REGISTRO DE TRABAJO DE GRADO

|  |  |  |  |
| --- | --- | --- | --- |
| **FECHA** |  |  | 2023 |

|  |
| --- |
| **DATOS DEL ESTUDIANTE (S)** |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| NOMBRES: Edwin Jair | | | | | | | APELLIDOS: Castiblanco Salas |
| TIPO IDENTIFICACIÓN: | T.I. |  | C.C. | x | C.E. |  | NÚMERO: 1014198740 |
| CORREO INSTITUCIONAL: [ecastiblancos@ucentral.edu.co](mailto:ecastiblancos@ucentral.edu.co) | | | | | | | TELÉFONO: 3102958284 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| NOMBRES: | | | | | | | APELLIDOS: |
| TIPO IDENTIFICACIÓN: | T.I. |  | C.C. |  | C.E. |  | NÚMERO: |
| CORREO INSTITUCIONAL: | | | | | | | TELÉFONO: |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| NOMBRES: | | | | | | | APELLIDOS: |
| TIPO IDENTIFICACIÓN: | T.I. |  | C.C. |  | C.E. |  | NÚMERO: |
| CORREO INSTITUCIONAL: | | | | | | | TELÉFONO: |

|  |
| --- |
| **MODALIDAD DE TRABAJO DE GRADO** (Seleccione una opción) |

|  |  |
| --- | --- |
| **I. Modalidad de investigación:** | **II. Modalidad de profundización:** □ |
| a. Proyecto final con resultados de nuevo conocimiento □ | a. Trabajo monográfico □ |
|  | b. Pasantía nacional o internacional □ |

|  |
| --- |
| **Línea de profundización** |
| **Modelos matemáticos** |

|  |
| --- |
| **AVAL DEL DOCENTE DIRECTOR** |

|  |  |
| --- | --- |
| NOMBRES: | DEPARTAMENTO: |
| CORREO INSTITUCIONAL: | TELÉFONO-EXT. : + |

|  |
| --- |
| **COMPONENTES** |

|  |
| --- |
| **1. TÍTULO DEL TRABAJO DE GRADO O PASANTÍA**  Aplicación de Modelos Mixtos y Algoritmos de Machine Learning para el Análisis Predictivo del Desempeño en las Pruebas Saber 11 y Sabre Pro en Colombia (2017–2022) |

