

Artículo Científico

Sistema de análisis de estudiantes SIATECH: Un enfoque integrado de análisis descriptivo, correlacional, multivariante y predictivo

SIATECH Student Analytics System: An integrated approach to descriptive, correlational, multivariate and predictive analytics

Alexander Haro Sarango¹ , Frankz Carrera Calderón² , Freddy Lalaleo Analuisa³ 

¹ Instituto Superior Tecnológico España, alexander.haro@iste.edu.ec

² Instituto Superior Tecnológico España, frankz.carrera@iste.edu.ec

³ Instituto Superior Tecnológico España, freddy.lalaleo@iste.edu.ec

Autor para correspondencia: alexander.haro@iste.edu.ec

Copyright

Los originales publicados en las ediciones impresa y electrónica de esta revista son propiedad del Instituto Superior Tecnológico Universitario Rumiñahui, por ello, es necesario citar la procedencia en cualquier reproducción parcial o total. Todos los contenidos de la revista electrónica se distribuyen bajo una licencia de Creative Commons Reconocimiento-NoComercial-4.0 Internacional.



Citas

Haro Sarango, A., Carrera Calderón, F., & Lalaleo Analuisa, F. (2025). Sistema de análisis de estudiantes SIATECH: Un enfoque integrado de análisis descriptivo, correlacional, multivariante y predictivo. CONECTIVIDAD, 6(2). <https://doi.org/10.37431/conectividad.v6i2.303>

RESUMEN

El estudio SIATECH aborda la necesidad de evaluar y mejorar la satisfacción y el rendimiento académico en la educación mediante la implementación de técnicas estadísticas avanzadas y de aprendizaje automático. Utilizando el modelo de análisis de importancia-desempeño (IPA) como base, el estudio tiene como objetivo desarrollar un sistema de análisis de estudiantes para realizar análisis descriptivos, correlacionales, multivariantes y predictivos, proporcionando así una herramienta integral para mejorar los procesos educativos. Este enfoque predictivo ofrece la posibilidad de personalizar intervenciones educativas, mejorando así los resultados académicos individuales, lo que proporciona a los instructores una comprensión clara de las expectativas estudiantiles y les permite ajustar sus métodos de enseñanza para mejorar el aprendizaje. Sin embargo, desde un punto de vista del conocimiento técnico estadístico, la falta de formación en análisis de datos entre algunos docentes puede limitar la capacidad de interpretar y utilizar eficazmente la información obtenida, subrayando la necesidad de integrar la capacitación estadística en los programas de formación docente, no obstante, SIATECH concibe una plataforma dinámica y de muy sencillo uso e interpretación. En síntesis, SIATECH demuestra ser una herramienta valiosa para mejorar las prácticas educativas, enfatizando

la importancia de un enfoque basado en datos en la toma de decisiones educativas.

Palabras clave: Aprendizaje automático, Predicción, Rendimiento académico, Satisfacción estudiantil, Educación, Docencia.

ABSTRACT

The SIATECH study addresses the need to assess and improve academic satisfaction and performance in education by implementing advanced statistical and machine learning techniques. Using the importance-performance analysis (IPA) model as a basis, the study aims to develop a student analytics system to perform descriptive, correlational, multivariate and predictive analyses, thus providing a comprehensive tool to improve educational processes. This predictive approach offers the possibility of customizing educational interventions, thus improving individual academic outcomes, providing instructors with a clear understanding of student expectations and allowing them to adjust their teaching methods to improve learning. However, from a statistical know-how point of view, the lack of training in data analysis among some teachers may limit the ability to interpret and effectively use the information obtained, highlighting the need to integrate statistical training into teacher training programs, however, SIATECH devises a dynamic platform that is very easy to use and interpret. In summary, SIATECH proves to be a valuable tool for improving educational practices, emphasizing the importance of a data-driven approach to educational decision-making.

Keywords: Machine learning, Prediction, Academic performance, Student satisfaction, Education, Teaching.

1. INTRODUCCIÓN

El análisis de estudiantes es esencial para evaluar y mejorar la satisfacción de los estudiantes en el ámbito de la educación superior, un estudio realizado por McLeay et al. (2017) aplicó el modelo de análisis de importancia-desempeño (IPA) en universidades privadas de Malasia para identificar brechas entre la importancia atribuida a ciertos atributos de satisfacción estudiantil y el desempeño percibido. Este enfoque permitió a los administradores de educación superior asignar recursos de manera más efectiva y desarrollar estrategias para mejorar la calidad y la satisfacción estudiantil.

Además, la predicción del desempeño académico de los estudiantes es otro aspecto fundamental del análisis educativo; Según Oyedeffi et al. (2020), el uso de herramientas de aprendizaje automático para analizar datos históricos de estudiantes, incluyendo atributos individuales como la edad, distribución demográfica, antecedentes familiares y actitud hacia el estudio, puede mejorar significativamente la capacidad de las instituciones para predecir y, en última instancia, mejorar el desempeño individual de los estudiantes.

