

Análisis del Rendimiento Académico de Estudiantes de las Carreras Economía y Turismo con Power BI en los Periodos (2021)

Analysis of the Academic Performance of Students of the Economics and Tourism Majors with Power BI in the Periods (2021)

Johnny Samuel Salto- Mero¹
Universidad Técnica de Manabí, Portoviejo - Ecuador
jsaltos5862@utm.edu.ec

Marely del Rosario Cruz-Felipe²
Universidad Técnica de Manabí, Portoviejo - Ecuador
marely.cruz@utm.edu.ec

doi.org/10.33386/593dp.2024.1.2162

V9-N1 (ene-feb) 2024, pp 762-772 | Recibido: 23 de septiembre del 2023 - Aceptado: 12 de diciembre del 2023 (2 ronda rev.)

1 ORCID: <https://orcid.org/0000-0002-9454-3633>

2 ORCID: <https://orcid.org/0000-0003-1937-1568>

Descargar para Mendeley y Zotero

RESUMEN

En el mundo universitario existe una gran relevancia por obtener datos precisos donde se pueda valorar el rendimiento académico de los estudiantes ya que es un tema de gran interés, muchas veces el proceso se vuelve un poco tedioso y complicado para llegar a un objeto de estudio. La ventaja de utilizar herramientas para el análisis del rendimiento académico de estudiantes es que proporciona datos objetivos y precisos lo que permite identificar áreas de mejoras, tomando medidas para abordarlas cumpliendo de una forma rápida y eficaz a los procesos necesarios para lidiar con grandes flujos de datos. La metodología seleccionada CRISP-DM incluye la ejecución de seis etapas: comprensión del negocio, comprensión de los datos, preparación de los datos, etapas del modelado, evaluación y despliegue, donde una vez analizada la base de datos se definen las variables y filtro de datos, y finalmente la ejecución de los algoritmos de aprendizaje automático (Árbol de decisión, Bosque Aleatorios, Redes Neuronales, Máquina de Soporte Vectorial) para obtener el rendimiento académico de los estudiantes mediante el lenguaje Python. El objetivo de este artículo es realizar un análisis del rendimiento académico en estudiantes de las carreras de Economía y Turismo de la Universidad Técnica de Manabí para lo cual se evaluó diferentes algoritmos, obteniendo que el algoritmo más eficiente es el Random Forest arrojando valores precisos permitiendo obtener un dashboard con las estadísticas de los estudiantes.

Palabras clave: universitario, rendimiento académico, herramientas, procesos, análisis.

ABSTRACT

In the university world, there is great relevance in obtaining precise data where academic performance of students can be evaluated since it is a topic of great interest. Many times, the process becomes a bit tedious and complicated to reach a study object. The advantage of using tools for the analysis of academic performance of students is that it provides objective and precise data, which allows identifying areas of improvement, taking measures to address them, quick and efficiently dealing with processes required to handle large flows of data. The selected methodology, CRISP-DM, includes the execution of six stages: understanding the business, understanding the data, data preparation, modeling stages, evaluation, and deployment. Once the database is analyzed, variables and data filters are defined, and finally, the execution of automatic learning algorithms (Decision Tree, Random Forests, Neural Networks, Support Vector Machine) is carried out to obtain academic performance of students using Python language. The objective of this article is to perform an analysis of academic performance in Economics and Tourism students at the Technical University of Manabi for which different algorithms were evaluated, obtaining that the most efficient algorithm is the Random Forest giving precise values, allowing to obtain a dashboard with student statistics.

Keywords: university; academic performance; tools; processes; analysis.

Introducción

El análisis del rendimiento académico de los estudiantes ha sido de gran interés en investigaciones a lo largo del tiempo, destaca su importancia en la educación y psicología porque evalúa la eficacia y calidad de los procesos educativos de los estudiantes ante el esfuerzo que realizan las instituciones y en especial el rol de los docentes, estos indicadores forman parte de la guía de los procesos y productos del sistema educativo, a pesar del continuo interés en el análisis del rendimiento académico, aún persisten desafíos en la comprensión integral de los factores que afectan la eficacia y calidad de los procesos educativos. La evaluación del rendimiento académico se considera crucial en los ámbitos de educación y psicología; sin embargo, la falta de un análisis exhaustivo de cómo diferentes variables, incluyendo el papel de los docentes y los indicadores de rendimiento, influyen en el desempeño estudiantil.

