



DOI: <http://dx.doi.org/10.23857/dc.v8i1.2481>

Ciencias de la Educación  
Artículo de Investigación

*Data Science para la Predicción del Rendimiento Académico*

*Data Science for Prediction of Academic Achievement*

*Ciência de dados para a previsão do desempenho acadêmico*

Zoila Jessenia Fabara-Sarmiento <sup>I</sup>  
[jessenia.fabars@ug.edu.ec](mailto:jessenia.fabars@ug.edu.ec)  
<https://orcid.org/0000-0002-9423-6174>

Janeth Pilar Diaz-Vera <sup>II</sup>  
[janeth.diazv@ug.edu.ec](mailto:janeth.diazv@ug.edu.ec)  
<https://orcid.org/0000-0001-8750-0216>

Alicia Karina Ruiz-Ramirez <sup>III</sup>  
[Alicia.ruizram@ug.edu.ec](mailto:Alicia.ruizram@ug.edu.ec)  
<https://orcid.org/0000-0002-3038-045X>

**Correspondencia:** [jessenia.fabars@ug.edu.ec](mailto:jessenia.fabars@ug.edu.ec)

\***Recibido:** 20 de noviembre de 2021 \***Aceptado:** 18 de diciembre de 2021 \* **Publicado:** 06 de enero de 2022

- I. Master Universitario en Tecnología Educativa y Competencias Digitales, Ingeniera en Diseño Gráfico, Licenciada en Ciencias de la Educación Mención Sistemas Multimedia, Diseñador Gráfico, Docente, Universidad de Guayaquil, Guayaquil, Ecuador.
- II. Magister en Educación Informática, Licenciada en Ciencias de la Educación Especialización Informática, Tecnóloga Pedagógica en Informática, Docente, Universidad de Guayaquil, Guayaquil, Ecuador.
- III. Magister en Educación Informática, Ingeniera en Sistemas Computacionales, Licenciada en Ciencias de la Educación Especialización Educadores de Párvulos, Docente, Universidad de Guayaquil, Guayaquil, Ecuador.

## Resumen

En el último año las instituciones educativas se han visto afectadas por la emergencia sanitaria, la educación se tornó netamente virtual, lo cual en muchos casos no ha sido favorecedor para todos los estudiantes. Esta investigación se realizó con el objetivo de proporcionar una aplicación de data science que permita realizar una predicción del rendimiento académico de los estudiantes de nivel básica, con el fin de evaluar la efectividad de los métodos de enseñanza virtual.

El diseño metodológico se ha desarrollado en la herramienta PMI que consiste en una serie de fases secuenciales en donde permitirá la obtención de un entregable. Por esta razón, el proyecto evalúa los factores que inciden en el rendimiento académico haciendo uso de técnicas actuales como son la data science y herramientas de machine learning.

**Palabras Clave:** data science; tecnología educativa; rendimiento académico.

## Abstract

In the last year educational institutions have been affected by the health emergency, education became purely virtual, which in many cases has not been favorable for all students. This project was carried out with the aim of providing a data science application that allows a prediction of the academic performance of students at the basic level, in order to evaluate the effectiveness of virtual teaching methods.

The methodological design has been developed in the PMI tool that consists of a series of sequential phases where it will allow the obtaining of a deliverable. For this reason, the project evaluates the factors that affect academic performance using current techniques such as data science and machine learning tools.

**Keywords:** data science; educational technology; academic performance.

## Resumo

No último ano as instituições de ensino foram afetadas pela emergência sanitária, a educação tornou-se puramente virtual, o que em muitos casos não tem sido favorável para todos os alunos. Este projeto foi realizado com o objetivo de fornecer um aplicativo de ciência de dados que permita uma previsão do desempenho acadêmico de alunos do nível básico, a fim de avaliar a eficácia dos métodos de ensino virtual.

