



EDUCATIONAL DATA MINING (EDM) CON RANDOM FOREST PARA CLASIFICACIÓN DE ESTUDIANTES SEGÚN ESTILOS DE APRENDIZAJE Y DESEMPEÑO ACADÉMICO EN LA UNIVERSIDAD AUTÓNOMA TOMÁS FRÍAS

EDUCATIONAL DATA MINING (EDM) COM RANDOM FOREST PARA CLASSIFICAÇÃO DE ESTUDANTES SEGUNDO ESTILOS DE APRENDIZAGEM E DESEMPENHO ACADÊMICO NA UNIVERSIDADE AUTÔNOMA TOMÁS FRÍAS.

EDUCATIONAL DATA MINING (EDM) WITH RANDOM FOREST FOR STUDENT CLASSIFICATION ACCORDING TO LEARNING STYLES AND ACADEMIC PERFORMANCE AT THE UNIVERSIDAD AUTÓNOMA TOMÁS FRÍAS

Teófilo Hermógenes Mamani Ruiz

Lic. en Estadística (UATF), Ingeniero Agrónomo (UMSA), M.Sc. en Educación Superior (UATF),
M.Sc. en Preparación y Evaluación de Proyectos (UAB-JB),
M.Sc. Ciencia de Datos (UPEA), Ph.D. en Educación Superior (UMSA).
Docente Universidad Católica Boliviana y Universidad Autónoma Tomás Frías.
Experto en Análisis de datos y Modelos de Pronóstico.
ORCID: <https://orcid.org/0009-0004-0741-8001>
Lattes: <https://lattes.cnpq.br/5215685034639508>
E-mail: hermopolitan@hotmail.com

Efraín Rojas

Lic. en Estadística (UATF), Técnico Superior en Sistemas Informática (INCOS),
M.Sc. En Ingeniería Estadística (UATF), Diplomado en Educación Superior (UPEA),
Docente Universidad Autónoma Tomas Frías, Docente INFOCAL,
Experto en Métodos y técnicas de Investigación, Analista de Datos.
ORCID: <https://orcid.org/0009-0002-9256-7567>
Lattes: <https://lattes.cnpq.br/7332185992404793>
E-mail: efrajojasm@gmail.com

Resumen

La presente investigación aplicó técnicas de Educational Data Mining mediante el algoritmo Random Forest para clasificar a estudiantes de la Universidad Autónoma Tomás Frías según sus estilos de aprendizaje y predecir su rendimiento académico, con el fin de contribuir al diseño de estrategias pedagógicas personalizadas desde un enfoque constructivista. Se adoptó un enfoque cuantitativo, aplicado, correlacional y predictivo, con diseño no experimental y transversal. La muestra estuvo compuesta por 220 estudiantes de cuatro carreras, a quienes se aplicó el Integrated Learning Styles Questionnaire (ILSQ), con alta confiabilidad ($\alpha = 0.92$), durante el semestre I-2025. Los resultados evidenciaron predominancia de los estilos pragmático (22.27%) y múltiple (27.27%), especialmente en Ingeniería Comercial e

Ingeniería de Sistemas, lo que refleja una tendencia hacia el aprendizaje aplicado y adaptativo. El modelo Random Forest Classifier alcanzó un desempeño satisfactorio (accuracy = 0.79; F1 = 0.77), mostrando alta eficacia en la clasificación de estilos observables como el múltiple, visual y kinestésico. Sin embargo, el modelo predictivo de rendimiento académico (Random Forest Regressor) obtuvo un $R^2 = 0.21$, evidenciando que los estilos de aprendizaje no son predictores sólidos de la calificación final, debido a la influencia multifactorial de variables cognitivas, motivacionales y contextuales. Pedagógicamente, estos hallazgos indican que la personalización del aprendizaje debe ir más allá de los estilos individuales, integrando factores emocionales y contextuales. Se concluye que la minería de datos educativos es útil para diagnosticar perfiles cognitivos y orientar estrategias didácticas diferenciadas, consolidando un modelo de enseñanza universitaria basado en evidencia, flexibilidad y equidad educativa.

Palabras clave: Educational data mining; random forest; estilos de aprendizaje; rendimiento académico; personalización del aprendizaje.

Resumo

A presente pesquisa aplicou técnicas de Educational Data Mining – EDM por meio do algoritmo Random Forest para classificar estudantes da Universidade Autônoma Tomás Frías de acordo com seus estilos de aprendizagem e prever seu desempenho acadêmico, com o objetivo de contribuir para o desenvolvimento de estratégias pedagógicas personalizadas a partir de uma abordagem construtivista. Adotou-se uma abordagem quantitativa, aplicada, correlacional e preditiva, com delineamento não experimental e transversal. A amostra foi composta por 220 estudantes de quatro cursos, aos quais foi aplicado o Integrated Learning Styles Questionnaire (ILSQ), com alta confiabilidade ($\alpha = 0,92$), durante o semestre I-2025. Os resultados evidenciaram predominância dos estilos pragmático (22,27%) e múltiplo (27,27%), especialmente nos cursos de Engenharia Comercial e Engenharia de Sistemas, refletindo uma tendência para a aprendizagem aplicada e adaptativa. O modelo Random Forest Classifier apresentou desempenho satisfatório (accuracy = 0,79; F1 = 0,77), mostrando alta eficácia na classificação de estilos observáveis como múltiplo, visual e cinestésico. Entretanto, o modelo preditivo de desempenho acadêmico (Random Forest Regressor) obteve $R^2 = 0,21$, evidenciando que os estilos de aprendizagem não são preditores sólidos da nota final, devido à influência multifatorial de variáveis cognitivas, motivacionais e contextuais. Pedagógicamente, esses achados indicam que a personalização da aprendizagem deve ir além dos estilos individuais, integrando fatores emocionais e contextuais. Conclui-se que a mineração de dados educacionais é útil para diagnosticar perfis cognitivos e orientar estratégias didáticas diferenciadas, consolidando um modelo de ensino universitário baseado em evidências, flexibilidade e equidade educacional.

Palavras-chave: Mineração de dados educacionais; random forest; estilos de aprendizagem; desempenho acadêmico; personalização da aprendizagem.

Abstract

This research applied Educational Data Mining techniques using the Random Forest algorithm to classify students from the Universidad Autónoma Tomás Frías according to their learning styles and to predict their academic performance, with the purpose of contributing to the design of personalized pedagogical strategies from a constructivist approach. A quantitative, applied, correlational, and predictive methodology was adopted, employing a non-experimental and cross-sectional design. The sample consisted of 220 students from four academic programs, who completed the Integrated Learning Styles Questionnaire (ILSQ), which demonstrated high reliability ($\alpha = 0.92$), during the first semester of 2025. The results revealed a predominance of the pragmatic (22.27%) and multiple (27.27%) learning styles, particularly among students in Business Engineering and Systems Engineering, reflecting a tendency toward applied and adaptive learning. The Random Forest Classifier achieved satisfactory performance (accuracy = 0.79; F1 = 0.77), showing high effectiveness in classifying observable styles such as multiple, visual, and kinesthetic. However, the predictive model of academic performance (Random Forest Regressor) obtained an $R^2 = 0.21$, indicating that learning styles are not strong predictors of final grades due to the multifactorial influence of cognitive, motivational, and contextual variables. Pedagogically, these findings suggest that learning personalization should go beyond individual styles by integrating emotional and contextual factors. It is concluded that Educational Data Mining is a valuable tool for diagnosing cognitive profiles and guiding differentiated instructional strategies, thereby strengthening a university teaching model based on evidence, flexibility, and educational equity.

Keywords: Educational data mining; learning styles; random forest; academic performance; personalized learning.

INTRODUCCIÓN

En el ámbito de la educación superior contemporánea, el constructivismo se erige como una de las bases epistemológicas más influyentes, porque concibe el aprendizaje como un proceso activo donde el estudiante construye conocimiento a partir de su experiencia, su interpretación e interacción con el entorno. Como argumenta Piaget (1970), “conocer no es copiar la realidad, sino actuar sobre ella y transformarla” (p. 14), mientras que Bruner (1997), enfatiza que el aprendizaje se fortalece con la participación activa del estudiante en los procesos de descubrimiento. En este sentido, los estilos de aprendizaje representan una expresión de la diversidad cognitiva, posibilitando la comprensión de la forma como los estudiantes perciben y procesan la información. Kolb (1984) y Alonso, Gallego y Honey (1994), estructuran tipologías que visibilizan esta heterogeneidad, favoreciendo el diseño de

estrategias pedagógicas diferenciadas en función de los intereses, formas de pensar y necesidades del estudiantado.

Paralelamente, la minería de datos educativos EDM, según Romero y Ventura (2020), emerge como un recurso estadístico innovador para el análisis de grandes volúmenes de información, al descubrir patrones que optimizan la enseñanza y el aprendizaje. Entre estas herramientas, el algoritmo Random Forest, sobresale según Breiman (2001), por su capacidad de clasificación y predicción, mostrando alta eficacia en la identificación de perfiles estudiantiles y la predicción del rendimiento académico. En la Universidad Autónoma Tomás Frías, este enfoque se articula con la perspectiva constructivista para generar diagnósticos pedagógicos precisos, que guíen intervenciones diferenciadas y fortalezcan la calidad educativa en un entorno caracterizado por la diversidad.