En las universidades es muy importante contar con datos precisos para medir el rendimiento académico de los estudiantes, sin embargo, el proceso es algo complicado hasta lograr el objetivo de la investigación. Existen herramientas especializadas en recopilar y analizar datos teniendo un proceso eficaz, como son las herramientas Tableau, Power BI, Google Data Studio entre otras (Martos, 2018), contando con características y enfoques adaptándose a las necesidades y obteniendo un buen análisis de estudio, su utilización permite conectar y transformar datos de diferentes fuentes, crear visualizaciones y reportes interactivos para ver el rendimiento académico, las ventajas de utilizar estas herramientas para el análisis del rendimiento académico de estudiantes de la Facultad de Ciencias Administrativas y Económicas, de las carreras de Economía y Turismo, es que cumple de una forma rápida y eficaz los procesos necesarios para lidiar con grandes flujos de datos, de tal manera que se puedan obtener resultados en períodos relativamente cortos de tiempo, básicamente teniendo la capacidad de analizar y procesar información, con el fin de llegar al objetivo de analizar el rendimiento académico de estudiantes

utilizando herramientas y estableciendo técnicas de Big Data para extraer la información de los estudiantes de las carreras Economía y Turismo utilizando la herramienta Power BI.

Materiales y Método

Para el desarrollo de esta investigación se empleó los siguientes métodos y técnicas:

El método inductivo-deductivo para identificar las características generales del proceso y sacar conclusiones de ellas durante la preparación del documento y del producto final.

El método análisis e interpretación de la información recopilada para analizar los datos y así interpretarlos, sintetizarlos y obtener una comprensión de los aspectos fundamentales.

La Revisión bibliográfica teniendo un respaldo teórico de las investigaciones que existen dentro de libros, revistas y documentos.

Además, en la investigación se realizó la evaluación de las técnicas de Aprendizaje Automático (Machine Learning) para predecir el rendimiento académico de los estudiantes de la Facultad de Ciencias Administrativas y Económicas, sobre la base de información de resultados académicos. Las técnicas de Aprendizaje Automático a considerar son: Árboles de decisión, Bosques aleatorios, Redes neuronales y Máquina de Soporte Vectorial, dentro de estos algoritmos se encontró a bosques aleatorios como el que representa mayor precisión en la predicción del rendimiento académico, según (Yañez Cabrejas, R. 2022) El aprendizaje automático es una rama de la inteligencia artificial que se centra en el desarrollo de algoritmos que pueden aprender de los datos sin ser explícitamente programados.

La metodología empleada CRISP-DM se ejecuta en 6 etapas: (1) Comprensión del negocio; (2) Comprensión de los datos; (3) Preparación de los datos; (4) Etapa de modelado; (5) Evaluación; (6) Despliegue. Al final de la investigación estos datos serán representados en la herramienta Power BI que permite visualizar y analizar grandes cantidades de información

de manera fácil y accesible (Parra, Ricón, & Romero, 2019). Según (Eckerson et al., 2000) CRISP-DM es una herramienta integral de extracción de datos. metodología y modelo de proceso que proporciona a cualquier persona, desde principiantes hasta expertos en minería de datos, con un plan completo para realizar un proyecto de minería de datos.

Comprensión del negocio

En esta primera fase, se recopiló información sobre los estudiantes de la Facultad de Ciencias Administrativas y Económicas, de las carreras de Economía y Turismo de la Universidad Técnica de Manabí, conociendo sus objetivos y desafíos, analizando las características de los estudiantes, su metodología de enseñanza, la infraestructura y los recursos que pueden afectar al rendimiento académico.

Comprensión de los datos

La comprensión de los datos es el proceso de extraer significado de los datos. Se trata de entender qué significan los datos, qué información contienen y cómo se pueden utilizar (O'Neil, 2016) En esta segunda fase, la investigación parte de un conjunto de datos brindado por la Universidad Técnica de Manabí, los cuales corresponden a los estudiantes de las carreras de Economía y Turismo en las diferentes asignaturas del período P1-P2, correspondientes al año 2021. Las técnicas evaluadas mediante la base de datos proporcionada por el área de Tecnologías de la Información y la Comunicación (TIC) se trabaja con los registros que contiene cada estudiante, seleccionando 14 variables que fueron analizadas mediante el lenguaje Python.