O desenho metodológico foi desenvolvido na ferramenta PMI que consiste em uma série de fases sequenciais onde permitirá a obtenção de um entregável. Por esse motivo, o projeto avalia os fatores que afetam o desempenho acadêmico usando técnicas atuais, como ciência de dados e ferramentas de aprendizado de máquina.

**Palavras-chave:** ciência de dados; tecnologia Educacional; performance acadêmica.

## Introducción

Las clases en línea hoy en día son un método de enseñanza que de manera fortuita la mayoría de las instituciones se tuvieron que adaptar, sin embargo, no es algo que ha sido fácil, ya que profesores, autoridades, padres de familia y los propios alumnos han sufrido por varios cambios, uno de los principales, el uso las diferentes herramientas tecnológicas que lleva como consecuencia el bajo rendimiento de los estudiantes.

El método de aprendizaje es muy distinto, si antes era algo complicado, ahora la situación es mucho más difícil, los estudiantes no están preparados para auto tareas, igual los maestros no capacitados para ser sus guías de aprendizajes, pues la educación pierde calidad. (Yuquilema, 2020)

Según (Espinoza, 2020) Cifras estadísticas, muestran que la mayor parte de la población joven en el país entre niñas, niños, y adolescentes son de 6 millones entre 0 a 17 años, el 33% de los jóvenes son de 5 a 11 años que requieren acompañamiento en el trabajo virtual, de esas cifras el 50% de la población joven se encuentra en la costa, el 33% está en la sierra y el 7 % se ubica en la Amazonia. La Unidad Educativa “José Elías Altamirano” se mantiene con un equipo de trabajo excelente que día a día buscan el bienestar de los estudiantes, sin embargo, sufre una gran problemática a consecuencia de contar con un año lleno de cambios, las bases de aprendizaje de los alumnos no son los más positivos, las autoridades de la institución buscan saber cuáles serían las mejores estrategias para mejorar el rendimiento de los estudiantes.

En la actualidad en nuestro país no existe un sistema predictivo que logre detectar el rendimiento académico de los estudiantes y así tener en cuenta diferentes estrategias de enseñanzas que se podrían aplicar para obtener mejores resultados.

En la Unidad Educativa “José Elías Altamirano” es muy importante brindar una excelente educación con sus diferentes métodos de enseñanzas, a pesar de ello, muchos docentes no se han adaptado a esta nueva modalidad y esto se ve reflejado en el rendimiento del último año lectivo.

Con el objetivo de cumplir sus horas de clases no se tiene un riesgo de usar las herramientas necesarias para brindar un mejor método de enseñanza. Esta problemática se puede solucionar haciendo una comparativa de cómo hay personal que toma un riesgo y mejora su técnica de estudio viendo como resultado una buena base de conocimiento para su siguiente año lectivo, teniendo en cuenta los resultados del análisis en data science.

La modalidad en línea son un método de enseñanza que de manera fortuita la mayoría de las instituciones se tuvieron que adaptar, sin embargo, no es algo que ha sido fácil, ya que profesores, autoridades, padres de familia y los propios alumnos han sufrido por varios cambios y ciertas dificultades en el proceso de aprendizaje.

## **Desarrollo**

La tecnología en el ámbito educativo en estos tiempos ha evolucionado de manera significativa, hoy en día se desarrollan diferentes métodos de enseñanza para poder brindar una mejor educación a los alumnos, sin embargo, una problemática en el país es la falta de conocimiento y las herramientas necesarias para poder sacar provecho de lo que se brinda la tecnología actualmente. Para el presente proyecto, se centró específicamente en el rendimiento académico de los estudiantes de la Unidad Educativa José Elías Altamirano año 2021, ya que se empieza un año lectivo la cual se tiene una experiencia previa en donde no se obtuvieron los mejores resultados generando que muchos estudiantes no tengan el mejor desempeño académico.