Por otro lado, la evaluación de la enseñanza desde la perspectiva de los estudiantes es crucial para identificar fortalezas y áreas de mejora en los métodos de enseñanza; Cladera (2020) propone el uso del análisis de importancia-desempeño junto con cuestionarios de evaluación de la enseñanza por parte de los estudiantes para obtener una representación visual de los atributos de enseñanza más valorados y cómo los instructores se desempeñan en estos aspectos. Esta metodología ayuda a los docentes a comprender mejor las expectativas de los estudiantes y ajustar sus métodos de enseñanza en consecuencia.

Un estudio realizado en Taiwán por Chen (2017) utilizó el análisis de importancia-desempeño para evaluar el rendimiento en empleabilidad de los estudiantes universitarios. Los resultados permitieron a las instituciones de educación superior formular e implementar políticas y estrategias de empleo más efectivas, basadas en una comprensión clara de las preferencias y el desempeño de los estudiantes en términos de empleabilidad.

Para Noronha et al. (2007), es crucial que los estudiantes de psicología comprendan y valoren sus competencias y conocimientos en evaluaciones psicológicas, la investigación mostró que, aunque los estudiantes valoran ciertas competencias, existen discrepancias entre la importancia atribuida y el conocimiento real, lo que subraya la necesidad de ajustes en los programas educativos para abordar estas brechas.

La aplicación de técnicas avanzadas de análisis de datos y minería de datos ha demostrado ser altamente efectiva en la predicción del éxito académico de los estudiantes. Bhumika Banswal (2023) utilizaron técnicas de aprendizaje automático como árboles de decisión, bosques aleatorios y redes neuronales para predecir el rendimiento académico basado en datos circundantes de los estudiantes. Este enfoque no solo mejora la precisión en la predicción del rendimiento, sino que también proporciona información valiosa para el desarrollo de estrategias

educativas personalizadas.

El análisis de datos en la educación es fundamental para evaluar el rendimiento académico de los estudiantes. Los datos pueden provenir de diversas fuentes, como resultados de exámenes, evaluaciones formativas, encuestas de satisfacción y otras métricas de desempeño; al analizar estos datos, las instituciones pueden identificar áreas donde los estudiantes están teniendo dificultades y desarrollar estrategias específicas para abordarlas. Por ejemplo, un estudio realizado por Pito et al. (2020) demostró que la implementación de un sistema de bodega de datos en la Universidad del Cauca permitió a los usuarios obtener modelos más efectivos y adaptar mejor sus estrategias de enseñanza, resultando en una alta satisfacción entre los usuarios del sistema.

A pesar de los beneficios del análisis de datos, muchos profesores carecen de la formación necesaria para interpretar y utilizar estos datos de manera efectiva, esto se debe en parte a que la formación docente tradicional no siempre incluye módulos sobre análisis de datos o estadística educativa, la falta de esta competencia puede llevar a una mala interpretación de los datos y a decisiones educativas que no están basadas en evidencia. Un estudio sobre la formación estadística de futuros profesores en matemática reveló que muchos de ellos mostraban dificultades significativas en la interpretación y análisis de datos, lo que subraya la necesidad de mejorar esta área en la formación docente (Arteaga et al., 2017).

Para mejorar la competencia de los profesores en análisis de datos, es fundamental que las instituciones educativas integren formación en análisis de datos y estadística en los programas de formación docente, esta formación debe incluir tanto aspectos teóricos como prácticos, permitiendo a los profesores adquirir habilidades para interpretar y aplicar datos en sus prácticas pedagógicas. Un enfoque efectivo podría ser la implementación de metodologías activas y herramientas tecnológicas que faciliten la comprensión y uso de los datos. Por ejemplo, un estudio realizado en la Universidad de Barcelona destacó la importancia de la colaboración entre docentes y bibliotecarios para el desarrollo de competencias informacionales, lo que resultó en una mejora significativa en la capacidad de los estudiantes para utilizar recursos y herramientas educativas de manera efectiva (Zaborras et al., 2020).

El objetivo del estudio es desarrollar el Sistema de Análisis de Estudiantes SIATECH para realizar análisis descriptivos, correlacionales, multivariantes y predictivos, proporcionando así una herramienta integral para mejorar los procesos educativos, este enfoque integrado tiene como meta proporcionar una visión holística del rendimiento estudiantil y otros aspectos relevantes.

2. MATERIALES Y MÉTODOS

Introducción al estudio

El estudio busca desarrollar y evaluar un modelo integrado que aplique métodos descriptivos,

correlacionales, multivariantes y predictivos para analizar una amplia variedad de datos de estudiantes, este enfoque tiene como objetivo identificar patrones y correlaciones entre diferentes variables educativas, como rendimiento académico, factores socioeconómicos, métodos de enseñanza, y apoyo institucional, para mejorar la toma de decisiones educativas. El estudio toma como premisa teórica-metodológica el estudio de McLeay et al. (2017) modelo de análisis de importancia-desempeño (IPA).