Preparación de los datos

En esta tercera fase, se utilizó el lenguaje Python que permite aplicar técnicas de minería de datos, realizando una visualización resumida de la estructura de la información e identificando los valores nulos, así como la información en detalle del conjunto de datos como número, etiquetas de las columnas, tipos de datos, número de celdas en cada columna, a través

de Python, para las variables con valores nulos se realizó la identificación y la eliminación y por último una separación del conjunto de datos en 80% para train (entrenamiento) y un 20% datos de test (evaluación). En el proceso de transformación se aplicó un método de normalización, transformando los tipos de variables seleccionados, este paso ayuda que los datos medidos tenga una escala diferente y sean más fácil de compararlos entre sí eliminando los efectos de influencia sobre la variable dependiente (Yuri Nieto, 2019).

Modelado

En esta cuarta fase, se utilizó los algoritmos de aprendizaje automáticos, que son técnicas que permiten combinar múltiples modelos al hacer predicciones, en lugar de usar uno solo. Se utilizó desde el más simple como el algoritmo Árbol de Decisión, un modelo de aprendizaje automático que utiliza un conjunto de reglas de decisión para construir un árbol de decisiones que se puede utilizar para predecir la clase de un objeto o evento. Este proceso implica la recopilación de datos, la selección de características, la construcción del árbol, la evaluación del árbol y la predicción (Díaz, Ahumada, & Melo, 2021), hasta el más sofisticado el algoritmo de Bosque Aleatorio (Random Forest), un método de aprendizaje automático supervisado que combina varios árboles de decisión independientes y no correlacionados para mejorar la precisión y la estabilidad del modelo, que implica un proceso de selección de datos, muestreo de datos, construcción de árboles de decisión, la evaluación del modelo y la predicción (Céspedes, 2022), en una amplia gama de problemas en la actualidad.

Evaluación

En esta quinta fase, se usó la evaluación de la factibilidad mediante el uso de técnicas de Aprendizaje Automático (Machine Learning) para así predecir el rendimiento académico de los alumnos, sobre la información brindada académica de los estudiantes (**Tabla 1**).

Tabla 1
Algoritmos de Aprendizaje Automático

Técnica	Algoritmos
Aprendizaje Automático (Machine Learning)	Arbol de Decision
	Bosques Aleatorios
	Redes Neuronales
	Máquina de Soporte Vectorial

Despliegue

En esta sexta fase, de despliegue y la visualización de los resultados se utiliza la herramienta Power Bi, la cual permite tratar con un gran volumen de datos, en el cual se carga el modelo generado y realiza la predicción a partir del dataset proporcionado acorde a la estructura establecida de los datos.

Resultados y Discusión

Resultados

Comprensión del negocio

Se hizo un análisis en las carreras de Economía y Turismo de la Universidad Técnica de Manabí obteniendo información sobre los estudiantes, con la ayuda del area de TIC se pudo obtener como estas carreras están definidas tanto de sus infraestructuras, mallas curriculares, su metodología de enseñanza y las características que tiene cada estudiante dentro de su facultad, para así definir con que características se podrá analizar el rendimiento académico y conocer cuáles son los factores que influyen.

Comprensión de los datos

En la metodología CRISP-DM una de las etapas es el levantamiento de información, por eso se cuenta con una base de datos académica de las carreras Economía y Turismo, donde se procede a realizar un diccionario para entender mejor las variables que ayudara mucho en el análisis.

