento de los datos e información generados por los distintos programas, sistemas informáticos y otro tipo de sistemas utilizados para automatizar el tratamiento de información en sus distintas áreas y departamentos. Este gran volumen de datos e información carece de importancia estratégica y no se extrae información de los datos para apoyo en la toma de decisiones. La plétora de sistemas informáticos dificulta a los directivos de las IES encontrar y consolidar la información requerida para informarse sobre el estado actual y prospectivo en la toma de decisiones en las instituciones de Educación Superior.

El creciente interés por parte de los responsables académicos y administrativos de las IES para contar con información de calidad y de forma oportuna para respaldar la toma de decisiones da a la apertura para el uso de tecnologías de exploración y análisis de grandes conjuntos de datos como la minería de datos educativos, análisis de aprendizaje, inteligencia empresarial, entre otras [1]. La inversión en tecnologías adecuadas provee a los directivos de información para monitorear el progreso a lo largo del tiempo de casi todos los aspectos como: rendimiento de los estudiantes, la productividad académica, la tasa de retención, la tasa de titulación, la pertinencia de las carreras, entre otros.

En la mayoría de IES en Latinoamérica, los indicadores tasa de empleabilidad y pertinencia de la carrera son obtenidos a partir de la comunicación entre los graduados y la universidad mediante los sistemas implementados para dicho fin, encuestas, eventos y otras estrategias, el proceso de seguimiento a graduados o en también conocido como seguimiento a egresados es el que facilita la información para el análisis y el seguimiento luego de que los graduados culminaron su proceso formativo y comienzan su actividad laboral, en el mejor de los casos en un área relacionada con su profesión. Este seguimiento debe estar estructurado, institucionalizado, de manera constante y flexible para que las IES tengan una retroalimentación y realicen las actualizaciones pertinentes en cada carrera o programa de formación.

Los primeros estudios de seguimiento a graduados se dieron en Europa, destacándose tres proyectos: 1. CHEERS (Careers after graduation-an European Research Study, EN 1999). 2. REFLEX (The Flexible professional in knowledge society, EN EL 2004) y PROFLEX (Profesional flexible en la sociedad del conocimiento, en el 2006) [2]. Cuyos objetivos se resumen a continuación: a) Una referencia para comparar el empleo con la transición laboral de los nuevos profesionales; b) Identificar los cambios más actuales en el empleo y nuevos desafíos que se le presentan; c) Establecer relaciones conceptuales entre educación superior y empleabilidad; d) La estructuración de una base de datos de la situación de empleo de los nuevos profesionales; e) las habilidades, destrezas y competencias que le mercado laboral demanda de los graduados; f) el nivel de satisfacción de las expectativas de los graduados con su campo laboral; g) rol de las competencias genéricas y específicas en cada fase del empleo y de la carrera profesional; h) Comparar el mercado laboral de los graduados latinoamericanos con sus pares en Europa.

Con los procesos de acreditación de la estructura universitaria desde inicio de siglo en Latino América, se incrementaron los procesos de seguimiento a graduados y análisis de esos resultados. De estudios realizados por el Consejo Superior Universitario Centroamericano (CSUCA y el Servicio Alemán de Intercambio – DAAD, 2004) en Centroamérica, el Instituto Tecnológico de Monterrey en 2008, en el año 2003, la universidad Autónoma de Baja California (UABC), en 2008 la Universidad Autónoma Metropolitana. En el 2005, el Ministerio de Educación de Colombia, fundó el Observatorio Nacional para la Educación. En el 2013, la Pontificia Universidad Javeriana, 2013 realizó un estudio con los graduados del 2009 al 2011, con investigaciones de tipo longitudinales y transversales realizaron en análisis de la situación de los graduados en cinco momentos: al graduarse, luego de un año, después de 3 años y después de 5 años de graduado; permitiendo que cada cohorte sea monitoreada en diferentes puntos clave en proceso de transformación, la indagación de guía y empleo, los primeros contacto con el mercado laboral, su primer empleo y la estabilidad laboral.

En Ecuador, la Constitución del 2008 inició un proceso de transformación de la educación en sus distintos niveles (Inicial, preparatoria, básica, secundaria y superior), reformas a la normativa legal vigente con cambios estructurales, además de nuevo roles y funciones en el Sistema Educativo Nacional. Para el nivel educativo de tercer y cuarto nivel, se promulgó Ley Orgánica de Educación Superior (LOES) y su reglamento en el 2011. En la LOES y en su reglamento se presenta la base legal para la ejecución de la evaluación de las IES en el país.

El modelo de evaluación clasificó a las IES por categorías aplicado por el entonces CONEA (Consejo Nacional de Evaluación y Acreditación de la Educación Superior del Ecuador), dividiendo la gestión institucional en dos grandes aspectos [3]:

- 1) Se entiende que la administración de recursos se valora de acuerdo con los siguientes parámetros: eficacia de gestión, bienes / estudiantes (patrimonio de la IES en dólares / número de estudiantes) y transparencia: difusión del importe que emite la IES en páginas Web institucional.
- 2) Una política organizacional que vincule la administración y los egresados mediante los siguientes mecanismos: a) Acción Afirmativa (reglas dirigidas a la inclusión de grupos minoritarios); b) Monitoreo a Egresados (sistema informático que posibilita la actualización de la propuesta del IES). c) Vinculación con la comunidad: Asociación de programas de enlace comunitario con números de carreras universitarias.

El artículo 142 de la Ley Orgánica de Educación Superior [4] estipula que todas las IES están obligadas a establecer un sistema de seguimiento después de la graduación. Esta disposición se encuentra reflejada en el artículo 26 del Reglamento LOES, donde la Secretaría de Nacional de Educación Superior, Ciencia, Tecnología e Innovación (SENESCYT) es la encargada de establecer los procesos para la organización sistemática de seguimiento de posgrado.

En la LOES se describe la cogestión de la LOES como uno de los principios básicos de las universidades ecuatorianas, regulando, entre otras cosas, la formación del Órgano Colegiado Académico Superior, incluyendo a los representantes de sus egresados. De acuerdo con la Norma 47; los programas de posgrado podrán elegir asesores y representantes para apoyar a los institutos tecnológicos y al Consejo Asesor de la Universidad. Según el artículo 15 de la LOES, los órganos públicos rectores del Sistema de Educación Superior del Ecuador son las Instituciones de Educación Superior y la Comisión de Evaluación, Acreditación y Aseguramiento de la Calidad de la Educación Superior.[5]

El CES tiene la responsabilidad de la coordinación de las actividades de educación superior a nivel nacional a través de una normativa educativa nacional que integre el modelo de desarrollo ecuatoriano, por otro lado, la CEEACES es responsable de monitorear la calidad educativa. De acuerdo con la dirección de la LOES en su artículo 95, el CEAACES es la entidad responsable de asegurar la calidad de la Educación Superior, sus disposiciones obligan a todos los organismos e instituciones del Sistema de Educación Superior del Ecuador [4].

Estudios de seguimiento a graduados en Ecuador se destaca el estudio realizado por la Escuela Superior Politécnica del Chimborazo (Espoch) publicado en 2008, con un universo de 5500 graduados durante los últimos seis cursos académicos, utilizando un muestreo aleatorio estratificado, careció de aplicabilidad por el tratamiento general realizado de los datos, la muestra utilizada es representativa de la universidad, pero no de las titulaciones ofertadas. Con respecto a esto, la Universidad Tecnológica Indoamérica realizó un reporte de monitoreo a graduados de su carrera de Ciencias Psicológicas en el año 2011, en el cual detalla que sus principales problemas fueron las escasas representatividad y capacidad de generalización de los resultados, al basarse sobre un universo de 19 graduados.

