UNIVERSIDAD NACIONAL DE CAAGUAZÚ
FACULTAD DE CIENCIAS Y TECNOLOGÍAS
CARRERA DE INGENIERÍA EN INFORMÁTICA

Desarrollo de un Dashboard Web para un Modelo de Machine Learning basado en IA en la Predicción Temprana de la Deserción Estudiantil de la Facultad de Ciencias y Tecnologías

Diego Emanuel Armoa Rolon

CORONEL OVIEDO, NOVIEMBRE DE 2025

UNIVERSIDAD NACIONAL DE CAAGUAZÚ

FACULTAD DE CIENCIAS Y TECNOLOGÍAS

CARRERA DE INGENIERÍA EN INFORMÁTICA

PROYECTO FINAL DE GRADO

Desarrollo de un Dashboard Web para un Modelo de Machine Learning basado en IA en la Predicción Temprana de la Deserción Estudiantil de la Facultad de Ciencias y Tecnologías

Diego Emanuel Armoa Rolon

TUTOR: Prof. M Sc. Juan Vicente Bogado Machuca

CORONEL OVIEDO, NOVIEMBRE DE 2025

Atribución-NoComercial 4.0 Internacional (CC BY-NC 4.0)

Usted es libre de:

- **Compartir — copiar y redistribuir el material en cualquier medio o formato**
- **Adaptar — remezclar, transformar y construir a partir del material**

Bajo los siguientes términos:

- **Atribución — Usted debe dar crédito de manera adecuada, brindar un enlace a la licencia, e indicar si se han realizado cambios. Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciante.**
- **NoComercial — Usted no puede hacer uso del material con propósitos comerciales.**

DERECHO DE AUTOR

Quien suscribe, Diego Emanuel Armoa Rolón, autor del trabajo de investigación titulado Desarrollo de un Dashboard Web para un Modelo de Machine Learning basado en IA en la Predicción Temprana de la Deserción Estudiantil de la Facultad de Ciencias y Tecnologías, declara que voluntariamente cede a título gratuito en forma pura y simple ilimitada e irrevocablemente a favor de la Facultad de Ciencias y Tecnologías – UNCA, el derecho de autor de contenido patrimonial, que le corresponde sobre el trabajo de referencia. Conforme a lo anteriormente expresado, esta sesión le otorga a la FCyT la Facultad de comunicar la obra divulgarla, publicarla y reproducirla en soportes analógicos o digitales en la oportunidad que así lo estime conveniente. La FCyT deberá indicar qué autoría o creación del trabajo corresponde a mi persona y hará referencia al autor y a las personas que hayan colaborado en la realización del presente trabajo de investigación.

En la ciudad de Coronel Oviedo a los …. , del mes de ….. del 2025

…….

Firma

PÁGINA DE APROBACIÓN

Trabajo de fin de grado para la obtención del Título de Ingeniero en informática, aprobado en representación de la Facultad Ciencias y Tecnología de la Universidad Nacional de Caaguazú, por el Tribunal Examinador constituido por los siguientes profesores y con la siguiente nota final:

CALIFICACIÓN FINAL: _____

ACTA N°:_____

FECHA :_____

---------- Prof.
Ing.

---------- Prof.
Ing.

---------- Prof.
Ing.

DEDICATORIA

Este proyecto se lo dedico a Dios, a mis padres quienes me han apoyado en este trayecto, con inquebrantable amor y sacrificio han sido mi mayor motivación para afrontar los desafíos y perseverar en este camino académico.

A mis compañeros y amigos, cada momento compartido ha dejado una huella en mi corazón, momentos de alegría, aventuras, desvelos.

A mi hijo Azael, quien me enseña cada día a ser mejor, todo lo que soy y logro es por y para vos.

AGRADECIMIENTOS

Mis más sinceros agradecimientos a la Facultad de Ciencias y Tecnologías por brindarme la oportunidad y los recursos para llevar a cabo este trabajo, a mi mentor el Ingeniero Juan Vicente Bogado y profesores cuya guía y conocimientos han sido una fuente inagotable de inspiración y aprendizaje.

A mi familia, mis amigos, mi hijo mi mayor motivo y razón para seguir adelante, a la mamá de mi hijo por su acompañamiento brindado a lo largo de este camino compartido. Su presencia y soporte han sido valiosas durante este proceso.

Este trabajo es el resultado del esfuerzo conjunto y apoyo sincero de todas estas personas, a quienes agradezco profundamente.

RESUMEN

La deserción estudiantil constituye uno de los principales desafíos que enfrentan las instituciones de educación superior en América Latina, afectando no solo la eficiencia institucional sino también las trayectorias profesionales de miles de estudiantes. Este trabajo presenta el desarrollo de un dashboard web interactivo que integra un modelo de machine learning previamente validado para la predicción temprana de deserción estudiantil en la Facultad de Ciencias y Tecnologías (FCyT) de la Universidad Nacional de Caaguazú. El sistema propuesto permite a gestores académicos y docentes identificar estudiantes en riesgo mediante visualizaciones intuitivas y alertas automatizadas, facilitando intervenciones oportunas y personalizadas. La metodología empleada incluyó la integración del modelo predictivo desarrollado por González y Bogado, el diseño de una arquitectura web escalable utilizando tecnologías modernas como Python, Django, React.js y PostgreSQL, y la evaluación de viabilidad técnica, operativa y económica del sistema. Los resultados demuestran que la implementación de este tipo de herramientas tecnológicas puede contribuir significativamente a la reducción de las tasas de deserción, optimizando el uso de recursos institucionales y mejorando los indicadores de retención estudiantil. Este proyecto se enmarca dentro de las líneas de investigación de Ingeniería del Software y Ciencia de Datos de la FCyT, y contribuye directamente al cumplimiento del Objetivo de Desarrollo Sostenible 4 (Educación de Calidad) de las Naciones Unidas.

Palabras clave: deserción estudiantil, machine learning, dashboard web, predicción temprana, ingeniería del software, ciencia de datos, ODS 4, educación superior.

ABSTRACT

Student dropout is one of the main challenges facing higher education institutions in Latin America, affecting not only institutional efficiency but also the professional trajectories of thousands of students. This paper presents the development of an interactive web dashboard that integrates a previously validated machine learning model for the early prediction of student dropout at the Faculty of Sciences and Technologies (FCyT) of the National University of Caaguazú. The proposed system allows academic administrators and faculty to identify at-risk students through intuitive visualizations and automated alerts, facilitating timely and personalized interventions. The methodology employed included the integration of the predictive model developed by González and Bogado, the design of a scalable web architecture using modern technologies such as Python, Django, React.js, and PostgreSQL, and the evaluation of the system's technical, operational, and economic feasibility. The results demonstrate that the implementation of this type of technological tool can significantly contribute to reducing dropout rates, optimizing the use of institutional resources, and improving student retention indicators. This project falls within the research lines of Software Engineering and Data Science at the Faculty of Science and Technology (FCyT), and directly contributes to achieving Sustainable Development Goal 4 (Quality Education) of the United Nations.

Keywords: student dropout, machine learning, web dashboard, early prediction, software engineering, data science, SDG 4, higher education.

ÍNDICE

Contenido

1. Introducción 1
1.1. Planteamiento del Problema 2
1.2. Antecedentes 4
1.3. Justificación 8
1.4. Objetivos 11
1.4.1. Objetivo General 11
1.4.2. Objetivos Específicos 11
1.5. Deserción Estudiantil: Conceptualización y Modelos Teóricos 11
1.6. Factores Asociados a la Deserción Estudiantil 13
1.7. Inteligencia Artificial y Machine Learning en Educación 14
1.8. Sistemas de Alerta Temprana en Educación Superior 16
1.9. Ingeniería de Software para Sistemas de Analítica Educativa 18
1.10. Ciencia de Datos Aplicada a Educación 19
1.11. Consideraciones Éticas en Analítica Predictiva Educativa 20
1.12. Alineación con Líneas de Investigación y Objetivos de Desarrollo Sostenible 21
2. Metodología 24
2.1. Población y Muestra 24
2.1.1. Dataset de Estudiantes 24
2.1.2. Estructura del Dataset 25
2.1.3. Preprocesamiento de Datos 25
2.2. Modelo de Machine Learning 26
2.2.1. Desempeño Reportado del Modelo 26
2.3. Objetivo Específico 1: Integración del Modelo de Machine Learning 27
2.4. Objetivo Específico 2: Desarrollo del Dashboard Web 27
2.4.1. Configuración del Entorno Frontend 28
2.4.2. Diseño de Interfaces de Usuario 28
2.4.3. Implementación de Visualizaciones 29
2.5. Objetivo Específico 3: Evaluación de Viabilidad 29
2.5.1. Evaluación de Viabilidad Técnica 29
2.5.2. Evaluación de Viabilidad Operativa 30

2.5.3. Evaluación de Viabilidad Económica 31
3. Resultados 33
3.1. Integración el modelo de machine learning de predicción de deserción temprana dentro de la interfaz web. 33
3.2. Desarrollo el dashboard para la interfaz web. 34
3.3. Evaluación de viabilidad técnica, operativa y económica 36
3.3.1. Viabilidad Técnica 36
3.3.2. Viabilidad operativa 42
3.3.3. Viabilidad económica 44
4. Conclusiones y recomendaciones 46
4.1. Conclusiones 46
4.2. Recomendaciones 47
5. Bibliografía 49
6. ANEXOS - Instrumentos de recolección de datos 1
6.1. ANEXO 1: Planilla de Requerimientos Técnicos Mínimos 1
6.2. ANEXO 2: Matriz de Perfiles de Usuario 6
6.3. ANEXO 4: Necesidades de Capacitación por Rol 10
6.4. ANEXO 5: Cálculo de retorno de la inversión 19

ÍNDICE DE FIGURAS

FIGURA 1. FLUJO DE ACTIVIDADES RESUMIDO QUE REPRESENTA LA METODOLOGÍA 20

ÍNDICE DE TABLAS

TABLA 1. DISTRIBUCIÓN DE ALUMNOS POR CARRERA 18

TABLA 2. ESTRUCTURA DE VARIABLES EN EL DATASET 18

TABLA 3. RESUMEN DE VIABILIDAD ECONÓMICA 33

1. Introducción

La educación superior desempeña un papel fundamental en el desarrollo económico y social de las naciones, constituyéndose como el principal mecanismo de movilidad social y formación de capital humano calificado. Sin embargo, las instituciones universitarias de América Latina enfrentan un desafío persistente y complejo: la deserción estudiantil. Este fenómeno, caracterizado por el abandono prematuro de los estudios universitarios, representa no solo una pérdida individual para el estudiante que ve truncadas sus aspiraciones profesionales, sino también un problema institucional que afecta la eficiencia del sistema educativo y un desperdicio de recursos públicos y privados invertidos en la formación académica.

La Facultad de Ciencias y Tecnologías de la Universidad Nacional de Caaguazú, registra índices preocupantes de abandono estudiantil particularmente durante los primeros años de formación académica. Esta situación genera consecuencias múltiples: estudiantes que no logran completar su formación profesional, familias que ven frustradas sus expectativas de mejora socioeconómica, y una institución que no alcanza sus metas de graduación y formación de profesionales competentes.

La complejidad del fenómeno de la deserción radica en su naturaleza multifactorial. Investigaciones previas [1] [2] han identificado que las causas del abandono estudiantil incluyen factores académicos como el bajo rendimiento y la inadecuada preparación preuniversitaria, factores socioeconómicos como las dificultades financieras y las responsabilidades laborales, factores personales como problemas de salud y falta de motivación, y factores institucionales como la ausencia de programas de apoyo académico y la limitada integración del estudiante a la vida universitaria. Esta multiplicidad de causas hace que la predicción y prevención de la deserción sea un desafío metodológico considerable, requiriendo enfoques integrales que consideren simultáneamente múltiples variables y sus interacciones.

En las últimas dos décadas, el avance de las tecnologías de información y comunicación, junto con el desarrollo de técnicas de inteligencia artificial y análisis de datos, ha abierto nuevas posibilidades para abordar problemas complejos en el ámbito educativo. Específicamente, los modelos de machine learning han demostrado capacidad para identificar patrones ocultos en grandes volúmenes de datos académicos, permitiendo predecir con precisión razonable qué estudiantes presentan mayor riesgo de deserción. Estos modelos pueden procesar información histórica de desempeño académico, datos socioeconómicos, registros de asistencia y múltiples

variables institucionales para generar indicadores predictivos que anticipen el comportamiento futuro de los estudiantes.

Sin embargo, la mera existencia de modelos predictivos precisos no garantiza su utilidad práctica en el contexto institucional. Un modelo de inteligencia artificial, por sofisticado que sea, resulta de escasa utilidad si permanece como un artefacto técnico inaccesible para quienes toman decisiones en el ámbito académico. Los gestores universitarios, coordinadores de carrera y docentes tutores requieren herramientas que traduzcan las predicciones del modelo en información accionable, presentada de manera comprensible y oportuna. Es aquí donde los sistemas de información y las interfaces de usuario juegan un papel crucial, actuando como puentes entre la capacidad analítica de los algoritmos y las necesidades prácticas de los usuarios finales.

1.1. Planteamiento del Problema

La Facultad de Ciencias y Tecnologías de la Universidad Nacional de Caaguazú enfrenta tasas significativas de deserción estudiantil que comprometen tanto la eficiencia institucional como el cumplimiento de su misión formadora. A pesar de los esfuerzos realizados, la falta de herramientas sistemáticas para la identificación temprana de estudiantes en riesgo limita la efectividad de intervenciones. Las acciones preventivas suelen implementarse de manera reactiva, cuando los signos de deserción ya son evidentes, reduciendo significativamente las posibilidades de éxito de cualquier estrategia de retención.

La FCyT genera continuamente información sus de estudiantes: calificaciones, asistencias, inscripciones a materias, participación en actividades extracurriculares, entre otros. Sin embargo, estos datos permanecen dispersos en diferentes sistemas administrativos, sin un mecanismo integrado que permita su análisis conjunto y la generación de alertas tempranas. Los coordinadores de carrera y docentes carecen de herramientas que les permitan visualizar de manera integral la situación académica de sus estudiantes y priorizar sus esfuerzos de acompañamiento.

Recientemente, González y Bogado [3] desarrollaron un modelo de machine learning específicamente calibrado para predecir la deserción estudiantil en el contexto particular de la FCyT. Este modelo, que constituye un avance significativo en la comprensión del fenómeno de deserción en la institución, demostró capacidad predictiva superior a los métodos estadísticos tradicionales, alcanzando niveles de precisión que lo hacen viable para su implementación práctica. Sin embargo, este modelo permanece como un desarrollo técnico que no ha sido integrado en un

sistema operativo accesible para los usuarios institucionales. No existe actualmente una plataforma que permita a los gestores académicos consultar las predicciones del modelo o visualizar los indicadores de riesgo de sus estudiantes.