Una vez conocidas todas las características se procede en la selección, mostrando aquellas que influyen en el desempeño académico:

Tabla 2
Variables tomadas del desempeño académico

Características	Descripción
Id_estudiante	Identificador unico que pertenece a cada estudiante
Estudiante	Nombres completos del estudiante
Facultad	Se refiere a la facultad que pertenece el dato
Departamento	Especifica la carrera que cursa cada estudiante
Periodo	Especifica el tiempo del semestre
Id_curso	Identificador unico del curso
Fullname	Asignatura que escogió el estudiante
Modalidad	Forma en la que el estudiante cursa las clases
Num_acesos	Se refiere a los accesos al aula virtual
Num_actividades_completadas	Número de tareas realizadas
Nota	Método utilizado para evaluar los trabajos académicos
Porc_final_nota	En la nota final del estudiante valorado en porcentaje
Rendimiento_académico	Se evalúa la capacidad del alumno mediante 1 y 0
Rendimiento_académico_	Si el alumno aprueba la materia 1 = Aprobado, caso contrario 0 = Reprobado

Preparación de los datos

En la ejecución de este módulo, luego de realizar la limpieza y transformación de los datos, se tuvo el registro de estudiantes de 9958, que corresponden a dos periodos académicos del año 2021 de la modalidad virtual, en dichos periodos las actividades se llevaron a cabo de manera virtual tomando en cuenta las materias

que se imparten por carrera Economía y Turismo en la UTM.

Modelado

Los datos que se utilizan son datos del periodo académico del año 2021, lo que procede a la utilización de los algoritmos de aprendizaje automático. Se mostrará el algoritmo más eficiente, en este caso el algoritmo Random Forest.

Bosque Aleatorio (Random Forest)

Con este modelo se busca normalizar la data, dividir la data en entrenamiento y testeo, crear el modelo Random Forest y entrenar con la data de entrenamiento en las siguientes funciones:

- Función rf_bch para:
- Calcular y mostrar la matriz de confusión.
- Calcular y mostrar la curva de ROC.
- Medir el nivel de precisión.

Figura 1

Declarando la Función rf_bch

```
# Definición de la función del arbol
def rf_bch(d,i,j,nom):
    var_dep = d[nom]
    var_ind = d.drop(nom,axis= 1)
    var_ind_norm = StandardScaler().fit_transform(var_ind)
    var_ind_ent, var_ind_test, var_dep_ent, var_dep_test = train
    #test_size=0.3 -- 30% de la data se usa para entrenar
    rf= RandomForestClassifier(n_estimators=10,max_depth=4)
    #activation="relu" -- activar una función para cada nodo. ma
    rf.fit(var_ind_ent,var_dep_ent)
    y_pred = rf.predict(var_ind_test)
    cm = confusion_matrix(var_dep_test, y_pred)
    print("Matriz de confusión")
    print(cm)
```

Función rf_arq para:

Medir el nivel de precisión.

Figura 2

Declarando la Función rf_arq

```
# Se define la siguiente función
def rf_arq(d,i,j,nom):
    var_dep = d[nom]
    var_ind = d.drop(nom,axis= 1)
    var_ind_norm = StandardScaler().fit_transform(var_ind)
    var_ind_ent, var_ind_test, var_dep_ent, var_dep_test = train
    #test_size=0.3 -- 30% de la data se usa para entrenar
    rf= RandomForestClassifier(n_estimators=1,max_depth=j)
    rf.fit(var_ind_ent,var_dep_ent)

    return rf.score(var_ind_test,var_dep_test)
```

Llamamos la función rf_arq dentro de un loop doble de 10 a101 en pasos de 10.

Figura 3

Ejecución de la sentencia Loop

```
# Se define la siguiente función
def rf_arq(d,i,j,nom):
    var_dep = d[nom]
    var_ind = d.drop(nom,axis= 1)
    var_ind_norm = StandardScaler().fit_transform(var_ind)
    var_ind_ent, var_ind_test, var_dep_ent, var_dep_test = train
    #test_size=0.3 -- 30% de la data se usa para entrenar
    rf= RandomForestClassifier(n_estimators=1,max_depth=j)
    rf.fit(var_ind_ent,var_dep_ent)

    return rf.score(var_ind_test,var_dep_test)
```

Usando la arquitectura que nos dio en mejor nivel de precisión llamamos a la función

Figura 4

Arquitectura de los datos de la función

```
A = rf_bch(d,n1,n2,'rendimiento_academico')
print("Arquitectura: %dx%d, Accuracy: %0.4f" % (n1, n2, A ))
```

Antes utilizaremos el algoritmo Random Forest, observando diferentes arquitecturas y mostrando el resultado del mayor nivel de precisión de su arquitectura, utilizando los datos de la carrera de Economía de la Universidad Tecnica de Manabi.

