



Formato de inscripción proyecto de investigación

NIT 900014966-5

CONSENTIMIENTO EXPRESO Con el envío y de acuerdo con la Ley Estatutaria 1581 de 2012 de Protección de Datos y con el Decreto 1377 de 2013, los responsables del proyecto, autorizamos como Titulares de los datos que plasmamos, que éstos sean incorporados en una base de datos de responsabilidad de la Red Colombiana de semilleros de Investigación RedCOLSI, siendo tratados con la finalidad de gestión administrativa, formativa, evaluación y de información institucional, de semilleristas y de proyectos entre otros, de conformidad con el aviso de privacidad publicado en www.fundacionredcolsi.org. De igual modo, los autores declaran haber sido informados que pueden ejercitar los derechos de acceso, corrección, supresión, revocación o reclamo por infracción sobre datos, mediante escrito dirigido a la dirección de correo electrónico coordinacion@fundacionredcolsi.org, indicando en el asunto el derecho que desea ejercitar, o mediante correo ordinario remitido a carrera 12 Número 53 – 42 Edificio Barcelona la Castellana Montería.

INFORMACION GENERAL			
País	Colombia		
Nodo	Antioquia		
Universidad	Iudigital de Antioquia		
Nombre del Semillero	Semillero de investigación en desarrollo de software		
Nivel de Formación	Pregrado		
Programa Académico	Ingeniería de software y datos		
Título del Proyecto	Predicción y análisis de asistencia de estudiantes en un proyecto educativo mediante técnicas de Machine Learning de un grupo de instituciones educativas de la ciudad de Medellín.		
Autor (es)	Juan David Ramírez García		
Identificación	CC 1128450244		
Ponente (s)	Juan David Ramírez García		
E-mail de Contacto	david.ramirez@est.iudigital.edu.co		
Teléfonos de Contacto	3002912276		
Categoría (seleccionar una)	Propuesta de Investigación <input checked="" type="checkbox"/>	Investigación en Curso <input type="checkbox"/>	Investigación Terminada <input type="checkbox"/>
Área de la investigación (Marque solo una opción)	Ciencias Agrarias <input type="checkbox"/> Biológicas y del mar <input type="checkbox"/> Ciencias de la Salud <input type="checkbox"/> Ciencias exactas y de la tierra <input type="checkbox"/> Ciencias humanas <input type="checkbox"/>	Ciencias sociales <input type="checkbox"/> Navales y de seguridad <input type="checkbox"/> Ingeniería <input checked="" type="checkbox"/> Lingüística artes y letras <input type="checkbox"/> Otra: (Mencione cuál)	

Formato de inscripción proyecto de investigación

CONTENIDO DE LA PROPUESTA DE INVESTIGACIÓN

1. TÍTULO: Predicción y análisis de asistencia de estudiantes en un proyecto educativo mediante técnicas de Machine Learning de un grupo de instituciones educativas de la ciudad de Medellín.

2. INTRODUCCIÓN: Se llevó a cabo un proyecto educativo en un grupo de instituciones educativas de la ciudad de Medellín, el cual consistía en realizar talleres experienciales en desarrollo de software, modelado y maquetado de materiales, y nuevas tecnologías; en 4 sesiones en los grados de escolaridad de sexto a undécimo. El tipo de taller se asignó a cada grado según su nivel de complejidad y las competencias esperadas, asegurando que las actividades fueran adecuadas al desarrollo académico y permitieran a los estudiantes aprovechar mejor la experiencia formativa, de la siguiente manera: Los estudiantes de 6 a 8 vivieron el taller de modelado de materiales, los estudiantes de 9 a 10 nuevas tecnologías y los grados 11 en desarrollo de software.

Mencionando algunos de los objetivos del proyecto educativo, que eran llegar a la mayor cantidad de estudiantes posible y aprovechar mejor los recursos invertidos. Se llegó a la conclusión de que predecir y analizar la asistencia era un punto clave para el éxito del programa.

La asistencia evidencia el interés de los estudiantes por esta clase de vivencias. Además, permite analizar si existe algún tipo de patrón de asistencia que permita anticipar la ausencia o el comportamiento de los estudiantes, en ambos casos, con el objetivo de mejorar la experiencia.

La información sobre la asistencia de los estudiantes a dichos talleres, con la que se dispone, es de tres instituciones educativas de la ciudad de Medellín, las cuales incluyen variables como **fecha de nacimiento, género, edad, grado, tipo de taller, fecha de la sesión y si asistió o no a la sesión.**

En conclusión, la finalidad de este proyecto es proponer técnicas de análisis de datos y Machine Learning para identificar patrones de asistencia y que estas sean una herramienta para la planeación de propuestas similares, o que den respaldo a proyectos donde se pueda replicar.