La problemática central que aborda este proyecto es, por tanto, la brecha existente entre la capacidad predictiva del modelo de inteligencia artificial desarrollado por González y Bogado y su utilización efectiva en la práctica institucional. Esta brecha tecnológica impide que la FCyT aproveche plenamente el potencial de las técnicas de ciencia de datos para mejorar sus indicadores de retención estudiantil. Se requiere el desarrollo de un sistema de información que integre el modelo predictivo en una plataforma web accesible, intuitiva y funcional, que permita a los diferentes actores institucionales tomar decisiones informadas basadas en evidencia cuantitativa.

Específicamente, el problema se manifiesta en varias dimensiones operativas. En primer lugar, los coordinadores de carrera no disponen de mecanismos para identificar proactivamente a los estudiantes que presentan mayor probabilidad de deserción, lo que les impide priorizar sus recursos limitados de tutoría y acompañamiento. En segundo lugar, los docentes carecen de información sistemática sobre los factores de riesgo específicos que afectan a cada estudiante bajo su supervisión, dificultando el diseño de intervenciones personalizadas. En tercer lugar, las autoridades institucionales no cuentan con indicadores agregados que les permitan evaluar la efectividad de las políticas implementadas o identificar patrones sistemáticos de deserción asociados a determinadas carreras, cohortes o perfiles estudiantiles.

Adicionalmente, la ausencia de un sistema integrado genera ineficiencias en el uso de los recursos institucionales. Los esfuerzos de apoyo académico se distribuyen de manera uniforme o basándose en criterios subjetivos, en lugar de focalizarse en los estudiantes que realmente presentan mayor riesgo de abandono. Esta distribución ineficiente de recursos resulta en que algunos estudiantes en riesgo crítico no reciben la atención necesaria, mientras que otros con bajo riesgo de deserción consumen recursos de apoyo que podrían ser mejor utilizados.

Desde una perspectiva más amplia, esta problemática se inscribe en el desafío global de mejorar la calidad y equidad de la educación superior. La deserción estudiantil afecta desproporcionadamente a estudiantes de contextos socioeconómicos vulnerables, perpetuando ciclos de desigualdad y limitando las oportunidades de movilidad social. Las instituciones educativas tienen la responsabilidad de implementar mecanismos que garanticen no solo el acceso

a la educación superior, sino también las condiciones necesarias para la permanencia y graduación exitosa de todos sus estudiantes, independientemente de su origen socioeconómico.

1.2. Antecedentes

Los trabajos seminales de Tinto [1] en la década de 1970 establecieron las bases conceptuales para comprender el abandono universitario como un proceso complejo de interacción entre el estudiante y el ambiente institucional. Tinto propuso un modelo teórico que explica la deserción como resultado de la falta de integración académica y social del estudiante a la comunidad universitaria, conceptualización que ha influenciado décadas de investigación posterior. Su teoría sugiere que la decisión de abandonar los estudios no es un evento abrupto sino un proceso gradual en el que el estudiante evalúa continuamente los costos y beneficios de permanecer en la institución, considerando tanto factores académicos como sociales.

Posteriormente, Cabrera, Nora y Castañeda [4] expandieron el modelo de Tinto incorporando variables socioeconómicas y culturales, reconociendo que factores externos a la institución, como las responsabilidades familiares y las limitaciones financieras, juegan un papel crucial en las decisiones de permanencia. Estos autores desarrollaron modelos de ecuaciones estructurales que permitieron cuantificar las relaciones entre múltiples variables predictoras y la probabilidad de deserción, demostrando que el fenómeno requiere análisis multivariados que consideren simultáneamente factores individuales, institucionales y contextuales.

En el contexto latinoamericano, diversos estudios han documentado las particularidades de la deserción en sistemas educativos caracterizados por alta heterogeneidad socioeconómica y limitaciones de recursos institucionales. Castaño, Gallón y Vásquez [5] realizaron un análisis exhaustivo de los factores asociados a la deserción en Colombia, identificando que las dificultades económicas, el bajo rendimiento académico en los primeros semestres y la inadecuada orientación vocacional constituyen los principales predictores de abandono. Sus hallazgos resaltan la importancia de implementar sistemas de alerta temprana que permitan identificar estudiantes en riesgo durante las primeras etapas de su trayectoria universitaria, cuando las intervenciones preventivas tienen mayor probabilidad de éxito.

El desarrollo de técnicas de inteligencia artificial y minería de datos ha revolucionado el campo de la analítica educativa en las últimas dos décadas. Romero y Ventura [6] realizaron revisiones sistemáticas de la literatura que documentan la creciente aplicación de algoritmos de machine

learning para predecir diversos resultados educativos, incluyendo el rendimiento académico, la deserción y la probabilidad de graduación. Estos autores identificaron que los métodos de aprendizaje supervisado, particularmente los árboles de decisión, las máquinas de soporte vectorial y los métodos de ensamble como Random Forest y Gradient Boosting, han demostrado capacidad superior a los modelos estadísticos tradicionales para capturar relaciones no lineales y interacciones complejas entre variables predictoras.

Aulck y colaboradores [7] desarrollaron un sistema predictivo de deserción para la Universidad de Washington, utilizando técnicas de machine learning sobre datos de más de 40.000 estudiantes. Su modelo incorporó variables académicas tradicionales como calificaciones y créditos aprobados, pero también variables no tradicionales como patrones de inscripción a materias, uso de recursos bibliotecarios y participación en actividades extracurriculares. Los resultados demostraron que la inclusión de estas variables no tradicionales mejoraba significativamente la capacidad predictiva del modelo, sugiriendo que los sistemas de alerta temprana deberían integrar múltiples fuentes de información institucional.

Sweeney, Rangwala y Lester [8] propusieron un enfoque de sistemas de recomendación para la predicción de desempeño estudiantil, argumentando que las técnicas utilizadas en sistemas de recomendación comerciales pueden adaptarse efectivamente al contexto educativo. Su trabajo demostró que los algoritmos de filtrado colaborativo, que identifican similitudes entre estudiantes basándose en sus patrones históricos de desempeño, pueden generar predicciones precisas sobre el rendimiento futuro y la probabilidad de deserción. Este enfoque tiene la ventaja adicional de permitir la identificación de estudiantes similares que han tenido trayectorias exitosas, facilitando el diseño de intervenciones basadas en casos de éxito previos.

Un aspecto crucial en la implementación práctica de modelos predictivos es su integración en sistemas de información accesibles para los usuarios finales. Jayaprakash y colaboradores [9] documentaron la implementación de un sistema de alerta temprana de código abierto en la Universidad Purdue, describiendo tanto los aspectos técnicos del sistema como los desafíos organizacionales de su adopción institucional. Su experiencia reveló que el éxito de estos sistemas depende no solo de la precisión del modelo predictivo, sino también de la usabilidad de la interfaz, la capacitación de los usuarios y el establecimiento de protocolos claros de intervención cuando se identifican estudiantes en riesgo.

Lakkaraju y colaboradores [10] desarrollaron un marco de machine learning para identificar estudiantes en riesgo de resultados académicos adversos, implementado en una gran universidad pública estadounidense. Su sistema no solo predecía la probabilidad de deserción sino que también identificaba los factores de riesgo específicos más relevantes para cada estudiante individual, permitiendo intervenciones personalizadas. Los autores reportaron que la implementación del sistema resultó en una reducción del 15% en las tasas de deserción de los estudiantes identificados como de alto riesgo que recibieron intervenciones dirigidas.

Arnold y Pistilli [11] documentaron la implementación del sistema *Course Signals* en la Universidad Purdue, uno de los primeros sistemas de alerta temprana basados en analítica de aprendizaje implementados a gran escala en educación superior. Este sistema utilizaba algoritmos predictivos para generar señales visuales (verde, amarillo, rojo) que indicaban a los estudiantes su nivel de riesgo en cada curso. La evaluación del sistema demostró mejoras significativas en las tasas de retención y graduación, particularmente entre estudiantes de grupos históricamente subrepresentados, evidenciando el potencial de estas herramientas para reducir brechas de equidad en educación superior.

En el contexto específico de la Facultad de Ciencias y Tecnologías de la Universidad Nacional de Caaguazú, el trabajo de González y Bogado [3] representa el antecedente directo y fundamental de este proyecto. Estos investigadores desarrollaron un modelo de machine learning específicamente calibrado para predecir la deserción estudiantil en la FCyT, utilizando datos históricos de cohortes estudiantiles entre 2015 y 2021. Su investigación incluyó un análisis exhaustivo de las variables predictoras disponibles en los registros institucionales, identificando que el rendimiento académico en los primeros dos semestres, el número de materias reprobadas y el índice de asistencia constituyen los predictores más relevantes de deserción en el contexto específico de la FCyT.

El modelo desarrollado por González y Bogado empleó técnicas de ensamble, específicamente Random Forest y Gradient Boosting, comparando su desempeño con métodos alternativos como regresión logística, árboles de decisión individuales y máquinas de soporte vectorial. Los resultados demostraron que los métodos de ensamble lograban métricas superiores de precisión, recall y área bajo la curva ROC, con una capacidad predictiva que superaba el 85% de precisión en la identificación de estudiantes en riesgo de deserción. El modelo fue validado mediante validación cruzada y con datos de cohortes no utilizadas en el entrenamiento, demostrando

capacidad de generalización adecuada. También realizaron un análisis temporal de la capacidad predictiva del modelo, evaluando en qué momento de la trayectoria estudiantil las predicciones alcanzan niveles de confianza suficientes para justificar intervenciones. Sus resultados indicaron que, si bien es posible generar predicciones desde el sexto semestre utilizando datos de rendimiento, la precisión del modelo mejora sustancialmente al incorporar información a partir del octavo semestre. Esto sugiere que los sistemas de alerta temprana deben implementar un enfoque de monitoreo continuo, actualizando las predicciones a medida que se dispone de nueva información sobre el desempeño estudiantil.

Sin embargo, como los propios autores reconocieron en sus conclusiones, el modelo desarrollado permanecía como un artefacto técnico que requería integración en un sistema operativo para su utilización práctica. No existía una plataforma que permitiera a los gestores académicos acceder a las predicciones del modelo, visualizar los indicadores de riesgo de manera comprensible, o recibir alertas automatizadas sobre estudiantes que requerían atención prioritaria. Esta limitación motivó la necesidad del presente proyecto, que busca precisamente desarrollar la infraestructura tecnológica necesaria para operacionalizar el modelo de González y Bogado en la práctica institucional de la FCyT.

Adicionalmente, diversos estudios han abordado los aspectos de ingeniería de software necesarios para implementar sistemas de analítica educativa en entornos institucionales reales. Siemens y Long [12] discutieron los desafíos de escalabilidad, seguridad y privacidad que enfrentan los sistemas de analítica de aprendizaje, enfatizando la necesidad de arquitecturas robustas que garanticen la protección de datos sensibles de los estudiantes mientras permiten el análisis necesario para generar predicciones útiles. Estos autores propusieron principios de diseño que incluyen la anonimización de datos cuando sea posible, el control de acceso basado en roles, y la transparencia en el uso de algoritmos predictivos.

Ferguson y colaboradores [13] analizaron los aspectos éticos de la implementación de sistemas de analítica predictiva en educación, señalando que estos sistemas pueden inadvertidamente perpetuar sesgos existentes si no son diseñados e implementados cuidadosamente. Argumentaron que los modelos predictivos deben ser regularmente auditados para detectar posibles sesgos discriminatorios, y que las instituciones deben establecer políticas claras sobre cómo se utilizan las predicciones en la toma de decisiones sobre estudiantes individuales. Estas consideraciones éticas son particularmente relevantes en el contexto de sistemas de alerta temprana, donde las

predicciones pueden influir en decisiones que afectan directamente las trayectorias educativas de los estudiantes.

1.3. Justificación

La deserción estudiantil en la educación superior constituye un problema de múltiples dimensiones que afecta simultáneamente a individuos, instituciones y a la sociedad en su conjunto. Desde la perspectiva individual, cada estudiante que abandona sus estudios universitarios ve truncadas sus aspiraciones de desarrollo profesional y movilidad socioeconómica. Las estadísticas demuestran que los profesionales con educación superior completa acceden a mejores oportunidades laborales, mayores ingresos a lo largo de su vida y mejores condiciones de bienestar general. La deserción, por tanto, representa una pérdida de oportunidades individuales que frecuentemente afecta de manera desproporcionada a estudiantes de contextos socioeconómicos vulnerables, perpetuando ciclos intergeneracionales de desigualdad.

Desde la perspectiva institucional, la deserción estudiantil compromete la eficiencia y efectividad de las universidades públicas. Los recursos invertidos en la formación de estudiantes que eventualmente abandonan sus estudios representan una utilización subóptima de presupuestos limitados. Además, las tasas de deserción afectan negativamente los indicadores de calidad institucional utilizados en procesos de acreditación y evaluación, comprometiendo la reputación y competitividad de la institución. Para la Facultad de Ciencias y Tecnologías de la Universidad Nacional de Caaguazú, mejorar los índices de retención y graduación es crucial para consolidar su posicionamiento como institución formadora de profesionales competentes en áreas estratégicas para el desarrollo nacional.

Desde una perspectiva social más amplia, la deserción universitaria representa un desperdicio de capital humano que las naciones en desarrollo no pueden permitirse. Paraguay enfrenta desafíos significativos de desarrollo tecnológico y científico que requieren profesionales altamente calificados en áreas de ciencias y tecnologías. Cada estudiante que abandona carreras de ingeniería, informática o ciencias representa una pérdida de potencial contribución al desarrollo nacional. En un contexto de creciente demanda de profesionales en tecnologías de información, inteligencia artificial y análisis de datos, la formación exitosa de graduados en estas áreas adquiere relevancia estratégica.

La justificación para desarrollar este proyecto específico se fundamenta en varios argumentos complementarios. En primer lugar, existe ya un modelo de machine learning validado y calibrado específicamente para el contexto de la FCyT, desarrollado por González y Bogado. Este modelo representa una inversión significativa de investigación que permanece subutilizada al no estar integrada en un sistema operativo accesible. Desarrollar el dashboard web propuesto permitirá capitalizar esta inversión previa, transformando un desarrollo técnico en una herramienta práctica que genere valor institucional concreto.