La Universidad Casa Grande publicó un estudio de medición del índice de satisfacción de egresados y empleadores sobre las competencias laborales de profesionales graduados en la UCG durante el año 2009, en tres de sus facultades, esto lo realiza haciendo uso de una herramienta para comprender la satisfacción de los graduados con la contribución de ciertas actividades educativas a su desempeño profesional, y otra para comprender la satisfacción de los empleadores en cuanto al desempeño de los graduados en diversas competencias transversales. Las universidades que realizan un seguimiento sistemático de sus graduados se apoyan en los recursos tecnológicos, implementando las encuestas de seguimiento en plataformas digitales integradas con servicios informativos de interés para los estudiantes, bolsas de empleo, asociaciones estudiantiles, redes sociales, entre otros servicios [6].

Los entornos de seguimiento a graduados implementados por varios IES en Ecuador se constituyen en un espacio virtual de comunicación entre la comunidad universitaria y el tejido socio-productivo. La Universidad Regional Autónoma de los Andes, Uniandes en sitio web tiene el acceso a la página de seguimiento a graduados con las opciones de: registros graduados, feria de empleabilidad, misión información y bolsa de trabajos. Mediante una encuesta a graduados de pregrado y postgrado, registra los datos de: Información personal, datos académicos, datos laborales, empleabilidad y satisfacción.

Minería de datos educativos (EDM). La Minería de Datos Educativos (Educational Data Mining – EDM) tiene por objetivo analizar los datos almacenados en los Sistemas Educativos Basados en la Web (WBES), para resolver problemas de investigación educativa: validación del sistema educativo, predicción de logros de aprendizaje, evaluación y retroalimentación a los docentes.

Varias investigaciones con técnicas de minería de datos aplicada a procesos educativos han tratado de determinar la relación entre los factores socioeconómicos y el desempeño en lectura (pruebas PISA), lo cual permite la predicción del rendimiento académico; también, la investigación del aprendizaje en Computer Based Learning Environments (CBLEs), entre otros. A pesar de la variedad de investigaciones, se presentan desafíos como la diferente granularidad o la necesidad de alineamiento temporal de los datos recogidos en las diferentes fuentes. En este sentido, el uso de técnicas de Data Fusion resulta necesario y prometedor en el campo de la educación en general y particularmente en el campo de EDM/LA, tal como se demuestran trabajos recientes en este sentido [7], [8]. Las técnicas usadas en minería de datos pueden ser clasificadas como supervisadas y no supervisadas, las técnicas más

comunes en EDM son árboles de decisión, redes bayesianas, regresión, correlación y análisis por clústeres (la técnica más usada).

La inclusión de la tecnología en entornos de aprendizaje presencia y virtual y entornos como: en El e-learning, b-learning, los entornos hipermedia, campus virtuales, Smart Learning Environments (SLEs), entre otros, son agentes determinantes del proceso de enseñanza - aprendizaje de la Educación Superior, dando lugar a entornos presenciales y virtuales enriquecidos capaces de generar una enorme cantidad de datos de diversas categorías, cuya combinación, puede generar un mejor conocimiento del proceso educativo. El análisis de aprendizaje multimodal (MMLA), un enfoque basado en la recopilación, integración y análisis de datos educativos de una variedad de fuentes que juntas brindan una comprensión holística del proceso de aprendizaje, es el resultado de esta idea de combinar fuentes de datos [9] La combinación de las técnicas de tratamiento de datos multimodales en su intersección con las áreas de EDM y LA ha demostrado ser una línea fructífera en los últimos años [9].

Machine Learning (ML)

También conocido como aprendizaje autónomo, es un subcampo de la inteligencia artificial (IA) que tiene como objetivo construir sistemas que puedan aprender de conjuntos de datos por sí mismos mediante la extracción de información y la detección de patrones recurrentes de comportamiento que respaldan el aprendizaje automático; tales sistemas necesitan algoritmos para poder aprender, los cuales pueden producir patrones de comportamiento, patrones predictivos y más [10, 54].

Los algoritmos de ML se dividen en cuatro categorías: aprendizaje supervisado, aprendizaje no supervisado, aprendizaje semisupervisado y aprendizaje por refuerzo. Comenzando con un conjunto de datos de entrenamiento, el aprendizaje supervisado entrena el modelo para entregar el resultado deseado, luego permite que el modelo aprenda con el tiempo, mide la precisión usando una función de pérdida y realiza ajustes hasta que se minimiza el error, comúnmente en el proceso de minería de datos, se puede dividir en dos tipos: clasificación y regresión [11].

#### **Aprendizaje supervisado**

Machine Learning (ML) permite que las máquinas se utilicen para extraer información y producir resultados que pueden ser útiles en el futuro utilizando datos históricos porque es difícil construir máquinas que puedan aprender en computación. En esta área especializada de aplicaciones de IA, la creación de un conjunto de datos etiquetados para clasificar o predecir con precisión los resultados aumenta la eficiencia, la competitividad y la resolución de problemas complejos.

Cada algoritmo de ML tiene base matemática y se puede cambiar si tiene suficientes datos haciendo cálculos para procesarlos porque la base de ML es que existe una relación matemática entre cualquier combinación de datos de entrada y salida. La regresión y la clasificación son dos tipos diferentes de aprendizaje supervisado que determinan la relación entre una variable dependiente y una o más variables independientes. Los algoritmos de clasificación dividen los datos en grupos discretos. En el aprendizaje supervisado, los algoritmos más conocidos son: regresión por mínimos cuadrados, árbol de decisión, clasificación Naive Bayes, regresión logística, Support Vector Machines (SVM), K-Nearest-Neighbor y métodos de Ensemble (conjunto de clasificadores).

El aprendizaje no supervisado, que se divide en tres categorías: agrupación de clústeres (uso de técnicas para crear grupos), asociación (búsqueda de relaciones entre variables) y reducción de dimensionalidad, utiliza algoritmos de aprendizaje automático para examinar y agrupar conjuntos de datos no etiquetados para encontrar patrones en los datos antes de construir un modelo entrenado. Los algoritmos de análisis de clustering, partición de valores singulares y agrupamiento son las tres categorías más populares de algoritmos de aprendizaje no supervisado.

#### **Técnicas de clasificación de machine learning**

K-nearest neighbors (KNN), para clasificar nuevas muestras y realizar una regresión de reconocimiento de patrones, el algoritmo de K vecinos más cercanos calcula la distancia al ejemplo de entrenamiento más cercano. Si hay similitudes, el objeto se clasifica utilizando el voto mayoritario de los vecinos que se utilizan con mayor frecuencia en los problemas de clasificación; los puntos también se pueden usar para problemas de clasificación o regresión y se pueden encontrar muy juntos. El método más popular para encontrar la distancia entre los puntos de datos más cercanos a un punto de consulta determinado es la distancia euclidiana ( $p = 2$ ), que solo se puede usar con vectores de valores reales y tiene la siguiente fórmula:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Bayesian networks, al expresar las relaciones entre conjuntos de variables aleatorias en términos de distribuciones condicionales, las redes bayesianas, también conocidas como redes de creencias, redes bayesianas o modelos bayesianos, nos permiten procesar eventos probabilísticos y resolver problemas desafiantes e inciertos. Estas redes representan conjuntos de variables que se generan aleatoriamente [12, 55]. A partir de los axiomas de probabilidad se deriva el Teorema de Bayes indicado a continuación:

$$P\left(\frac{X}{Y}\right) = \frac{P\left(\frac{Y}{X}\right) x P(X)}{P(Y)}$$

Decision Tree, las tareas de regresión y clasificación emplean árboles de decisión, que parten de un conjunto de condiciones organizadas jerárquicamente y tienen como objetivo desarrollar un modelo que prediga el valor de una variable por medio del aprendizaje de reglas de decisión derivadas de las propiedades de los datos. La decisión final se puede tomar desde la raíz hasta la hoja[10].