Con el fin de conocer las falencias en los métodos impartidos por los docentes, es importante contar con un sistema de predicción que se ejecute las primeras semanas del año lectivo, para realizar un comparativo de las posibles mejorías que los docentes pueden ejecutar y así los alumnos tendrían un mejor rendimiento académico.

Esta propuesta se planteó con el fin de poder brindar una solución tecnológica que permita medir el rendimiento académico de los estudiantes de la Unidad Educativa José Elías Altamirano y predecir si los métodos de enseñanzas utilizados hoy en día son los adecuados tomando en cuenta que en la actualidad el sistema educativo ha optado a una modalidad virtual, dando como resultados inconvenientes en el proceso de aprendizaje.

Para esto se va a llevar a cabo el desarrollo de un aplicativo usando algoritmos que prediga el rendimiento de los estudiantes, tomando como entradas variantes relevantes para el estudio como

por ejemplo su entorno familiar, el tiempo que le dedican a los estudios y sus promedios parciales. Basado en estos parámetros podrá predecir la calificación final del estudiante y de esta forma evaluaremos si el método de enseñanza aplicado es efectivo.

De acuerdo con lo investigado, se utilizó el algoritmo regresión lineal ya que es el mecanismo más eficiente para realizar predicciones basadas en la relación entre dos o más variables. Así mismo, dicho algoritmo fue codificado en el framework Google Colab ya que es una plataforma que no requiere configuración o instalación y está diseñada específicamente para el análisis de datos.

Además, se creó un aplicativo que mostrará el resultado de la predicción realizada de forma gráfica. Dicha aplicación es amigable al usuario y se puede acceder desde un ordenador o desde un dispositivo móvil, facilitando a los docentes el acceso a la información.

La plataforma permite predecir el rendimiento académico de los estudiantes según las notas ingresadas por los docentes, con los resultados obtenidos el docente puede mejorar las técnicas de enseñanzas.

## Metodología

Con el fin de lograr el objetivo principal planteado en este proyecto, se determinó que la metodología más propicia es PMI (Project Management Institute), ya que establece una serie de fases secuenciales por las que atraviesa el proyecto, donde cada fase permitirá la consecución de un objetivo o entregable. Según indica el (Project Management Institute., 2013) “una fase del proyecto es un conjunto de actividades del proyecto, relacionadas de manera lógica, que culmina con la finalización de uno o más entregables.

Las fases del proyecto se utilizan cuando la naturaleza del trabajo a realizar en una parte del proyecto es única y suelen estar vinculadas al desarrollo de un entregable específico importante.”

Aunque no existe una estructura ideal de fases que se puedan aplicar a todos los proyectos, (Project Management Institute., 2013) indica que, para facilitar la planificación y control del proyecto, se deben establecer los siguientes procesos:

**Inicio:** Aquellos procesos realizados para definir un nuevo proyecto o nueva fase de un proyecto existente al obtener la autorización para iniciar el proyecto o fase.

**Planificación:** Aquellos procesos requeridos para establecer el alcance del proyecto, refinar los objetivos y definir el curso de acción requerido para alcanzar los objetivos propuestos del proyecto.

**Ejecución:** Aquellos procesos realizados para completar el trabajo definido en el plan para la dirección del proyecto a fin de satisfacer las especificaciones del mismo.

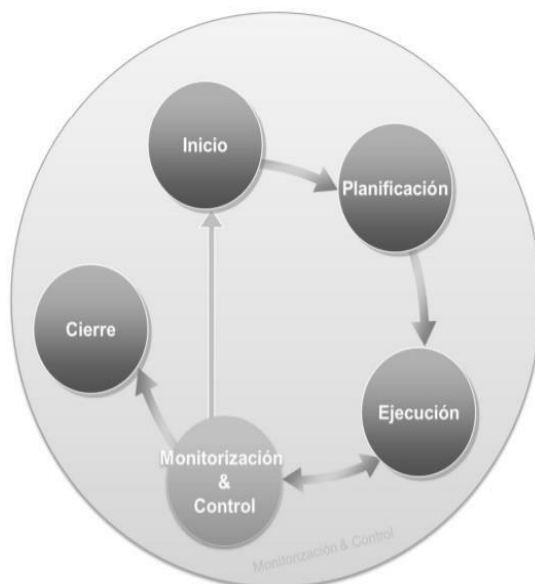
**Monitoreo y Control:** Aquellos procesos requeridos para rastrear, revisar y regular el progreso y el desempeño del proyecto, para identificar áreas en las que el plan requiera cambios y para iniciar los cambios correspondientes.