Objetivos específicos de la investigación

- Implementar y validar métodos estadísticos y de aprendizaje automático para el análisis de datos educativos.
- Identificar factores clave que influyen en el rendimiento académico de los estudiantes.
- Desarrollar modelos predictivos que permitan anticipar trayectorias académicas en función de diversas variables de entrada.

Diseño de la investigación

El estudio se estructura en un diseño mixto, combinando técnicas cuantitativas y cualitativas para abordar la complejidad de los datos educativos y las interacciones entre diversas variables.

Recolección de datos

- **Fuentes de datos:**
 - Datos académicos históricos y actuales.
 - Encuestas sobre factores socioeconómicos y ambientales.
 - Evaluaciones de desempeño de estudiantes y profesores.
- **Instrumentos:**
 - Cuestionarios estructurados para recolectar datos socioeconómicos y opiniones sobre métodos pedagógicos.
 - Bases de datos institucionales para obtener información académica.
 - Entrevistas semiestructuradas con educadores y administradores para obtener insights cualitativos.

Limitaciones del estudio

- **Generalización de resultados:** Los hallazgos pueden no ser generalizables a todos los contextos educativos, especialmente debido a las diferencias culturales y estructurales entre instituciones.
- **Possibles sesgos:** Sesgos en la selección de la muestra y en las respuestas de las encuestas pueden influir en la interpretación de los datos.

3. RESULTADOS Y DISCUSIÓN

- **Repository:** GitHub es un espacio de almacenamiento donde se va a alojar y gestionar

proyectos de código. GitHub utiliza Git, un sistema de control de versiones distribuido, para rastrear y gestionar cambios en los archivos del proyecto (Cosentino et al., 2016). Para quienes deseen replicar este proyecto *Sistema de análisis de estudiantes SIATECH* podrán aplicarlo y adminístralo de forma libre y, lo podrán conseguir en el siguiente enlace:

- <https://gist.github.com/aharo8014/a6bd37ebc51c0a944c30c8a33f96fc0c/revisions>
- **Página de presentación y carga de datos:** Se menciona que existe un límite de carga de hasta 200 MB, el formato admisible es CSV (comma-separated values / Valores separados por comas).

Figura. 1. Página de presentación



Fuente: Visualizado en navegador; estructurado en PyCharm, gestionado con Streamlit

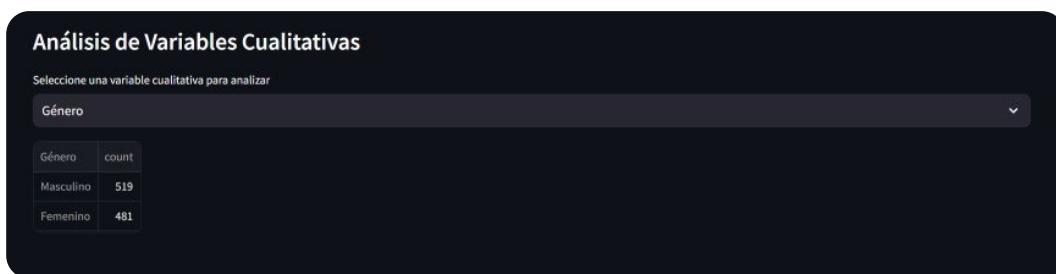
- **Análisis descriptivo - Cuantitativo:** resumen estadístico de los datos cargados. Esto incluye medidas como la media, mediana, desviación estándar, valores mínimos y máximos, entre otros.
- **Análisis descriptivo - Cualitativo:** Conteo por categorías.

Figura. 2. Análisis descriptivo cuantitativo

	ID	Edad	Género	Horas de Estudio	Calificación Final	Asistencia (%)	Participación en Clase	Promedio Anterior	Horas de Sueño	Ingresos Familiares	Tipo de Estudio
count	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
unique	None	None	2	None	None	None	None	None	None	None	2
top	None	None	Masculin	None	None	None	None	None	None	None	Tiempo Comp
freq	None	None	519	None	None	None	None	None	None	None	526
mean	500.5	20.96	None	19.362	50.7474	75.5432	4.593	49.1809	6.9824	54,141.8116	None
std	288.8194	2.0036	None	11.5222	29.1498	14.21	2.8102	28.8982	1.7385	25,694.9125	None
min	1	18	None	0	0.0188	50.1009	0	0.0227	4.0235	10,021.6904	None
25%	250.75	19	None	9.75	25.0957	63.571	2	24.7464	5.4615	31,749.4726	None

Fuente: Visualizado en navegador; estructurado en PyCharm, gestionado con Streamlit

Figura. 3. Análisis descriptivo cualitativo



Fuente: Visualizado en navegador, estructurado en PyCharm, gestionado con Streamlit

- **Histograma:** Un histograma es una representación gráfica que muestra la distribución de un conjunto de datos numéricos, permite visualizar la frecuencia de los valores dentro de intervalos específicos (bins), ayuda a identificar la forma de la distribución (normal, sesgada, bimodal, etc.) y es útil para detectar outliers, tendencias y patrones en los datos (Jaramillo et al., 2023).