Figura 5
 Métrica de evaluación Random Forest de la carrera Economía

```
Arquitectura: 10x10, Accuracy: 0.8909
Arquitectura: 20x10, Accuracy: 0.8920
Arquitectura: 20x10, Accuracy: 0.8920
Arquitectura: 20x10, Accuracy: 0.8920
Arquitectura: 50x10, Accuracy: 0.8925
Arquitectura: 50x10, Accuracy: 0.8925
Arquitectura: 70x10, Accuracy: 0.8927
Arquitectura: 70x10, Accuracy: 0.8927
Arquitectura: 70x10, Accuracy: 0.8927
Arquitectura: 70x10, Accuracy: 0.8927
```

Mostramos el resultado del nivel de precisión, matriz de confusión y curva de ROC.

La matriz de confusión se utiliza para describir el rendimiento de un modelo de clasificación en un conjunto de datos de prueba para los cuales los valores verdaderos son conocidos.

Figura 6
 Matriz de confusión de la carrera Economía

```
Matriz de confusión
[[ 282  519]
 [   37 3441]]
AUROC = 0.847
```

El modelo ha acertado 3441 de los casos de aprobar el semestre, y 282 de los casos de no aprobar el semestre. Sin embargo, podemos ver 37 falsos positivos (error tipo I) en donde aprobó el semestre, pero el modelo clasificó que no aprobó y 519 falsos negativos (error tipo II) en donde no aprobó el semestre, pero el modelo clasificó que sí aprobó. Si comparamos con los datos obtenidos podemos observar que en la carrera de Economía el total de los datos aprobados fue de 3475 y el total de reprobados es de 803, comprobando que el algoritmo Random Forest con su modelo acertó en un 0.85 por ciento con la data brindada. Así mismo podemos observar la curva ROC y el nivel de precisión (Accuracy) calculado el cual es de: 0.8701

Figura 7
 Gráfica del nivel de precisión Random Forest con datos de la carrera Economía

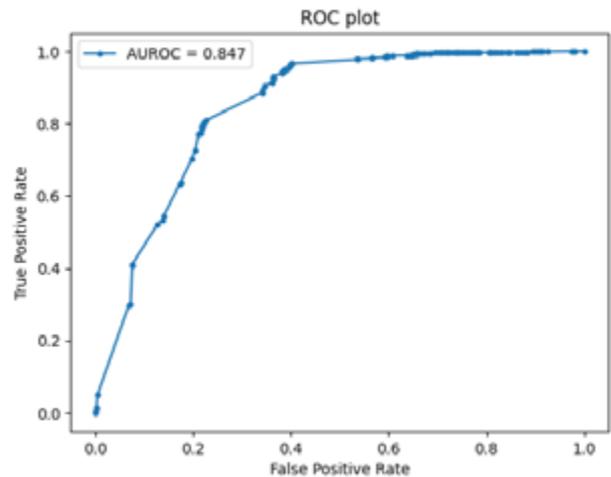


Figura 8
 Resultados de métrica de evaluación de Random Forest con datos de la carrera Economía

```
Arquitectura: 70x10, Accuracy: 0.8701
```

Utilizaremos también el algoritmo Random Forest para observar las diferentes arquitecturas y mostrar el resultado del mayor nivel de precisión, utilizando ahora los datos de la carrera de Turismo y Gastronomía de la Universidad Técnica de Manabí.

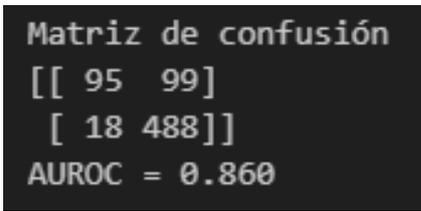
Figura 9
 Métrica de evaluación Random Forest de la carrera Turismo y Gastronomía

```
Arquitectura: 10x10, Accuracy: 0.8381
Arquitectura: 60x10, Accuracy: 0.8429
```

Mostramos el resultado del nivel de precisión, matriz de confusión y curva de ROC. Se muestra también la matriz de confusión.