**Cierre:** Aquellos procesos realizados para finalizar todas las actividades a través de todos los Grupos de Procesos, a fin de cerrar formalmente el proyecto o una fase del mismo.

### Etapas de la metodología del proyecto

En el presente proyecto se utilizó la metodología PMI, dicha metodología define el “ciclo de vida de un proyecto como la serie de fases que un proyecto pasa desde su inicio hasta su cierre. La estructuración en fases proporciona una base formal para el control. Cada fase se inicia formalmente con la especificación de lo que se permite y se espera de la misma.” (García, 2016)  
En el siguiente gráfico se detalla la interacción de las fases del proyecto establecidas por la metodología PMI.

Gráfico 1 Metodología PMI



Fuente: (García, 2016)

## **Inicio**

Se estableció el contexto del problema a investigar, su alcance y los recursos que serán utilizados para su ejecución. Así mismo, se obtuvo la autorización de la directora de la unidad educativa José Elías Altamirano Lcda. Jenny Herrera Gómez para llevar a cabo el desarrollo del presente proyecto dentro del plantel educativo.

## **Planificación**

En esta fase definimos las actividades necesarias para ejecutar el proyecto y lograr los objetivos planteados. Se define una lista ordenada de todo lo que se realizará durante el proyecto, es decir, enumerar todas las características, funcionalidades, requisitos, mejoras y correcciones del proyecto para entregas futuras. Por lo que se fijará un cronograma con dichas actividades y su tiempo de duración, el cual se presenta en el anexo X.

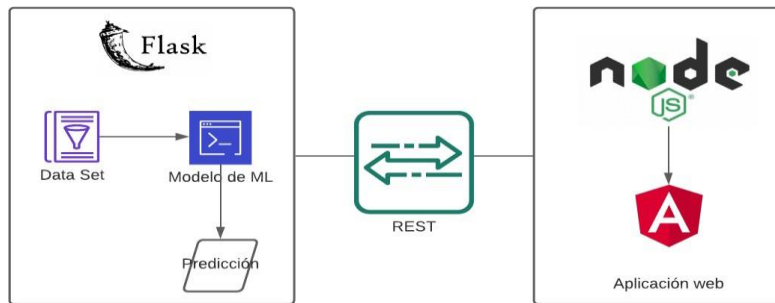
Las actividades que se realizarán son:

- Investigación bibliográfica y levantamiento de información, con el objetivo de definir la base teórica sobre la cual está sustentado el presente proyecto.
- Recolección de datos para determinar las variantes relevantes para el estudio de la predicción del rendimiento académico.
- Realización de encuestas al personal docente implicados en el proceso, estudio y análisis de los datos obtenidos.
- Programación del algoritmo de machine learning utilizado para predicción de datos.
- Diseño y desarrollo del aplicativo web para visualización de la predicción obtenida.
- Pruebas de funcionamiento del aplicativo.
- Elaboración de manual de usuario.

## **Ejecución**

Se llevan a cabo las actividades definidas en la fase anterior para realizar el diseño del proyecto. El diseño propuesto consiste en una aplicación web diseñada en Flask, la cual muestra el resultado del algoritmo de regresión lineal elaborado en Python mediante una arquitectura REST desarrollada en Angular y Node Js. En el siguiente gráfico se detalla el diseño de la aplicación:

Gráfico 2 Diseño de la aplicación



Fuente: Trabajo de Investigación

El data set utilizado está constituido por una matriz csv que contiene las variables a estudiar para realizar la predicción, el cual será entrenado posteriormente mediante el algoritmo de regresión lineal, como se visualiza en la siguiente figura.