3. PLANTEAMIENTO DEL PROBLEMA Y JUSTIFICACIÓN: La asistencia es un elemento clave dentro de los proyectos educativos porque a partir de ella se evidencia el impacto de estos en la comunidad educativa en términos de cantidad, calidad y pertinencia de los temas impartidos. De acuerdo con Barreno-Freine, Haro-Jácome y Florés-Yandún (2018), la asistencia en el aula es un elemento fundamental dentro del proceso de aprendizaje, junto con la participación activa de los estudiantes, porque tiene como resultado fortalecer el proceso de aprendizaje mediante el desarrollo de pedagogías significativas y el trabajo continuo. A partir de aquí lo oportuno de analizar y aplicar métodos de Machine Learning para identificar patrones de asistencia y deserción, fortaleciendo la planeación y el impacto de nuevas experiencias.

Dentro del proyecto educativo realizado no se encontraron causas puntuales de inasistencia, pero como se mencionó anteriormente, se vuelve importante analizar este tema. Sin embargo, no existen dentro del proyecto las herramientas que permitan analizar y predecir los datos de asistencia y actuar de forma preventiva. De no hacerlo, se podría afectar: la continuidad de los estudiantes, el aprovechamiento de recursos y la cobertura estudiantil.

Ante esta problemática surge la necesidad de implementar un análisis de datos e implementación de técnicas de machine learning para identificar áreas de riesgo que puedan afectar la continuidad estudiantil.

4. OBJETIVO GENERAL: Aplicar modelos de Machine learning basados en análisis de datos para predecir y encontrar patrones de asistencia de estudiantes en proyectos educativos realizados en algunos colegios de la ciudad de Medellín.

ESPECIFICOS:

- Analizar el comportamiento de asistencia de los estudiantes en los diferentes momentos del proyecto, considerando variables demográficas, sociales y educativas.
- Diseñar y documentar una guía metodológica para la construcción, implementación y evaluación de modelos de machine learning aplicados a proyectos educativos, con el fin de facilitar su replicabilidad en otras instituciones.

5. REFERENTE TEORICO:

Colombia

- En la corporación universitaria Remington en el año 2024 se realizó un proyecto donde se aplicaron técnicas de Machine Learning para la predicción de la deserción escolar a partir de variables académicas a estudiantes de varias cohortes de pregrado de la misma universidad. Si bien su foco es la deserción escolar, resulta valioso, para mi proyecto, la aplicabilidad de modelos de Machine Learning en un contexto educativo.
<https://repositorio.uniremington.edu.co/entities/publication/3a2e1b04-a3fe-4add-af0d-ea3370ab1a65>
- En la Universidad de Córdoba, en un proyecto realizado se implementó un modelo de machine learning con el objetivo de identificar acciones de recomendación para la inasistencia escolar de algunas instituciones educativas participantes.
<https://repositorio.unicordoba.edu.co/entities/publication/a0d9e45e-29d1-4119-bfb4-ff5c6c16e48a/full>

Extranjero

- En Estados Unidos se desarrolló un modelo de Deep Learning para predecir la asistencia de estudiantes con trastorno del espectro autista y brindar recomendaciones preventivas para el ausentismo.
<https://pubmed.ncbi.nlm.nih.gov/35082310/>
- En Albania se realizó un proyecto para predecir la asistencia de estudiantes usando variables demográficas, sociales y académicas usando modelos de Machine Learning.
<https://edumag.mrsu.ru/index.php/en/articles-en/112-22-2/974-10-15507-1991-9468-107-026-202202-2>

6. METODOLOGIA:

Para este proyecto, la metodología utilizada para identificar los **patrones más relevantes** en la asistencia estudiantil se desarrollará a través de los siguientes pasos:

1. **Definición del tipo de investigación:** El tipo de investigación que se utilizará es la cuantitativa, la cual pretende recolectar y analizar datos numéricos y categóricos para calcular promedios, identificar tendencias y relaciones, y realizar predicciones en los datos estudiados. El proyecto se implementará bajo un diseño no experimental, observacional y longitudinal, ya que vamos a analizar datos de variables educativas y de asistencia, y no a manipularlas o transformarlas.
2. **Identificación de la población y muestra:** La población será el total de estudiantes de sexto a once de las tres instituciones educativas participantes en el programa. Y la muestra será el total de estudiantes que participaron en al menos una de las sesiones de los talleres del programa.
3. **Recolección y limpieza de los datos:** Los datos recolectados son los datos de asistencia durante los cuatro encuentros del programa, los cuales incluyen variables como fecha de nacimiento, fecha del encuentro, género, tipo de taller, fecha de nacimiento y asistencia. Estos datos tendrán un proceso de limpieza como depuración de datos inconsistentes, datos nulos y repetidos. Asimismo, como el cálculo de nuevas variables como **Edad**, **Es_grado_9_o_superior**, **Es_mayor_o_igual_a_15**, **Talleres**, **asistencia_previa**, **numero_sesion** y **promedio_asistencia_previa**.
4. **Análisis exploratorio de datos:** En esta parte se realizará un análisis exploratorio de datos para identificar relacionamiento entre los diferentes tipos de datos, identificar valores atípicos, descubrir patrones y tendencias, y generar más preguntas e hipótesis.
5. **Aplicación de modelo de machine learning:** Una vez se tengan los datos organizados y limpios e identificadas las variables más relevantes, se aplicarán varias técnicas de machine learning como regresión logística, árboles de decisión y Random forest para predecir las futuras asistencias a proyectos similares.
6. **Evaluación del modelo:** Se aplicarán técnicas de evaluación del modelo para verificar qué tan efectivos y precisos son los modelos de machine learning implementados. Asimismo, como técnicas de desempeño como precisión, recall o F1-score.
7. **Diseño de guía para replicar en otros proyectos:** Se elaborará un paso a paso del procedimiento seguido a lo largo del proyecto que documente y detalle todos los análisis y técnicas implementadas para mostrar de manera clara los resultados obtenidos en el proyecto de investigación. Con el fin de que pueda ser replicado en proyectos educativos similares.