Figura 10

Matriz de confusión de la carrera Turismo y Gastronomía



El modelo ha acertado 488 de los casos de aprobar el semestre, y 95 de los casos de no aprobar el semestre. Sin embargo, podemos ver 18 falsos positivos (error tipo I) en donde aprobó la asignatura, pero el modelo clasifico que no aprobó y 99 falsos negativos (error tipo II) en donde no aprobó el semestre, pero el modelo clasifico que si aprobó.

Si comparamos con los datos obtenido podemos observar que en la carrera de Turismo el total de los datos aprobado fue de 522 y el total de reprobado es de 179, comprobando que el algoritmo Random Forest con su modelo acertó en un 0.86 por ciento con la data brindada. Así mismo podemos observar la curva ROC y el nivel de precisión (Accuracy) calculado el cual es de: 0.8329.

Figura 11

Gráfica del nivel de precisión Random Forest con datos de la carrera Gastronomía y Turismo

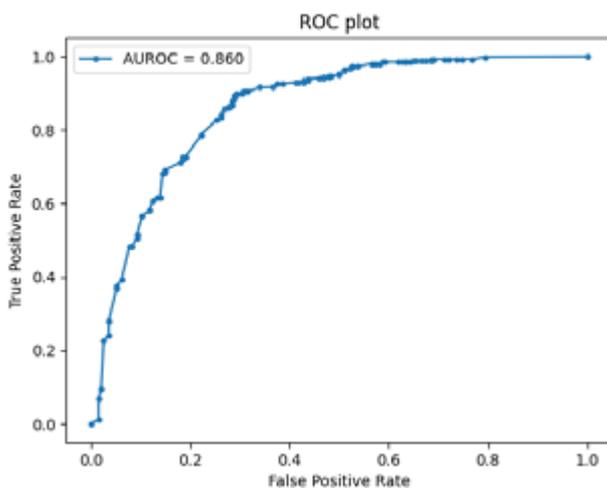


Figura 12

Resultados de métrica de evaluación de Random Forest con datos de la carrera Gastronomía y Turismo



Evaluación

Una vez que se realizó la evaluación de las técnicas de Aprendizaje Automático (Machine Learning) de la base de datos, con las variables escogida en la tabla 2, para predecir el rendimiento académico, utilizando los algoritmos de Bosque Aleatorio, Árbol de Decisión, Redes Neuronales y Maquinas de Decisiones Vectoriales, del conjunto de datos brindado, los cuales fueron analizadas mediante el lenguaje Python.

Observaremos los porcentajes de precisión que brindo cada uno de los algoritmos en ambas carreras, observando cual fue el más eficiente.

Figura 13

Resultados de los algoritmos con datos de la carrera Economía

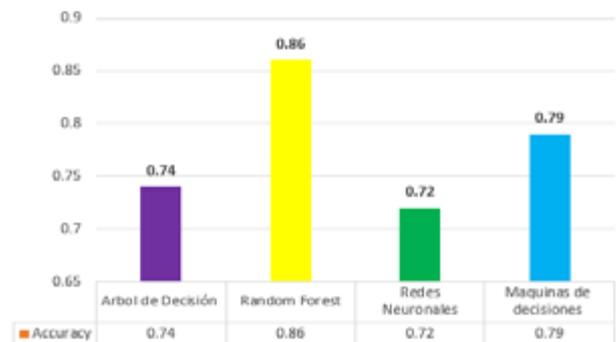
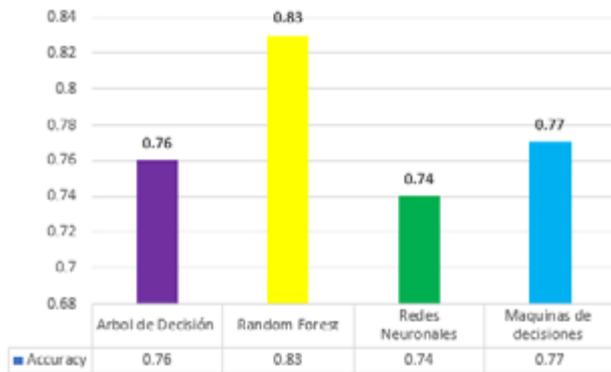


Figura 14

Resultados de los algoritmos con datos de la carrera Gastronomía y Turismo



Despliegue

En esta investigación de análisis del rendimiento académico de los estudiantes de la carrera Economía y Turismo de la Universidad Técnica de Manabí, se implementó el modelo con la utilización de los algoritmos de aprendizaje automático y la creación del dashboard, lo que permitiría su utilización por un tiempo determinado dentro de las carreras escogidas de la institución.