|  |
| --- |
| **2. INTRODUCCIÓN Y JUSTIFICACIÓN**  En el contexto educativo colombiano, las pruebas Saber 11 y Saber Pro son evaluaciones fundamentales para medir el rendimiento académico de los estudiantes en distintas etapas de su formación. Estas pruebas sirven como herramienta de diagnóstico, medio de certificación e insumo clave para la toma de decisiones en políticas educativas. Saber 11 evalúa las competencias al finalizar la educación media, mientras que Saber Pro mide las habilidades adquiridas en la educación superior, sirviendo como puente entre la formación académica y el mercado laboral. Sin embargo, la interpretación de los resultados de estas evaluaciones sigue siendo un desafío complejo.  El principal problema radica en la falta de análisis exhaustivos que consideren las variables contextuales que influyen en el desempeño de los estudiantes. Factores como la edad, el contexto de residencia (urbano o rural) y el puntaje obtenido en Saber 11 pueden afectar significativamente los resultados de Saber Pro. Esto es especialmente relevante en Colombia, un país con marcada diversidad geográfica, cultural y socioeconómica. Según García (2018) y Vargas (2018), el contexto socioeconómico y geográfico tiene un peso importante en el rendimiento académico. Sin un análisis riguroso de estas variables, las interpretaciones de los resultados pueden ser parciales, llevando a decisiones de política educativa mal informadas y potencialmente ineficaces.  En este contexto, es crucial aplicar técnicas avanzadas de análisis de datos, como los modelos mixtos, que permiten una comprensión más profunda de las relaciones entre variables contextuales y rendimiento académico. Los modelos mixtos combinan efectos fijos y aleatorios, lo que permite analizar tanto tendencias generales como variaciones individuales y grupales. Esta metodología es particularmente relevante en sistemas educativos como el colombiano, donde existen diferencias significativas entre regiones, instituciones y poblaciones. Por ejemplo, un análisis con modelos mixtos puede identificar cómo el contexto de residencia influye en el rendimiento académico, diferenciando entre estudiantes urbanos y rurales. También permite analizar cómo el puntaje en Saber 11, como medida de desempeño previo, se relaciona con los resultados en Saber Pro. Esta aproximación puede desentrañar complejas interacciones entre variables, proporcionando una visión más detallada del desempeño educativo.  Según López (2021) y Rodríguez (2019), el uso de modelos mixtos en estudios educativos no solo ayuda a entender mejor las desigualdades en el sistema educativo, sino que también genera evidencia para orientar políticas más equitativas. Integrar estos modelos permite a los formuladores de políticas diseñar estrategias específicas para atender las necesidades de diferentes grupos de estudiantes, promoviendo así una educación más inclusiva y efectiva.  Colombia enfrenta desafíos únicos en su sistema educativo debido a la marcada desigualdad socioeconómica y las diferencias regionales. Las zonas rurales, que a menudo carecen de recursos adecuados, presentan un rendimiento académico generalmente más bajo en comparación con las zonas urbanas. Estas disparidades afectan no solo los resultados inmediatos en las pruebas Saber, sino también las oportunidades laborales y la movilidad social de los estudiantes a largo plazo. En este escenario, resulta esencial analizar cómo variables como la edad, asociada con la madurez y la experiencia académica, y el contexto de residencia influyen en los resultados de Saber Pro. Asimismo, el desempeño previo, medido a través del puntaje en Saber 11, es un indicador clave que puede ofrecer información valiosa sobre la preparación académica de los estudiantes.  El análisis de estos factores tiene implicaciones no solo académicas, sino también sociales y económicas. Una comprensión más profunda de las variables que afectan el rendimiento académico puede ayudar a diseñar políticas que reduzcan las brechas de desigualdad, permitiendo que más estudiantes accedan a una educación de calidad y, en última instancia, a mejores oportunidades de vida.  Este estudio tiene una relevancia especial para diversos actores dentro del sistema educativo colombiano. Las instituciones educativas, tanto de nivel medio como superior, pueden usar los hallazgos para ajustar sus estrategias pedagógicas y de apoyo a los estudiantes. Los formuladores de políticas educativas, por su parte, pueden emplear la evidencia generada para diseñar intervenciones más precisas y eficaces que aborden las desigualdades en el sistema. Además, los resultados de este análisis son útiles para padres y estudiantes, al ofrecer una visión más clara de los factores que contribuyen al éxito académico. Por último, el estudio también puede interesar a investigadores y académicos en el campo de la educación y la analítica de datos, proporcionando un marco metodológico robusto para futuros estudios en contextos similares.  En conclusión, la aplicación de modelos mixtos en el análisis de los resultados de las pruebas Saber 11 y Saber Pro tiene el potencial de transformar nuestra comprensión del sistema educativo colombiano. Al identificar y analizar las variables contextuales que influyen en el rendimiento académico, este estudio puede contribuir significativamente a la formulación de políticas educativas más equitativas y efectivas, ayudando a cerrar las brechas de desigualdad y a mejorar la calidad de vida de los estudiantes en Colombia. |

|  |
| --- |
| **3. OBJETIVO GENERAL**  Evaluar el impacto de variables contextuales como la edad, el contexto de residencia y el puntaje obtenido en las pruebas Saber 11 sobre el rendimiento académico en las pruebas Saber Pro, mediante la aplicación de modelos mixtos en analítica de datos. |

|  |
| --- |
| **4. OBJETIVOS ESPECÍFICOS**   Aplicar modelos mixtos para analizar la influencia de variables contextuales como la edad y el contexto de residencia en el rendimiento académico de las pruebas Saber Pro.   Comparar los resultados del análisis de modelos mixtos entre los puntajes de Saber 11 y Saber Pro para identificar patrones y diferencias significativas.   Formular recomendaciones basadas en los hallazgos del análisis para informar la toma de decisiones en políticas educativas en Colombia. |