7. RESULTADOS:

- Los modelos de machine learning aplicados mostraron un desempeño sólido al predecir la asistencia estudiantil, con niveles de exactitud entre 77% y 79%. Los algoritmos árbol de decisión y random forest alcanzaron una precisión cercana al 91% para la clase de asistencia, mientras que la regresión logística logró un 89%, reflejando consistencia en la

identificación de patrones de participación regular.

- En contraste, los tres modelos tuvieron un rendimiento más bajo al predecir las inasistencias, con precisiones entre 43% y 47%. Esto sugiere que existen factores personales o contextuales no considerados en el conjunto de datos que influyen en la no asistencia.
- El análisis descriptivo mostró un 84% de participación general en los encuentros, con mayor asistencia en los grados superiores y entre los estudiantes mayores de 15 años. Estas variables, junto con el comportamiento previo de asistencia, fueron las más relevantes para explicar la participación.
- Las correlaciones revelaron que los estudiantes con buena asistencia en sesiones anteriores tienden a mantener su compromiso, consolidando un patrón de continuidad. Además, se observó una leve disminución en la asistencia en las últimas sesiones, lo que sugiere una pérdida progresiva de motivación o un aumento de la carga académica.

8. CONCLUSIONES:

- Los modelos desarrollados demostraron ser útiles para comprender y anticipar los patrones de asistencia, permitiendo identificar variables estructurales clave como la edad, el nivel educativo y la asistencia previa.
- Aunque el rendimiento fue alto en la predicción de asistencia, la detección de inasistencias requiere mayor profundidad analítica, incorporando nuevas variables que reflejen el entorno y las motivaciones del estudiante.
- La alta participación en general confirma el compromiso de los estudiantes con el programa, especialmente en los niveles educativos superiores, donde la madurez y el sentido de responsabilidad se reflejan en una mayor constancia.
- En conjunto, los modelos pueden emplearse en proyectos futuros como herramientas predictivas y de seguimiento, orientadas a diseñar estrategias que fortalezcan la permanencia y la asistencia continua a los programas educativos.

9. BIBLIOGRAFIA:

- Barreno-Freire, S. N., Haro-Jácome, O. F., & Flores-Yandún, P. F. (2019). Relación entre rendimiento académico y asistencia como factores de promoción estudiantil. *Revista Cátedra*, 2(1), 44-59. <https://dialnet.unirioja.es/servlet/articulo?codigo=9101145>
- Dúo Terrón, P., Moreno Guerrero, A. J., López Belmonte, J., & Marín Marín, J. A. (2023). Inteligencia Artificial y Machine Learning como recurso educativo desde la perspectiva de docentes en distintas etapas educativas no universitarias. *RiiTE Revista interuniversitaria de investigación en Tecnología Educativa*, 15, 58–78. <https://doi.org/10.6018/riite.579611>
- Moodley, R., Chiclana, F., Carter, J., & Caraffini, F. (2020). Using Data Mining in Educational Administration: A Case Study on Improving School Attendance. *Applied Sciences*, 10(9), 3116. [doi:10.3390/app10093116]
- Jarbou, M., Won, D., Gillis-Mattson, J., & Romanczyk, R. (2022). Deep learning-based school attendance prediction for autistic students. *Scientific Reports*, 12, Article 1431. <https://doi.org/10.1038/s41598-022-05258-z>
- Universidad de Córdoba. (2024). Modelo de recomendación de acciones para la prevención de la inasistencia escolar basado en un sistema predictivo de aprendizaje automático. Universidad de Córdoba. <https://repositorio.unicordoba.edu.co/entities/publication/a0d9e45e-29d1-4119-bfb4-ff5c6c16e48a/full>

- Guiral, H., Ospina López, J. E., Reina Morales, H. D., Y López Bedoya, G. (2024). Investigación sobre la deserción escolar en Colombia mediante el uso de técnicas profesionales de machine learning [Trabajo de grado, Corporación Universitaria Remington]. Repositorio Institucional Remington.
<https://repositorio.uniremington.edu.co/entities/publication/3a2e1b04-a3fe-4add-af0d-ea3370ab1a65>