Gráfico 3 Carga del dataframe

```
# carga de Dataset
from google.colab import files
import io
uploaded = files.upload()
dataset = pd.read_csv('student-mat.csv', sep=';')
```

Fuente: Google Colab

A continuación, se realiza la partición del nuevo dataframe para el entrenamiento y validación de datos.

Gráfico 4 Partición del dataframe

```
#particion de datos para entrenamiento
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)

from sklearn.preprocessing import StandardScaler

X_scaler = StandardScaler()
X_train = X_scaler.fit_transform(X_train)
X_test = X_scaler.transform(X_test)

y_scaler = StandardScaler()
y_train = np.array(y_train).reshape(-1, 1)
y_train = y_scaler.fit_transform(y_train)
```

Fuente: Google Colab



Posteriormente se realiza el entrenamiento de los datos mediante el modelo de regresión lineal.

Gráfico 5 Aplicación del modelo Regresión Lineal

```
from sklearn.linear_model import LinearRegression
lr = LinearRegression()
lr.fit(X_train, y_train)
y_pred = lr.predict(X_test)
y_pred = y_scaler.inverse_transform(y_pred)

rmse = mean_squared_error(y_test, y_pred, squared=False)
print("porcentaje de error:", rmse)
```

Fuente: Google Colab

Luego de haberse realizado este procedimiento, se muestra el dataframe de los resultados del modelo de machine learning y una gráfica para facilitar la interpretación de los valores obtenidos.

Gráfico 6 Dataframe de predicción

```
df = pd.DataFrame( np.c_[y_test,y_pred], columns = ["original", "predicted"] )
print(df)

plt.plot(range(len(y_pred)), y_pred, color = 'red')
plt.plot(range(len(y_test)), y_test, color = 'blue')
plt.title('Predicted G3 VS Real G3')
plt.xlabel('Students')
plt.ylabel('Grades')
plt.show()
```

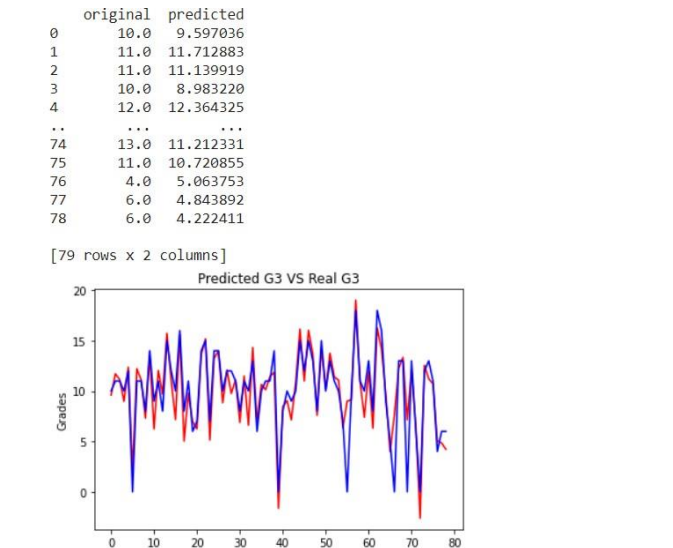
Fuente: Google Colab

## Monitoreo y control

Luego de realizar el diseño de la aplicación, se verifica el correcto funcionamiento del algoritmo propuesto, así como también la correcta visualización de los datos en la plataforma web. Durante esta fase se explicó también el uso al personal docente de la Unidad Educativa “José Elías Altamirano”.