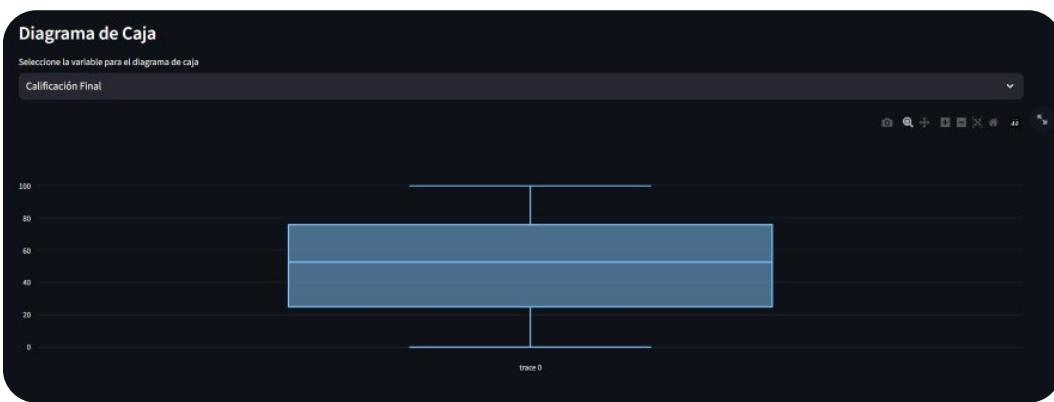
Figura. 4. Estructuración de Histograma



Fuente: Visualizado en navegador, estructurado en PyCharm, gestionado con Streamlit

- **Diagrama de caja:** Un diagrama de caja es una representación gráfica que resume la distribución de un conjunto de datos a través de sus cuartiles, muestra la mediana, los cuartiles (Q1 y Q3), y los valores extremos (mínimo y máximo), identifica outliers y la dispersión de los datos, comparar distribuciones entre diferentes grupos o categorías (Flores & Flores, 2018).

Figura. 5. Estructuración de diagrama de caja



Fuente: Visualizado en navegador, estructurado en PyCharm, gestionado con Streamlit

- **Gráfico de barras:** Un gráfico de barras es una representación gráfica que muestra la relación entre diferentes categorías a través de barras de longitud proporcional a los valores que representan, comparan cantidades entre diferentes categorías, visualiza cambios en los datos a lo largo del tiempo o entre diferentes grupos, destacan las diferencias y similitudes entre distintas categorías, bastante utilizado en análisis descriptivo para comparar datos categóricos, visualizar frecuencias, proporciones y resumir información en categorías discretas (Fernández et al., 2019).

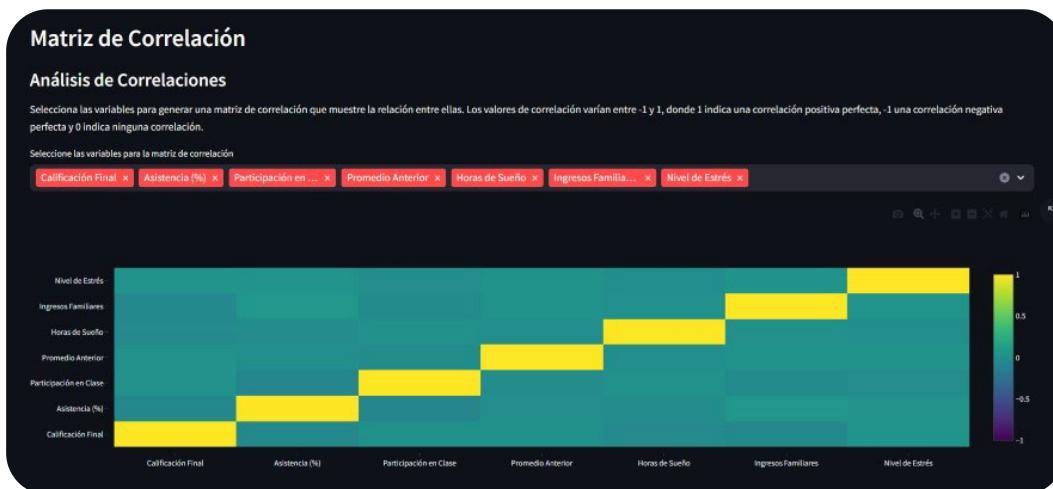
Figura. 6. Estructuración de gráfico de barras



Fuente: Visualizado en navegador, estructurado en PyCharm, gestionado con Streamlit

- **¿Qué mide?** La correlación mide la relación lineal entre dos variables. Por ejemplo, podría ser entre el tiempo de estudio y las calificaciones de los estudiantes.
- **Valores:** El coeficiente de correlación varía entre -1 y 1. Un valor cercano a 1 indica una fuerte correlación positiva (a medida que una variable aumenta, la otra también lo hace), un valor cercano a -1 indica una fuerte correlación negativa (a medida que una aumenta, la otra disminuye), y un valor cercano a 0 indica ninguna correlación lineal (Martínez Ortega et al., 2009).