La ganancia de información y la impureza de Gini son dos técnicas que se utilizan frecuentemente para elegir las mejores características en cada nodo. La entropía, que se deriva de la teoría de la información y mide el grado de impureza en los valores de muestra, es un concepto clave en el contexto de la extracción de información, se define con la fórmula siguiente:

$$Entropy(S) = \sum_{i=1}^c -p(C) \log_2 p(C),$$

S representa el conjunto de dato sen el que se calcula entropía.

c representa las clases en conjunto, S

p(C) es la proporción de puntos de datos que pertenecen la clase C al número total de puntos de datos en el conjunto, S.

La ganancia de información generalmente se representa con la fórmula siguiente:

$$Information\ Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

A es un atributo o una etiqueta de clase.

Entropy(S) es la entropía del conjunto de datos, S.

|S<sub>v</sub>|/|S| es la porción de valores en S al número de valores en el conjunto de datos, S.

SVM (Máquina de Soporte Vectorial). Con la ayuda de categorías preestablecidas, este algoritmo de aprendizaje supervisado ayuda a agrupar datos similares y es capaz de operar en grandes dimensiones y no requerir un conocimiento profundo de los datos lo cual lo convierte en uno de los algoritmos de aprendizaje autónomo ML más populares [13]. El algoritmo utiliza un hiperplano para calcular la distancia entre cada punto de datos con el fin de clasificar la categoría a la que pertenece cada uno. Las categorías se separan utilizando la idea de un límite máximo, y la distancia entre los puntos de datos se puede calcular y ajustar en el hiperplano hasta que esté altamente distribuida. El algoritmo hace uso de motores kernel (Lineal, Polinómico, Función de base radial RBF y sigmoide) con el fin de ajustar los hiperplanos [14, 48]. Los kernel más utilizados se resumen en la tabla 1.

**Tabla 1.** Kernels populares empelados con SVM.

Tipo de SVM	Kernel Mercer	Descripción
Función de base radial (RBF) o Gaussiana	$K(x_1, x_2) = \exp\left(-1 \frac{\ x_1 - x_2\ }{2\sigma^2}\right)$	Aprendizaje de una sola clase; $\sigma$ es la anchura del kernel
Lineal	$K(x_1, x_2) = x_1^T x_2$	Aprendizaje de dos clases
Polinómica	$K(x_1, x_2) = (x_1^T x_2 + 1)^p$	P es el orden del polinomio
Sigmoid	$K(x_1, x_2) = \tan h(\beta_0 x_1^T x_2 + \beta_1)$	Es un kernel Mercer solo para determinados valores $\beta_0$ y $\beta_1$

Red neuronal: Sirve como una simulación de un número considerable de unidades de procesamiento interconectadas que se asemejan a versiones abstractas de neuronas y funcionan de manera similar a cómo el cerebro humano procesa la información. Esta técnica de aprendizaje automático, también conocida como aprendizaje profundo, emplea nodos o neuronas conectadas en una estructura jerárquica similar al cerebro humano, lo cual desarrolla un sistema adaptativo permitiendo que las computadoras aprendan de sus errores cuando intentan abordar problemas desafiantes, como armar documentos o reconocer rostros.

## 2 Modelo de agregación

La presente sección describe el modelo de agregación [15], [16, 56]. Metodología KDD (Descubrimiento del conocimiento basado en datos)

Es un proceso automático basado en requerimientos del usuario final, en el que se combinan descubrimiento y análisis, a partir de los cuales se extrae patrones en forma de reglas o funciones para que el usuario los analice. Después de identificar el conocimiento relevante y prioritario desde el punto de vista del usuario final, se pasa a la etapa de selección, en la cual se realiza el proceso de descubrimiento. Se realizan operaciones fundamentales como eliminar datos ruidosos (datos fuera del rango de valores esperado), datos desconocidos (perdidos y vacíos), datos nulos, datos duplicados y técnicas estadísticas para su reemplazo (valor por omisión, valor más cercano, media, moda, mínimos y máximos), esto durante el preprocesamiento de datos o el paso de data cleaning.

En la etapa de transformación (reducción) se utilizan métodos de reducción de dimensiones para disminuir el número efectivo de variables bajo consideración o para encontrar representaciones invariantes de los datos [17]. La reducción horizontal, se realiza al: eliminar tuplas idénticas, discretización de valores continuos (ejemplo edad por rango), en la reducción vertical se elimina atributos redundantes o poco significativos para el problema. Los métodos de reducción utilizados incluyen segmentación, discretización basada en entropía, escaneo, histogramas, agregación, compresión de datos y compresión de datos [18]. En la etapa de minería de datos, se busca descubrir patrones de interés mediante tareas de descubrimiento como: Clasificación, clustering, patrones, secuenciales, asociación, entre otras, dando como resultado modelos predictivos (estimar valores futuros o desconocidos de las variables objetivo) con tareas de clasificación y regresión. Los modelos descriptivos (explorar las propiedades de los datos examinados) que permitan por ejemplo identificar grupos de personas con gustos similares, patrones de compra, mediante reglas de asociación, patrones secuenciales, los clustering y las correlaciones [6].

En la etapa de interpretación o evaluación de datos, se descifran los patrones descubiertos o se realizan iteraciones en las etapas anteriores, se puede aplicar: visualización de los patrones extraídos, remoción de patrones irrelevantes o redundantes, traducción de patrones útiles para ser entendibles por los usuarios; además se puede consolidar el nuevo conocimiento en otros sistemas o documentarlo para las partes interesadas.

### Metodología CRISP-DM (Cross -industry Standard Process)

Metodología creada en 1993 por líderes de la industria DaimlerChrysler, SPSS, OHRA y NCR, tiene como finalidad proporcionar nuevas ideas a los mineros de datos, basada en experiencias reales de quienes hacen proyectos. Fases, tareas generales, tareas especializadas e instancias de proceso son los cuatro niveles de abstracción en los que se clasifican estas tareas [4, 52]. Su proceso se compone de seis fases: Comprensión del negocio, comprensión de datos (Formular hipótesis), preparación de los datos (selección, limpieza y transformación), modelado (técnicas de modelado y calibración de parámetros), evaluación (Verificar cumplimiento de objetivos del proyecto) e Implantación (difusión del conocimiento).

### Metodología SEMMA

Es un conjunto de herramientas funcionales enfocadas al desarrollo propio de un modelo de minería orientado al descubrimiento en una gran cantidad de datos de patrones desconocidos de negocio [19]. Permite aplicar la estadística exploratoria, técnicas de visualización, selección y transformación de las variables más significativas, para crear modelos predictivos de apoyo a la toma de decisiones. Consta de cinco fases: Muestreo (muestra de un conjunto de datos grande y relevante), exploración (Buscar relaciones y tendencias desconocidas para formular hipótesis), modificación (limpieza de valores atípicos), Modelo (Técnicas de predicción: árboles de decisión, redes neuronales, análisis discriminante), evaluación (utilidad y exactitud de los modelos).

## 3 Materiales y métodos

El modelo está compuesto por estructura, componentes, cualidades y principios con el objetivo de modelar la problemática existente [20-23]. El modelo basa su funcionamiento a partir de la modelación mediante la lógica neutrosófica para medir para medir la influencia de la depresión en la salud oral en adolescentes mediante Mapa

Cognitivo Neutrosófico [24-26].