Por otra parte, en la educación superior, es persistente la escasa personalización de los procesos de enseñanza-aprendizaje frente a la creciente diversidad cognitiva, cultural y académica del estudiantado. En la Universidad Autónoma Tomás Frías, los modelos pedagógicos tradicionalmente homogéneos dificultan la atención efectiva a las particularidades individuales, principalmente respecto a los estilos de aprendizaje y las trayectorias de rendimiento académico. De acuerdo con Bruner (1997) y Piaget (1970), desde la perspectiva constructivista, aprender implica construir activamente el conocimiento a partir de las experiencias, intereses y capacidades de cada estudiante. De esta forma, resulta imprescindible contar con herramientas para identificar y clasificar perfiles de aprendizaje sustentados en datos objetivos y sistemáticos. En esa línea, el desarrollo de estrategias pedagógicas adaptativas requiere una comprensión profunda de cómo aprenden los estudiantes y del uso de tecnologías analíticas que permitan descubrir patrones útiles para la toma de decisiones didácticas.

La minería de datos educativos y los algoritmos de aprendizaje automático, como Random Forest, se erigen en una valiosa alternativa para la clasificación automatizada de estudiantes en función de múltiples variables. Sin embargo, en la UATF no existen estudios que integren la evaluación de estilos de aprendizaje con el análisis predictivo de rendimiento académico utilizando modelos de machine learning. Esta carencia limita la posibilidad de

diseñar intervenciones pedagógicas diferenciadas, dificultando la mejora sostenida de la calidad educativa. Es así que, se plantea como problema de investigación: ¿De qué manera la aplicación de técnicas de Educational Data Mining con el algoritmo Random Forest permite clasificar a los estudiantes de la Universidad Autónoma Tomás Frías según sus estilos de aprendizaje y predecir su desempeño académico, para contribuir al diseño de estrategias pedagógicas personalizadas desde un enfoque constructivista?

Planteándose como objetivo general aplicar técnicas de Educational Data Mining con el algoritmo Random Forest para clasificar a estudiantes de la Universidad Autónoma Tomás Frías según sus estilos de aprendizaje y predecir su rendimiento, con el fin de contribuir al diseño de estrategias pedagógicas personalizadas desde un enfoque constructivista. Asimismo, se han planteado las hipótesis H_{o1} : Existe relación significativa entre los estilos de aprendizaje identificados y el rendimiento académico de los estudiantes; H_{o2} : El modelo Random Forest predice con alta precisión los perfiles de estudiantes combinando variables de estilo de aprendizaje.

MARCO TEÓRICO

a) Estilos de aprendizaje- Los estilos de aprendizaje han sido abordados desde diversas perspectivas teóricas, reconociéndose como un campo multidisciplinario que vincula la psicología cognitiva, la pedagogía y las neurociencias. Según Keefe (1988), los estilos de aprendizaje son “las características cognitivas, afectivas y fisiológicas que sirven como indicadores relativamente estables de cómo los alumnos perciben, interactúan y responden a sus ambientes de aprendizaje” (p. 43), es decir que, los estudiantes difieren en cuanto al contenido que aprenden y en cómo lo aprenden. En este sentido, los estilos de aprendizaje constituyen una herramienta importante para comprender la diversidad estudiantil y orientar el diseño de estrategias pedagógicas diferenciadas. Es así que, diversos autores han propuesto clasificaciones de estilos de aprendizaje basadas en criterios sensoriales, cognitivos o actitudinales.

Uno de los modelos más influyentes y aplicados en contextos universitarios para comprender las preferencias cognitivas del estudiantado es el Modelo de Aprendizaje Experiencial, propuesto por David Kolb, basado en la premisa de que el aprendizaje es un proceso continuo, donde el conocimiento se crea mediante la transformación de la experiencia, distinguiéndose dos dimensiones principales: la percepción (experiencia concreta vs. conceptualización abstracta) y el procesamiento (experimentación activa vs. observación reflexiva), de cuya combinación surgen cuatro estilos: convergente, divergente, asimilador y acomodador. Según Kolb (1984), “el estilo de aprendizaje se determina por la forma en que una persona prefiere percibir y procesar la información” (p. 41).

Por otro lado, el modelo de Fleming y Mills (1992), conocido como VARK, posee un enfoque sensorial del aprendizaje, identificando cuatro canales preferentes: Visual, Aural (auditivo), Read/Write (lectura/escritura) y Kinestésico. A diferencia de Kolb, que se centra en procesos cognitivos profundos, VARK se enfoca en las formas sensoriales mediante las cuales el estudiante prefiere recibir información. Para Fleming (2001), “reconocer los estilos sensoriales permite adaptar la presentación del contenido para mejorar la retención y comprensión” (p. 9). Esta herramienta es muy utilizada por su aplicabilidad práctica.

Por su parte, Alonso, Gallego y Honey, adaptaron el modelo de Honey y Mumford propuesto en 1986, dando lugar al instrumento CHAEA (Cuestionario Honey-Alonso de Estilos de Aprendizaje), el cual clasifica a los estudiantes en cuatro categorías: activo, reflexivo, teórico y pragmático, estos estilos no son excluyentes, sino que coexisten en cada individuo en distintas proporciones. Según Alonso et al. (1994), el conocimiento del estilo predominante permite proponer actividades acordes al modo de aprender de estudiantado, favoreciendo la personalización del proceso de enseñanza-aprendizaje. A su vez, Gregorc (1982), clasificó los estilos según dos ejes: abstracto-concreto y secuencial-aleatorio, dando lugar a cuatro tipos: abstracto secuencial, abstracto aleatorio, concreto secuencial y concreto aleatorio, esta clasificación se enfoca en los modos de procesamiento de la información. A su vez, Felder y Silverman (1988), propusieron una tipología útil para la enseñanza en ingeniería, que considera dimensiones como activo-reflexivo, sensorial-intuitivo, visual-

verbal y secuencial–global. Según Felder (1993), los estudiantes aprenden más eficazmente cuando la enseñanza se alinea con su estilo preferido.

En la misma línea, Howard Gardner introdujo la teoría de las inteligencias múltiples, aunque no es estrictamente un modelo de estilos de aprendizaje, ha influido significativamente en la clasificación y adaptación pedagógica. Gardner (1983), identifica ocho inteligencias distintas: lingüística, lógico-matemática, musical, espacial, corporal-kinestésica, interpersonal, intrapersonal y naturalista, sugiriendo que cada estudiante tiene un perfil intelectual único. Esta teoría refuerza la necesidad de personalizar la enseñanza y reconocer la diversidad como una fortaleza educativa.

Sintéticamente, estos modelos de estilos de aprendizaje coinciden en señalar que no existe un estilo único que garantice el aprendizaje óptimo para todos los estudiantes, esto implica un reto para el docente. La educación superior, particularmente la Universidad Autónoma Tomás Frías, debe avanzar hacia enfoques pedagógicos diferenciados, donde los estilos de aprendizaje se diagnostiquen y se integren activamente en la planificación didáctica. Esta comprensión se convierte en un paso fundamental hacia la personalización del aprendizaje, principio clave del enfoque constructivista.

b) Rendimiento académico- El rendimiento académico es el indicador más utilizado para evaluar el progreso del estudiante y la eficacia del sistema educativo, el cual, Según Álvarez Méndez (2001), se entiende como el grado de cumplimiento de los objetivos de aprendizaje, medido con las calificaciones, tasa de aprobación, promedio ponderado o indicadores estandarizados. El rendimiento refleja el dominio de contenidos curriculares y las habilidades cognitivas, socioemocionales y organizativas, desarrolladas durante el proceso educativo y se constituye en una variable crítica para predecir la permanencia y éxito del estudiante. Sin embargo, el rendimiento está influenciado por múltiples factores personales, familiares, institucionales y contextuales. Robbins et al. (2004), señalan como factores determinantes del rendimiento a la motivación, el compromiso, las habilidades de estudio, el apoyo social y la autoeficacia. En la misma línea, Astin (1999), sostiene que uno de los predictores más sólidos es el involucramiento activo del estudiante en el proceso

educativo. González-Pienda et al. (2002), añaden a esto, el nivel socioeconómico, el entorno familiar, la salud mental y el acceso a recursos tecnológicos.

La relación entre estilos de aprendizaje y desempeño académico ha sido objeto de numerosos estudios, Felder y Brent (2005), detectaron que la adecuación entre el enfoque pedagógico docente y las preferencias cognitivas estudiantiles mejora significativamente los resultados. A su vez Cano y Justicia (1994), encontraron correlaciones positivas entre estilos reflexivos y teóricos con un rendimiento académico más elevado, especialmente en contextos de alta exigencia cognitiva. Por su parte, Kolb (1984), argumenta que los estudiantes que logran integrar distintos modos de aprendizaje tienden a alcanzar mejores niveles de comprensión y aplicación del conocimiento.