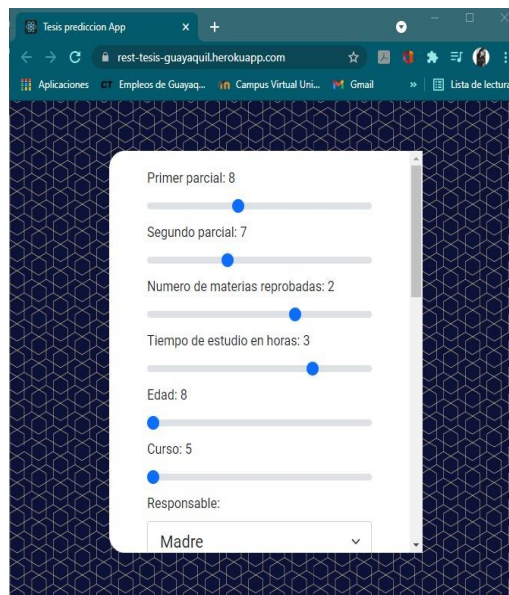
Data Science para la Predicción del Rendimiento Académico

Gráfico 7 Visualización de datos obtenidos en la predicción



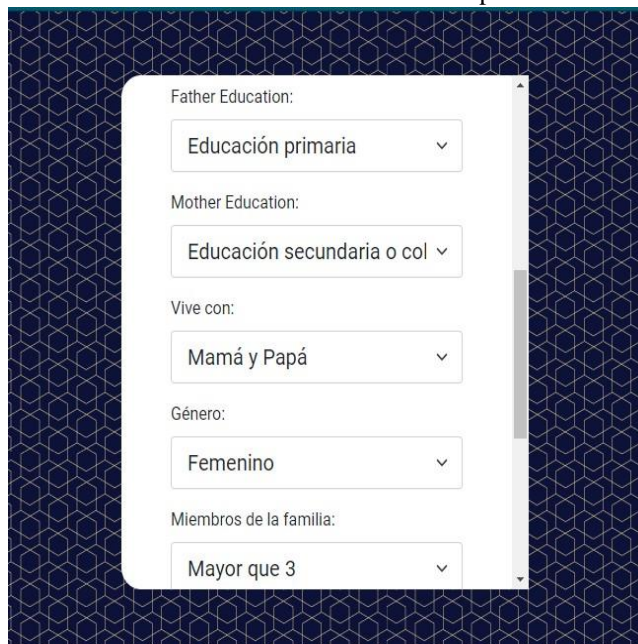
Fuente: Google Colab

Gráfico 9 Visualización de datos en la plataforma web



Fuente: Framework Flask

Gráfico 10 Visualización de datos en la plataforma web



Father Education:  
Educación primaria

Mother Education:  
Educación secundaria o col

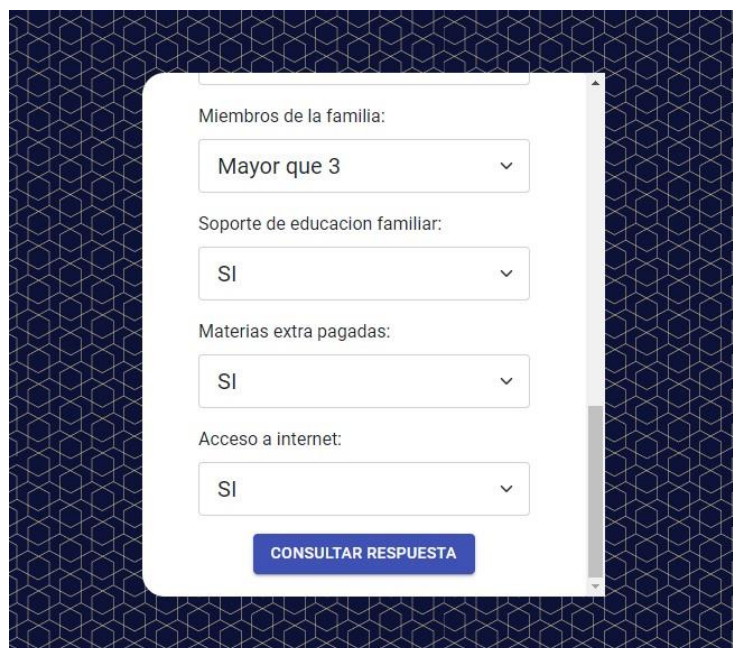
Vive con:  
Mamá y Papá

Género:  
Femenino

Miembros de la familia:  
Mayor que 3

Fuente: Framework Flask

Gráfico 11 Visualización de datos en la plataforma web



Miembros de la familia:  
Mayor que 3

Soporte de educacion familiar:  
SI

Materias extra pagadas:  
SI

Acceso a internet:  
SI

CONSULTAR RESPUESTA

Elaborado por : Mirna Moreno Herrera, Melanie Reyes Maldonado (2021)

Fuente: Framework Flask

Tabla 1 Criterios de validación

<b>CRITERIOS DE VALIDACIÓN DE LA PROPUESTA</b>					
<b>PROYECTO: “DISEÑO DE UNA APLICACIÓN DE DATA SCIENCE PARA LA PREDICCIÓN DEL RENDIMIENTO ACADÉMICO DE LOS ESTUDIANTES DE LA UNIDAD EDUCATIVA “JOSE ELÍAS ALTAMIRANO”, AÑO 2021”</b>					
<b>DESCRIPCION</b>	<b>TA</b>	<b>DE</b>	<b>I</b>	<b>ED</b>	<b>TD</b>
La aplicación es amigable con el usuario y fácil de utilizar.	X				
La aplicación es confiable en la predicción del rendimiento académico	X				
El software presentado cumple con las funciones previamente establecidas.	X				
El diseño propuesto cumple con los objetivos propuestos.	X				

**Fuente:** Trabajo de Investigación

### **Criterios de aceptación del producto o servicio**

Luego de haber finalizado la etapa de ejecución del proyecto planteado, se realizaron pruebas de funcionamiento con el fin de verificar el correcto funcionamiento de la aplicación y así mismo el cumplimiento de los objetivos propuestos en el presente trabajo