El modelo posee una estructura abierta a partir de lo cual todas las informaciones son posibles gestionar [27, 28, 53], [29-31]. Se diseña a partir de componentes que rigen el flujo de trabajo. Presentan un conjunto de cualidades que garantizan la integración flexible de los principales componentes [32, 33]. Sustenta la propuesta sobre un conjunto de principios como la neutralidad, el procesamiento de la inferencia mediante Mapa Cognitivo Neutrosófico (MCN) y una autonomía propia que facilita el resultado de la toma de decisiones sobre una inferencia propia [34-36], [37], [38, 39].

La definición original de valor de verdad en la lógica neutrosófica es formalizada como [40], [41], [42, 50]: sean  $N = \{(T, I, F) : T, I, F \subseteq [0, 1]\}$ , una evaluación neutrosófica es un mapeo de un grupo de fórmulas proporcionales a  $N$ , esto es que por cada sentencia  $p$  se tiene:

$$v(p) = (T, I, F) \quad (1)$$

#### 4 Resultados y discusión

Según Natek junto con Zwilling [43, 51], los pasos para el análisis de la minería de datos son: crear el conjunto de datos, definir las herramientas de minería de datos (MD), evaluar las técnicas de MD y analizar los datos por cada modelo para determinar el mejor.

Se realizó una revisión bibliográfica de las metodologías para minería de datos (Tabla 2) tales como: CRISP-DM, SEMMA y KDD, se optó por la metodología KDD en función de que esta hace referencia al proceso completo de descubrimiento del conocimiento [44, 49, 58]. La metodología KDD propuesta por Fayyad et al, 1996 consta de seis fases: análisis previo, selección, preprocesamiento - limpieza, transformación-reducción, minería de datos, interpretación-evaluación de datos. También se realizó un resumen de los trabajos relacionados con el tema de estudio y publicados en revistas de alto impacto. [62]

Mediante un análisis previo de aproximación al problema partiendo del sistema de seguimiento a graduados de la institución de educación superior para determinar las variables de las cuales se recopila la información y el sistema que utilizan, el resultado fue el cuestionario de seguimiento con la clasificación de las variables en: Información personal, datos académicos, laborales, empleabilidad y satisfacción. [59, 60, 61]

En la fase de selección se identificaron las variables relevantes para la creación de un conjunto de datos objetivo en un archivo de Excel con 14557 registros, el resultado de la fase el archivo de datos a analizar. En la fase de preprocesamiento - limpieza, se analizó la calidad de los datos, eliminar datos personales, remoción de columnas de datos incompletas, obtención de datos faltantes, datos nulos, para generar un nuevo archivo de Excel.

En transformación - reducción, se codificaron columnas de la data como: provincia, sexo, discapacidad, modalidad, estado del graduado, tipo de empresa y otros. Dando como resultados un nuevo archivo con los datos de valores de codificación de los valores de las columnas. En minería se realizó la búsqueda de patrones insospechados y de interés aplicando análisis de datos con Python, correlaciones, gráfico de histogramas, entre otros.

Tabla 2. Revisión bibliográfica.

Autores	Título	Problema / objetivo	Técnicas utilizadas	Conclusiones
Twinkle Mae C. Gatbonton Betchie E. Aguinaldo	Employability Predictive Model Evaluator Using PART and JRip Classifier	Desarrollar un Evaluador de Modelo predictivo de Empleabilidad aplicando el proceso de minería de datos utilizando el algoritmo clasificador PART y JRip [45].	Descubrimiento de conocimiento en base de datos (KDD). Evaluador de Modelo Predictivo de Empleabilidad aplicando el proceso de minería de datos utilizando el algoritmo clasificador PART y JRip.	Para que el graduado consiga trabajo, no debe tener un promedio que va de 2.13 a 2.62 que puede ser categorizado como regular y bastante bueno
Juan C. Cevallos, Martha C. Escobar, Jesseña E. Falcones y Wilson J. Cevallos	Modelado laboral de los egresados de la Facultad de Ciencias Informáticas de la Universidad Técnica de Manabí (Ecuador)	desarrollar un modelo de minería de datos para identificar patrones de empleabilidad de egresados, que contribuya a mejorar la tasa de inserción laboral afín a la formación profesional [44].	Revisión bibliográfica. Descubrimiento de conocimiento en base de datos (KDD). Selección de variables relevantes en los datos, buscando su simetría, normalidad y correlación existente en la información.	predictores como CBET (carrera), BTTC (Búsqueda de trabajo), UCHTA (Uso de conocimiento), IM (ingresos mensuales), PT (Primer trabajo), IL(Institución), APF (Actividad afín), IF(Instrucción formal), MG(modalidad)

Autores	Título	Problema / objetivo	Técnicas utilizadas	Conclusiones
			Data en Excel con un total de 929 graduados, 22 958 registros, de las variables objetivo. Algoritmos: árbol de decisión (C4.5), Multilayer Perceptron, regresión lineal.	
Hanan Aldowah Hosam Al-Samarraie Wan Mohamad Fauzy	Educational data mining and learning analytics for 21st century higher education: A review and synthesis [46].	¿Cómo podemos usar EDM y LA para resolver desafíos prácticos en educación?  ¿Qué técnicas de minería de datos se adaptan mejor a estos problemas?	Se realizó una revisión exhaustiva de la utilización de EDM (Education Data Mining) y LA (Learning Analytics). El proceso de revisión se derivó de consideraciones teóricas previas que permiten extraer conclusiones sobre la literatura revisada.	EDM y LA se usan comúnmente para brindar oportunidades y soluciones a varios problemas de aprendizaje relacionados con CSLA, CSPA, CABA y CSVA.  Las principales técnicas de minería de datos de agrupamiento, regla de asociación, minería de datos visuales, estadísticas y regresión se usan comúnmente en estas dimensiones
Chavarry Chanchay Mariana. Aquino Trujillo Jury Yessenia. Li Vega Fiorella Vanessa. Germán Reyes Nilton César	La analítica académica y la minería de datos educacional en el nivel universitario [47, 57].	Indagar por medio de una revisión sistemática las diversas formas de aplicación la analítica académica y minería de datos educacional en las universidades.  ¿Cómo ha progresado en el tiempo las investigaciones en la cual se aplica la analítica académica y la minería de datos educacional en el nivel universitario?	Revisión sistemática por medio de la declaración prisma, con 187 artículos identificados, finalmente fueron seleccionados 22.	La gran variedad de técnicas de minería de datos implementadas en las investigaciones analizadas no permite establecer con precisión el uso específico de algunas de ellas. Temas del contexto enseñanza – aprendizaje con muy pocas evidencias de estudio o temas

## Dataset

La información de la situación laboral de los graduados se extrajo de una encuesta en línea que luego se obtuvo un dataset digital en formato CSV, que luego de un primer proceso de depuración de registros nulos e incompletos directamente en el archivo, se redujo a 14557 registros de graduados entre los años 2018-2021. En la figura 1 se representa el proceso de obtención, procesamiento, análisis y resultados del proceso de minería de datos al dataset.

Atributos de la data, procesados con Python. Utilizando el lenguaje de programación Python se obtuvo las características del conjunto de datos, como se muestra en la figura 2, está compuesto por 14557 filas y 39 columnas. El tipo y rango de datos de cada columna es variable, algunas dicotómicas (Si, no), otras de opción múltiple en función de tipo de empresa, tipo de contrato y las 12 preguntas de satisfacción con la carrera que tienen valores según la escala de Likert.