Finalmente, la comprensión de los factores asociados al rendimiento académico y su medición rigurosa es esencial para implementar políticas educativas efectivas. Como indica Cabrera et al. (2006), su seguimiento sistemático permite identificar perfiles de riesgo, diseñar intervenciones pedagógicas diferenciadas y mejorar los índices de retención y titulación. Así, la incorporación de modelos predictivos como Random Forest, contribuirá a mejorar la precisión del diagnóstico, fundamentando la toma de decisiones basada en datos para optimizar los procesos de enseñanza-aprendizaje.

c) Educational Data Mining EDM- La minería de datos educativos, es una disciplina emergente que aplica técnicas de minería de datos, estadística y aprendizaje automático, para analizar grandes volúmenes de datos generados en contextos educativos. Romero y Ventura (2020), la definen como "el proceso de descubrir patrones y conocimientos útiles a partir de grandes cantidades de datos educativos" (p. 601). La EDM se ha consolidado como una herramienta esencial en la educación superior, donde se genera grandes cantidades de datos heterogéneos, para transformarlos en información significativa, según Baker y Siemens (2014), esto que permite mejorar sistemáticamente la calidad educativa. Asimismo, Bienkowski, Feng y Means (2012), afirman que permite identificar tendencias de rendimiento, predecir resultados y adaptar contenidos a las necesidades individuales de los estudiantes.

El desarrollo de la EDM se encuentra fuertemente vinculado con la evolución de la inteligencia artificial y los sistemas de gestión del aprendizaje (LMS). La minería de datos educativos es un campo interdisciplinario que contribuye de significativamente a la toma de decisiones educativas basada en evidencia, fortaleciendo la planificación institucional y la intervención pedagógica personalizada. En contraste con la analítica educativa tradicional, la EDM analiza indicadores de rendimiento y también comportamientos digitales, patrones de interacción y estilos de aprendizaje, puesto que como indica Peña-Ayala (2014), la EDM puede integrarse a enfoques pedagógicos constructivistas para generar ambientes de aprendizaje más inclusivos y adaptativos.

Por otro lado, las herramientas empleadas en la minería de datos educativos varían desde entornos estadísticos tradicionales hasta plataformas especializadas en aprendizaje automático, como: WEKA para aplicar algoritmos de clasificación, regresión, agrupamiento y reglas de asociación; RapidMiner, para análisis predictivo, el desarrollo de modelos para predicción del abandono, clasificación de estudiantes y análisis de participación en LMS; Softwares de código abierto, como Python y R, con bibliotecas poderosas como Scikit-learn, Caret, XGBoost y TensorFlow, permiten desarrollar modelos avanzados de minería de datos, como afirma VanderPlas (2016), Python es un lenguaje popular para EDM por su versatilidad, comunidad activa y capacidad de automatización; Tableau y Power BI, en la visualización de datos, permiten transformar resultados complejos en paneles gráficos comprensibles, facilitando la toma de decisiones; Finalmente, plataformas educativas como Moodle, Canvas o Blackboard, integran funciones para la extracción de datos para capturar interacciones, tiempos de conexión, actividad en foros y progreso en tareas.

d) Aplicaciones en educación- Entre las aplicaciones de la minería de datos educativos en el ámbito universitario destacan: la predicción del rendimiento académico, para identificar estudiantes en riesgo de abandono, Kotsiantis et al. (2010), reportan modelos predictivos con una precisión superior al 85% en la detección temprana de bajo rendimiento; la clasificación de perfiles de aprendizaje, para adaptar contenidos y metodologías a las características individuales del estudiante, Romero et al. (2008) utilizaron árboles de decisión para categorizar estudiantes según su participación en foros y evaluaciones; la detección de

patrones de interacción, para estudiar navegación en plataformas virtuales; el diseño de cursos y currículos, para analizar la efectividad de actividades, recursos y secuencias didácticas; y la educación personalizada, para crear sistemas de adaptativos, con contenidos, actividades y evaluaciones según las características de cada estudiante.

e) Algoritmo Random Forest.- Random Forest fue propuesto por Breiman en 2001, es un método de aprendizaje supervisado basado en conjuntos de árboles de decisión, combina sus resultados para mejorar la precisión y estabilidad del modelo. De acuerdo con Breiman (2001), cada árbol se entrena con una muestra aleatoria del conjunto de datos y utiliza un subconjunto aleatorio de variables para dividir los nodos, esto introduce diversidad y reduce el sobreajuste. Este algoritmo se basa en el principio del bagging o bootstrap aggregating, que consiste según James et al. (2013), en entrenar cada árbol sobre una muestra con reemplazo del conjunto original y luego promediar sus predicciones en tareas de regresión o usar el voto mayoritario en tareas de clasificación. La robustez de este enfoque ha sido ampliamente validada en diversos contextos educativos y científicos.

Una de las fortalezas del algoritmo es su capacidad para manejar grandes conjuntos de datos con alta dimensionalidad, así como para captar relaciones no lineales entre variables. Hastie, Tibshirani y Friedman (2009), destacan su utilidad en contextos donde existen muchas variables explicativas, incluso si muchas de ellas no tienen relevancia directa, característica que lo convierte en una herramienta idónea para contextos educativos complejos. En concordancia con Lantz (2019), Random Forest también permite estimar la importancia de las variables predictoras, permitiendo identificar los factores que más influyen en el comportamiento del estudiante. Esta capacidad explicativa permite diseñar intervenciones pedagógicas basadas en evidencia. En resumen, Random Forest es un modelo versátil, preciso y explicativo, ideal para clasificación educativa como la predicción del rendimiento académico o la identificación de estilos de aprendizaje.

Entre las ventajas de Random Forest se encuentra: Su capacidad de generalización, permite obtener modelos más robustos y confiables incluso en presencia de datos ruidosos o incompletos, según Breiman (2001), utilizar múltiples árboles, reduce el riesgo de sobreajuste, un problema frecuente en los modelos de decisión simples; Permite manejar

variables categóricas y numéricas sin transformaciones complejas, como indican Kuhn y Johnson (2013), esta flexibilidad es ideal para entornos educativos donde las variables son muy heterogéneas; Proporciona una medida interna de validación, conocida como out-of-bag error, el cual, de acuerdo con James et al. (2013), estima el error de generalización sin necesidad de dividir el conjunto de datos en entrenamiento y prueba, esto es útil en estudios con muestras reducidas; Es capaz de detectar interacciones complejas entre variables y manejar valores faltantes, como señala Han, Kamber y Pei (2012), esto lo hace superior a otros métodos lineales en contextos donde los datos no cumplen los supuestos clásicos. Estas cualidades lo posicionan como una técnica poderosa en la minería de datos educativos, siendo ampliamente aplicada en varios lenguajes y entornos de programación.

Random Forest se ha utilizado en variedad de investigaciones educativas, Ramesh et al. (2014), lo aplicaron en la predicción del rendimiento académico, para detectar estudiantes en riesgo de reprobación. Baradwaj y Pal (2011), lo utilizaron para identificar estilos de aprendizaje, segmentar a los estudiantes y adaptar la instrucción a sus preferencias. Dekker, Pechenizkiy y Vleeshouwers (2009), lo emplearon en la detección y predicción temprana de la deserción universitaria, encontrando que el modelo superaba a otros algoritmos como SVM y redes neuronales. Asimismo, Ahmed y Elaraby (2014), en la evaluación automatizada de tareas y en el análisis del comportamiento en entornos virtuales de aprendizaje, para predecir el nivel de participación en Moodle.

f) Investigaciones previas relevantes- Los estudios sobre la aplicación de minería de datos y algoritmos de clasificación en educación han crecido exponencialmente en la última década. Romero, Ventura y García (2008), utilizaron árboles de decisión y redes bayesianas para analizar datos de Moodle, mejorando la toma de decisiones. Por su parte, Peña-Ayala (2014), revisó exhaustivamente más de 240 artículos sobre EDM y concluyó que los algoritmos como Random Forest son efectivos para clasificación de rendimiento y deserción. Por otro lado, en un estudio longitudinal, Tempelaar, Rienties y Giesbers (2015), predijeron con EDM el desempeño estudiantil en base a variables conductuales, cognitivas y emocionales. Thai-Nghe, Drumond y Schmidt-Thieme (2011), desarrollaron un sistema de recomendación para personalizar actividades de aprendizaje en función del perfil de cada

estudiante, mejorando significativamente la retención de contenido. También Ramesh et al. (2014), utilizaron Random Forest para clasificar patrones de participación y predicción de abandono en plataformas MOOC. Asimismo, García, Romero y Ventura (2011), predijeron estilos a partir de los registros de actividad de los estudiantes. Estos estudios coinciden en que el uso de EDM y Random Forest permiten mejorar la calidad educativa mediante estrategias pedagógicas basadas en evidencia, personalización del aprendizaje y toma de decisiones informada.