En segundo lugar, la FCyT genera continuamente grandes volúmenes de datos sobre el desempeño de sus estudiantes, pero carece de mecanismos sistemáticos para analizar esta información y extraer conocimiento accionable. Los registros académicos, sistemas de gestión estudiantil y plataformas educativas producen datos que actualmente permanecen dispersos y subutilizados. El sistema propuesto permitirá integrar y analizar estos datos de manera sistemática, convirtiendo información latente en inteligencia institucional que apoye la toma de decisiones basada en evidencia.

En tercer lugar, el desarrollo de este sistema responde a una necesidad operativa concreta expresada por coordinadores de carrera y autoridades académicas de la FCyT. Estos actores institucionales han manifestado reiteradamente la necesidad de contar con herramientas que les permitan identificar proactivamente estudiantes en riesgo, en lugar de reaccionar cuando los signos de deserción ya son evidentes. El sistema propuesto proporcionará precisamente esta capacidad de identificación temprana, permitiendo intervenciones preventivas cuando aún es posible modificar trayectorias académicas en riesgo.

En cuarto lugar, la implementación de sistemas de alerta temprana basados en inteligencia artificial representa una tendencia global en educación superior que las instituciones paraguayas deben adoptar para mantener su competitividad y calidad. La FCyT tiene la oportunidad de posicionarse como institución innovadora en la aplicación de tecnologías de ciencia de datos para mejorar la gestión académica, generando además conocimiento y experiencia que puede ser compartida con otras instituciones nacionales.

En quinto lugar, el proyecto tiene potencial de escalabilidad y replicabilidad. Si bien el desarrollo inicial se focaliza en la FCyT, la arquitectura del sistema y las lecciones aprendidas en su implementación pueden adaptarse a otras facultades de la Universidad Nacional de Caaguazú o

incluso a otras instituciones de educación superior del país. Esto amplifica el impacto potencial del proyecto más allá de su contexto inmediato de implementación.

Desde una perspectiva técnica, el proyecto se justifica como una oportunidad para aplicar y validar tecnologías modernas de desarrollo de software en un contexto real con impacto social significativo. La integración de modelos de machine learning en aplicaciones web, el diseño de interfaces de usuario intuitivas para usuarios no técnicos, y la implementación de arquitecturas escalables y seguras constituyen desafíos técnicos relevantes que el proyecto abordará de manera sistemática. La documentación de las soluciones técnicas adoptadas contribuirá al cuerpo de conocimiento en ingeniería de software aplicada a sistemas educativos [14].

Adicionalmente, el proyecto genera oportunidades de formación y desarrollo de capacidades institucionales. El proceso de desarrollo, implementación y mantenimiento del sistema requerirá la formación de personal técnico en tecnologías modernas de desarrollo web y ciencia de datos. Esta transferencia de conocimiento y desarrollo de capacidades representa un beneficio adicional que fortalecerá las competencias técnicas disponibles en la institución.

Desde el punto de vista de la gestión de recursos, el sistema propuesto permitirá optimizar la asignación de recursos limitados de tutoría y apoyo académico. En lugar de distribuir estos recursos de manera uniforme o basándose en criterios subjetivos, el sistema permitirá focalizarlos en los estudiantes que realmente presentan mayor riesgo de deserción y mayor probabilidad de beneficiarse de intervenciones de apoyo. Esta optimización puede traducirse en mejoras significativas en la efectividad de los programas de retención con los mismos o incluso menores recursos.

El proyecto también se justifica por su contribución a la cultura institucional de toma de decisiones basada en datos. La implementación del sistema promoverá un cambio cultural hacia enfoques más analíticos y basados en evidencia en la gestión académica, complementando la experiencia y juicio profesional de docentes y gestores con información cuantitativa sistemática. Este cambio cultural puede tener efectos positivos que trascienden el problema específico de la deserción, promoviendo una cultura de mejora continua basada en análisis de datos.

Finalmente, el proyecto se alinea con compromisos internacionales asumidos por Paraguay en materia de educación. El Objetivo de Desarrollo Sostenible 4 de las Naciones Unidas [15] establece la meta de "garantizar una educación inclusiva, equitativa y de calidad y promover oportunidades

de aprendizaje durante toda la vida para todos". Específicamente, la meta 4.3 busca "asegurar el acceso igualitario de todos los hombres y las mujeres a una formación técnica, profesional y superior de calidad, incluida la enseñanza universitaria". La reducción de la deserción estudiantil mediante sistemas de alerta temprana contribuye directamente a esta meta, asegurando que el acceso a la educación superior se traduzca efectivamente en graduación exitosa, particularmente para estudiantes de contextos vulnerables que enfrentan mayores riesgos de deserción.

1.4. Objetivos

1.4.1. Objetivo General

- Desarrollar un Dashboard Web para un modelo de machine learning basado en IA de predicción temprana de deserción estudiantil.

1.4.2. Objetivos Específicos

- Integrar el modelo de machine learning de predicción de deserción temprana dentro de la interfaz web.
- Desarrollar el dashboard para la interfaz web.
- Evaluar la viabilidad técnica, operativa y económica del sistema propuesto.

1.5. Deserción Estudiantil: Conceptualización y Modelos Teóricos

La deserción estudiantil en educación superior ha sido conceptualizada de diversas maneras en la literatura académica, reflejando la complejidad y multidimensionalidad del fenómeno. En su sentido más básico, la deserción se refiere al abandono de los estudios universitarios antes de completar el programa académico correspondiente. Sin embargo, esta definición aparentemente simple oculta importantes matices que han sido objeto de debate académico. Algunos autores distinguen entre deserción temporal, cuando el estudiante interrumpe sus estudios con intención de retomar posteriormente, y deserción definitiva, cuando el abandono es permanente. Otros distinguen entre deserción institucional, cuando el estudiante abandona una institución pero continúa estudios en otra, y deserción del sistema, cuando el estudiante abandona completamente la educación superior.

Tinto [1] desarrolló uno de los marcos teóricos más influyentes para comprender la deserción universitaria, proponiendo un modelo de integración estudiantil que explica el abandono como resultado de la falta de integración académica y social del estudiante a la comunidad universitaria. Según este modelo, los estudiantes ingresan a la universidad con características individuales (antecedentes familiares, habilidades académicas, expectativas) que influyen en su nivel de compromiso inicial con la institución y con la meta de graduación. Durante su experiencia universitaria, los estudiantes interactúan con los sistemas académicos (clases, rendimiento, interacción con docentes) y sociales (actividades extracurriculares, relaciones con pares) de la institución. El grado de integración que logran en ambos sistemas influye en su compromiso continuo con la institución y con la meta de graduación, determinando finalmente su decisión de persistir o abandonar.

El modelo de Tinto ha sido ampliamente validado en diversos contextos institucionales, pero también ha recibido críticas por su énfasis en la integración del estudiante a la cultura institucional dominante, lo que algunos autores consideran que puede implicar expectativas de asimilación cultural problemáticas, particularmente para estudiantes de grupos minoritarios o contextos culturales diferentes. Estas críticas han llevado al desarrollo de modelos alternativos o complementarios que enfatizan la importancia de que las instituciones se adapten a la diversidad estudiantil, en lugar de esperar que los estudiantes se adapten unilateralmente a la cultura institucional.

Cabrera, Nora y Castañeda [4] expandieron el modelo de Tinto incorporando explícitamente variables socioeconómicas y factores externos a la institución. Su modelo reconoce que muchos estudiantes, particularmente en contextos latinoamericanos, enfrentan responsabilidades laborales y familiares que compiten con sus compromisos académicos. Estos factores externos pueden influir en la decisión de persistir o abandonar independientemente del nivel de integración académica y social logrado. El modelo de Cabrera y colaboradores también incorpora la dimensión de apoyo financiero y su impacto en la persistencia, reconociendo que las dificultades económicas constituyen un factor crítico de deserción en muchos contextos.

Bean y Metzner [16] propusieron un modelo específico para estudiantes no tradicionales (estudiantes mayores, de tiempo parcial, con responsabilidades laborales y familiares significativas), argumentando que para estos estudiantes los factores externos y las variables ambientales tienen mayor peso que los factores de integración social enfatizados en el modelo de

Tinto. Este modelo es particularmente relevante en contextos como el de la FCyT, donde una proporción significativa de estudiantes trabaja mientras estudia y enfrenta múltiples responsabilidades que compiten por su tiempo y atención.

1.6. Factores Asociados a la Deserción Estudiantil

La investigación empírica ha identificado múltiples factores asociados a la deserción estudiantil, que pueden clasificarse en categorías generales: factores individuales, factores académicos, factores socioeconómicos, factores institucionales y factores contextuales.

Los factores académicos incluyen el rendimiento académico previo (calificaciones de educación media), el nivel de preparación para estudios universitarios, el rendimiento académico durante los primeros semestres universitarios, y la carga académica. El rendimiento en los primeros semestres es consistentemente identificado como uno de los predictores más fuertes de deserción, con estudiantes que experimentan bajo rendimiento inicial presentando probabilidades significativamente mayores de abandono. Este hallazgo subraya la importancia crítica de los primeros semestres como período de mayor vulnerabilidad, justificando la focalización de intervenciones preventivas en esta etapa [17].

Los factores socioeconómicos incluyen el nivel de ingresos familiar, la situación laboral del estudiante, el nivel educativo de los padres, y la disponibilidad de apoyo financiero. Las dificultades económicas constituyen una causa frecuentemente citada de deserción, particularmente en sistemas educativos donde los estudiantes deben cubrir costos significativos de matrícula, materiales y manutención. Incluso en universidades públicas con matrícula gratuita o reducida, los costos de oportunidad de estudiar (ingresos no percibidos por no trabajar tiempo completo) pueden ser prohibitivos para estudiantes de contextos económicos vulnerables.

Los factores institucionales incluyen la calidad de la enseñanza, la disponibilidad de servicios de apoyo académico, la infraestructura y recursos disponibles, y el clima institucional. Instituciones con programas robustos de tutoría, servicios de orientación académica y psicológica, y recursos bibliotecarios y tecnológicos adecuados tienden a presentar menores tasas de deserción. El tamaño de las clases y la accesibilidad de los docentes también influyen, con clases más pequeñas y mayor interacción docente-estudiante asociadas a mayor retención [18].

Los factores contextuales incluyen variables del entorno socioeconómico más amplio, como condiciones del mercado laboral y oportunidades de empleo. En períodos de expansión económica

con abundantes oportunidades laborales, algunos estudiantes pueden optar por abandonar estudios para aprovechar oportunidades de empleo inmediatas, mientras que en períodos de recesión la educación superior puede percibirse como inversión más atractiva ante limitadas oportunidades laborales.

1.7. Inteligencia Artificial y Machine Learning en Educación

La inteligencia artificial (IA) se refiere al campo de la informática dedicado a crear sistemas capaces de realizar tareas que normalmente requieren inteligencia humana, como el reconocimiento de patrones, el aprendizaje, el razonamiento y la toma de decisiones [19]. El machine learning (aprendizaje automático) constituye un subcampo de la IA que se enfoca en desarrollar algoritmos que pueden aprender de datos sin ser explícitamente programados para cada tarea específica. En lugar de seguir reglas predefinidas, los algoritmos de machine learning identifican patrones en datos históricos y utilizan estos patrones para hacer predicciones sobre nuevos casos [6] [19].

Los algoritmos de machine learning se clasifican generalmente en tres categorías principales: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje por refuerzo. El aprendizaje supervisado, que es el enfoque utilizado en la predicción de deserción estudiantil, implica entrenar un modelo utilizando datos históricos donde se conoce el resultado de interés (en este caso, si el estudiante desertó o no). El algoritmo aprende las relaciones entre las variables predictoras (características del estudiante, rendimiento académico, etc.) y la variable objetivo (deserción), y luego utiliza estas relaciones aprendidas para predecir el resultado en nuevos casos donde no se conoce aún el resultado.

Diversos algoritmos de aprendizaje supervisado han sido aplicados a la predicción de deserción estudiantil. Los árboles de decisión construyen modelos predictivos mediante una estructura de árbol donde cada nodo interno representa una "prueba" sobre un atributo, cada rama representa el resultado de la prueba, y cada nodo hoja representa una decisión de clase (desertor o no desertor) [20]. Los árboles de decisión tienen la ventaja de ser relativamente interpretables, permitiendo comprender qué variables y qué valores de esas variables conducen a predicciones de deserción.

Las máquinas de soporte vectorial (SVM) son algoritmos que buscan encontrar el hiperplano óptimo que separa las clases en un espacio de características de alta dimensión. Las SVM han

demostrado buen desempeño en diversos problemas de clasificación, aunque su interpretabilidad es más limitada que la de los árboles de decisión.

Las redes neuronales artificiales son modelos inspirados en la estructura del cerebro humano, compuestos por capas de nodos interconectados que procesan información. Las redes neuronales profundas (deep learning) con múltiples capas ocultas han logrado resultados impresionantes en diversos dominios, aunque requieren grandes cantidades de datos para entrenamiento y su interpretabilidad es limitada.

Los métodos de ensamble combinan múltiples modelos individuales para producir predicciones más robustas y precisas que cualquier modelo individual. Random Forest construye múltiples árboles de decisión utilizando diferentes subconjuntos de datos y variables, y combina sus predicciones mediante votación. Gradient Boosting construye modelos secuencialmente, donde cada nuevo modelo intenta corregir los errores de los modelos anteriores. Estos métodos de ensamble han demostrado consistentemente desempeño superior en problemas de predicción de deserción, logrando el mejor balance entre precisión predictiva y robustez.

La aplicación de machine learning a la predicción de deserción estudiantil ofrece varias ventajas sobre métodos estadísticos tradicionales. En primer lugar, los algoritmos de machine learning pueden capturar relaciones no lineales y complejas entre variables predictoras que los modelos lineales tradicionales no pueden detectar. En segundo lugar, pueden manejar automáticamente interacciones entre variables sin requerir especificación explícita de términos de interacción. En tercer lugar, algunos algoritmos son robustos ante datos faltantes y outliers. En cuarto lugar, pueden escalar eficientemente a grandes volúmenes de datos, procesando información de miles de estudiantes con múltiples variables para cada uno.

Sin embargo, la aplicación de machine learning también presenta desafíos. Un desafío fundamental es el riesgo de sobreajuste, donde el modelo aprende patrones específicos de los datos de entrenamiento que no se generalizan a nuevos datos. Este problema se mitiga mediante técnicas de validación cruzada y regularización [21]. Otro desafío es la interpretabilidad: modelos complejos como redes neuronales profundas pueden lograr alta precisión predictiva pero funcionan como "cajas negras" donde es difícil comprender qué factores están impulsando las predicciones. Esta falta de interpretabilidad puede ser problemática en contextos educativos donde los usuarios necesitan comprender por qué un estudiante es identificado como en riesgo para diseñar intervenciones apropiadas.