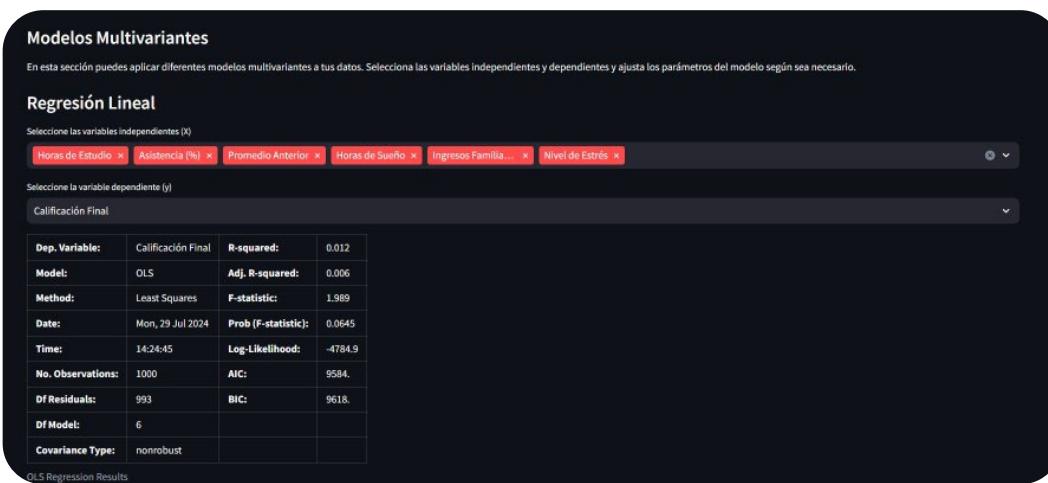
Figura. 7. Matriz de correlación



Fuente: Visualizado en navegador, estructurado en PyCharm, gestionado con Streamlit

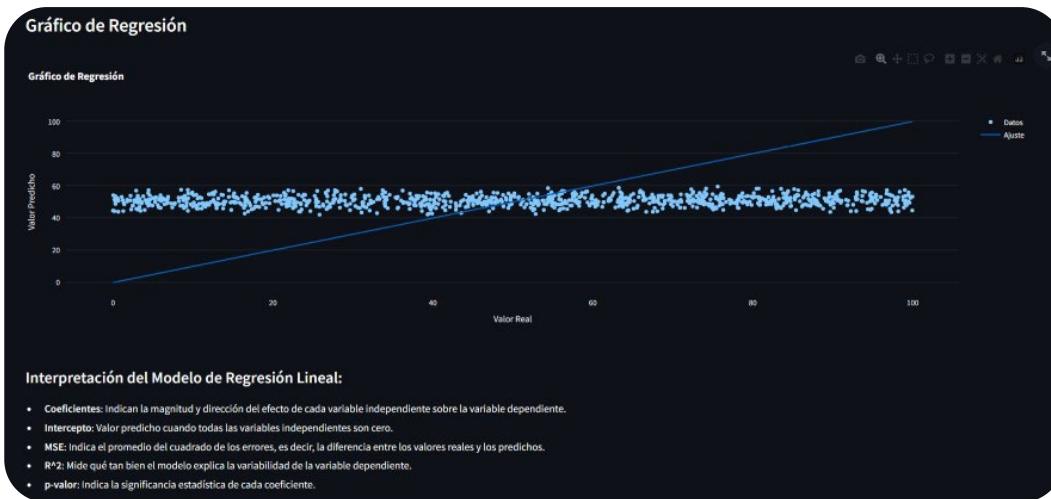
- **¿Qué mide?** La regresión simple busca modelar la relación entre una o varias variables independientes y una variable dependiente, estimando cómo una o unas variables afectan a la otra.
- **Modelo:** El modelo proporciona una ecuación de la forma $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon$, donde y es la variable dependiente (e.g., calificaciones), x_i es la variable independiente (e.g., horas de estudio, nivel de estrés, entre otras), β_i es la pendiente de la línea (cómo cambia y en respuesta a x_i), y ϵ es la intersección (Carrasquilla-Batista et al., 2016)

Figura. 8. Regresión lineal simple y múltiple



Fuente: Visualizado en navegador, estructurado en PyCharm, gestionado con Streamlit