En el ámbito latinoamericano y particularmente en Bolivia, los estudios sobre EDM y Random Forest son aún incipientes. En México, García Peñalvo y Rebaque (2016), aplicaron EDM con árboles de decisión y Random Forest, para clasificar estudiantes según su nivel de compromiso académico e intervenir a tiempo con tutorías personalizadas. En Perú, Vargas y Yupanqui (2019), utilizaron algoritmos de clasificación para identificar factores asociados al bajo rendimiento. En Bolivia, Mamani y López (2022), utilizaron EDM para identificar factores de riesgo académico y con Random Forest detectaron que la falta de orientación vocacional y la escasa participación en actividades extracurriculares eran buenos predictores del rendimiento académico. Por su parte, Chura, Velásquez y Callisaya (2023), evaluaron la relación entre la asistencia a clases, tipo de evaluación y desempeño académico, detectando que la participación activa y exposiciones grupales generaban mejor rendimiento. Esto muestra que la aplicación de EDM y Random Forest en contextos educativos tienen alto potencial para la toma de decisiones pedagógicas, la mejora del rendimiento y el diseño de estrategias educativas personalizadas.

METODOLOGÍA

a) Enfoque, tipo y diseño de investigación- El presente estudio se enmarca en el paradigma positivista, dado que busca generar evidencia empírica mediante la aplicación del Cuestionario Integrado de Estilos de Aprendizaje (Integrated Learning Styles Questionnaire ILSQ) y el análisis de datos mediante técnicas estadísticas y algoritmos de minería de datos. Adopta un enfoque cuantitativo porque trabaja con información numérica derivada de los puntajes del instrumento y las calificaciones de los estudiantes, para establecer relaciones

entre estilos de aprendizaje y rendimiento académico. Este enfoque posibilita la medición objetiva de las variables, la validación de instrumentos a través del coeficiente alfa de Cronbach y la construcción de modelos predictivos.

La tipología de esta investigación se clasifica como aplicada, correlacional y predictiva. Es aplicada porque busca resolver un problema educativo concreto, la clasificación de estilos de aprendizaje para diseñar estrategias pedagógicas personalizadas. Es correlacional porque analiza la relación existente entre los estilos de aprendizaje y el rendimiento académico. Finalmente, es predictiva porque incorpora técnicas de Educational Data Mining y el algoritmo Random Forest, para modelar y pronosticar el rendimiento futuro de los estudiantes, en función de su perfil de aprendizaje, evaluando las métricas de precisión del modelo. El diseño adoptado es no experimental, transversal y de tipo explicativo-predictivo, ya que no se manipulan variables, sino que se observan y analizan en un único momento, con el propósito de comprender y anticipar patrones educativos relevantes.

b) Población y muestra- Se ha considerado estudiantes de la Universidad Autónoma Tomás Frías (UATF), específicamente en las carreras de Ingeniería Comercial, Sistemas, Diseño y Programación y Estadística, esta selección responde a un criterio metodológico para capturar la diversidad cognitiva, metodológica y profesional existente en el ámbito universitario, es decir, la representatividad de distintos perfiles académicos. Ingeniería Comercial representa un campo eminentemente aplicativo, centrado en la gestión, toma de decisiones y la resolución de problemas en contextos reales; Estadística, con una base teórica y analítica sólida, propia de las ciencias formales y cuantitativas; e Ingeniería de Sistemas como un punto intermedio, integrando el pensamiento lógico, la abstracción matemática y la aplicación tecnológica, por último, Diseño y Programación, ligado más a la creatividad y tecnología.

Esto permitió analizar cómo varían los estilos de aprendizaje y el rendimiento académico en función de las exigencias cognitivas y procedimentales de cada área. La heterogeneidad de estos programas garantiza la representatividad de distintos perfiles estudiantiles, fortaleciendo la validez externa del modelo de clasificación, al aplicar el enfoque constructivista en contextos académicos contrastantes dentro de la misma

institución. Se aplicó un muestreo estratificado con una muestra de 220 estudiantes (151 de Ingeniería Comercial, 41 en Sistemas, 9 en Diseño y Programación y 19 en Estadística), durante el periodo académico I-2025, esto proporcionó datos suficientes y balanceados para la construcción del modelo de clasificación y predicción mediante Random Forest.

c) Técnicas e instrumentos de recolección de datos- Se utilizó el test de estilos de aprendizaje Integrated Learning Styles Questionnaire ILSQ, estructurado a partir de la integración conceptual y metodológica de la prueba VARK de Fleming y Mills y el CHAEA, de Alonso, Gallego y Honey. Este test, conformado por 48 ítems en escala de Likert de cinco opciones captura las preferencias sensoriales y cognitivas de los estudiantes, se validó mediante el coeficiente alfa de Cronbach, cuyo valor α de 0.92, indica alta consistencia interna. En tanto que, el test-retest reportó un coeficiente de correlación r de 0.701, entre ambas mediciones, con alta significación estadística ($p < 0.01$), reflejando la estabilidad de las respuestas en ambos momentos de aplicación. Esto avala la pertinencia psicométrica del test como un instrumento confiable para la evaluación del constructo estilos de aprendizaje en el contexto universitario. Por otro lado, el desempeño académico se obtuvo a partir de los registros oficiales de calificaciones disponibles en las actas de las asignaturas de cada carrera, garantizando la validez y el rigor metodológico del estudio.

d) Procesamiento de datos- Se desarrolló en las siguientes fases: Limpieza y depuración de la base de datos, eliminando registros incompletos, duplicados o inconsistentes, descartándose los casos con datos faltantes; Normalización, los valores de rendimiento académico se transformaron a escalas comparables, para reducir la influencia de magnitudes extremas y mejorar la estabilidad del algoritmo de clasificación, se aplicó la técnica min-max, que ajusta los valores a un rango de 0 a 1, facilitando la convergencia del modelo de aprendizaje; Codificación, la variable categórica estilo de aprendizaje, fue codificada mediante variables dummy para ser utilizada como predictor, puesto que los algoritmos de aprendizaje supervisado requieren representaciones numéricas para procesar correctamente las categorías, esto permitió a Random Forest Regressor manejar la información categórica en el entrenamiento.

En cuanto a la aplicación del modelo Random Forest, se realizó en Python en el entorno Jupyter Notebook de Anaconda e incluyó las siguientes fases: Entrenamiento, el modelo se entrenó utilizando la base de datos preprocesada, combinando múltiples clasificadores débiles en un clasificador fuerte, concretamente se generaron 500 árboles de decisión con muestreo aleatorio Bootstrap; Validación cruzada, para evitar el sobreajuste y asegurar la generalización, se utilizó la técnica k-fold (k=10), el cual permitió evaluar la estabilidad del modelo a lo largo de diferentes subconjuntos de datos; Selección de hiperparámetros, para su optimización, como la profundidad máxima de los árboles, el número mínimo de muestras por nodo y la cantidad de características seleccionadas en cada división, eso según Hastie, Tibshirani y Friedman (2009), permite maximizar el equilibrio entre sesgo y varianza en el aprendizaje del modelo, la optimización se realizó mediante búsqueda en cuadrícula grid search.

e) Métricas de evaluación del modelo- Se consideraron las siguientes métricas: Accuracy o precisión, que representa la proporción de predicciones correctas sobre el total de observaciones evaluadas; Recall, sensibilidad o tasa de verdaderos positivos, mide la capacidad del modelo para identificar correctamente a las instancias positivas dentro del conjunto de datos; F1-Score, referida a la media armónica entre la precisión y el recall, proporciona un balance entre ambas métricas; y la Matriz de confusión, que muestra la distribución de predicciones correctas e incorrectas en función de las clases reales, permitiendo identificar errores de tipo I y II. El modelo Random Forest, según Breiman (2001), permite calcular la importancia de cada variable predictora en el proceso de clasificación, basada en la reducción de impureza de Gini o en la disminución del error de predicción. Louppe (2014), afirma que una variable es considerada relevante si su peso supera el 5% de la varianza explicada respecto al total de predictores, mientras que aquellas con valores cercanos a cero se consideran prescindibles en la construcción del modelo.

RESULTADOS

a) Clasificación por estilo de aprendizaje- Se evidencian diferencias significativas que reflejan la naturaleza disciplinar de cada carrera y los estilos de aprendizaje preferentes

de sus estudiantes, donde los estilos pragmático y múltiple predominan en casi todas las carreras. Desde una perspectiva pedagógica, esto denota la necesidad de diseñar estrategias formativas basadas en proyectos, simulaciones y aprendizaje experiencial, pues como señalan Honey y Mumford (1992) y Kolb (1984), responden mejor a la diversidad cognitiva evidenciada en la muestra. Resalta en Estadística que el 36.84% tiene un estilo pragmático, mientras que en Ingeniería Comercial el 29.14% tiene estilo múltiple. Este patrón sugiere una inclinación hacia el aprendizaje aplicado, contextual y adaptativo, propio de estudiantes que prefieren resolver problemas reales, integrar diversos enfoques y vincular la teoría con la práctica. Por otro lado, los estilos visual, auditivo y kinestésico, exhiben presencia moderada, pero diferenciada por carrera.