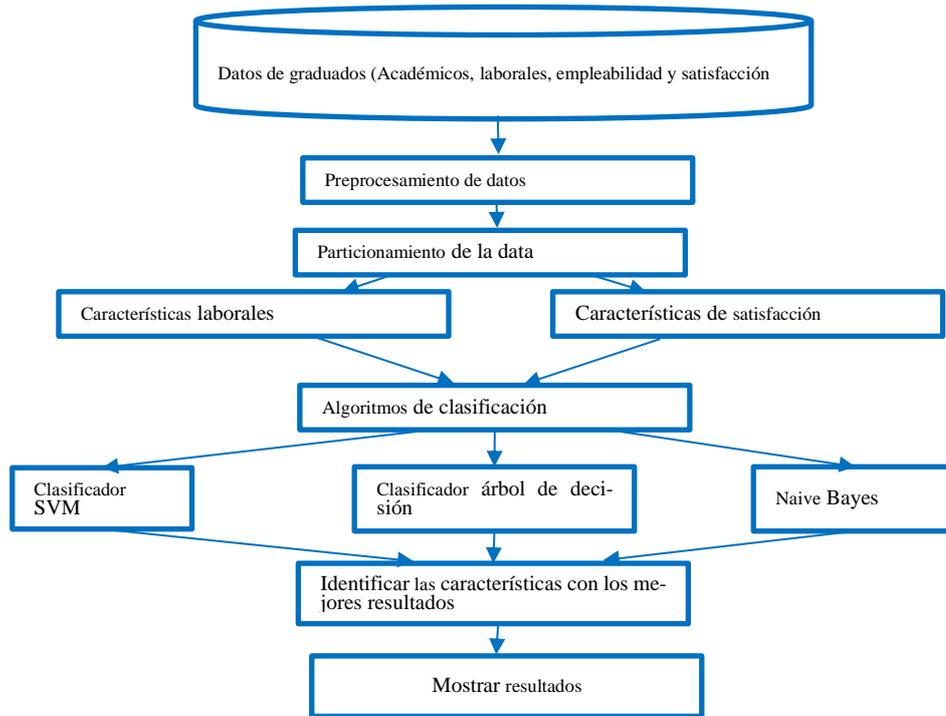


Figura 1. Metodología aplicada para el desarrollo del modelo.

Limpieza y procesamiento. Para el análisis mediante los algoritmos de minería de datos, fue necesario un proceso de depuración de datos perdidos o incorrectos, eliminar registros duplicados, limpieza de ruido, tratamiento de datos faltantes y anonimizar datos, la mayor parte del proceso se la realizó directamente en archivo en Excel, pero se verificó utilizando los comandos de Python para eliminar las columnas con datos pedidos y luego la confirmación de forma visual. De igual manera el proceso de reducción se lo realizó en MS Excel, se redujo 3 columnas y se redijeron dimensiones, además de renombrar las variables (ver tabla 2).

División. Para el análisis de correlación de las variables se realizó dos grupos, el primer grupo asociado con las variables de datos del graduado y la empleabilidad (fig. 2) y el segundo grupo con las variables satisfacción de graduado con la carrera.

Entrenamiento de testing. Se asignó el 80% de los datos para entrenamiento y el resto para el test de forma aleatoria, Fig. 3, para en lo posterior verificar el tamaño del train y del test.

Minería de datos. Según las fases de la metodología KDD, se construyó el modelo predictivo de clasificación o segmentación con los algoritmos indicados anteriormente. Estos modelos predictivos pretendieron estimar valores de empleabilidad de los graduados según su área de estudio, siendo esta la variable objetivo, las variables predictivas fueron: trabaja actualmente, sexo, modalidad de estudio en la que se graduó y el año en que se graduó.

	Num	pais	provincia	ciudad	Sexo	Discapacidad	Extension	Carrera	Modal
0	1	1	23	Santo Domingo	2	0	SANTO DOMINGO	NEGOCIOS	
1	2	1	12	Babahoyo	1	0	BABAHOYO	NEGOCIOS	
2	3	1	12	Babahoyo	2	0	BABAHOYO	SISTEMAS	
3	4	1	23	Santo Domingo	2	0	SANTO DOMINGO	DERECHO	
4	5	1	23	Santo Domingo	2	0	SANTO DOMINGO	NEGOCIOS	
...	...	...	...	...	...	...	...	...	...
14552	14553	1	23	Santo Domingo	2	0	SANTO DOMINGO	CONTABILIDAD	
14553	14554	1	18	Ambato	2	0	AMBATO	NEGOCIOS	
14554	14555	1	18	Santiago de Pillaro	1	0	AMBATO	DERECHO	
14555	14556	1	5	Latacunga	1	0	AMBATO	DERECHO	
14556	14557	1	16	Puyo	1	0	PUYO	DERECHO	

14557 rows x 39 columns

Figura 2. Detalle de la data en formato de archivo de Excel.

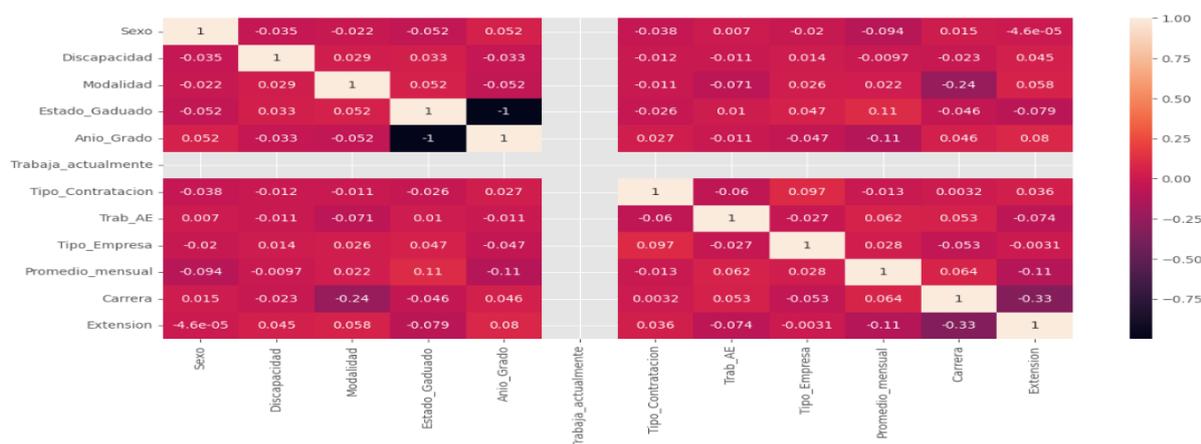


Figura 3. Matriz de correlación de las variables de empleabilidad.

Modelo SVM y evaluación. El primer algoritmo evaluado es el de máquina a de vectores de soporte (Support Vector Machines SVM), con la función kernel lineal, Fig 5, Esto al considerar que la separación de los datos es sencilla.

```
print(X_train.shape, y_train.shape)
print(X_test.shape, y_test.shape)

(5898, 4) (5898, 1)
(1475, 4) (1475, 1)
```

Figura 4. División de los datos en train y test.

```
# Creación del modelo SVM lineal
# =====
modelo = SVC(C = 100, kernel = 'linear', random_state=123)
modelo.fit(X_train, y_train)
```

SVC

SVC(C=100, kernel='linear', random\_state=123)

Figura 5. Creación del modelo SVM lineal.

Para la evaluación del modelo se ejecutó las predicciones con los datos de test, obteniendo una exactitud del modelo de 77, 084.