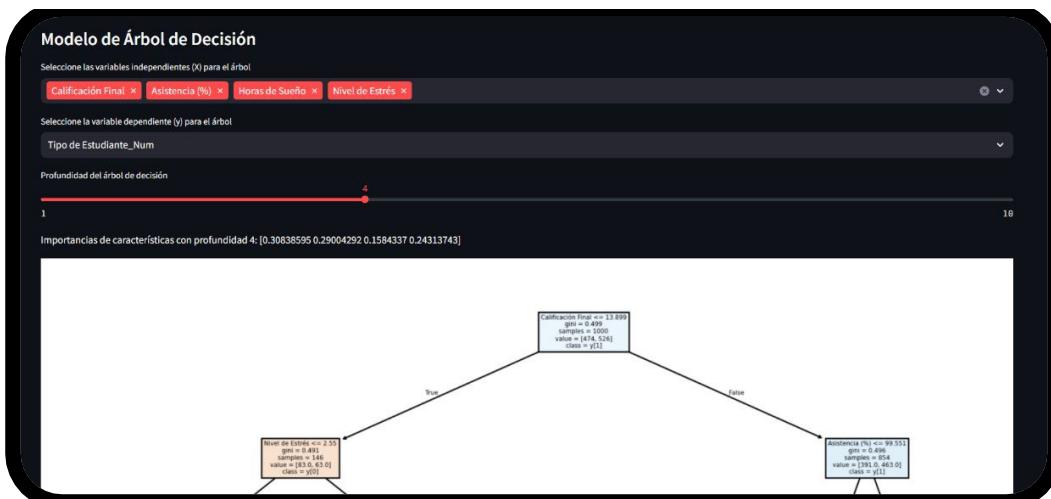
Figura. 9. Gráfico de regresión



Fuente: Visualizado en navegador; estructurado en PyCharm, gestionado con Streamlit

- **¿Qué mide?** Los árboles de decisión son modelos predictivos que dividen los datos en ramas para llegar a decisiones o predicciones.
- **Estructura:** Consiste en nodos de decisión y hojas. Los nodos representan una pregunta basada en una característica, y las hojas representan una decisión o resultado. Puede utilizarse para determinar qué factores contribuyen más al éxito académico, como el entorno familiar, los métodos de estudio, o la participación en actividades extracurriculares.

Figura. 10. Modelo de árbol de decisión

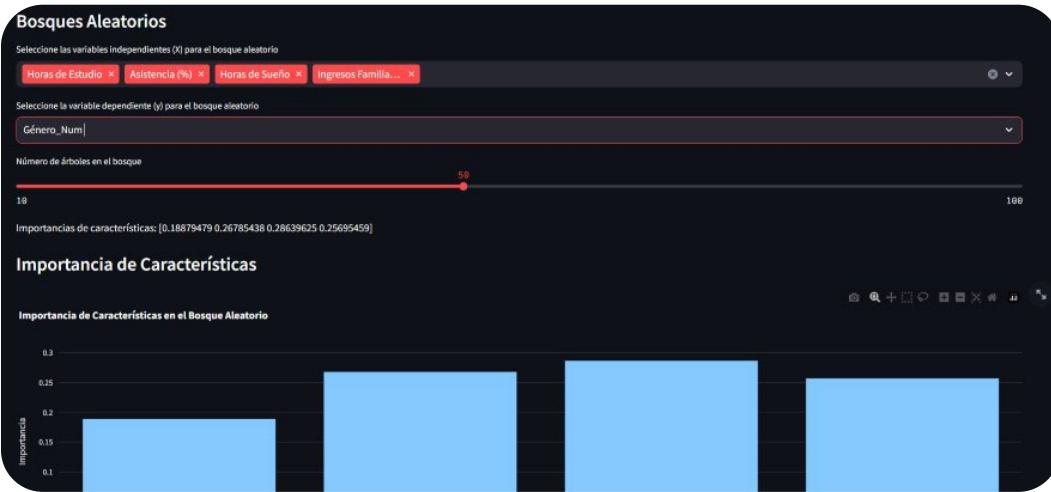


Fuente: Visualizado en navegador; estructurado en PyCharm, gestionado con Streamlit

- **¿Qué mide?** Los bosques aleatorios son un ensamble de árboles de decisión diseñados para mejorar la precisión en la clasificación y la predicción (Carrasquilla-Batista et al., 2016).
- **Importancia de las variables:** Este método también proporciona un indicativo de qué variables son más importantes en la predicción. Útil para analizar la efectividad de diferentes intervenciones educativas, identificando qué variables impactan más

significativamente el rendimiento de los estudiantes.

Figura. 11. Bosques aleatorios y determinación de variables influyentes



Fuente: Visualizado en navegador, estructurado en PyCharm, gestionado con Streamlit

- **¿Qué mide?** El PCA es una técnica de reducción de dimensiones que transforma las variables originales en un nuevo conjunto de variables que son combinaciones lineales de las originales, llamadas componentes principales.
- **Componentes principales:** Estos componentes se seleccionan de modo que el primero captura la mayor varianza posible, y cada componente subsiguiente tiene la mayor varianza posible bajo la restricción de que sea ortogonal a los componentes anteriores. Puede utilizarse para identificar patrones y simplificar los datos educativos, facilitando la visualización y comprensión de las relaciones entre múltiples factores académicos y comportamentales (Gozá-León et al., 2020).

Figura. 12. Análisis de Componentes Principales (PCA)



Fuente: Visualizado en navegador, estructurado en PyCharm, gestionado con Streamlit

- **¿Qué mide la prueba T de Student?** evalúa si las medias de dos grupos son estadísticamente diferentes entre sí. Es aplicable cuando se comparan dos grupos bajo

supuestos de normalidad y varianzas homogéneas. Permite comparar el rendimiento promedio de dos grupos de estudiantes que usan diferentes métodos de enseñanza, también permite evaluar la eficacia de un nuevo currículo al comparar los resultados de pruebas de estudiantes antes y después de su implementación. El resultado de la prueba T es un valor p, que si es menor que el nivel de significancia (usualmente 0.05), indica que hay una diferencia estadísticamente significativa entre las medias de los dos grupos (Lugo-Armenta & Pino-Fan, 2022).