Ingeniería Comercial concentra 68.64% del total, mostrando alta diversidad y predominancia de los estilos múltiple 29.14%, pragmático 23.18%, kinestésico 11.26% y visual 11.26%, es decir que, en esta carrera los estudiantes se inclinan hacia el aprendizaje aplicado, contextual, adaptativo, la experimentación y observación, característico de estudiantes que prefieren resolver problemas reales, integrar diversos enfoques y vincular la teoría con la práctica, lo cual concuerda con el enfoque aplicado y estratégico del campo empresarial. En el caso de Ingeniería de Sistemas, que representa el 18.64% de la muestra, predomina el estilo múltiple con 29.27%, seguido por el estilo reflexivo 14.63% y pragmático 12.20%, esta configuración revela una sinergia entre pensamiento analítico, reflexión estructurada y la aplicación práctica del conocimiento, rasgos que caracterizan los procesos cognitivos implicados en la programación, la modelación de sistemas y la resolución de problemas técnicos complejos. Desde una perspectiva epistemológica, esta combinación indica que estos estudiantes tienden a integrar diversas modalidades cognitivas, adaptando su forma de aprender según las demandas del contexto y la naturaleza del desafío tecnológico que enfrentan.

La carrera de Estadística representa el 8.64% de la muestra, tiene mayor proporción en los estilos visual con 10.53%, auditivo 10.53% y teórico 10.53%, evidenciando una marcada preferencia por el razonamiento abstracto, el análisis conceptual y la estructuración lógica del conocimiento. Este patrón de aprendizaje es coherente con la naturaleza de la

disciplina, donde la comprensión visual de datos, la interpretación analítica y la formulación teórica son esenciales para el tratamiento estadístico de la información. Por su parte, la Ingeniería de Diseño y Programación, que representa el 4.09%, presenta predominio de los estilos activo con 33.33%, visual 22.22% y pragmático 22.22%, reflejando un perfil de aprendizaje orientado a la experimentación, la creatividad aplicada y la resolución práctica de problemas, atributos fundamentales en los procesos de diseño digital y desarrollo de software. Desde una perspectiva cognitiva, este grupo evidencia una fuerte tendencia hacia el aprendizaje experiencial y la acción constructiva, donde el conocimiento se genera mediante la práctica, la exploración y la adaptación continua al entorno tecnológico.

Tabla 1. Clasificación de los estilos de aprendizaje por carreras

Estilo de Aprendizaje	Carrera				Media General
	Estadística	Ing. Diseño y Progr.	Ing. Comercial	Ing. Sistemas	
E. Visual	10.53	22.22	11.26	4.88	10.45
E. Auditivo	10.53	0.00	5.96	9.76	6.82
E. Lectura-Escritura	0.00	11.11	2.65	7.32	3.64
E. Kinestésico	5.26	11.11	11.92	7.32	10.45
E. Activo	5.26	33.33	1.32	7.32	4.09
E. Reflexivo	0.00	0.00	8.61	14.63	8.64
E. Teórico	10.53	0.00	5.96	7.32	6.36
E. Pragmático	36.84	22.22	23.18	12.20	22.27
E. Múltiple	21.05	0.00	29.14	29.27	27.27
Media General	8.64	4.09	68.64	18.64	100

Fuente: Elaboración propia, 2025.

Comparativamente, las cuatro carreras muestran perfiles de aprendizaje diferenciados. En Estadística, predominan los estilos visual, auditivo y teórico, propios de aprendices asimiladores y teóricos, orientados al análisis conceptual y la abstracción. En Ingeniería de Diseño y Programación, destacan los estilos activo, visual y pragmático, vinculados con aprendices convergentes y pragmáticos, que aprenden con la experimentación y la aplicación creativa del conocimiento. En Ingeniería Comercial, los estilos múltiple, pragmático y kinestésico reflejan un perfil acomodador, caracterizado por la acción, la

adaptabilidad y la orientación a resultados. Finalmente, en Ingeniería de Sistemas, la combinación de estilos múltiple, reflexivo y pragmático sugiere un perfil divergente y reflexivo, que integra observación analítica y aplicación técnica. Estos resultados evidencian la necesidad de estrategias pedagógicas diferenciadas, que articulen la reflexión teórica con la acción práctica, promoviendo un aprendizaje integral y contextualizado.

Con respecto a las tendencias globales de estilos de aprendizaje, existe prevalencia del estilo pragmático con 22.27%, múltiple 27.27%, kinestésico 10.45% y visual con 10.45%. Esto señala que la mayoría de los estudiantes tiende a aprender mediante la práctica y la aplicación directa del conocimiento, así como integrando múltiples modalidades cognitivas, en efecto, el 27.27% con estilo múltiple sugiere una alta flexibilidad cognitiva, es decir, estudiantes capaces de adaptarse a diferentes entornos y estrategias pedagógicas.

Es evidente la diversidad cognitiva significativa entre carreras, siendo la pragmaticidad y el aprendizaje múltiple los rasgos más destacados. Este hallazgo respalda la necesidad de modelos educativos flexibles y personalizados, coherentes con el enfoque del aprendizaje centrado en el estudiante, promovido por autores como Kolb (1984) y Tomlinson (2001). La enseñanza universitaria debe diversificar sus estrategias didácticas, combinando métodos expositivos, actividades prácticas y recursos visuales, para atender la heterogeneidad cognitiva. Particularmente en Ingeniería Comercial, debe aplicarse un enfoque basado en casos, simulaciones y aprendizaje experiencial, mientras que, en Estadística, una enseñanza con énfasis en la visualización de datos y razonamiento analítico. En Sistemas y Programación, estrategias mixtas, que combinen reflexión y aplicación práctica. Como señalan Felder y Silverman (1988), el aprendizaje mejora cuando los métodos de enseñanza se alinean con los estilos predominantes del estudiante.

En cuanto al rendimiento del modelo Random Forest Classifier, se evidencia un desempeño general bueno en la clasificación de estilos de aprendizaje, con un accuracy global de 0.791, que refleja un desempeño general sólido y equilibrado, el modelo clasificó correctamente el 79,1% de los casos, lo cual, de acuerdo con Han, Pei y Kamber (2022), es notable en contextos educativos donde las variables psicológicas, cognitivas y contextuales, tienden a introducir ruido y ambigüedad en los datos. Este valor refleja la capacidad general

del modelo, aunque no proporciona información sobre cómo se comporta para cada estilo de aprendizaje específico. Respecto a la precisión ponderada, se obtuvo un valor de 0.827, señalando que el modelo predice correctamente un estilo de aprendizaje en 82.7% de las veces, consiguiendo la emisión de predicciones es confiable, especialmente para los estilos con mayor número de estudiantes. Esto evidencia que el modelo mantiene un control adecuado sobre los falsos positivos y clasifica correctamente la mayoría de los casos, sin atribuir erróneamente estilos que no corresponden. Esto es esencial en el Educational Data Mining, donde según Romero y Ventura (2020), las predicciones erróneas pueden conducir a diagnósticos pedagógicos inapropiados o estrategias de enseñanza mal dirigidas.

Por otro lado, el recall que mide la proporción de verdaderos positivos correctamente identificados, indica que el modelo posee buena sensibilidad para detectar los patrones auténticos de los estilos de aprendizaje, evitando la omisión de casos relevantes, el valor de recall ponderado es de 0.791, revelando que el modelo identifica correctamente al 79.1% de los estudiantes que realmente pertenecen a cada estilo. El modelo es bastante preciso, aunque podría no detectar a todos los estudiantes de ciertas clases con menos casos. Finalmente, el F1-Score ponderado es de 0.767, confirmando un adecuado equilibrio entre exactitud y cobertura de las predicciones, es decir, entre la capacidad del modelo para identificar correctamente los estilos y su habilidad para evitar clasificaciones erróneas. Este nivel de acierto indica que el modelo ha aprendido patrones relevantes y no triviales en los datos, logrando buena generalización en la predicción de los estilos de aprendizaje predominantes. En conjunto, estas métricas revelan que el modelo Random Forest Classifier constituye una herramienta fiable para la predicción de estilos de aprendizaje y el apoyo a la toma de decisiones educativas basadas en datos, consolidando la pertinencia del Educational Data Mining en la investigación pedagógica contemporánea.

Por otra parte, el modelo de clasificación presenta mejor rendimiento en el estilo múltiple (precision = 0.85, recall = 0.90, F1 = 0.87), esto es coherente con su mayor representación en la muestra y su relevancia dentro del conjunto de datos. Otros estilos, como el kinestésico y el aural, también muestran resultados favorables, mientras que los estilos teórico y reflexivo tienen menor desempeño, reflejando cierta dificultad del modelo para

distinguir patrones en estos grupos. En conjunto, los indicadores señalan que el clasificador logra una predicción robusta para las categorías dominantes, aunque requiere ajustes para mejorar la discriminación en estilos con menor frecuencia o con características más difusas.