|  |
| --- |
| **5. ANTECEDENTES**  En el ámbito del análisis educativo, los modelos mixtos se han establecido como herramientas fundamentales para entender cómo las variables contextuales influyen en el rendimiento académico. En este contexto, el análisis de datos educativos en Colombia se ha enfocado en las pruebas Saber 11 y Saber Pro, las cuales son utilizadas para medir el rendimiento de estudiantes en distintos niveles educativos. A continuación, se describen los antecedentes más relevantes para el uso de modelos mixtos en este ámbito. Consideraciones preliminares Los estudios previos han demostrado la importancia de integrar métodos cualitativos y cuantitativos en el análisis social y educativo. Según Bryman (2016a), esta combinación permite una visión más completa y profunda de los fenómenos estudiados. Creswell y Creswell (2018a) también resaltan la necesidad de enfoques mixtos en el diseño de investigaciones, lo que proporciona una base sólida para el análisis educativo. Tabachnick y Fidell (2019a) destacan la relevancia de los modelos estadísticos multivariados, particularmente los modelos mixtos, para el análisis jerárquico de datos educativos. Metodologías CRISP-DM La metodología CRISP-DM (Cross-Industry Standard Process for Data Mining) es ampliamente utilizada en proyectos de minería de datos. Según Wirth y Hipp (2000), esta metodología proporciona un marco estructurado que guía desde la comprensión del problema hasta la implementación de los resultados. En el contexto educativo, permite un análisis integral y sistemático de las pruebas Saber, asegurando que todos los aspectos del proceso de análisis sean abordados rigurosamente. Modelado y Evaluación El modelado mediante modelos mixtos permite capturar tanto efectos fijos como aleatorios. Esto es particularmente útil para analizar datos jerárquicos, como los obtenidos de instituciones educativas. Field (2018) y Hair et al. (2019) subrayan la capacidad de estos modelos para manejar variabilidad a nivel individual e institucional. Además, la evaluación de los modelos asegura que las relaciones entre las variables sean representadas adecuadamente, garantizando la validez y confiabilidad de las conclusiones. Implementación Los resultados obtenidos del análisis mediante modelos mixtos pueden informar políticas educativas y estrategias de mejora. Según Yin (2018a), estos resultados permiten identificar factores clave que influyen en el rendimiento académico, proporcionando información valiosa para la toma de decisiones. Las implementaciones basadas en estos hallazgos pueden incluir la redistribución de recursos educativos, la mejora de prácticas pedagógicas y la adaptación de currículos para atender mejor las necesidades de los estudiantes. |

|  |
| --- |
| **6. FUENTE DE LOS DATOS** (Cómo se obtendrán los datos)  Los datos utilizados provienen de la plataforma del ICFES, la cual recopila información sobre las pruebas Saber 11 y Saber Pro en Colombia. Estas bases de datos incluyen resultados por competencias, puntajes globales, y datos sociodemográficos de los estudiantes y su entorno familiar. Se consideraron los datos comprendidos entre 2017 y 2021​. |

|  |
| --- |
| **7. APLICACIÓN Y/O APORTE ESPECÍFICO AL CAMPO**  El análisis realizado aplica modelos mixtos para entender la influencia de variables contextuales en el rendimiento académico. Este enfoque permite desentrañar cómo factores como la edad, el contexto de residencia y el desempeño previo impactan los resultados de las pruebas Saber Pro. Los hallazgos pueden orientar la formulación de políticas educativas más equitativas​ |

|  |
| --- |
| **8. METODOLOGÍA O ACTIVIDADES ESPECIFICAS**  La metodología empleada se basa en CRISP-DM, que estructura el proceso en fases como comprensión del negocio, comprensión de datos, preparación de datos, modelado, evaluación e implementación. Los datos fueron procesados y analizados utilizando Python, con técnicas de filtrado, limpieza, y unión de bases de datos mediante métodos estadísticos avanzados​. |

|  |
| --- |
| **9. RECURSOS**  El documento no detalla un presupuesto específico, pero los recursos incluyen el acceso a bases de datos del ICFES, herramientas de análisis de datos (principalmente Python y Google Colab), y bibliografía sobre modelos estadísticos avanzados y minería de datos​. |