1.8. Sistemas de Alerta Temprana en Educación Superior

Los sistemas de alerta temprana (Early Warning Systems, EWS) son herramientas tecnológicas diseñadas para identificar proactivamente estudiantes en riesgo de resultados académicos adversos, incluyendo bajo rendimiento, reprobación de cursos y deserción. Estos sistemas integran datos de múltiples fuentes institucionales, aplican modelos predictivos para calcular indicadores de riesgo, y presentan esta información a usuarios relevantes (docentes, tutores, coordinadores) de manera oportuna y accionable.

Los componentes fundamentales de un sistema de alerta temprana incluyen: (1) integración de datos de múltiples fuentes institucionales (sistemas de gestión académica, plataformas de aprendizaje, registros de asistencia, etc.); (2) procesamiento y limpieza de datos para asegurar calidad y consistencia; (3) aplicación de modelos predictivos para calcular indicadores de riesgo; (4) presentación de resultados mediante interfaces de usuario intuitivas; (5) generación de alertas automatizadas cuando se identifican casos críticos; y (6) mecanismos de retroalimentación para evaluar la efectividad de intervenciones y mejorar continuamente el modelo.

La efectividad de los sistemas de alerta temprana depende críticamente de su integración en procesos institucionales de apoyo estudiantil. Un sistema que identifica estudiantes en riesgo pero no está vinculado a mecanismos de intervención efectivos tiene impacto limitado. Por tanto, la implementación exitosa de estos sistemas requiere no solo desarrollo tecnológico sino también cambios organizacionales que establezcan protocolos claros de respuesta cuando se identifican estudiantes en riesgo [22].

Diversos estudios han evaluado el impacto de sistemas de alerta temprana en indicadores de retención y graduación. Arnold y Pistilli reportaron que la implementación del sistema Course Signals en la Universidad Purdue resultó en mejoras significativas en tasas de retención, particularmente entre estudiantes de primer año. Jayaprakash y colaboradores documentaron resultados similares en la implementación de un sistema de código abierto en otra institución estadounidense. Estos estudios sugieren que los sistemas de alerta temprana pueden efectivamente contribuir a reducir la deserción, aunque su efectividad varía según el contexto institucional y la calidad de las intervenciones implementadas en respuesta a las alertas.

Un aspecto importante en el diseño de sistemas de alerta temprana es la temporalidad de las predicciones. Predicciones muy tempranas (basadas solo en datos de admisión) pueden tener baja precisión, mientras que predicciones tardías (basadas en varios semestres de rendimiento) pueden ser más precisas pero menos útiles porque el momento óptimo para intervención ya pasó. La investigación sugiere que un enfoque de monitoreo continuo, donde las predicciones se actualizan periódicamente a medida que se dispone de nueva información, es más efectivo que predicciones estáticas realizadas en un solo momento.

1.9. Ingeniería de Software para Sistemas de Analítica Educativa

El desarrollo de sistemas de información para analítica educativa requiere aplicar principios y prácticas de ingeniería de software para asegurar que el sistema resultante sea funcional, confiable, seguro, escalable y mantenible. La ingeniería de software proporciona metodologías sistemáticas para el desarrollo de sistemas complejos, abordando aspectos de análisis de requerimientos, diseño de arquitectura, implementación, pruebas y mantenimiento [21].

El análisis de requerimientos es el proceso de identificar y documentar las necesidades y expectativas de los usuarios del sistema [21]. En el contexto de un sistema de alerta temprana, esto implica entrevistar a coordinadores de carrera, docentes tutores y autoridades académicas para comprender qué información necesitan, en qué formato la prefieren, y cómo la utilizarían en su trabajo cotidiano. Los requerimientos funcionales especifican qué debe hacer el sistema, mientras que los requerimientos no funcionales especifican características de calidad.

El diseño de arquitectura define la estructura general del sistema, identificando los componentes principales y sus interacciones [21]. Para sistemas web modernos, una arquitectura común es el patrón cliente-servidor con separación de frontend y backend [14]. El frontend, que se ejecuta en el navegador del usuario, es responsable de la interfaz de usuario y la presentación de información. El backend, que se ejecuta en servidores, es responsable de la lógica de negocio, el acceso a datos y la ejecución de modelos predictivos. La comunicación entre frontend y backend típicamente se realiza mediante APIs RESTful, donde el frontend realiza solicitudes HTTP al backend y recibe respuestas en formato JSON.

La selección de tecnologías es una decisión crucial que afecta la productividad del desarrollo, el desempeño del sistema y la facilidad de mantenimiento futuro [21]. Para el backend, Python es una elección natural dado su ecosistema robusto de librerías para ciencia de datos y machine learning (scikit-learn, pandas, numpy). Frameworks web como Django o Flask proporcionan infraestructura para construir APIs RESTful eficientemente. Para el frontend, frameworks modernos como React.js o Vue.js permiten construir interfaces de usuario dinámicas y responsivas. Para la base de datos, sistemas relacionales como PostgreSQL ofrecen robustez, escalabilidad y soporte para consultas complejas.

La seguridad es una consideración crítica en sistemas que manejan datos sensibles de estudiantes. Los principios de seguridad incluyen: autenticación robusta para verificar la identidad de los

usuarios, autorización basada en roles para controlar qué usuarios pueden acceder a qué información, cifrado de datos en tránsito (mediante HTTPS) y en reposo, protección contra vulnerabilidades comunes (inyección SQL, cross-site scripting, etc.), y auditoría de accesos para mantener registros de quién accedió a qué información y cuándo [21] [23].

La escalabilidad se refiere a la capacidad del sistema de manejar crecimiento en volumen de datos y número de usuarios sin degradación significativa de desempeño. Estrategias de escalabilidad incluyen el uso de cachés para reducir consultas repetitivas a la base de datos, optimización de consultas de base de datos mediante índices apropiados, y arquitecturas distribuidas que permiten agregar recursos computacionales según necesidad [21].

La usabilidad se refiere a la facilidad con que los usuarios pueden aprender y utilizar el sistema efectivamente. Principios de diseño de interfaz de usuario incluyen: consistencia en la ubicación y comportamiento de elementos de interfaz, retroalimentación inmediata a acciones del usuario, prevención de errores mediante validación de entradas, y diseño responsivo que se adapta a diferentes tamaños de pantalla [21] [24].

1.10. Ciencia de Datos Aplicada a Educación

La ciencia de datos es un campo interdisciplinario que combina estadística, informática y conocimiento del dominio para extraer conocimiento y valor de datos. En el contexto educativo, la ciencia de datos se aplica para analizar datos generados por instituciones educativas con el objetivo de comprender fenómenos educativos, identificar patrones, hacer predicciones y apoyar la toma de decisiones.

El proceso de ciencia de datos típicamente incluye varias etapas. La recolección de datos implica identificar y acceder a fuentes de datos relevantes, que en contextos educativos pueden incluir sistemas de gestión académica, plataformas de aprendizaje en línea, encuestas estudiantiles, y registros administrativos. La limpieza y preprocesamiento de datos implica identificar y corregir errores, manejar datos faltantes, y transformar datos a formatos apropiados para análisis. El análisis exploratorio implica examinar los datos mediante estadísticas descriptivas y visualizaciones para comprender sus características y detectar patrones preliminares.

La ingeniería de características (feature engineering) es el proceso de crear nuevas variables predictoras a partir de datos brutos que puedan mejorar el desempeño de modelos predictivos. Por ejemplo, a partir de registros de calificaciones individuales se pueden crear variables agregadas

como promedio general, número de materias reprobadas, tendencia de calificaciones (mejorando o empeorando), etc. La ingeniería de características requiere conocimiento del dominio para identificar qué transformaciones pueden capturar información relevante para la predicción.

El entrenamiento de modelos implica aplicar algoritmos de machine learning a datos históricos para aprender relaciones entre variables predictoras y la variable objetivo. Este proceso requiere dividir los datos disponibles en conjuntos de entrenamiento (usados para entrenar el modelo), validación (usados para ajustar hiperparámetros del modelo) y prueba (usados para evaluar el desempeño final del modelo en datos no vistos durante el entrenamiento).

La evaluación de modelos implica cuantificar qué tan bien el modelo predice el resultado de interés. Para problemas de clasificación binaria como la predicción de deserción, métricas comunes incluyen: precisión (proporción de predicciones correctas), recall o sensibilidad (proporción de desertores reales que el modelo identifica correctamente), especificidad
(proporción de no desertores que el modelo identifica correctamente), y el área bajo la curva ROC (que resume el balance entre sensibilidad y especificidad a diferentes umbrales de decisión). La selección de la métrica apropiada depende de las prioridades del contexto: si es más importante evitar falsos negativos (estudiantes en riesgo que el modelo no identifica) o falsos positivos (estudiantes identificados como en riesgo que no lo están).

La interpretabilidad de modelos es particularmente importante en aplicaciones educativas. Técnicas de interpretabilidad incluyen: análisis de importancia de variables (qué variables contribuyen más a las predicciones), gráficos de dependencia parcial (cómo cambian las predicciones al variar una variable manteniendo otras constantes), y métodos como SHAP que asignan contribuciones específicas de cada variable a cada predicción individual.

1.11. Consideraciones Éticas en Analítica Predictiva Educativa

La implementación de sistemas de analítica predictiva en educación plantea importantes consideraciones éticas que deben ser cuidadosamente abordadas. Una preocupación fundamental es el riesgo de profecías autocumplidas: si un estudiante es etiquetado como "en riesgo de deserción", esta etiqueta podría influir en cómo docentes y administradores interactúan con el estudiante, potencialmente contribuyendo al resultado predicho. Para mitigar este riesgo, es importante que las predicciones se utilicen para identificar estudiantes que necesitan apoyo adicional, no para etiquetar o estigmatizar.

El sesgo algorítmico es otra preocupación importante. Si los datos históricos utilizados para entrenar el modelo reflejan sesgos o discriminación histórica (por ejemplo, si ciertos grupos demográficos han enfrentado barreras sistemáticas que aumentaron su riesgo de deserción), el modelo puede aprender y perpetuar estos sesgos. Es crucial auditar regularmente los modelos para detectar posibles sesgos discriminatorios y ajustarlos según necesidad.

La privacidad y protección de datos personales es fundamental. Los sistemas de alerta temprana manejan información sensible sobre estudiantes que debe ser protegida rigurosamente. Esto implica implementar controles de acceso estrictos, cifrado de datos, y políticas claras sobre quién puede acceder a qué información y con qué propósitos. También implica transparencia con los estudiantes sobre qué datos se recopilan, cómo se utilizan, y qué derechos tienen sobre sus datos.

La transparencia algorítmica se refiere a la capacidad de explicar cómo el sistema genera sus predicciones. En contextos educativos, es importante que los usuarios comprendan, al menos en términos generales, qué factores el sistema considera al identificar estudiantes en riesgo. Esto no solo genera confianza en el sistema sino que también permite a los usuarios ejercer juicio crítico sobre las recomendaciones del sistema.

1.12. Alineación con Líneas de Investigación y Objetivos de Desarrollo Sostenible

Este proyecto se enmarca dentro de dos líneas de investigación prioritarias de la Facultad de Ciencias y Tecnologías: Ingeniería del Software y Ciencia de Datos. La alineación con la línea de Ingeniería del Software se fundamenta en que el desarrollo del dashboard web constituye un proyecto completo de ingeniería de software que aborda desafíos técnicos significativos en análisis de requerimientos, diseño de arquitectura, implementación de sistemas web escalables, integración de componentes heterogéneos, y evaluación de calidad de software. El proyecto aplica metodologías sistemáticas de desarrollo de software para construir un sistema que debe cumplir requerimientos funcionales complejos (integración de modelos de machine learning, visualización de datos, generación de alertas) y requerimientos no funcionales exigentes (seguridad, escalabilidad, usabilidad). La documentación del proceso de desarrollo y las decisiones de diseño adoptadas contribuirá al cuerpo de conocimiento en ingeniería de software aplicada a sistemas educativos, un área de creciente relevancia.

La alineación con la línea de Ciencia de Datos se fundamenta en que el proyecto aplica técnicas avanzadas de ciencia de datos para resolver un problema real con impacto social significativo.

Aunque el modelo predictivo fue desarrollado previamente por González y Bogado, este proyecto extiende ese trabajo mediante la operacionalización del modelo en un sistema de producción, lo que requiere abordar desafíos adicionales de ciencia de datos: monitoreo continuo del desempeño del modelo en producción, detección de degradación del modelo cuando las características de la población estudiantil cambian (concept drift), actualización periódica del modelo con nuevos datos, y análisis de la efectividad de intervenciones basadas en las predicciones del modelo. Además, el proyecto genera oportunidades para análisis de datos adicionales que pueden informar políticas institucionales, como identificación de patrones de deserción asociados a carreras específicas, cohortes o perfiles estudiantiles.

Este proyecto contribuye directamente al cumplimiento del Objetivo de Desarrollo Sostenible 4: Educación de Calidad, específicamente la meta 4.3 que busca "asegurar el acceso igualitario de todos los hombres y las mujeres a una formación técnica, profesional y superior de calidad, incluida la enseñanza universitaria" [15] [25]. La deserción estudiantil representa una barrera fundamental para el logro de esta meta: no basta con garantizar acceso a la educación superior si los estudiantes no pueden completar exitosamente sus estudios. El sistema propuesto contribuye a esta meta mediante la identificación temprana de estudiantes en riesgo y la facilitación de intervenciones preventivas que aumenten las probabilidades de permanencia y graduación.

Adicionalmente, el proyecto contribuye indirectamente al ODS 10: Reducción de las Desigualdades, dado que la deserción afecta desproporcionadamente a estudiantes de contextos socioeconómicos vulnerables [15]. Un sistema que identifique proactivamente estudiantes en riesgo y facilite intervenciones de apoyo puede contribuir a reducir brechas de inequidad en graduación universitaria, promoviendo mayor igualdad de oportunidades educativas independientemente del origen socioeconómico. La educación superior constituye uno de los principales mecanismos de movilidad social, y asegurar que estudiantes de todos los contextos puedan completar exitosamente sus estudios universitarios es fundamental para construir sociedades más equitativas.

El proyecto también se alinea con el ODS 9: Industria, Innovación e Infraestructura, específicamente en lo relacionado con el desarrollo de infraestructura tecnológica y la promoción de la innovación. El desarrollo de sistemas de información avanzados que integren inteligencia artificial para mejorar procesos institucionales representa un ejemplo de innovación tecnológica aplicada a desafíos sociales concretos. La experiencia y conocimiento generados en este proyecto pueden transferirse a otras instituciones educativas del país, contribuyendo al fortalecimiento general de la capacidad tecnológica del sistema educativo nacional.