Tabla 2. Clasificación del modelo Random Forest para predicción de estilos de aprendizaje

Estilo de Aprendizaje	Precision	Recall	F1-Score	Support
I. Estilo Visual	0.62	0.71	0.67	23
II. Estilo Aural/Auditivo	1.00	0.50	0.67	15
III. Estilo Lectura/Escritura	1.00	0.38	0.55	8
IV. Estilo Kinestésico	0.90	0.64	0.75	23
V. Estilo Activo	1.00	35.00	1.94	9
VI. Estilo Reflexivo	0.40	1.00	0.57	19
VII. Estilo Teórico	0.33	0.67	0.44	14
VIII. Estilo Pragmático	0.56	0.67	0.61	49
IX. Estilo Múltiple	0.85	0.90	0.87	60
Accuracy			0.79	220
Macro avg	0.61	0.58	0.55	220
Weighted avg	0.83	0.79	0.77	220

Fuente: Elaboración propia, 2025.

En términos pedagógicos, de acuerdo a la tabla 2, destaca el estilo múltiple, que tiene los mejores indicadores de desempeño (precision = 0.85, recall = 0.90, F1 = 0.87), reflejando su naturaleza integradora y la facilidad con que el modelo identifica patrones asociados a combinaciones de modalidades sensoriales y cognitivas. Igualmente, los estilos con componentes conductuales observables, como el visual (F1 = 0.67), kinestésico (F1 = 0.75) y pragmático (F1 = 0.61), exhiben buen nivel de clasificación, indicando que estos patrones son más fácilmente detectables por los algoritmos de aprendizaje automático. En contraste, el estilo reflexivo y teórico, de carácter más abstracto e introspectivo, presentan menor desempeño, indicando la necesidad de modelos con mayor sensibilidad semántica o híbridos que incorporen análisis textual, contextual o de comportamiento digital. Esta tendencia coincide con los reportes de Romero y Ventura (2020), respecto a que la minería de datos educativos logra mejor ajuste en la identificación de estilos basados en comportamientos

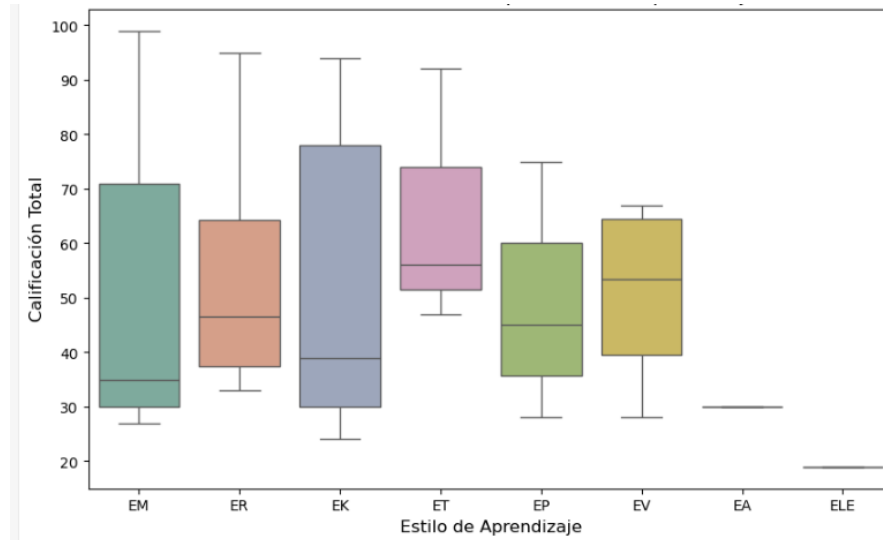
observables, mientras que los estilos de naturaleza cognitiva compleja requieren estrategias analíticas más adaptativas y profundas.

b) Modelo predictivo de la calificación final en función del estilo de aprendizaje.-

A través del modelo Random Forest Regressor, se estructuró un modelo estadístico para predecir la calificación total en función del estilo de aprendizaje, para lo cual la variable estilo de aprendizaje es codificado mediante variables dummy para ser utilizada como predictor. Luego, el conjunto de datos se divide en entrenamiento y prueba, ajustándose el modelo de Random Forest, que combina múltiples árboles de decisión para capturar relaciones complejas y no lineales entre los estilos y la calificación.

La evaluación del desempeño del modelo de Random Forest Regressor, se hizo mediante métricas de precisión que cuantifican la capacidad del modelo para explicar la variabilidad de las calificaciones y la magnitud de los errores de predicción, detectándose limitaciones en la capacidad predictiva, con un R^2 de 0.21, indica que el modelo explica de forma insuficiente la variabilidad de los datos y es menos eficiente que un modelo ingenuo basado en la media. Asimismo, los valores de Root Mean Squared Error RMSE de 23.16 y Mean Absolute Error MAE de 21.26 señalan desviaciones promedio considerables respecto a los valores observados, reflejando errores de predicción importantes. Por tanto, con la configuración actual y la estructura de los datos, Random Forest no proporciona inferencias robustas ni confiables para la predicción de calificaciones en función de los estilos de aprendizaje, señalando la necesidad de explorar alternativas metodológicas o incorporar variables adicionales que aumenten la explicación de la variabilidad observada.

Figura 2. Distribución de las calificaciones por estilo de aprendizaje



Fuente: Elaboración propia, 2025.

Pedagógicamente, los resultados de importancia de los estilos de aprendizaje (tabla 3), señalan que no todos los estilos influyen de la misma manera en el rendimiento académico. Los estilos de lectura-escritura y teórico, con mayor importancia en el modelo, podrían estar asociados con estrategias de aprendizaje más estructuradas o introspectivas que favorecen la adquisición de conocimientos de forma más consistente, mientras que los estilos pragmático y reflexivo, de menor peso, podrían representar enfoques de aprendizaje más flexibles o situacionales que impactan menos directamente en la calificación. Esto señala que, para mejorar el desempeño estudiantil, las intervenciones pedagógicas deberían enfocarse en fortalecer los estilos de aprendizaje más determinantes o en adaptar las estrategias de enseñanza para compensar la menor influencia de estilos menos predictivos, promoviendo así una educación más personalizada y efectiva.

Tabla 3. *Importancia de cada estilo en la predicción*

Estilo	Importancia
E. Lectura-Escritura	0.220847
E. Teórico	0.211505
E. Múltiple	0.161463
E. Kinestésico	0.156595
E. Reflexivo	0.119796
EP	0.073849

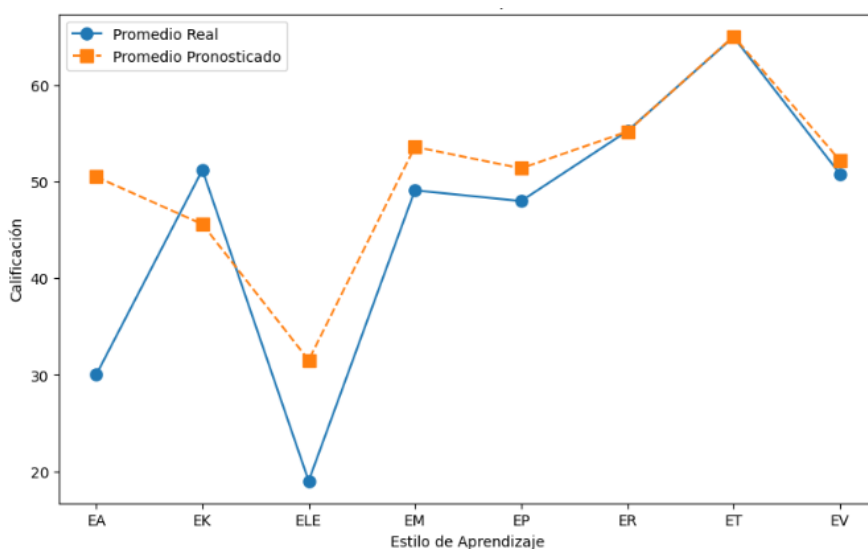
Fuente: Elaboración propia, 2025.

El análisis de importancia de variables del modelo Random Forest Regressor muestra que los estilos de aprendizaje contribuyen de forma heterogénea a la predicción de la calificación total. El estilo de lectura-escritura tiene mayor influencia individual con 22.08%, seguido por el estilo teórico 21.15% y múltiple con 16.15%, indicando que estos estilos son los más determinantes en la estimación de la calificación. En contraste, el etilo kinestésico con 15.66%, reflexivo con 11.98% y pragmático 7.38%, tienen menor contribución, aunque aún relevante. Esta distribución evidencia que ciertos estilos tienen un efecto predictivo más fuerte sobre el desempeño académico, mientras que otros ejercen un impacto relativamente limitado, esto puede orientar futuras investigaciones sobre la relación entre estilos de aprendizaje y rendimiento académico, así como la selección de variables para modelos predictivos más robustos.