- **¿Qué mide ANOVA?** se utiliza para determinar si existen diferencias significativas entre las medias de tres o más grupos independientes. A diferencia de la prueba T, que solo puede comparar dos grupos, ANOVA puede manejar múltiples grupos al mismo tiempo. Permite comparar las calificaciones de los estudiantes en diferentes tipos de escuelas (públicas, privadas, charter) para determinar si el tipo de escuela influye en el rendimiento académico. También permite evaluar el rendimiento en diferentes asignaturas para determinar cuáles presentan mayores desafíos para los estudiantes. Al igual que en la prueba T, el resultado principal de un ANOVA es un valor p. Un valor p bajo (menor que 0.05) sugiere que al menos uno de los grupos difiere significativamente de los otros en términos de la media (Liu & Wang, 2021).

Figura. 13. T-Student y ANOVA

The screenshot shows a Streamlit application interface. At the top, there is a navigation bar with tabs: Análisis Descriptivo, Visualización de Datos, Correlaciones, Modelos Multivariantes, Proyecciones, and Análisis de Diferencias entre Grupos (which is highlighted in red). Below the navigation bar, the main title is "Análisis de Diferencias entre Grupos". Under this title, there is a sub-section titled "Análisis de Diferencias entre Grupos" with the following instructions: "Utiliza este análisis para comparar las medias de dos grupos y determinar si son significativamente diferentes." There are four dropdown menus for selecting variables: "Selección la categoría para analizar" (set to "Género"), "Selección la variable cuantitativa para comparar" (set to "Calificación Final"), "Selección el grupo 1" (set to "Masculino"), and "Selección el grupo 2" (set to "Femenino"). At the bottom of the section, there is a text output: "t-stat: 0.8056191616358357, p-valor: 0.42065437593745714". Below this, another sub-section is titled "Análisis de Varianza (ANOVA)" with the text output: "F-stat: 0.649022335948234, p-valor: 0.42065437593718225".

Fuente: Visualizado en navegador, estructurado en PyCharm, gestionado con Streamlit

- **Sistema en línea de Acceso Abierto:** <https://siatech.streamlit.app>

Figura. 14. Sistema en Línea



Fuente: Cargado en línea con conexión a Github

4. CONCLUSIONES

El estudio SIATECH ha proporcionado una valiosa visión holística sobre el impacto de diferentes factores en el rendimiento y la satisfacción estudiantil dentro del ámbito de la educación, pero no solo dentro de la superior, sino de todos los otros niveles educativos existentes; a través de la aplicación de técnicas estadísticas avanzadas y métodos de aprendizaje automático, hemos forjado una herramienta que permite identificar y analizar patrones y correlaciones significativas entre variables clave como el rendimiento académico, los atributos socioeconómicos, los métodos de enseñanza y el apoyo institucional.

La implementación del modelo de análisis de importancia-desempeño (IPA) en universidades privadas en Malasia ha revelado brechas cruciales entre la importancia que los estudiantes atribuyen a ciertos aspectos de su educación y el desempeño percibido de estas áreas. Este conocimiento ha permitido a los administradores educativos asignar recursos de manera más efectiva y desarrollar estrategias para potenciar la calidad y la satisfacción estudiantil; además, la integración del análisis de importancia-desempeño en la evaluación de la enseñanza ha proporcionado a los docentes una mejor comprensión de las expectativas estudiantiles, facilitando ajustes pertinentes en los métodos pedagógicos.

Por otra parte, la predicción del desempeño académico mediante herramientas de aprendizaje automático ha demostrado ser particularmente efectiva, los modelos desarrollados han permitido predecir trayectorias académicas con mayor precisión, basándose en una diversidad de datos históricos y atributos individuales de los estudiantes. Este enfoque predictivo en la nueva era ha abierto nuevas posibilidades para intervenciones educativas personalizadas, dirigidas a mejorar el rendimiento individual de los estudiantes.

Sin embargo, el estudio también ha destacado desafíos importantes, particularmente en relación

con la capacitación de los docentes en análisis de datos, la falta de competencias estadísticas entre el personal docente puede conducir a una interpretación y utilización ineficaz de los datos educativos, lo que a su vez afecta la toma de decisiones basada en evidencias; la integración de formación en análisis de datos en los programas de formación docente es, por tanto, esencial para superar este obstáculo y mejorar la eficacia educativa a largo plazo.