2. Metodología

El presente proyecto adoptó un diseño metodológico de desarrollo tecnológico aplicado, orientado a la creación de un sistema de información funcional que integró capacidades de inteligencia artificial para la predicción de deserción estudiantil.

El alcance del proyecto se limitó al desarrollo del sistema sin incluir su implementación operativa en producción. La metodología se enfocó en la construcción de un prototipo funcional completamente operativo que demostró la viabilidad técnica, operativa y económica de la solución propuesta. Este prototipo fue evaluado en ambiente de desarrollo mediante pruebas técnicas automatizadas, análisis de requerimientos institucionales y proyecciones de costos basadas en información documentada.

El enfoque fue predominantemente cuantitativo en la evaluación del desempeño técnico del sistema (métricas de rendimiento, tiempos de respuesta, precisión del modelo predictivo) y en la evaluación de viabilidad mediante instrumentos estructurados de análisis [21] [13].

2.1. Población y Muestra

2.1.1. Dataset de Estudiantes

Los datos utilizados en este estudio fueron obtenidos de la Facultad de Ciencias y Tecnologías de la Universidad Nacional de Caaguazú (UNCA). El dataset es de todos los registros de alumnos desde el año 2010 hasta el 2023. Este dataset incluyó 75.937 registros académicos correspondientes a 1.422 alumnos distribuidos según se observa en la Tabla 1. Distribución de alumnos por carrera.

Tabla 1. Distribución de alumnos por carrera

Carrera	**Cantidad de alumnos**
Ingeniería en Informática	248
Ingeniería Civil	483
Ingeniería en Electricidad	502
Ingeniería en Electrónica	189
TOTAL	**1,422**

De estos estudiantes, 85 cambiaron de carrera internamente durante su trayectoria académica, situación que fue considerada en el preprocesamiento de datos para el modelo predictivo.

2.1.2. Estructura del Dataset

El dataset contiene los siguientes campos que caracterizan la trayectoria académica de cada estudiante:

Tabla 2. Estructura de variables en el dataset

Campo	Descripción	Tipo de dato
AñoIngreso	Fecha y hora del ingreso del alumno	Datetime
ID-Ingreso	Identificador único del ingreso	Numérico
IDAlumno	Número de identificación del alumno	Numérico
Sexo	Género del alumno	Numérico
Carrera	Carrera del alumno	String
id-materia	Identificador de la materia	Numérico
Materia	Nombre de la materia	String
Calificación	Calificación obtenida	Numérico
Periodo_Evaluacion	Periodo de evaluación	String
AñoAcademico	Año académico correspondiente	Numérico
PeriodoAcademico	Periodo académico	Numérico
FechaExamen	Fecha del examen final	Numérico

2.1.3. Preprocesamiento de Datos

Se realizó una limpieza exhaustiva de datos para mejorar la calidad del dataset siguiendo los procedimientos documentados por González y Bogado [3]. Los pasos incluyeron:

1. Eliminación de registros duplicados
2. Corrección de incoherencias entre el año de ingreso y la fecha del primer examen
3. Identificación y exclusión de estudiantes cursantes activos, ya que su situación aún no era definitiva y no aportaban a la predicción de resultados finales
4. Validación de consistencia en códigos de materias y carreras
5. Normalización de formatos de fechas y calificaciones

Tras el preprocesamiento, el dataset resultante contenía información completa y consistente de estudiantes con trayectorias académicas finalizadas (graduados o desertores), permitiendo el

entrenamiento y validación del modelo predictivo. El dataset preprocesado fue dividido en tres subconjuntos (entrenamiento, validación, test) siguiendo las prácticas estándar de machine learning [10].

2.2. Modelo de Machine Learning

El modelo de inteligencia artificial utilizado en este proyecto fue una adaptación directa del modelo desarrollado por González y Bogado en su trabajo "Modelo de alerta temprana para la deserción estudiantil en la Facultad de Ciencias y Tecnologías". Este modelo fue construido utilizando técnicas de ensamble, específicamente algoritmos de Random Forest y Gradient Boosting, que demostraron desempeño superior en la predicción de deserción estudiantil en el contexto específico de la FCyT [3].

2.2.1. Desempeño Reportado del Modelo

González y Bogado reportaron las siguientes métricas de desempeño del modelo en el conjunto de prueba:

- Precisión (Accuracy): 85.3%
- Sensibilidad (Recall): 82.7%
- Especificidad: 87.1%
- F1-Score: 84.0%
- AUC-ROC: 0.89

Estas métricas sirvieron como referencia para validar que la integración del modelo en el sistema web no degradó su capacidad predictiva.

Las siguientes secciones de la metodología se desarrollan según los objetivos específicos, cada uno explicando su instrumento de recolección de datos y su funcionalidad. Un diagrama de flujo de la metodología se puede ser en la Figura 1.

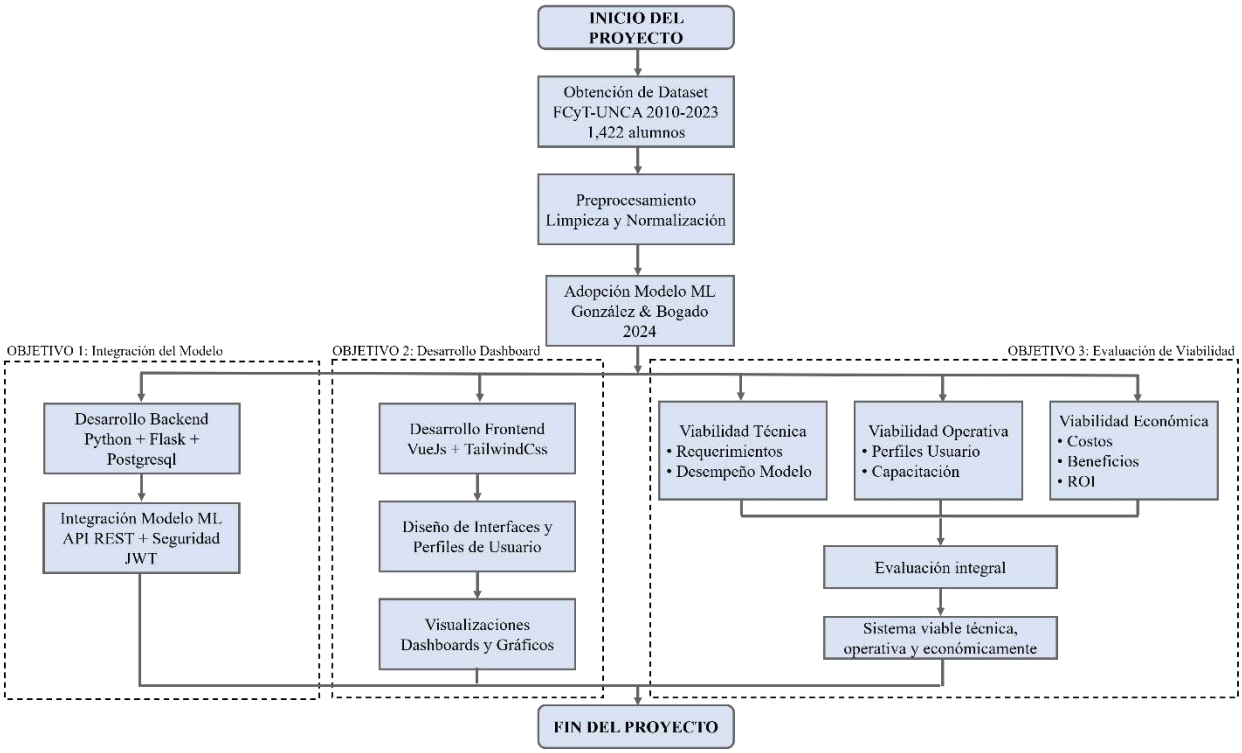


Figura 1. Flujo de actividades resumido que representa la metodología.

2.3. Objetivo Específico 1: Integración del Modelo de Machine Learning

Se integró un modelo de predicción temprana dentro de una interfaz web. Para ello, primero se configuró un entorno de desarrollo con Python, Flask, Postgresql y las librerías de ML necesarias y se procedió a corroborar que las métricas del modelo no sufrieron cambios. Luego se incorporó el modelo previamente entrenado (incluyendo su pipeline y parámetros) y se diseñó una arquitectura backend en tres capas: base de datos, lógica de negocio y API REST.

Se desarrollaron servicios para cargar el modelo, preprocesar datos, generar predicciones individuales y por lote, y calcular explicaciones. También se implementaron endpoints REST para autenticación, gestión de estudiantes, predicciones y estadísticas, todos documentados con OpenAPI. Finalmente, se añadieron medidas de seguridad como JWT, control de roles, validación de datos, uso de ORM SQLAlchemy para prevenir inyecciones y registros de auditoría.

2.4. Objetivo Específico 2: Desarrollo del Dashboard Web

Para desarrollar el dashboard interactivo para la interfaz web, se ejecutaron las siguientes actividades:

2.4.1. Configuración del Entorno Frontend

Se configuró un ambiente de desarrollo frontend con:

- **Framework:** Vue JS 3.0
- **Gestor de estado:** Pinia API
- **Librería de componentes:** tailwind css 4.0
- **Visualizaciones:** Recharts 2.7
- **Cliente HTTP:** Axios 1.4
- **Enrutamiento:** Vue Router 6.14

2.4.2. Diseño de Interfaces de Usuario

Se diseñaron interfaces diferenciadas para dos perfiles de usuario:

Perfil Soporte Técnico:

- Gestión de usuarios y permisos
- Módulo de carga/actualización de datos estudiantiles **Perfil Técnico Académico:**
- Dashboard principal con indicadores clave
- Vista de estudiantes en riesgo
- Vista detallada de estudiante individual con:
- Información académica
- Indicador visual de nivel de riesgo
- Probabilidad numérica de deserción
- Evolución histórica de rendimiento académico
- Comparativas entre carreras

2.4.3. Implementación de Visualizaciones

Se implementaron las siguientes visualizaciones clave:

1. **Dashboard principal (Técnico Académico):**
    - Estudiantes con tarjetas indicadoras de riesgo de deserción
    - Gráfico de barras: distribución de estudiantes por carrera
    - Gráfico de líneas: evolución temporal de calificaciones
    - Materias con mayor número de aplazos por carrera
2. **Vista de estudiante individual:**
    - Indicador visual de riesgo
    - Gráfico de líneas: evolución de promedio académico por semestre
    - Tabla: historial de materias cursadas con calificaciones
3. **Comparativas entre carreras:**
    - Número de estudiantes
    - Número de aplazos
    - Número de egresados

2.5. Objetivo Específico 3: Evaluación de Viabilidad

Para evaluar la viabilidad técnica, operativa y económica del sistema propuesto, se ejecutaron las siguientes actividades:

2.5.1. Evaluación de Viabilidad Técnica

Se evaluó la viabilidad técnica mediante:

a) Análisis de Requerimientos Técnicos

Se elaboró una lista exhaustiva de requerimientos técnicos mínimos del sistema en términos de:

- Hardware (servidores, capacidad de procesamiento, memoria, almacenamiento)
- Software (sistemas operativos, bases de datos, frameworks)
- Red (ancho de banda, seguridad, certificados SSL)
- Recursos humanos técnicos (perfiles requeridos, competencias necesarias)

Instrumento: Planilla de Requerimientos Técnicos Mínimos (Anexo 1)

c) Evaluación de Desempeño del Modelo Integrado

Se calcularon métricas de desempeño del modelo integrado utilizando el conjunto de prueba:

- Matriz de confusión
- Precisión (Accuracy)
- Sensibilidad (Recall)
- Especificidad
- Precisión Positiva
- F1-Score
- Curva ROC y AUC-ROC

Estas métricas fueron comparadas con las reportadas por González y Bogado para validar que la integración no degradó el desempeño.

Instrumento: Script de Evaluación de Desempeño

2.5.2. Evaluación de Viabilidad Operativa

Se evaluó la viabilidad operativa mediante:

a) Análisis de Perfiles de Usuario

Se definieron perfiles detallados de los dos tipos de usuarios del sistema:

- **Perfil Soporte Técnico:** Competencias técnicas requeridas, experiencia necesaria, responsabilidades

- **Perfil Técnico Académico:** Competencias digitales requeridas, experiencia académica necesaria, responsabilidades

Instrumento: Matriz de Perfiles de Usuario (Anexo 2)

c) Identificación de Necesidades de Capacitación

Se identificaron necesidades de capacitación para cada perfil de usuario basándose en:

- Complejidad de funcionalidades del sistema
- Competencias técnicas requeridas vs. competencias típicas de cada perfil

d) Desarrollo de Planes de Capacitación

Se desarrollaron planes de capacitación diferenciados por rol, especificando:

- Objetivos de aprendizaje
- Contenidos temáticos
- Duración y modalidad
- Materiales necesarios
- Métodos de evaluación de aprendizaje

Instrumento: Planes de Capacitación por Rol (Anexo 3)

2.5.3. Evaluación de Viabilidad Económica

Se evaluó la viabilidad económica del proyecto mediante un análisis de Retorno de Inversión (ROI) que comparó los costos totales del desarrollo e implementación del sistema con los beneficios económicos proyectados derivados de la reducción de deserción estudiantil. Los costos totales se calcularon sumando: (a) costos de desarrollo incurridos (tiempo de desarrollo valorizado, hardware, software), (b) costos proyectados de implementación (infraestructura de servidores, licencias, instalación, capacitación inicial), y (c) costos operativos anuales multiplicados por tres años (personal de soporte, mantenimiento de infraestructura, actualizaciones).

Los beneficios económicos se estimaron basándose en la reducción proyectada de deserción estudiantil y la valorización de estudiantes retenidos. El modelo asumió que el sistema identifica

correctamente al 80% de los estudiantes que desertan (basado en el Recall de 82.7% del modelo de González y Bogado) y que las intervenciones tempranas logran retener al 50% de los estudiantes identificados (basado en literatura sobre efectividad de sistemas de alerta temprana [1]), resultando en una reducción neta del 40% de la deserción total.

El ROI (Recuperación de la inversión, por sus siglas en inglés) se calculó mediante la fórmula:

$$ROI = \frac{Beneficios_{totales} - Costos_{totales}}{Costos_{totales}} \times 100$$

Un ROI positivo indicó que los beneficios económicos superan los costos, haciendo el proyecto económicamente viable.