investigativo. Mediante la realización de dichas pruebas, se verificó que el producto cumple con los requerimientos establecidos y que la aplicación web permite una correcta visualización de las predicciones realizadas mediante el algoritmo de machine learning, así mismo se demostró que la aplicación web cuenta con una interfaz amigable e intuitiva que será de gran utilidad para el personal docente de la Unidad Educativa “José Elías Altamirano”.

A través de las encuestas realizadas, se evidenció la viabilidad de uso de la aplicación de data science ya que el 58% de los docentes consideran que mediante el uso de un análisis predictivo del rendimiento de sus estudiantes se puede fortalecer la metodología de enseñanza.

A continuación, se muestra los criterios cumplidos mediante la realización del presente proyecto:

Tabla 2 Criterios de aceptación del producto

<b>Requerimiento</b>	<b>Porcentaje de Cumplimiento</b>
Ingreso a la aplicación web mediante un navegador.	100%
Predicción de notas académicas de los estudiantes.	100%
Visualización de la predicción realizada mediante ML	100%
Interfaz web intuitiva y amigable.	100%

**Fuente:** Datos obtenidos de encuesta

## Conclusiones

Una de las ventajas del uso de data science es que al ser una disciplina que incluye técnicas para el tratamiento y modelamiento de los datos, permite utilizar grandes cantidades de información para realizar análisis o predicciones, como se demostró en el presente trabajo, lo cual resulta de gran utilidad para la creación de modelos predictivos que puedan ser utilizados con fines educativos.

Existen diversos modelos de machine learning utilizados en data science, por lo que es importante tomar en cuenta los criterios necesarios como la calidad de datos que se analizarán, al momento de elegir el modelo de ML que pueda brindar un resultado correcto, de forma clara y comprensible al usuario.