A partir de la figura 2, se aprecia que el modelo Random Forest Regressor logra aproximarse bastante bien a las calificaciones reales, el modelo tiene una buena capacidad predictiva, sin embargo, se observan diferencias notables en algunos puntos, como en estilo activo y lectura-escritura, donde el modelo subestima o sobreestima considerablemente la calificación real. En los estilos múltiple y reflexivo, la predicción está muy cercana al valor real, mostrando buena precisión, inclusive para el estilo visual, el modelo predice casi exactamente el valor real, lo que indica que el comportamiento de este grupo fue bien aprendido por el modelo. Esto sugiere que algunos estilos presentan mayor variabilidad o menor cantidad de datos representativos, dificultando al modelo generalizar correctamente,

Consiguientemente, el modelo de regresión Random Forest, logra predecir adecuadamente las calificaciones promedio según los estilos de aprendizaje, pero existen errores localizados en estilos específicos. Puede considerarse fiable para estimar tendencias o promedios de rendimiento académico en función del estilo de aprendizaje, especialmente si se mejora con más datos o ajuste de hiperparámetros.

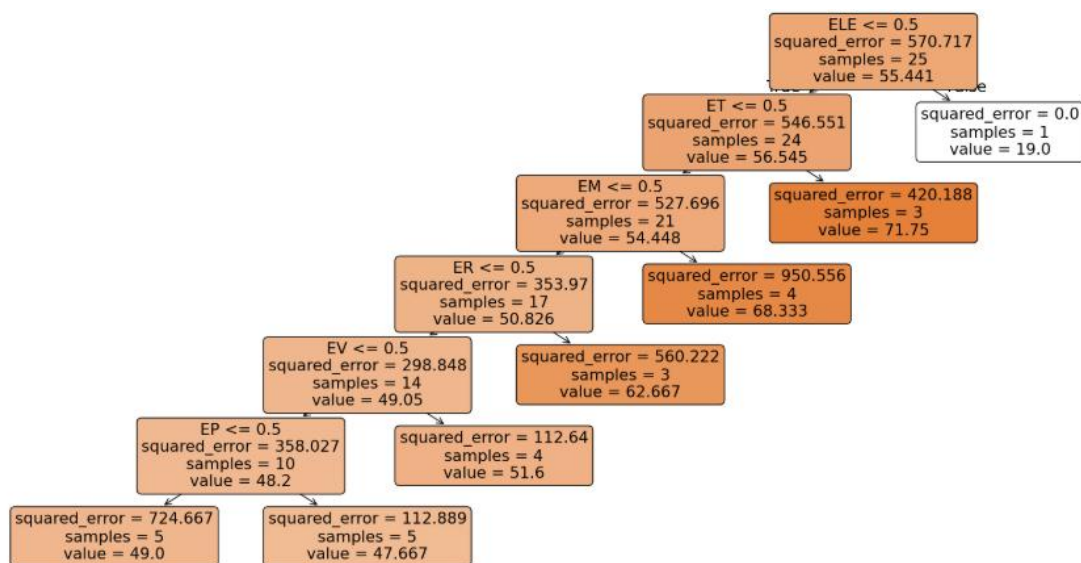
Figura 2. Calificación real y pronosticada según estilo



Fuente: Elaboración propia, 2025.

Por otra parte, la figura 3 representa uno de los árboles de decisión generados por el modelo Random Forest Regressor, para predecir la calificación promedio de los estudiantes en función de sus estilos de aprendizaje, donde cada nodo muestra: la condición de división del estilo pragmático EP; el error cuadrático medio (squared error), que indica el grado de dispersión o error dentro de ese grupo; la cantidad de muestras (samples), es decir, la cantidad de estudiantes que pertenecen a esa rama; y el valor promedio (value), correspondiente a la calificación pronosticada para ese conjunto de estudiantes.

Figura 3. Árboles de decisión generados por el modelo Random Forest Regressor



Fuente: Elaboración propia, 2025.

El modelo busca dividir los datos según los estilos de aprendizaje que mejor expliquen la variación de las calificaciones, en la raíz o inicio del árbol, el modelo podría dividir primero según un estilo que genera mayor diferencia en las calificaciones, es decir, si el estudiante pertenece o no al estilo pragmático EP, luego, en cada rama se aplican nuevas divisiones como estilo reflexivo $ER \leq 0.5$ o teórico $ET \leq 0.5$, hasta agrupar a los estudiantes con calificaciones similares. Cada hoja del árbol o bloque final representa una predicción promedio, por ejemplo, la hoja con $value = 71.75$ indica que para los estudiantes que cumplen con esa combinación de estilos, el modelo predice una calificación promedio de 71.75 puntos, en cambio, la hoja con $value = 47.66$ muestra que otro grupo, con distinto perfil de aprendizaje, tiene rendimientos más bajos.

El modelo diferencia entre estilos con mayor tendencia a calificaciones altas como estilo múltiple EM, teórico ET y lectura-escritura ELE y otros con rendimientos más

moderados o bajos estilo activo EA o kinestésico EK. Además, los valores de squared error son bajos en las hojas finales, lo que indica que el modelo logra una buena precisión local, es decir, predice de manera consistente dentro de cada grupo y tiene una buena capacidad de ajuste y generalización. En conjunto, el Random Forest muestra que los estilos de aprendizaje influyen significativamente en las calificaciones y que el modelo puede aprender relaciones no lineales entre estos factores.

Sin embargo, es preciso señalar que los resultados obtenidos indican que los estilos de aprendizaje por sí solos no constituyen predictores confiables de la calificación final. Las métricas de precisión negativas del modelo Random Forest y la elevada dispersión de los errores, RMSE y MAE, señalan que incluso los estilos con mayor importancia relativa, como lectura-escritura y teórico, no logran explicar sustancialmente la variabilidad en las calificaciones. Esto sobre todo porque la calificación final está influenciada por variables complejas y multifactoriales, como la motivación, hábitos de estudio, nivel de comprensión, contexto socioeducativo y estrategias pedagógicas implementadas, que no se capturan mediante la categorización de estilos de aprendizaje. Además, determinado estilo de aprendizaje no siempre conjunciona positivamente con las demás dimensiones didácticas, consiguientemente, no es posible generalizar patrones significativos.

Estos hallazgos coinciden con estudios previos que sugieren asociaciones moderadas entre estilos de aprendizaje y rendimiento académico, como los reportes de Coffield et al. (2004) y Pashler et al. (2008), que afirman que los estilos pueden influir en la preferencia de estrategias de aprendizaje, pero no necesariamente predicen consistentemente en calificaciones. Esto refuerza la idea de que los estilos son constructos descriptivos más que deterministas, útiles para orientar la enseñanza personalizada y diversificar metodologías, pero insuficientes como variables únicas de predicción. Consecuentemente, los modelos predictivos de rendimiento académico deben incorporar un conjunto más amplio de factores cognitivos, motivacionales y contextuales para alcanzar mayor precisión y relevancia práctica en entornos educativos.

Por último, la predominancia de los estilos pragmático y múltiple, especialmente en Ingeniería Comercial y Sistemas, sugiere una orientación hacia el aprendizaje aplicado y

flexible, fortalecido con metodologías activas como el aprendizaje basado en proyectos, simulaciones y el trabajo colaborativo, que favorecen la integración práctica del conocimiento y la resolución de problemas reales. Asimismo, los estilos visual y kinestésico, con presencia moderada en todas las carreras, resaltan la necesidad de incorporar recursos gráficos, digitales e interactivos que estimulen la observación y la experimentación. En conjunto, estos hallazgos respaldan la adopción de un modelo pedagógico multimodal y adaptativo, capaz de atender la diversidad cognitiva del estudiantado.

CONCLUSIONES

Existen diferencias significativas en la distribución de los estilos de aprendizaje según las carreras universitarias, la clasificación con Random Forest Classifier indica que en Ingeniería Comercial e Ingeniería de Sistemas predominan los estilos pragmático y múltiple, orientados a la aplicación práctica del conocimiento, mientras que en Estadística destacan los estilos visual y teórico, asociados al razonamiento analítico. Cada carrera desarrolla patrones cognitivos acordes con sus demandas epistemológicas y metodológicas, es evidente que los entornos formativos condicionan el estilo de aprendizaje predominante a través de las experiencias disciplinares.

El modelo de regresión Random Forest Regressor evidencia baja capacidad predictiva de los estilos de aprendizaje sobre la calificación total, los estilos de aprendizaje por sí solos no constituyen predictores sólidos, porque pedagógicamente la calificación final depende de un conjunto multifactorial de variables que tienen mayor incidencia.

El modelo Educational Data Mining y Random Forest permiten clasificar los estilos de aprendizaje de los estudiantes, las métricas satisfactorias accuracy, recall y F1-score, demuestran que el modelo Random Forest Classifier es eficaz para identificar patrones cognitivos latentes y segmentar al estudiantado con base en sus características de aprendizaje. Sin embargo, su capacidad predictiva respecto al rendimiento académico es limitada, esto indica que el modelo funciona adecuadamente en tareas de clasificación, pero no en predicción cuantitativa del desempeño.

Existen patrones diferenciados de estilos de aprendizaje según carrera, con predominancia de los estilos pragmático y múltiple, especialmente en Ingeniería Comercial e Ingeniería de Sistemas. En contraste, los estilos auditivo y lectura-escritura tienen menor frecuencia, lo que refuerza la necesidad de un diseño curricular que diversifique las estrategias didácticas mediante enfoques multimodales y constructivistas, alineando la docencia con la diversidad cognitiva del estudiantado.