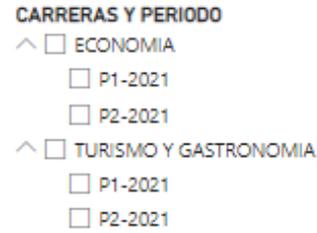
Los datos obtenidos se representan en la herramienta Power BI a través del modelo que se obtuvo con el algoritmo Random Forest, demostrando que fue el mejor para predecir el rendimiento académico de los estudiantes.

Descripción del Dashboard en Power BI

En la primera visualización dentro se podrá observar las carreras donde se realizó el análisis del rendimiento, en este caso de las carreras de Economía y Turismo de la Universidad Técnica de Manabí, además de los periodos que corresponde a cada semestre.

Figura 15

Tomando medidas de acorde al diseño del material lúdico



En la segunda visualización se podrá observar las materias y los docentes que la imparten en cada periodo, esto ayudara para conocer si en ese periodo puede haber un bajo rendimiento académico que tiene los estudiantes, identificando las fortalezas y debilidades que tienen en dicha materia con sus docentes.

Figura 17

Lista de estudiantes y su nota respectiva de cada materia

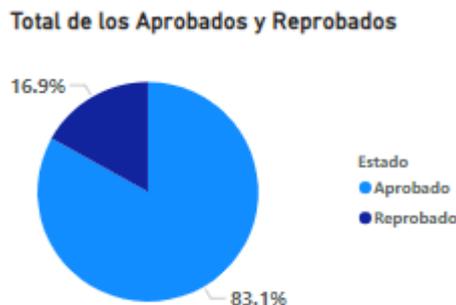
Estudiante	Suma de nota
ABAD BETANCOURT ANA CRISTINA	12
ACEBO PIN LUIS ADRIAN	448
ACERO LANCHIMBA NANCY PATRICIA	89
ACITIMBAY MERO KARLA MAYIYA	100
ACOSTA ESPINOZA DAYANA MERCEDES	99
ACOSTA INTRIAGO ARIANA LISBETH	93
ACOSTA LINO JULEXY DORILA	78
ACOSTA MACIAS KARLA XIOMARA	201
ACOSTA RODRIGUEZ EYMI LEONORIS	93
ACOSTA ZAMBRANO LUIS DAVID	100
AGUAIZA SANTOS FREDDY JOEL	82
AGUAYO CEDEÑO WALTHER STEVEN	198
AGUAYO ESPINALES NAIROBITH MAYERLINE	93
AGUAYO LOOR GISELLA LISSETTE	94
AGUAYO MENENDEZ JOSSELYN KATHERINE	378
AGUAYO MOREIRA JEAN CARLOS	88
AGUAYO RIVERA JOHAN ARIEL	99
AGUAYO VELIZ DAYANA YAMILETH	91
AGUAYO ZAMBRANO KARLA MARIBEL	93
AGUILAR BARAHONA MARIA DE LOS ANGELES	10
AGUIRRE CHAVEZ MILENKA ALEJANDRA	85
AGUIRRE COVEÑA RONALD KELLY	69
AGUIRRE GARCIA MANUEL STEVEN	460
AJILA VILLEGAS XAVIER ALEJANDRO	91
ALARCON ACOSTA SHERY MAIBETH	563
ALARCON ACOSTA SHEYLA DAYANA	96
ALARCON ALVARADO EMILY ANELY	85

En la cuarta visualización mostrara el porcentaje de aprobado y los no aprobado que tiene cada materia, además el total global de todos los estudiantes de ambas carreras visualizando

cuantos aprobado y no aprobado tuvo en ese año de estudio.