En función de la pregunta científica a contestarse mediante el presente proyecto: “¿Es posible que mediante el uso de data science se logre predecir el rendimiento académico de los estudiantes de la Unidad Educativa “Jose Elías Altamirano”, para así identificar y comprender los problemas académicos de forma temprana y tomar medidas correctivas que beneficien a la comunidad educativa?” se logró validarla mediante el modelo de data science propuesto, ya que se evidencia la correcta predicción de las notas de los estudiantes y con esta información se pueden detectar las falencias en los métodos de enseñanza.

## Referencias

1. Alpaydin, E. (2020). Introduction to Machine Learning (4th ed.). MIT Press.  
[https://books.google.com.ec/books?hl=es&lr=&id=tZnSDwAAQBAJ&oi=fnd&pg=PR7&ots=F3TUc-6oxc&sig=JzmrZtSgJTGO5A0ogcfX4x-Co2A&redir\\_esc=y#v=onepage&q&f=false](https://books.google.com.ec/books?hl=es&lr=&id=tZnSDwAAQBAJ&oi=fnd&pg=PR7&ots=F3TUc-6oxc&sig=JzmrZtSgJTGO5A0ogcfX4x-Co2A&redir_esc=y#v=onepage&q&f=false)
2. Código Orgánico Integral Penal. (2014). CODIGO ORGANICO INTEGRAL PENAL.
3. Constitución de la República del Ecuador. (2008). Constitución del Ecuador.
4. García, L. A. (2016). Gestión de proyectos según el PMI. Ingeniería Técnica de Informática de Gestión.
5. Gutiérrez, J. J. (2018). ¿Qué es un framework web?
6. Heredia-Mayorga, H. (2020). Big data en la educación. Horizontes. Revista de Investigación En Ciencias de La Educación, 4(16), 523–535.  
<https://doi.org/10.33996/REVISTAHORIZONTES.V4I16.134>

7. Holguín, C., Díaz-Ricardo, Y., & Antonio Becerra-García, R. (2014). Ciencias Holguín, Revista trimestral, Año XX, abril-junio 2014 El lenguaje de programación Python/The programming language Python Ivett Challenger-Pérez. <http://www.linuxjournal.com/article/2959>
8. Mora, J. P. (2016). Que es la Ciencia de Datos, el aprendizaje automático (ML), el Big Data y cuales son sus usos?
9. Mueller, J. P., & Massaron, L. (2016). Machine Learning for dummies (I. John Wiley & Sons (ed.)).
10. Naqa, I. El, & Murphy, M. J. (2015). What Is Machine Learning? Machine Learning in Radiation Oncology, 3–11. [https://doi.org/10.1007/978-3-319-18305-3\\_1](https://doi.org/10.1007/978-3-319-18305-3_1)
11. Oracle Enterprise. (2015, April). Improving Higher Education Performance with Big Data. <https://silo.tips/download/improving-higher-education-performance-with-big-data>
12. Project Management Institute. (2013). Fundamentos para la Dirección de Proyectos (Guía del PMBOK) (Quinta Edi).
13. Reglamento para la Adquisición de Software por parte de las entidades contratantes del Sector Público. (2017). Reglamento para la Adquisición de Software por parte de las entidades contratantes del Sector Público.
14. Sosa, M., & Herrera, S. I. (2015). La informática como disciplina científica. Ensayo de mapeo disciplinar. <https://www.researchgate.net/publication/264868897>
15. Universitat Politècnica de Catalunya. (2021, March 21). ¿Qué es un Data Scientist? | inLab FIB. <https://inlab.fib.upc.edu/es/blog/que-es-un-data-scientist>
16. Utilización de Software Libre en la Administración Pública. (2011). Utilización de Software Libre en la Administración Pública.

©2022 por los autores. Este artículo es de acceso abierto y distribuido según los términos y condiciones de la licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0)

[\(https://creativecommons.org/licenses/by-nc-sa/4.0/\)](https://creativecommons.org/licenses/by-nc-sa/4.0/).