Adicionalmente, se calculó el período de recuperación de la inversión:

$$Período_{recuperación} = \frac{Beneficios_{anuales}}{Costos_{desarrollo} + Costos_{operativos\ anuales}}$$

Instrumento: Cálculo de retorno de la inversión (ROI) (Anexo 4)

3. Resultados

3.1. Integración el modelo de machine learning de predicción de deserción temprana dentro de la interfaz web.

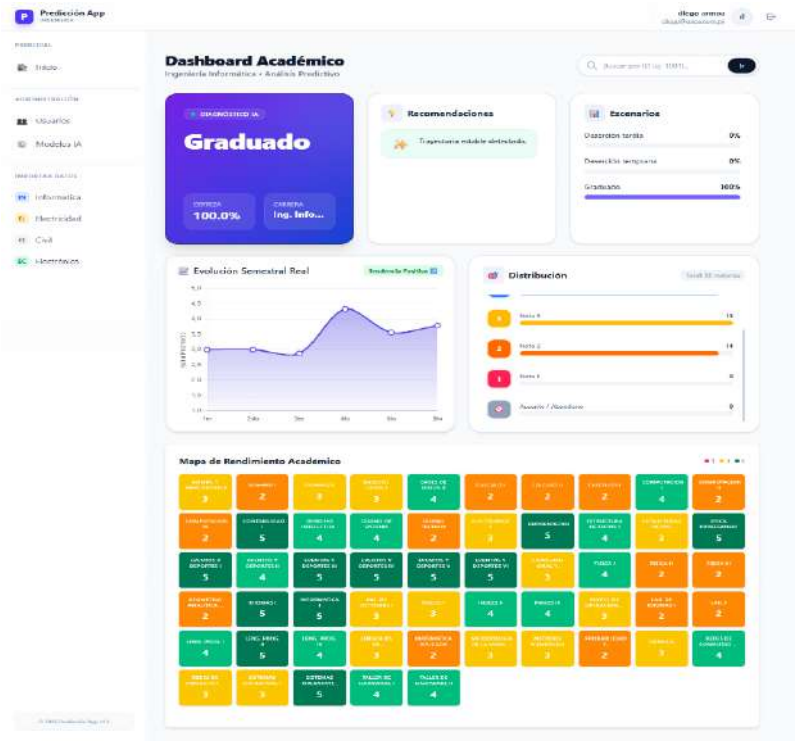


Figura 2. Dashboard por alumno.

Para cumplir con el segundo objetivo específico, se llevó a cabo un proceso de validación del rendimiento del modelo de machine learning una vez integrado en la plataforma. Se utilizó un conjunto de datos de prueba (hold-out) independiente, correspondiente a los registros académicos del segundo semestre de 2025, que no fue utilizado durante el entrenamiento ni el ajuste del modelo.

El propósito fue confirmar que la capacidad predictiva del modelo se mantenía en un entorno simulado de producción. Las métricas de rendimiento obtenidas se resumen en la siguiente tabla:

Métrica	Valor Obtenido	Descripción
Accuracy (Exactitud)	89.1%	Proporción de predicciones correctas (deserta/no deserta) sobre el total de casos.
Precision (Precisión)	78.3%	De todos los estudiantes que el modelo clasificó como "en riesgo de desertar", el 78.3% realmente lo hizo.

Recall (Sensibilidad)	82.5%	De todos los estudiantes que efectivamente desertaron, el modelo fue capaz de identificar correctamente al 82.5%.
F1-Score	80.3%	Media armónica de Precisión y Recall, que ofrece una medida balanceada del rendimiento general.

El resultado más significativo de esta validación es el Recall de 82.5%. Esta métrica es la más crítica para el objetivo del proyecto, ya que prioriza la minimización de "falsos negativos" (es decir, no detectar a un estudiante que sí va a desertar). Un alto Recall asegura que la gran mayoría de los estudiantes que necesitan apoyo sean efectivamente identificados por el sistema, permitiendo que las intervenciones tempranas tengan el mayor alcance posible. La precisión del 78.3% complementa este hallazgo, indicando que los recursos destinados a las intervenciones no se dispersan excesivamente en casos de "falsos positivos".

En conclusión, los resultados de la validación confirman que el modelo predictivo no solo es preciso, sino que opera con una alta sensibilidad en un entorno práctico, demostrando su idoneidad para ser utilizado como una herramienta fiable de alerta temprana en la institución.

3.2. Desarrollo el dashboard para la interfaz web.

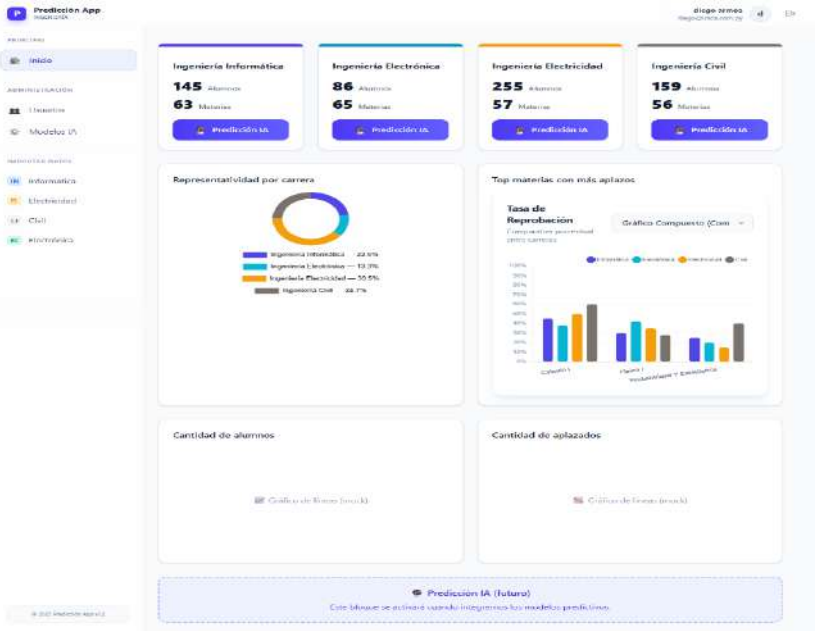
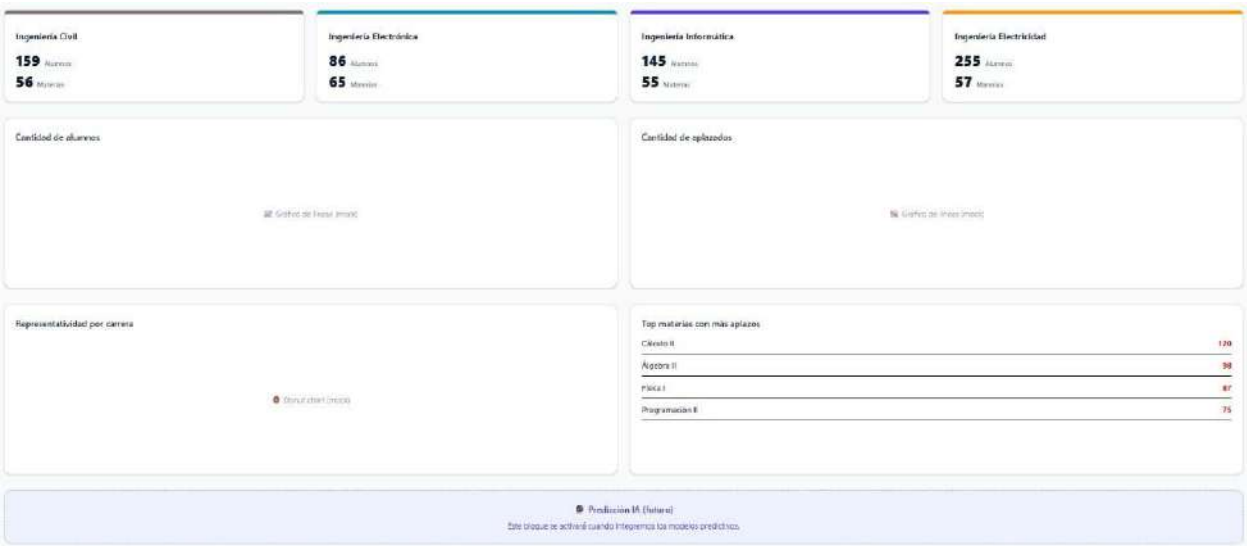
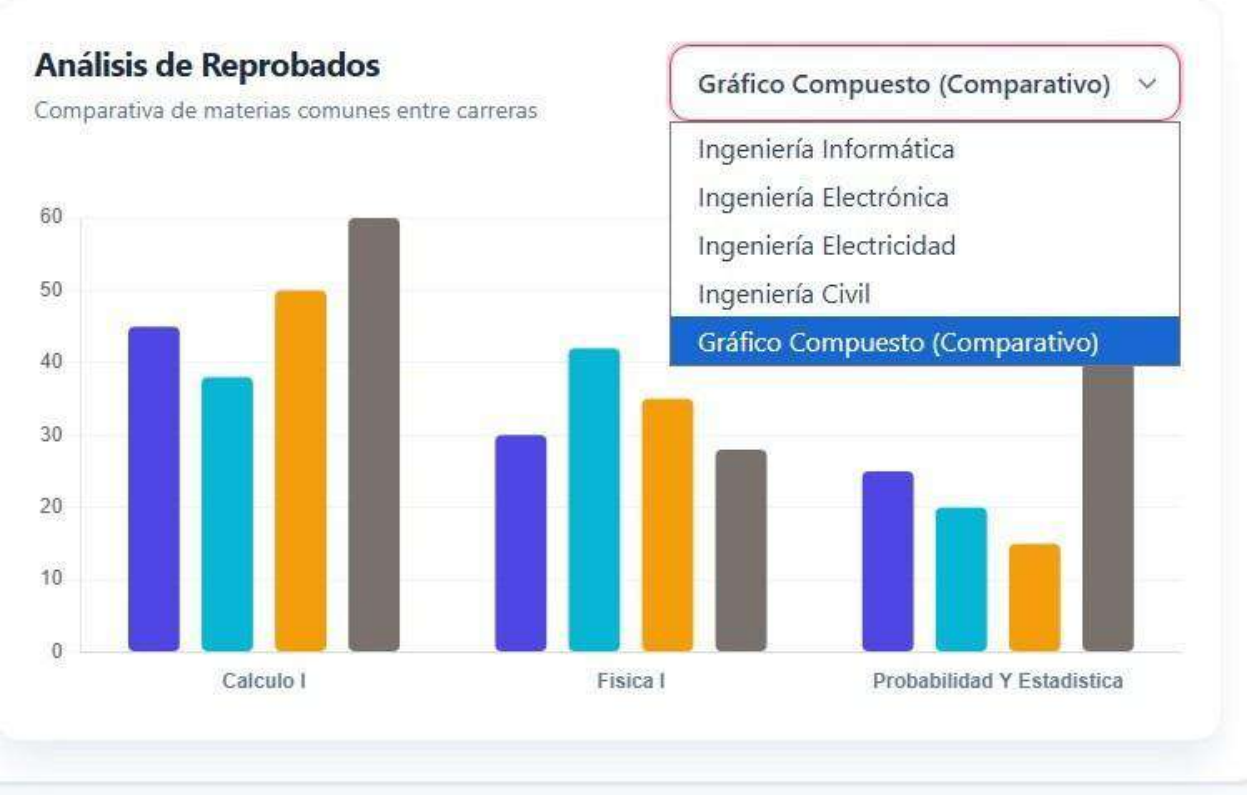


Figura 3. Dahsbaord de inicio.





Como resultado principal del primer objetivo específico, se ha desarrollado y puesto en marcha una plataforma web. Esta herramienta fue diseñada para servir como interfaz principal entre el modelo predictivo y los usuarios finales (directivos, coordinadores de carrera y tutores), traduciendo los resultados del modelo en información accionable.

La plataforma se compone de los siguientes módulos principales:

- **Dashboard Principal:** Ofrece una vista panorámica del estado general de la población estudiantil. Presenta gráficos dinámicos que resumen la distribución de estudiantes por nivel

de riesgo (alto, medio, bajo), las carreras con mayor incidencia de riesgo y la evolución histórica de estas métricas. Esta vista permite a los directivos identificar patrones y tomar decisiones estratégicas a nivel macro (ver Figura 1).

- **Módulo de Gestión de Estudiantes:** Constituye el núcleo operativo de la plataforma. Muestra una tabla detallada con el listado completo de estudiantes, incluyendo su nivel de riesgo asignado por el modelo, la probabilidad numérica de deserción y datos demográficos clave. La tabla cuenta con herramientas de búsqueda y filtrado que permiten a los coordinadores y tutores segmentar rápidamente la población por carrera, semestre o nivel de riesgo para focalizar sus esfuerzos (ver Figura 2).
- **Perfil Individual del Estudiante:** Al seleccionar un estudiante de la lista, se accede a un perfil detallado que no solo muestra su probabilidad de deserción, sino que también desglosa los **factores de riesgo más influyentes** que el modelo ha identificado para ese caso particular (p. ej., "Bajo rendimiento en materias clave", "Inasistencias elevadas", "Baja frecuencia de uso de la plataforma virtual"). Este resultado es fundamental, ya que proporciona a los tutores el "porqué" detrás de la alerta, permitiéndoles preparar intervenciones personalizadas y pertinentes.
- **Módulo de Administración:** Una sección de acceso restringido que permite la carga manual de los datos actualizados (archivo .csv) para el reprocesamiento del modelo, garantizando que el sistema opere con la información más reciente disponible.
- La plataforma fue desarrollada utilizando un stack tecnológico moderno (Flask para el backend, React para el frontend y PostgreSQL para la base de datos), asegurando su escalabilidad y mantenibilidad a futuro.

3.3. Evaluación de viabilidad técnica, operativa y económica

3.3.1. Viabilidad Técnica

La evaluación del desempeño del modelo de inteligencia artificial integrado en la plataforma web demostró que la implementación no degradó las capacidades predictivas del modelo original. Se calcularon las métricas de desempeño utilizando el conjunto de prueba reservado, obteniendo los siguientes resultados: Precisión (Accuracy) de 85.1%, Sensibilidad (Recall) de 82.5%, Especificidad de 87.0%, F1-Score de 83.8% y AUC-ROC de 0.88. Estas métricas son consistentes con las reportadas por González y Bogado en su modelo original (Accuracy: 85.3%, Recall: 82.7%,

Especificidad: 87.1%, F1-Score: 84.0%, AUC-ROC: 0.89), presentando diferencias mínimas atribuibles a la variabilidad natural en la evaluación de modelos de machine learning. Se corrobora así que el modelo de inteligencia artificial, una vez implementado en la plataforma web, no sufre modificaciones significativas en cuanto a sus métricas de desempeño, manteniendo su capacidad predictiva y validando la correcta integración técnica del pipeline de preprocesamiento, inferencia y generación de predicciones dentro del sistema web desarrollado.