Pedagógicamente existe la necesidad de estrategias diferenciadas por estilo y carrera. En Ingeniería Comercial y Sistemas se aplican metodologías basadas en proyectos, simulaciones y resolución de problemas reales, mientras que en Estadística conviene promover el razonamiento analítico y la visualización de datos. Estas estrategias deben consolidar un modelo de enseñanza centrado en el estudiante, flexible y adaptativo, donde la tecnología educativa y la analítica de aprendizaje, conduzcan a la toma de decisiones basadas en evidencia.

La minería de datos educativos describe patrones de aprendizaje, pero además puede retroalimentar el diseño curricular y optimizar las estrategias didácticas desde una visión constructivista. Al identificar perfiles cognitivos diversos, la universidad puede personalizar la enseñanza, promover la autorregulación del aprendizaje y fortalecer la equidad educativa. El uso del Random Forest en este contexto representa un avance metodológico significativo en la integración entre ciencia de datos y pedagogía, contribuyendo a una educación universitaria más inteligente e inclusiva.

RECOMENDACIONES

Las instituciones educativas universitarias, deberían implementar estrategias de enseñanza que vayan más allá de la identificación de los estilos de aprendizaje individuales, integrando factores emocionales, motivacionales y contextuales en el diseño de experiencias de aprendizaje, asegurando que los estudiantes reciban apoyo que considere tanto sus características cognitivas como su entorno académico y personal. Esto permitirá una atención más integral y efectiva, potenciando el desarrollo de competencias adaptativas y aplicadas.

Se sugiere emplear sistemáticamente herramientas de Educational Data Mining para el diagnóstico de perfiles cognitivos y de aprendizaje. La utilización de algoritmos como Random Forest puede orientar la creación de intervenciones pedagógicas diferenciadas, identificar necesidades específicas de los estudiantes y mejorar la planificación académica. No obstante, se debe complementar el análisis con otras fuentes de información, como evaluaciones formativas y observaciones docentes, para lograr un enfoque más holístico y confiable.

Es preciso capacitar a los docentes en el manejo de datos educativos y en la interpretación de resultados de modelos predictivos, promoviendo competencias para diseñar estrategias didácticas basadas en evidencia. Asimismo, las universidades deberían promover una cultura institucional de investigación aplicada en educación, donde los hallazgos de estudios como este se traduzcan en políticas académicas y programas de mejora continua, favoreciendo la equidad y la eficacia del proceso de enseñanza-aprendizaje.

REFERENCIAS

- AHMED, A. & ELARABY, I. **Data mining: a prediction for student's performance using classification method.** *World Journal of Computer Application and Technology*, 2(2), 43–47, 2014.
- ALONSO, C. M., GALLEGU, D. J., & HONEY, P. **Los estilos de aprendizaje: procedimientos de diagnóstico y mejora.** Ediciones Mensajero, 1994.
- ÁLVAREZ MÉNDEZ, J. M. **Evaluar para comprender y mejorar la educación.** Morata, 2001.
- ASTIN, A. W. **Student involvement: a developmental theory for higher education.** *Journal of College Student Development*, 40(5), 518–529, 1999.
- BAKER, R. S., & SIEMENS, G. Educational data mining and learning analytics. In Sawyer, R. K. (Ed.), **The Cambridge Handbook of the Learning Sciences** (2nd ed., pp. 253–272). Cambridge University Press, 2014.
- BARADWAJ, B. K., & PAL, S. **Mining educational data to analyze students' performance.** *International Journal of Advanced Computer Science and Applications*, 2(6), 63–69, 2011.
- BIENKOWSKI, M., FENG, M., & MEANS, B. **Enhancing teaching and learning through educational data mining and learning analytics: an issue brief.** U.S. Department of Education, 2012.

- BREIMAN, L. **Random Forests**. Machine Learning, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>, 2001.
- Bruner, J. S. **La educación, puerta de la cultura**. Gedisa, 1997.
- CABRERA, A. F., NORA, A., & CASTAÑEDA, M. B. The role of finances in the persistence process: a structural model. **Research in Higher Education**, 33(5), 571–593, 2006.
- CANO, E., & JUSTICIA, F. Estilos de aprendizaje y rendimiento académico. **Revista de Psicología de la Educación**, 9(1), 35–47, 1994.
- CHURA, R., VELÁSQUEZ, D., & CALLISAYA, M. Aplicación de algoritmos de clasificación para el análisis del desempeño académico. **Revista Boliviana de Educación Superior**, 12(1), 45–59, 2023.
- DEKKER, G. W., PECHENIZKIY, M., & VLEESHOUWERS, J. M. **Predicting students drop out: a case study**. International Working Group on Educational Data Mining, 41–50, 2009.
- FELDER, R. M., & BRENT, R. **Understanding student differences**. Journal of Engineering Education, 94(1), 57–72, 2005.
- GARCÍA PEÑALVO, F. J., & REBAQUE, G. Data mining en sistemas de tutoría universitaria. **Revista Iberoamericana de Tecnología Educativa**, 11(3), 57–65, 2016.
- GARCÍA, E., ROMERO, C., & VENTURA, S. **Predicting student learning styles in learning management systems using data mining techniques**. International Journal of Information and Education Technology, 1(2), 102–108, 2011.
- GONZÁLEZ-PIENDA, J. A., NÚÑEZ, J. C., GONZÁLEZ-PUMARIEGA, S., ÁLVAREZ, L., ROCES, C., & GARCÍA, M. Autoconcepto, autoconfianza y rendimiento escolar. **Psicothema**, 14(1), 209–214, 2002.
- HAN, J., KAMBER, M., & PEI, J. **Data mining: concepts and techniques** (3rd ed.). Morgan Kaufmann 2012.
- HASTIE, T., TIBSHIRANI, R., & FRIEDMAN, J. **The elements of statistical learning: data mining, inference, and prediction** (2nd ed.). Springer, 2009.
- JAMES, G., WITTEN, D., HASTIE, T., & TIBSHIRANI, R. **An introduction to statistical learning** (Vol. 112). Springer, 2013.
- KOLB, D. A. **Experiential learning: experience as the source of learning and development**. Prentice-Hall, 1984.
- KOTSIANTIS, S. B., PIERRAKEAS, C., & PINTELAS, P. Predicting students' performance in distance learning using machine learning techniques. **Applied Artificial Intelligence**, 18(5), 411–426, 2010.
- KUHN, M., & JOHNSON, K. **Applied predictive modeling**. Springer, 2013.
- LANTZ, B. **Machine learning with R: expert techniques for predictive modeling**. Packt Publishing, 2019.

- MAMANI, H., & LÓPEZ, R. Predicción de factores de riesgo académico en estudiantes de primer año. **Revista Científica de la UMSA**, 8(2), 78–90, 2022.
- PEÑA-AYALA, A. Educational data mining: a survey and a data mining-based analysis of recent works. **Expert Systems with Applications**, 41(4), 1432–1462, 2014.
- PIAGET, J. **La psicología de la inteligencia**. Editorial Crítica, 1970.
- RAMESH, A., GOLDWASSER, D., HUANG, B., DAUMÉ III, H., & GETOOR, L. **Learning latent engagement patterns of students in online courses**. AAAI Conference on Artificial Intelligence, 1272–1278, 2014.
- ROBBINS, S. B., LAUVER, K., LE, H., DAVIS, D., LANGLEY, R., & CARLSTROM, A. **Do psychosocial and study skill factors predict college outcomes? A meta-analysis**. *Psychological Bulletin*, 130(2), 261–288. <https://doi.org/10.1037/0033-2909.130.2.261>, 2004.
- ROMERO, C., & VENTURA, S. **Educational data mining: a review of the state of the art**. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 40(6), 601–618, 2020.
- ROMERO, C., VENTURA, S., & GARCIA, E. Data mining in course management systems: moodle case study and tutorial. **Computers & Education**, 51(1), 368–384, 2008.
- SIEMENS, G., & LONG, P. Penetrating the fog: analytics in learning and education. **EDUCAUSE Review**, 46(5), 30–40, 2011.
- TEMPELAAR, D. T., RIENTIES, B., & GIESBERS, B. In search for the most informative data for feedback generation: learning analytics in a data-rich context. **Computers in Human Behavior**, 47, 157–167, 2015.
- THAI-NGHE, N., DRUMOND, L., & SCHMIDT-THIEME, L. Recommender system for predicting student performance. **Procedia Computer Science**, 1(2), 2811–2819, 2011.
- TOMLINSON, C. A. **How to differentiate instruction in mixed-ability classrooms** (2nd ed.). ASCD, 2001.
- VANDERPLAS, J. **Python data science handbook: essential tools for working with data**. O'Reilly Media, 2016.
- VARGAS, J., & YUPANQUI, J. Minería de datos para la predicción del rendimiento académico en estudiantes universitarios. **Revista de Investigación en Educación**, 7(1), 33–47, 2019.