Ajuste del modelo. Se repite el ajuste del modelo, esta vez empleando un kernel radial y utilizando validación cruzada para identificar el valor óptimo de penalización C. y se obtienen los mejores hiper parámetros.

Árbol de decisión. Para el procesamiento con el algoritmo de árbol de decisión se seleccionaron los campos numéricos: país, Sexo, Discapacidad, Extension, Carrera, Modalidad, Estado\_Gaduado, Anio\_Grado, Trabaja\_actualmente, Tipo\_Empresa, Clasificacion\_empresa, Nivel\_Cargo, Tipo\_Contratacion, Tiempo\_trabajo, Empresa\_propia, Trab\_AE, Jornada\_trabajo, Tipo\_trabajo. El resultado gráfico del árbol se muestra en la figura 7. Con una precisión de 87.02

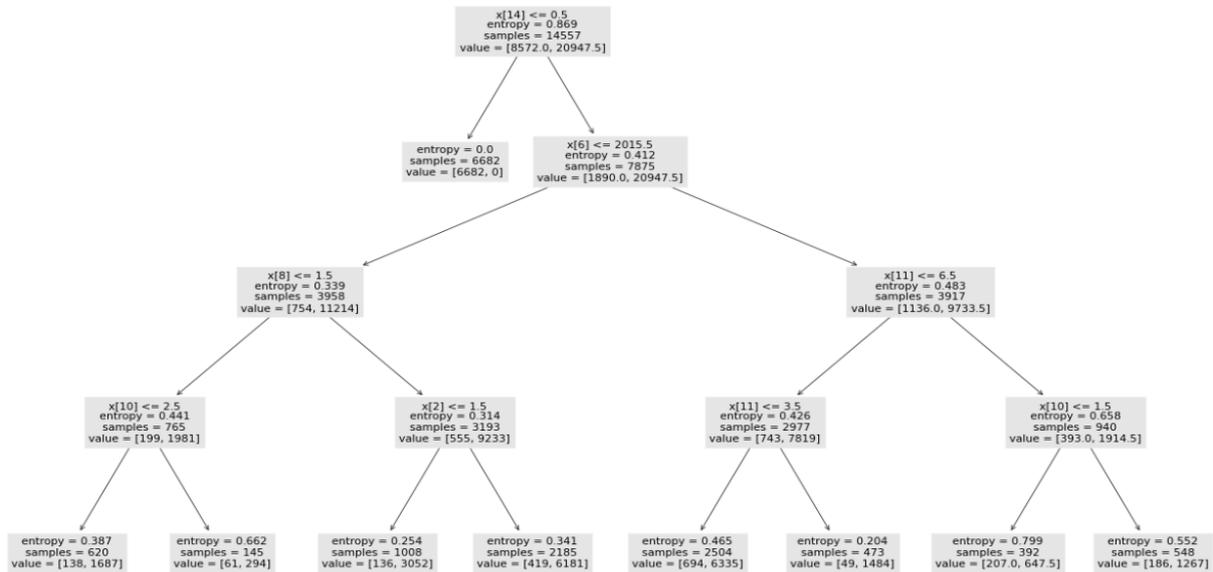


Figura 6. Gráfico del árbol de decisión

Algoritmo Naive Bayes. El algoritmo es un clasificador probabilístico simple con fuerte suposición de independencia (a menudo suposición pobre para conjunto de datos verdaderos), aprende de los datos de entrenamiento y luego predice la clase de la instancia de prueba con la mayor probabilidad posterior, los pasos se describen a continuación:

Utilización de la clase SelectKBest de SkLearn para elegir las 5 mejores características a usar. (Figura 7), la correlación de Pearson no es muy alta entre las 5 variables.

```

X = datos2[['Trabaja_actualmente', 'Carrera', 'Modalidad', 'Extension']]
y = datos2[['Trab_AE']]

best=SelectKBest(k=4)
X_new = best.fit_transform(X, y)
X_new.shape
selected = best.get_support(indices=True)
print(X.columns[selected])
  
```

Index(['Trabaja\_actualmente', 'Carrera', 'Modalidad', 'Extension'], dtype='object')

Figura 7. Uso de la clase SelectKBest de SkLearn para elegir las 5 mejores características a usar.

Creamos el modelo de Gaussian Naive Bayes. Dividimos en un set de entrenamiento y otro de Test con el 20% de las entradas, para luego proceder al entrenamiento, figura 8. La Precisión en el set de Entrenamiento: 0.87 y la precisión en el set de Test: 0.87. También se generó la matriz de confusión. Figura 9.

```

gnb = GaussianNB()
# Train classifier
gnb.fit(
    X_train[used_features].values,
    y_train
)
y_pred = gnb.predict(X_test[used_features])
  
```

Figura 8. Entrenamiento del modelo Naive Bayes.

```
print(confusion_matrix(y_test, y_pred))
print(classification_report(y_test, y_pred))
```

	[[1389 372]				
	[ 0 1151]]				
		precision	recall	f1-score	support
	0	1.00	0.79	0.88	1761
	1	0.76	1.00	0.86	1151
	accuracy			0.87	2912
	macro avg	0.88	0.89	0.87	2912
	weighted avg	0.90	0.87	0.87	2912

Figura 9. Matriz de confusión

## Conclusión

Las técnicas de minería de datos de SVM y árbol de decisión han demostrado ser efectivas en un 77,1% y 87,01% al igual que el algoritmo Naive Bayes con el 87%; la mayoría de las variables de satisfacción no representaron un papel importante en la clasificación. La carrera, la extensión, la modalidad y el sexo del graduado sí influyen en el resultado de obtener trabajo en el área de formación académica. El tratamiento de los datos juega un papel importante en el nivel de precisión de los modelos, por lo que es necesario que los programas de seguimiento a graduados realicen un mejor control en la toma de datos y busquen estrategias que permitan una constante actualización del estado laboral de sus graduados.

La pertinencia de las carreras universitarias para satisfacer las demandas del mercado laboral es un tema constantemente investigado a partir de encuestas dirigidas a los estudiantes de bachillerato y a empresarios, con los modelos propuestos se suma una herramienta que brinda información oportuna y confiable para la toma de decisiones en cuanto al rediseño curricular, programas de maestría y otras decisiones relacionadas con la calidad de las carreras universitarias y los estudiantes en formación.

No se consideró a las variables como: país, provincia, ciudad, estado del graduado, año de graduación, la empresa y tipo de empresa donde labora, nivel de cargo, tipo de contratación, tiempo de trabajo, si tiene empresa propia, la jornada laboral y el promedio mensual del salario, además de las variables de satisfacción, que pueden ser analizadas con otros modelos en próximos estudios. Una de las particularidades de la institución de educación superior de la cual se obtuvieron los datos, es la de contar con una matriz, seis sedes y una extensión en diferentes ciudades, motivo por el cual una variable considerada fue la extensión, por la inequidad de género aún existente en Ecuador, se consideró la variable sexo en el modelo, además de la carrera y la modalidad de estudio (presencial y a distancia al momento de la captura de los datos).