10. CRONOGRAMA DE ACTIVIDADES

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Semanas de ejecución de cada actividad** | | | | | | | | | | | | | | | |
| **ACTIVIDADES A REALIZAR** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** | **16** |
| Actividad 1: Consolidación de bases iniciales |  | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Actividad 2: Construcción y revisión de bases iniciales como pivote |  |  | X |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Actividad 3: Solicitud de nuevas bases para comparación y segregación de variables. |  |  |  |  |  |  |  |  |  |  |  |  | x |  |  |  |
| Actividad 4 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Actividad 5 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Actividad 6 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Actividad 7 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Actividad 8 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

|  |
| --- |
| **11. PRESUPUESTO** (En caso de modalidad Investigación) **Y FUENTES DE FINANCIACIÓN** (En caso de modalidad Profundización)  **11.1. Presupuesto monetario:**   1. **Acceso y procesamiento de datos**: Las bases de datos del ICFES. 2. **Infraestructura tecnológica**: Uso de herramientas de análisis como Python y plataformas colaborativas como Google Colab. 3. **Formación y capacitación**: Lectura de materiales y guías sobre modelos mixtos y minería de datos.   **11.2. Presupuesto temporal:**  El desarrollo del proyecto incluye varias fases que podrían distribuirse temporalmente de la siguiente manera:   1. **Preparación y limpieza de datos**: 1 mes. 2. **Desarrollo y evaluación de modelos estadísticos**: 2-3 meses. 3. **Validación y ajuste de modelos**: 1 mes. 4. **Desarrollo de aplicación para predicción**: 2 meses. 5. **Documentación y presentación de resultados**: 1 mes.  Metodología o Actividades Específicas La metodología CRISP-DM guía todo el proceso, dividiéndolo en fases clave:   1. **Comprensión del negocio y de los datos**. 2. **Preparación de datos**: Limpieza y transformación de grandes bases de datos para el análisis. 3. **Modelado y evaluación**: Aplicación de modelos mixtos para analizar los datos de rendimiento académico. 4. **Implementación y seguimiento**: Creación de una herramienta de predicción para instituciones educativas basada en los resultados analíticos​ |

|  |
| --- |
| **12. RESULTADOS ESPERADOS**  Se espera identificar los factores contextuales que influyen significativamente en el rendimiento de los estudiantes, proporcionando información clave para desarrollar políticas y estrategias educativas personalizadas. Además, se busca promover la formación continua de docentes y ajustar currículos en base a evidencias empíricas. |

|  |
| --- |
| **13. BIBLIOGRAFÍA**   Bryman, A. (2016a). *Social research methods* (5th ed.). Oxford University Press.   Creswell, J. W., & Creswell, J. D. (2018a). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.   Field, A. (2018). *Discovering statistics using IBM SPSS statistics*. SAGE Publications.   Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis*. Pearson.   Tabachnick, B. G., & Fidell, L. S. (2019a). *Using multivariate statistics* (7th ed.). Pearson.   Wirth, R., & Hipp, J. (2000). *CRISP-DM: Towards a standard process model for data mining*. Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining.   Yin, R. K. (2018a). *Case study research and applications: Design and methods* (6th ed.). SAGE Publications. |

|  |  |  |
| --- | --- | --- |
| **14. DATOS DE LA ORGANIZACIÓN** (En caso de pasantía) | | |
| NOMBRE DE LA ORGANIZACIÓN: | | |
| RESPONSABLE DE LA ORGANIZACIÓN: | | |
| CORREO: | TELÉFONO: | DURACIÓN VINCULACIÓN (horas): |

15. FIRMAS

|  |  |
| --- | --- |
| **FIRMA DEL ESTUDIANTE:** | **FIRMA DEL DOCENTE DIRECTOR O TUTOR** |
|  |  |

16. DATOS DE TRÁMITE COMITÉ DE INVESTIGACIÓN DEL PROGRAMA (Espacio para diligenciar por el Comité del Programa)

|  |  |
| --- | --- |
| **No. CONSECUTIVO** |  |
| **No. ACTA** |  |
| **FECHA** |  |