Figura 18

Total, de aprobados y no aprobados de los estudiantes



Discusión

Los resultados que arrojaron cada algoritmo, sugieren que el Bosque Aleatorio (Random Forest) es el mejor en analizar dicho rendimiento ya que da más precisión al momento de analizar el conjunto de datos, dándonos una estimación aceptable de 0.86 especificado en la figura 13, con los datos de la carrera de Economía, mientras que en la carrera de Gastronomía y Turismo con una estimación aceptable de 0.83 en la figura 14, en cambio el algoritmo Árbol de decisión nos arrojó una estimación de 0.74 (figura 13) en la carrera de Economía, mientras que en la carrera de Gastronomía y Turismo la estimación de 0.76 (figura 14), mientras los algoritmos Redes neuronales con una estimación de 0.72 (figura 13) en la carrera de Economía y en la carrera de Gastronomía y Turismo la estimación de 0.74 (figura 14) y por último en la Máquina de decisiones vectorial con 0.79 (figura 13) en la carrera de Economía, mientras en la carrera de Turismo y Gastronomía con 0.77 (figura 14), deduciendo que el Random Forest presenta una mayor precisión en la predicción del rendimiento académico. Con estos resultados se puede observar que los estudiantes tuvieron un buen desempeño en la carrera de Economía, y que muy pocos no lo tuvieron, lo que ayuda a proponer estrategias de estudios para conocer cuáles fueron los factores más esenciales

que dichos estudiantes no tuvieron el buen rendimiento.

Conclusión

Los algoritmos de aprendizaje automático pueden analizar grandes conjuntos de datos para identificar los factores que tienen un impacto significativo en el rendimiento académico de los estudiantes incluyendo las variables como el historial académico, la asistencia a clases, promedios, entre otros.

Al utilizar diversas técnicas y algoritmos en análisis de datos con la metodología CRISP-DM permitió analizar y evaluar el rendimiento académico de los estudiantes ya que la metodología proporciona un marco estructurado y flexible para guiar el proceso de análisis e implementación de soluciones basadas en datos.

Con la visualización del dashboard en Power BI permitió observar y analizar datos de manera clara y efectiva recopilando datos como asistencia, notas, calificaciones y otros resultados académicos de los estudiantes, y presentarlos en un panel de control fácil de entender, identificando patrones y tendencias en los datos, lo que permite a los educadores y administradores tomar decisiones informadas sobre cómo mejorar el desempeño académico y maximizar el éxito de los estudiantes.

Referencias Bibliografía

- Martos, S. (2018). *Visualización y análisis de métricas de desarrollo y mantenimiento en productos software*. Etsinf, 64.
- Arteaga, W., & Sandoval, J. P. (2018). *Factores que intervienen en el rendimiento académico en la Universidad*. Acta Nova.
- Basto, R. (2017). *La función docente y el rendimiento académico: Una aportación al estado del conocimiento*. Comie.
- Bedregal, N., Tupacyupanqui, D., & Cornejo, V. (2020). *Análisis del rendimiento académico de los estudiantes de Ingeniería de Sistemas, posibilidades*

de deserción y propuestas para su retención. Ingeniare.

- Céspedes, B. (2022). *Aplicación del algoritmo del bosque aleatorio a un modelo de clasificación de la anemia en niños peruanos.* Mediciego.
- Díaz, M., Ahumada, M. d., & Melo, J. (2021). *Árboles de Decisión como Metodología para Determinar el Rendimiento Académico en Educación Superior.* Unilasallista.
- Galo Flores, J. C. (2019). *Minería de datos como herramienta estratégica.* Revista Científica Mundo de la Investigación y el Conocimiento.
- Parra, J., Rincon, M., & Romero, D. (2019). *Propuesta de inteligencia de negocios mediante la herramienta microsoft Power Bi como Soporte para la toma de decisiones del área comercial de la empresa Abc Manufacturera de Productos Plásticos.* Poli.
- Recarey, R. (2020). *Métodos de ensamblado en Machine Learning.* Universidad de Santiago de Compostela.
- Rodriguez, J., & Rodriguez, S. (2022). *Uso de Python para el análisis de datos aplicado en la investigación.* Incaing.
- Eckerson, W. W., Hanlon, N., & Barquin, R. (2000). *DIRECTOR OF EDUCATION AND RESEARCH.* 5(4).
- Yañez Cabrejas, R. (2022). *Aprendizaje automático: estado de la cuestión y casos de estudio.* Universidad Politécnica de Madrid.
- O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy.* Crown Publishing Group.