## Referencias

- [1] S. Gaftandzhieva, S. Hussain, S. Hilcenko, R. Doneva, and K. Boykova, "Data-driven decision making in higher education institutions: State-of-play," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 6, pp. 397-405, 2023.
- [2] R. Y. Palloroso Granizo, and I. García Rondón, "Seguimiento a graduados: importancia y principales experiencias internacionales. Análisis del tópico en la República del Ecuador," *Caribeña de Ciencias Sociales*, no. julio, 2019.
- [3] C. Conea, "Evaluación de desempeño institucional de las Universidades y Escuelas Politécnicas del Ecuador," *Mandato constituyente*, no. 14, 2009.
- [4] D. T. Larose, and C. D. Larose, *Discovering knowledge in data: an introduction to data mining*: John Wiley & Sons, 2014.
- [5] C. de Educación Superior, "Reglamento de régimen académico," *Recuperado de [http://www.ces.gob.ec/doc/regimen\\_academico/proyecto% 20de% 20reglamento% 20de% 20rgimen% 20acadmico% 20para% 20segundo% 20debate. pdf](http://www.ces.gob.ec/doc/regimen_academico/proyecto%20de%20reglamento%20de%20rgimen%20acadmico%20para%20segundo%20debate.pdf)*, 2013.
- [6] S. Timarán-Pereira, I. Hernández-Arteaga, S. Caicedo-Zambrano, A. Hidalgo-Troya, and J. Alvarado-Pérez, "El proceso de descubrimiento de conocimiento en bases de datos," *Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional*, pp. 63-86, 2016.
- [7] B. Alfano, L. Barretta, A. Del Giudice, S. De Vito, G. Di Francia, E. Esposito, F. Formisano, E. Massera, M. L. Miglietta, and T. Polichetti, "A review of low-cost particulate matter sensors from the developers' perspectives," *Sensors*, vol. 20, no. 23, pp. 6819, 2020.

- [8] M. Sultana, F. A. Jibon, and A. Kashem, "An improved model of virtual classroom using information fusion and NS-DBSCAN," *Global Journal of Computer Science and Technology*, vol. 20, no. 5, pp. 26-33, 2020.
- [9] K. Sharma, and M. Giannakos, "Multimodal data capabilities for learning: What can multimodal data tell us about learning?," *British Journal of Educational Technology*, vol. 51, no. 5, pp. 1450-1484, 2020.
- [10] J. M. O. Candel, *Big data, machine learning y data science en Python*: RA-MA Editorial, 2022.
- [11] K. B. S. Burgos, "FACULTAD DE INGENIERÍA EN CIENCIAS DE LA COMPUTACIÓN Y TELECOMUNICACIONES," UNIVERSIDAD AUTÓNOMA "GABRIEL RENE MORENO, 2022.
- [12] M. B. Rao, and C. Rao, "Bayesian networks," *Handbook of Statistics*, pp. 357-385: Elsevier, 2014.
- [13] I. K. Nti, J. A. Quarcoo, J. Aning, and G. K. Fosu, "A mini-review of machine learning in big data analytics: Applications, challenges, and prospects," *Big Data Mining and Analytics*, vol. 5, no. 2, pp. 81-97, 2022.
- [14] M.-P. Hosseini, A. Hosseini, and K. Ahi, "A review on machine learning for EEG signal processing in bioengineering," *IEEE reviews in biomedical engineering*, vol. 14, pp. 204-218, 2020.
- [15] B. KOSKO, "Fuzzy cognitive maps," *International Journal of Man-Machine Studies*, vol. 24, no. 1, pp. 65-75, 1986.
- [16] J. Salmeron, "Augmented fuzzy cognitive maps for modeling LMS critical success factors," *Knowledge-Based Systems*, vol. 22 no. 4, pp. 275-278, 2009.
- [17] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "The KDD process for extracting useful knowledge from volumes of data," *Communications of the ACM*, vol. 39, no. 11, pp. 27-34, 1996.
- [18] J. Han, J. Pei, and H. Tong, *Data mining: concepts and techniques*: Morgan kaufmann, 2022.
- [19] A. A. Aquino, G. Molero-Castillo, and R. Rojano, "Hacia un nuevo proceso de minería de datos centrado en el usuario," *Pistas Educativas*, vol. 36, no. 114, 2018.
- [20] M. Y. L. Vázquez, J. E. Ricardo, and N. B. Hernández, "Investigación científica: perspectiva desde la neutrosfía y productividad," *Universidad y Sociedad*, vol. 14, no. S5, pp. 640-649., 2022.
- [21] E. G. Caballero, M. Leyva, J. E. Ricardo, and N. B. Hernández, "NeutroGroups Generated by Uninorms: A Theoretical Approach," *Theory and Applications of NeutroAlgebras as Generalizations of Classical Algebras*, pp. 155-179: IGI Global, 2022.
- [22] M. L. Vázquez, J. Estupiñán, and F. Smarandache, "Neutrosfía en Latinoamérica, avances y perspectivas," *Revista Asociación Latinoamericana de Ciencias Neutrosóficas. ISSN 2574-1101*, vol. 14, pp. 01-08, 2020.
- [23] J. Torales, I. Barrios, and I. González, "Problemáticas de salud bucodental en personas con trastornos mentales," *Medwave*, vol. 17, no. 08, 2017.
- [24] J. E. Ricardo, M. Y. L. Vázquez, A. J. P. Palacios, and Y. E. A. Ojeda, "Inteligencia artificial y propiedad intelectual," *Universidad y Sociedad*, vol. 13, no. S3, pp. 362-368, 2021.
- [25] I. A. González, A. J. R. Fernández, and J. E. Ricardo, "Violación del derecho a la salud: caso Albán Comejo Vs Ecuador," *Universidad Y Sociedad*, vol. 13, no. S2, pp. 60-65, 2021.
- [26] G. Á. Gómez, J. V. Moya, J. E. Ricardo, and C. V. Sánchez, "La formación continua de los docentes de la educación superior como sustento del modelo pedagógico," *Revista Conrado*, vol. 17, no. S1, pp. 431-439, 2021.
- [27] J. E. Ricardo, J. J. D. Menéndez, and R. L. M. Manzano, "Integración universitaria, reto actual en el siglo XXI," *Revista Conrado*, vol. 16, no. S 1, pp. 51-58, 2020.
- [28] J. E. Ricardo, N. B. Hernández, R. J. T. Vargas, A. V. T. Suntaxi, and F. N. O. Castro, "La perspectiva ambiental en el desarrollo local," *Dilemas contemporáneos: Educación, Política y Valores*, 2017.
- [29] V. V. Falcón, B. S. Martínez, J. E. Ricardo, and M. Y. L. Vázquez, "Análisis del Ranking 2021 de universidades ecuatorianas del Times Higher Education con el Método Topsis," *Revista Conrado*, vol. 17, no. S3, pp. 70-78, 2021.
- [30] J. Ricardo, A. Fernández, and M. Vázquez, "Compensatory Fuzzy Logic with Single Valued Neutrosophic Numbers in the Analysis of University Strategic Management," *International Journal of Neutrosophic Science*, pp. 151-159, 2022.
- [31] J. E. Ricardo, M. Y. L. Vázquez, and N. B. Hernández, "Impacto de la investigación jurídica a los problemas sociales postpandemia en Ecuador," *Universidad y Sociedad*, vol. 14, no. S5, pp. 542-551., 2022.
- [32] S. D. Álvarez Gómez, A. J. Romero Fernández, J. Estupiñán Ricardo, and D. V. Ponce Ruiz, "Selección del docente tutor basado en la calidad de la docencia en metodología de la investigación," *Conrado*, vol. 17, no. 80, pp. 88-94, 2021.
- [33] J. E. Ricardo, V. M. V. Rosado, J. P. Fernández, and S. M. Martínez, "Importancia de la investigación jurídica para la formación de los profesionales del Derecho en Ecuador," *Dilemas Contemporáneos: Educación, Política y Valores*, 2020.