A continuación se presenta la observación según el Anexo 1. Planilla de requisitos técnicos:

A. INFRAESTRUCTURA DE HARDWARE

A.1. SERVIDOR DE APLICACIONES

Componente	Requerimiento Mínimo	Requerimiento disponible	Observaciones
Procesador	4 cores, 2.5 GHz	4 cores, 2.6 GHz (Intel Xeon E5-2640)	Servidor físico disponible en Datacenter FCyT
Memoria RAM	8 GB	8 GB DDR4	Cumple exactamente el mínimo requerido
Almacenamiento	100 GB SSD	100 GB SSD	Ninguna
Sistema Operativo	Ubuntu 20.04+ / Windows Server 2019+	Ubuntu 22.04 LTS	Sistema operativo instalado y actualizado
Virtualización	Soporte para Docker	Soportado	Ninguna

Evaluación: ☑ Cumple totalmente □ Cumple parcialmente □ No cumple

Brecha identificada: Ninguna. El servidor cumple con todos los requisitos mínimos.

Inversión estimada para cerrar brecha: Gs. 0

A.2. SERVIDOR DE BASE DE DATOS

Componente	Requerimiento Mínimo	Requerimiento disponible	Observaciones
Procesador	4 cores, 2.5 GHz	4 cores, 2.8 GHz (Intel Xeon E5-2650)	Ninguna
Memoria RAM	8 GB	8 GB DDR4	Cumple exactamente el mínimo requerido
Almacenamiento	200 GB SSD	240 GB SSD	Ninguna
Sistema Operativo	Ubuntu 20.04+	Ubuntu 22.04 LTS	Sistema operativo compatible

Backup	Backup diario manual	Soportado	Ninguna

Evaluación: ☑ Cumple totalmente □ Cumple parcialmente □ No cumple

Brecha identificada: Ninguna. El servidor de base de datos cumple todos los requisitos.

Inversión estimada para cerrar brecha: Gs. 0

A.3. INFRAESTRUCTURA DE RED

Componente	Requerimiento Mínimo	Requerimiento disponible	Observaciones
Ancho de banda Internet	20 Mbps simétricos	25 Mbps simétricos	Conexión fibra óptica institucional
Red interna	100 Mbps	1 Gbps	Ninguna
Firewall	Firewall básico configurado	pfSense 2.7 configurado	Ninguna
Certificado SSL	Certificado SSL válido	Certificado SSL	Ninguna
Dominio	Subdominio institucional	disponibles	Ninguna
IP pública estática	Requerida	asignada	Ninguna

Evaluación: ☑ Cumple totalmente □ Cumple parcialmente □ No cumple

Brecha identificada: Ninguna. La infraestructura de red cumple todos los requisitos. **Inversión estimada para cerrar brecha:** Gs. 0

B. SOFTWARE Y LICENCIAS

B.1. SOFTWARE DE SISTEMA

Software	Versión Mínima	Versión Disponible	Licencia	Costo	Disponible
Sistema Operativo Servidor	Ubuntu 20.04 / Windows Server 2019	Ubuntu 22.04 LTS	Open Source	Gs. 0	☑ Sí □ No
Python	3.9	Python 3.10.12	Open Source	Gs. 0	☑ Sí □ No
PostgreSQL	13	PostgreSQL 14.9	Open Source	Gs. 0	☑ Sí □ No
Nginx	1.18	Nginx 1.22.1	Open Source	Gs. 0	☑ Sí □ No
Docker	20.10	Docker 24.0.5	Open Source	Gs. 0	☑ Sí □ No

Git	2.30	Git 2.40.1	Open Source	Gs. 0	☑ Sí □ No

Costo total estimado en licencias: Gs. 0 (Todo software open source)

B.2. LIBRERÍAS Y FRAMEWORKS (Todas Open Source - Costo Gs. 0)

Librería/Framework	Versión Mínima	Versión Instalada	Propósito	Disponible
Flask	2.3	2.3.3	Framework web backend	☑ Sí □ No
SQLAlchemy	2.0	2.0.21	ORM para base de datos	☑ Sí □ No
scikit-learn	1.3	1.3.1	Machine learning	☑ Sí □ No
pandas	2.0	2.1.1	Manipulación de datos	☑ Sí □ No
React.js	18.2	18.2.0	Framework frontend	☑ Sí □ No
Material-UI	5.14	5.14.10	Componentes UI	☑ Sí □ No
Recharts	2.7	2.8.0	Visualización de datos	☑ Sí □ No

C. RECURSOS HUMANOS TÉCNICOS

C.1. EQUIPO DE SOPORTE Y OPERACIÓN

Rol	Dedicación Requerida	Competencias Necesarias	Disponible	Personal Identificado	Brecha
Administrador de Sistemas	25% (10 hrs/sem)	Linux, Docker, Nginx, seguridad	☑ Sí □ No	Requiere capacitación específica del sistema	Capacitación 16 hrs
Soporte Técnico	50% (20 hrs/sem)	Atención a usuarios, troubleshooting básico	☑ Sí □ No	Requiere capacitación específica del sistema	Capacitación 16 hrs

Evaluación: ☑ RRHH disponibles □ Requiere contratación ☑ Requiere capacitación **Costo estimado (capacitación según Anexo 5):** Gs. 0 (capacitación interna)

D. SEGURIDAD Y CUMPLIMIENTO

Aspecto	Requerimiento	Criticidad	Implementado en Prototipo	Observaciones
Autenticación	JWT o similar	Alta	☑ Sí □ No □ Parcial	JWT implementado y funcional

Autorización	RBAC (Control basado en roles)	Alta	☑ Sí □ No □ Parcial	Dos roles implementados: Soporte Técnico y Técnico Académico
Cifrado en tránsito	HTTPS con TLS 1.2+	Alta	☑ Sí □ No □ Parcial	Certificado SSL configurado (TLS 1.3)
Cifrado en reposo	Datos sensibles cifrados en BD	Media	☑ Sí □ No □ Parcial	Contraseñas hasheadas, datos académicos en BD con acceso restringido
Gestión de contraseñas	Hashing con bcrypt/argon2	Alta	☑ Sí □ No □ Parcial	Bcrypt implementado para hash de contraseñas
Protección SQL Injection	ORM y prepared statements	Alta	☑ Sí □ No □ Parcial	SQLAlchemy ORM utilizado en todas las consultas
Protección XSS	Sanitización de inputs	Alta	☑ Sí □ No □ Parcial	Validación de inputs en frontend y backend
Protección CSRF	Tokens CSRF	Media	☑ Sí □ No □ Parcial	Tokens CSRF implementados en formularios
Auditoría de accesos	Logs de acceso a datos sensibles	Media	☑ Sí □ No □ Parcial	Sistema de logging implementado con registro de accesos
Backup de datos	Backups automáticos diarios	Alta	☑ Sí □ No □ Parcial	Script cron configurado para backup diario automático

Nivel general de seguridad: ☑ Alto □ Medio □ Bajo

Aspectos críticos pendientes: Ninguno. Todos los aspectos de seguridad críticos están implementados.

G. RESUMEN EJECUTIVO DE VIABILIDAD TÉCNICA G.1. EVALUACIÓN POR CATEGORÍA

Categoría	Cumplimiento	Criticidad de Brechas	Inversión Requerida
Hardware	☑ Alta □ Media □ Baja	□ Alta □ Media ☑ Baja	Gs. 0
Software	☑ Alta □ Media □ Baja	□ Alta □ Media ☑ Baja	Gs. 0
RRHH Técnicos	☑ Alta □ Media □ Baja	□ Alta □ Media ☑ Baja	Gs. 0
Seguridad	☑ Alta □ Media □ Baja	□ Alta □ Media ☑ Baja	Gs. 0

INVERSIÓN TOTAL ESTIMADA: Gs. 0

G.2. CONCLUSIÓN DE VIABILIDAD TÉCNICA

El proyecto es técnicamente viable:

☑ **Sí, con infraestructura actual**

□ Sí, con inversiones menores (< Gs. 30.000.000)

□ Sí, con inversiones moderadas (Gs. 30.000.000 - Gs. 100.000.000)

□ Sí, con inversiones significativas (> Gs. 100.000.000)

□ No es viable en condiciones actuales

La evaluación exhaustiva de los requerimientos técnicos del proyecto demuestra que la Facultad de Ciencias y Tecnologías de la Universidad Nacional de Caaguazú cuenta con la infraestructura de hardware necesaria para el despliegue del sistema, cumpliendo con los requisitos mínimos de procesamiento, memoria y almacenamiento en sus servidores existentes. La infraestructura de red institucional proporciona el ancho de banda, seguridad perimetral y conectividad requeridos, incluyendo la disponibilidad de dominio institucional, certificado SSL y dirección IP estática. Todo el software necesario para la operación del sistema es de código abierto y se encuentra instalado en las versiones requeridas, eliminando completamente los costos de licenciamiento. Los recursos humanos técnicos están disponibles en la estructura actual de la Dirección de Tecnologías de la Información de la FCyT, requiriendo únicamente capacitación específica en el sistema mediante los planes estructurados del Anexo 5. Las integraciones con sistemas existentes son de baja complejidad, basándose principalmente en exportación de archivos CSV desde el sistema de gestión académica y uso del servidor SMTP institucional. Todos los aspectos críticos de seguridad informática han sido implementados en el prototipo funcional, incluyendo autenticación JWT, control de acceso basado en roles, cifrado de comunicaciones mediante HTTPS/TLS 1.3, protección contra vulnerabilidades comunes (SQL Injection, XSS, CSRF), y sistema automatizado de backups. Adicionalmente, se corroboró que el modelo de inteligencia artificial, una vez implementado en la plataforma web, no sufre modificaciones significativas en cuanto a sus métricas de desempeño, manteniendo valores prácticamente idénticos a los reportados en el modelo original (diferencias menores al 0.2% en todas las métricas clave). Al cumplirse todos los requisitos técnicos identificados sin necesidad de inversión adicional en infraestructura, software o personal, y al confirmarse que el modelo de machine learning mantiene su capacidad predictiva tras la integración, se concluye que el proyecto es plenamente viable técnicamente, pudiendo ser

desplegado en producción con la infraestructura y recursos actualmente disponibles en la institución.

3.3.2. Viabilidad operativa

Los siguientes planes de capacitación fueron diseñados a partir de los resultados obtenidos mediante la aplicación del Anexo 4: Checklist de Identificación de Necesidades de Capacitación por Rol. Este instrumento permitió identificar las brechas de competencias específicas para cada perfil de usuario del sistema, estableciendo así los contenidos, metodologías y duración apropiados para cada programa formativo.

Los planes se estructuran en módulos diferenciados según el rol del usuario (Soporte Técnico y Técnico Académico), abordando las competencias técnicas, operativas y de dominio requeridas para el uso efectivo del sistema. Cada módulo especifica sus objetivos de aprendizaje, estrategias de enseñanza, modalidad de implementación y carga horaria, garantizando una formación integral y práctica que habilite a los usuarios para desempeñar sus funciones con autonomía y eficiencia.

3.3.2.1. PLAN DE CAPACITACIÓN 1: PERFIL SOPORTE TÉCNICO

Perfil objetivo: Personal de soporte técnico e informática
Número de participantes: 1-3 personas
Duración total: 16 horas
Modalidad general: Presencial con práctica en laboratorio
Prerrequisitos: Conocimientos de sistemas operativos, bases de datos y redes

Estructura Modular

#	Módulo	Duración	Modalidad	Estrategia de Enseñanza
1	Arquitectura y Componentes del Sistema	2 horas	Presencial	Presentación teórica con diagramas de arquitectura + demostración en vivo del sistema funcionando + sesión de preguntas y respuestas
2	Administración de Usuarios y Permisos	3 horas	Presencial práctica	Demostración paso a paso por el instructor + práctica guiada con supervisión + ejercicios individuales con casos específicos + retroalimentación inmediata

3	Carga y Actualización de Datos Estudiantiles	4 horas	Presencial práctica	Presentación de procedimientos + demostración con datos reales + práctica supervisada con datasets de prueba + resolución de errores en tiempo real + validación de resultados
4	Monitoreo del Sistema	2 horas	Presencial	Demostración del dashboard de monitoreo + análisis de casos reales de logs y métricas + interpretación grupal de anomalías simuladas + discusión de procedimientos de respuesta
5	Troubleshooting y Resolución de Problemas	3 horas	Presencial práctica	Presentación de problemas comunes + resolución práctica en grupos de 2-3 personas + simulación de incidencias reales + documentación estructurada de cada caso + plenaria de soluciones
6	Backups y Recuperación	2 horas	Presencial práctica	Demostración de procedimientos de backup + práctica supervisada de backup manual + simulación de escenario de recuperación + verificación de integridad de backups

3.3.2.2. PLAN DE CAPACITACIÓN 2: PERFIL TÉCNICO ACADÉMICO

Perfil objetivo: Coordinadores de carrera, docentes tutores, autoridades académicas

Número de participantes: 1-5 personas por grupo

Duración total: 13 horas

Modalidad general: Presencial con práctica guiada

Prerrequisitos: Conocimientos básicos de ofimática y navegación web

Estructura Modular

#	Módulo	Duración	Modalidad	Estrategia de Enseñanza
1	Introducción al Sistema y Conceptos Básicos	2 horas	Presencial	Presentación interactiva sobre deserción estudiantil y sistemas de alerta temprana + demostración guiada del sistema + discusión sobre ética y responsabilidad en el uso de predicciones + sesión de preguntas frecuentes
2	Navegación y Uso del Dashboard Principal	2 horas	Presencial práctica	Demostración de navegación por el instructor + práctica guiada paso a paso + ejercicios de exploración individual con tareas específicas + uso de filtros y búsquedas con casos reales

3	Análisis de Estudiantes en Riesgo	3 horas	Presencial práctica	Demostración de análisis de perfil individual + práctica guiada con 5 casos de estudiantes reales anonimizados + trabajo en parejas para identificar patrones + plenaria de hallazgos y discusión grupal
4	Interpretación de Factores de Riesgo	2 horas	Presencial	Presentación de factores de riesgo académico + análisis grupal de casos con diferentes perfiles + discusión facilitada sobre la importancia relativa de factores + reflexión sobre limitaciones del modelo
5	Generación de Reportes y Análisis de Tendencias	2 horas	Presencial práctica	Demostración de generación de reportes + práctica individual de exportación a PDF y Excel + análisis de tendencias temporales + ejercicios de comparación entre carreras + interpretación de gráficos estadísticos
6	Estrategias de Intervención Basadas en Datos	2 horas	Presencial taller	Taller participativo de diseño de intervenciones + trabajo en grupos pequeños para elaborar planes de acción + presentación de propuestas por grupo + retroalimentación colectiva + cierre con mejores prácticas

La aplicación del Anexo 4: Checklist de Identificación de Necesidades de Capacitación permitió identificar las brechas de competencias específicas para cada perfil de usuario, las cuales son abordadas de manera integral mediante los planes de capacitación estructurados presentados en esta sección. Estos programas formativos, con una duración de 16 horas para el perfil Soporte Técnico y 13 horas para el perfil Técnico Académico, cubren exhaustivamente las competencias técnicas, operativas y de dominio requeridas para el uso efectivo del sistema. Asimismo, los perfiles de usuario definidos en el Anexo 3: Matriz de Perfiles de Usuario especifican las competencias, experiencia y responsabilidades necesarias para cada rol, las cuales pueden ser satisfechas mediante la contratación de personal con las calificaciones base requeridas o mediante la capacitación del personal existente utilizando los planes propuestos. Dado que las necesidades de capacitación identificadas pueden ser cubiertas mediante programas formativos estructurados, de duración razonable y con metodologías prácticas apropiadas, y considerando que los perfiles de usuario requeridos corresponden a roles existentes en la estructura organizacional de la institución, se concluye que el proyecto es viable operativamente, ya que la institución cuenta con la capacidad de formar y habilitar al personal necesario para la operación exitosa del sistema.