REFERENCIAS

- Arteaga, P., Batanero, C., & Gea, M. (2017). La componente mediacional del conocimiento didáctico-matemático de futuros profesores sobre Estadística: Un estudio de evaluación exploratorio. 1, 54-75. <https://doi.org/10.24116/EMD25266136V1N12017A03>
- Bhumika Banswal, A. (2023). Analysing and Predicting Student's Performance Using their Surrounding Data. 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), 1-7. <https://doi.org/10.1109/ICCCNT56998.2023.10307525>
- Carrasquilla-Batista, A., Chacón-Rodríguez, A., Núñez-Montero, K., Gómez-Espinoza, O., Valverde, J., & Guerrero-Barrantes, M. (2016). Regresión lineal simple y múltiple: Aplicación en la predicción de variables naturales relacionadas con el crecimiento microalgal. *Revista Tecnología en Marcha*, 29, 33-45. <https://doi.org/10.18845/tm.v29i8.2983>
- Chen, Y. C. (2017). Applying importance-performance analysis to assess student employability in Taiwan. *Journal of Applied Research in Higher Education*, 10, 76-86. <https://doi.org/10.1108/JARHE-10-2017-0118>
- Cladera, M. (2020). An application of importance-performance analysis to students' evaluation of teaching. *Educational Assessment, Evaluation and Accountability*, 33, 701-715. <https://doi.org/10.1007/s11092-020-09338-4>
- Cosentino, V., Luis, J., & Cabot, J. (2016). Findings from GitHub: Methods, datasets and limitations. *Proceedings of the 13th International Conference on Mining Software Repositories*, 137-141. <https://doi.org/10.1145/2901739.2901776>
- Fernández, A., García García, J. I., Arredondo, E. H., & López Calvario, C. (2019). Comprensión de una tabla y un gráfico de barras por estudiantes universitarios. Areté: *Revista Digital del Doctorado en Educación de la Universidad Central de Venezuela*, 5(10), 145-162.
- Flores, J., & Flores, R. (2018). La Enseñanza del Diagrama de Caja y Bigotes para Mejorar su Interpretación. *Revista Bases de la Ciencia*, 3(1), Article 1. https://doi.org/10.33936/rev_bas_de_la_ciencia.v3i1.1107
- Gozá-León, O., Fernández-Águila, M., Rodríguez-Garcel, R. H., Ojito-Magaz, E., Gozá-León, O., Fernández-Águila, M., Rodríguez-Garcel, R. H., & Ojito-Magaz, E. (2020). Aplicación del Análisis de Componentes Principales en el proceso de purificación de un biofármaco. *Vaccimonitor*, 29(1), 5-13.
- Jaramillo, H. A. L., Pinos, C. A. E., Sarango, A. F. H., & Román, H. D. O. (2023). Histograma y distribución normal: Shapiro-Wilk y Kolmogorov Smirnov aplicado en SPSS: Histogram and normal distribution: Shapiro-Wilk and Kolmogorov Smirnov applied in SPSS. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*, 4(4),

Article 4. <https://doi.org/10.56712/latam.v4i4.1242>

Liu, Q., & Wang, L. (2021). T-Test and ANOVA for data with ceiling and/or floor effects. *Behavior Research Methods*, 53(1), 264-277. <https://doi.org/10.3758/s13428-020-01407-2>

Lugo-Armenta, J. G., & Pino-Fan, L. R. (2022). Niveles de Razonamiento Inferencial para el Estadístico t-Student. *Bolema: Boletim de Educação Matemática*, 35, 1776-1802. <https://doi.org/10.1590/1980-4415v35n71a25>

Martínez Ortega, R. M., Tuya Pendás, L. C., Martínez Ortega, M., Pérez Abreu, A., & Cánovas, A. M. (2009). El Coeficiente de Correlacion de los Rangos de Spearman Caracterizacion. *Revista Habanera de Ciencias Médicas*, 8(2), 0-0.

McLeay, F., Robson, A., & Yusoff, M. (2017). New Applications for Importance-Performance Analysis (IPA) in Higher Education: Understanding Student Satisfaction. *Journal of Management Development*, 36, 780-800. <https://doi.org/10.1108/JMD-10-2016-0187>

Noronha, A., Nunes, M. F. O., & Ambiel, R. A. M. (2007). Importance and knowledge in psychological assessments: A study with Psychology students. 17, 231-244. <https://doi.org/10.1590/S0103-863X2007000200007>

Oyedeffi, A., Salami, A. M., Folorunsho, O., & Abolade, O. R. (2020). Analysis and Prediction of Student Academic Performance Using Machine Learning. *JITCE (Journal of Information Technology and Computer Engineering)*. <https://doi.org/10.25077/jitce.4.01.10-15.2020>

Pito, D. U., Moreno, E. C., & Becerra, M. M. (2020). Bodega de datos con alta capacidad de análisis para el desempeño académico de Universidades. 8, 102-118. <https://doi.org/10.17081/INVINNO.8.3.4707>

Zaborras, R., Martín, C. R., & Castellà, C. O. i. (2020). Análisis del comportamiento informacional de los estudiantes posgraduados de la Facultad de Educación de la Universidad de Barcelona. *Revista Interuniversitaria de Formación del Profesorado*, 34, 167-186. <https://doi.org/10.47553/rifop.v34i2.79612>