- [34] B. B. Fonseca, and O. Mar, "Implementación de operador OWA en un sistema computacional para la evaluación del desempeño," *Revista Cubana de Ciencias Informáticas*, 2021.
- [35] C. Marta Rubido, and O. M. Cornelio, "Práctica de Microbiología y Parasitología Médica integrado al Sistema de Laboratorios a Distancia en la carrera de Medicina," *Revista de Ciencias Médicas de Pinar del Río*, vol. 20, no. 2, pp. 174-181, 2016.
- [36] O. Mar, and B. Bron, "Procedimiento para determinar el índice de control organizacional utilizando Mapa Cognitivo Difuso," *Serie Científica*, pp. 79-90.
- [37] M. L. Vázquez, and F. Smarandache, *Neutrosophia: Nuevos avances en el tratamiento de la incertidumbre: Infinite Study*, 2018.
- [38] M. Glykas, and P. Groumpos, "Fuzzy Cognitive Maps: Basic Theories and Their Application to Complex Systems Fuzzy Cognitive Maps " *Springer Berlin / Heidelberg.*, pp. 1-22, 2010.
- [39] Gonzalo Nápoles, Elpiniki Papageorgiou, Rafael Bello, and K. Vanhoof, "Learning and convergence of fuzzy cognitive maps used in pattern recognition," *Neural Processing Letters*, vol. 45, no. 2, pp. 431-444, 2017.
- [40] F. Smarandache, *A Unifying Field in Logics: Neutrosophic Logic. Neutrosophy, Neutrosophic Set, Neutrosophic Probability: Neutrosophic Logic. Neutrosophy, Neutrosophic Set, Neutrosophic Probability: Infinite Study*, 2005.
- [41] H. Wang, F. Smarandache, R. Sunderraman, and Y. Q. Zhang, *Interval Neutrosophic Sets and Logic: Theory and Applications in Computing: Theory and Applications in Computing: Hexis*, 2005.
- [42] F. Smarandache, "Neutrosophia y Plitogenia: fundamentos y aplicaciones," *Serie Científica de la Universidad de las Ciencias Informáticas*, vol. 17, no. 8, pp. 164-168, 2024.
- [43] S. Natek, and M. Zwillig, "Student data mining solution–knowledge management system related to higher education institutions," *Expert systems with applications*, vol. 41, no. 14, pp. 6400-6407, 2014.
- [44] J. C. Cevallos, M. C. Escobar, J. E. Falcones, and W. J. Cevallos, "Modelado laboral de los egresados de la Facultad de Ciencias Informáticas de la Universidad Técnica de Manabí (Ecuador)," *Información tecnológica*, vol. 32, no. 6, pp. 111-122, 2021.
- [45] T. M. C. Gatbonton, and B. E. Aguinaldo, "Employability predictive model evaluator using part and JRIP classifier." pp. 307-310.
- [46] H. Aldowah, H. Al-Samraie, and W. M. Fauzy, "Educational data mining and learning analytics for 21st century higher education: A review and synthesis," *Telematics and Informatics*, vol. 37, pp. 13-49, 2019.
- [47] M. C. Chankay, J. Y. A. Trujillo, F. V. L. Vega, and N. C. G. Reyes, "La Analítica Académica y la Minería de Datos Educativo en el nivel universitario: revisión sistemática," *Universidad y Sociedad*, vol. 14, no. S6, pp. 377-390., 2022.
- [48] Velázquez-Soto, O. E., Muñoz, E. E. C., Vazquez, M. Y. L., Chieng, L. Y. D., & Ricardo, J. E. "Analysis of Scientific Production on Neutrosophy: A Latin American Perspective". *Neutrosophic Sets and Systems*, vol 67, pp 285-306, 2024.
- [49] Macas-Acosta, G., Márquez-Sánchez, F., Vergara-Romero, A., & Ricardo, J. E. "Analyzing the Income-Education Nexus in Ecuador: A Neutrosophic Statistical Approach". *Neutrosophic Sets and Systems*, vol 66, pp 196-203, 2024.
- [50] Esparza-Pijal, F. I., Sandoval-Loyo, J. A., Zuña-Anilema, L. H., & Estupiñán-Ricardo, J. "Incidencia del consumo de sustancias sujetas a fiscalización en el rendimiento académico de los adolescentes". *CIENCIAMATRIA*, vol 10 núm 1, pp 795-805, 2024.
- [51] Vázquez, M. Y. L., Hernández, N. B., Ricardo, J. E., & García, J. F. G. "Aplicación de análisis de sentimientos y enfoques neutrosóficos para la comprensión de información textual en la investigación". *Revista Conrado*, vol 19 núm 94, pp 294-300, 2023.
- [52] Hernández, N. B., Yelandi, L. V. M., Ricardo, J. E., & Manzano, R. L. M. "Análisis prospectivo del estado actual de la carrera de derecho en la sede Babahoyo de la universidad UNIANDES". *Revista Conrado*, vol 19 núm S2, pp 505-513, 2023.
- [53] Feigenblatt, O. F. V., & Estupiñán Ricardo, J. "El reto de la sostenibilidad en los países en desarrollo: el caso de Tailandia". *Revista Universidad y Sociedad*, vol 15 núm 4, pp 394-402.
- [54] Jirón Abad, Á. D. "Aplicación de las atenuantes establecidas en los artículos 44 y 45 del Coip y su incidencia en contravenciones penales en la administración de justicia de Pastaza año 2021" (Master's thesis), 2022.
- [55] Estupiñán Ricardo, J., Leyva Vázquez, M., & Romero Fernández, A. "Evaluación del aprendizaje basado en proyectos". *Investigación Operacional*, vol 43 núm 3, pp 409-419, 2022.
- [56] Ricardo, J. E., Vásquez, Á. B. M., Herrera, R. A. A., Álvarez, A. E. V., Jara, J. I. E., & Hernández, N. B. "Management System of Higher Education in Ecuador. Impact on the Learning Process". *Dilemas Contemporáneos: Educación, Política y Valore, (Special)*, 2018.

- [57] López Gutiérrez, J. C., Pérez Ones, I., Estupiñán Ricardo, J., & Ramos Carpio, J. P. "Integrating Neutrosophic Theory into Regression Models for Enhanced Prediction of Uncertainty in Social Innovation Ecosystems". *Neutrosophic Sets and Systems*, vol 71 núm 1, pp 18, 2024.
- [58] Sánchez, F. M., Santos, C. E. O., & Sentí, V. E. "La gestión del conocimiento y el aprendizaje. Aspectos metodológicos. UCE Ciencia". *Revista de postgrado*, vol 7 núm 2, 2019.
- [59] Pozo-Estupiñán, C., Sorhegui Ortega, R., Márquez-Sánchez, F., & Vergara Romero, A. "Soberanía alimentaria desde la política pública y sus argumentos". *Revista Científica ECOCIENCIA*, núm 8, pp 79-93. 2021.
- [60] Amores, E. R., Vega, L. L., Sánchez, F. M., & León, V. L. "Modelo econométrico de los gastos operativos de la banca en el Ecuador: Periodo 2012–2019: Econometric model of banking operating expenses in Ecuador: Period 2012-2019". *REVISTA CIENTÍFICA ECOCIENCIA*, vol 8 núm 3, pp 80-99, 2021.
- [61] Sanchez, F. M. "Linkage between poverty and environmental degradation: a profound opinion", 1998.
- [62] Márquez-Carriel, D. C., Márquez-Sánchez, F., Oña-Garcés, L., & Vergara-Romero, A. "Feminist Foreign Policy's Prospects in The Context of Global Cooperation". *Kurdish Studies*, vol 12 núm 1, pp 3038-3050, 2024.

Recibido: Agosto 31, 2024. Aceptado: Septiembre 19, 2024