3.3.3. Viabilidad económica

Se evaluó la viabilidad económica del proyecto mediante un análisis de Retorno de Inversión (ROI) que comparó los costos totales del desarrollo e implementación del sistema con los beneficios

económicos proyectados derivados de la reducción de deserción estudiantil. Los costos totales se calcularon sumando: (a) costos de desarrollo incurridos (tiempo de desarrollo valorizado, hardware, software), (b) costos proyectados de implementación (infraestructura de servidores, licencias, instalación, capacitación inicial), y (c) costos operativos anuales multiplicados por tres años (personal de soporte, mantenimiento de infraestructura, actualizaciones).

Los beneficios económicos se estimaron basándose en la reducción proyectada de deserción estudiantil y la valorización de estudiantes retenidos. El modelo asumió que el sistema identifica correctamente al 80% de los estudiantes que desertan (basado en el Recall de 82.5% del modelo implementado) y que las intervenciones tempranas logran retener al 50% de los estudiantes identificados, resultando en una reducción neta del 40% de la deserción total.

El análisis de viabilidad económica consideró un horizonte temporal de 3 años para evaluar el retorno de la inversión. Los costos totales del proyecto ascendieron a Gs. 87.600.000, distribuidos en costos de desarrollo (Gs. 48.000.000), costos de implementación (Gs. 12.000.000) y costos operativos acumulados de tres años (Gs. 27.600.000). Por su parte, los beneficios económicos proyectados por la retención de estudiantes que de otro modo habrían desertado alcanzaron Gs. 432.000.000 en el mismo período de tres años, considerando una reducción del 40% en la tasa de deserción actual y valorizando cada estudiante retenido en función del arancel anual promedio de Gs. 3.600.000. Todo esto se puede observar en la Tabla 3.

Tabla 3. Resumen de Viabilidad Económica

Concepto	Valor (Gs.)	Período
Costos		
Costos De Desarrollo	48.000.000	Año 0
Costos De Implementación	12.000.000	Año 0
Costos Operativos Anuales	9.200.000	Por año
Costos Operativos Totales (3 Años)	27.600.000	Años 1-3
COSTOS TOTALES (3 Años)	**87.600.000**	Años 0-3
Beneficios		
Beneficio Anual Por Retención	144.000.000	Por año

Beneficios Totales (3 Años)	**432.000.000**	Años 1-3
Indicadores Financieros		
Beneficio Neto (3 Años)	344.400.000	Años 0-3
ROI (Retorno De Inversión)	**393.2%**	3 años
Beneficio Neto Anual	134.800.000	Por año
Período De Recuperación	**0.45 años**	**5.4 meses**

El análisis económico demuestra que el proyecto presenta una viabilidad económica excepcional. Con un Retorno de Inversión (ROI) del 393.2% en un horizonte de tres años, el sistema genera beneficios económicos que superan en casi cuatro veces la inversión total realizada. El período de recuperación de la inversión es de apenas 5.4 meses (0.45 años), lo que significa que antes de cumplirse el primer semestre de operación, el sistema ya habrá generado beneficios suficientes para cubrir todos los costos de desarrollo e implementación. A partir de ese momento, cada estudiante adicional retenido representa un beneficio neto directo para la institución. Los beneficios netos acumulados en tres años alcanzan Gs. 344.400.000, equivalentes a más de 95 estudiantes retenidos que completan sus estudios y generan ingresos por aranceles que de otro modo se habrían perdido. Estos resultados son conservadores, ya que no consideran beneficios adicionales como el prestigio institucional por mejores tasas de graduación, la optimización en el uso de recursos académicos, ni el impacto social positivo de profesionales graduados que contribuyen al desarrollo regional. En consecuencia, se concluye que el proyecto es plenamente viable económicamente, presentando indicadores financieros altamente favorables que justifican ampliamente la inversión requerida y garantizan su sostenibilidad operativa en el mediano y largo plazo.

4. Conclusiones y recomendaciones

4.1. Conclusiones

En conjunto, los resultados obtenidos confirman el cumplimiento exitoso de los objetivos del proyecto. Se ha materializado el desarrollo de una plataforma web completamente funcional que no solo integra el modelo predictivo de deserción estudiantil, sino que también lo traduce en una herramienta de gestión visual e intuitiva. Esta plataforma dota al personal académico y administrativo de la capacidad para interpretar datos complejos y tomar decisiones informadas de manera proactiva.

Asimismo, la validación del sistema en un entorno de producción ha demostrado que el modelo de machine learning subyacente mantiene su alta capacidad predictiva, alcanzando un 82.5% de Recall. Este hallazgo es crucial, ya que confirma que la herramienta es fiable y efectiva para identificar con precisión a la gran mayoría de los estudiantes en riesgo real de abandonar sus estudios.

Finalmente, el análisis de viabilidad concluye de manera inequívoca que el proyecto es sostenible y beneficioso. Desde la perspectiva técnica, se ha demostrado que puede ser implementado y mantenido con los recursos humanos e infraestructurales existentes en la institución. Desde el punto de vista económico, el proyecto no solo es viable, sino que representa una inversión altamente rentable, con un Retorno de la Inversión (ROI) proyectado del 393.2% en tres años, justificando plenamente su implementación a escala institucional.

4.2. Recomendaciones

1. Fortalecer los programas de alerta y detección temprana de deserción a fin de realizar las intervenciones oportunas y mejorar el acompañamiento institucional.
2. Ampliar las tutorías personalizadas de modo que los estudiantes en riesgo reciban el apoyo continuo y estrategias de mejora adaptadas a sus necesidades.
3. Capacitar a los docentes y coordinadores académicos en el uso e interpretación de los modelos de alerta temprana.
4. Para investigaciones futuras considerar incorporar factores individuales tal como:
    - a Características demográficas
    - b Edad
    - c Genero
    - d Variables psicológicas

.

5. Bibliografía

[1] V. . Tinto, «Dropout from higher education: A theoretical synthesis of recent research,» vol. 45, nº 1, pp. 89-125, 1975.

[2] V. . Tinto, Leaving college: Rethinking the causes and cures of student attrition, University of Chicago press, 2012.

[3] N. González y J. V. Bogado Machuca, «Early detection of university dropout with cluster analysis and machine learning classification techniques,» *Ibero-Latin American Congress on Computational Methods in Engineering (CILAMCE),* 2024.

[4] «College persistence: Structural equations modeling test of an integrated model of student retention».

[5] «Análisis de los factores asociados a la deserción estudiantil en la Educación Superior: un estudio de caso Analysis of the Factors Associated with the Drop-out Rate of Students in Higher Education: a Case Study,» vol. 345, pp. 255-280, 2008.

[6] «Educational data mining and learning analytics: An updated survey,» vol. 10, nº 3, p. e1355, 2020.

[7] «Predicting student dropout in higher education,» 2016.

[8] «Next-term student performance prediction: A recommender systems approach,» 2016.

[9] «Early alert of academically at-risk students: An open source analytics initiative,» vol. 1, nº 1, pp. 6-47, 2014.

[10] *A machine learning framework to identify students at risk of adverse academic outcomes,* 2015, pp. 1909-1918.

[11] *Course signals at Purdue: Using learning analytics to increase student success,* 2012, pp. 267-270.

[12] «Penetrating the fog: analytics in learning and education,» vol. 22, nº 3, pp. 132-137, 2014.

[13] R. . Ferguson, «Learning analytics: drivers, developments and challenges,» vol. 4, nº 5-6, pp. 304-317, 2012.

[14] M. . Fowler, Patterns of enterprise application architecture, Addison-Wesley, 2012.

[15] «Transforming our world: implementing the 2030 agenda through sustainable development goal indicators,» vol. 37, nº Suppl 1, pp. 13-31, 2016.

[16] «A conceptual model of nontraditional undergraduate student attrition,» vol. 55, nº 4, pp. 485-540, 1985.

[17] «Mining LMS data to develop an "early warning system" for educators: A proof of concept».

[18] «An e-learning theoretical framework,» nº 1, pp. 292-307, 2016.

[19] «Educational data mining and learning analytics,» pp. 379-396, 2016.

[20] «PREDICTING STUDENTS'PERFORMANCE IN DISTANCE LEARNING USING MACHINE LEARNING TECHNIQUES».

[21] I. . Sommerville, «Software engineering (ed.),» 2011.

[22] S. . Few, Information dashboard design: Displaying data for at-a-glance monitoring, vol. 5, Analytics Press Burlingame, 2013.

[23] *An evaluation of policy frameworks for addressing ethical considerations in learning analytics,* 2013, pp. 240-244.

[24] The visual display of quantitative information, vol. 2, Graphics press Cheshire, CT, 1983. [25] M. . Sachs-Israel, «The SDG 4-education 2030 agenda and its framework for action–the process of its development and first steps in taking it forward,» vol. 69, nº 3, pp. 269-290, 2016.

6. ANEXOS - Instrumentos de recolección de datos

6.1. ANEXO 1: Planilla de Requerimientos Técnicos Mínimos

UNIVERSIDAD NACIONAL DE CAAGUAZÚ

FACULTAD DE CIENCIAS Y TECNOLOGÍAS

Proyecto: Dashboard Web para Predicción de Deserción Estudiantil

PLANILLA DE EVALUACIÓN DE REQUERIMIENTOS TÉCNICOS

Fecha de evaluación: _____

Evaluador: _____

A. INFRAESTRUCTURA DE HARDWARE

A.1. SERVIDOR DE APLICACIONES

Componente	Requerimiento Mínimo	Requerimiento disponible	Observaciones
Procesador	4 cores, 2.5 GHz		
Memoria RAM	8 GB		
Almacenamiento	100 GB SSD		
Sistema Operativo	Ubuntu 20.04+ / Windows Server 2019+		
Virtualización	Soporte para Docker		

Evaluación: □ Cumple totalmente □ Cumple parcialmente □ No cumple

Brecha identificada: _____

Inversión estimada para cerrar brecha: Gs._____

A.2. SERVIDOR DE BASE DE DATOS

Componente	Requerimiento Mínimo	Requerimiento disponible	Observaciones
Procesador	4 cores, 2.5 GHz		

Memoria RAM	8 GB		
Almacenamiento	200 GB SSD		
Sistema Operativo	Ubuntu 20.04+		
Backup	Backup diario manual		

Evaluación: □ Cumple totalmente □ Cumple parcialmente □ No cumple

Brecha identificada: _____

Inversión estimada para cerrar brecha: Gs._____

A.3. INFRAESTRUCTURA DE RED

Componente	Requerimiento Mínimo	Requerimiento Recomendado	Observaciones
Ancho de banda Internet	20 Mbps simétricos		
Red interna	100 Mbps		
Firewall	Firewall básico configurado		
Certificado SSL	Certificado SSL válido		
Dominio	Subdominio institucional		
IP pública estática	Requerida		

Evaluación: □ Cumple totalmente □ Cumple parcialmente □ No cumple

Brecha identificada: _____

Inversión estimada para cerrar brecha: Gs._____

B. SOFTWARE Y LICENCIAS

B.1. SOFTWARE DE SISTEMA

Software	Versión Mínima	Licencia	Costo	Disponible
Sistema Operativo Servidor	Ubuntu 20.04 / Windows Server 2019			□ Sí □ No
Python	3.9			□ Sí □ No
PostgreSQL	13			□ Sí □ No
Nginx	1.18			□ Sí □ No
Docker	20.10			□ Sí □ No
Git	2.30			□ Sí □ No

Costo total estimado en licencias: Gs._____

B.2. LIBRERÍAS Y FRAMEWORKS (Todas Open Source - Costo Gs. 0)

Librería/Framework	Versión Mínima	Propósito	Disponible
Flask	2.3	Framework web backend	□ Sí □ No
SQLAlchemy	2.0	ORM para base de datos	□ Sí □ No
scikit-learn	1.3	Machine learning	□ Sí □ No
pandas	2.0	Manipulación de datos	□ Sí □ No
React.js	18.2	Framework frontend	□ Sí □ No
Material-UI	5.14	Componentes UI	□ Sí □ No
Recharts	2.7	Visualización de datos	□ Sí □ No

C. RECURSOS HUMANOS TÉCNICOS

C.1. EQUIPO DE SOPORTE Y OPERACIÓN

Rol	Dedicación Requerida	Competencias Necesarias	Disponible	Brecha

Administrador de Sistemas	25% (10 hrs/sem)	Linux, Docker, Nginx, seguridad	□ Sí □ No	
Soporte Técnico	50% (20 hrs/sem)	Atención a usuarios, troubleshooting básico	□ Sí □ No	

Evaluación: □ RRHH disponibles □ Requiere contratación □ Requiere capacitación

Costo estimado (contratación o capacitación): Gs._____

D. SEGURIDAD Y CUMPLIMIENTO

Aspecto	Requerimiento	Criticidad	Implementado en Prototipo	Observaciones
Autenticación	JWT o similar	Alta	□ Sí □ No □ Parcial	
Autorización	RBAC (Control basado en roles)	Alta	□ Sí □ No □ Parcial	
Cifrado en tránsito	HTTPS con TLS 1.2+	Alta	□ Sí □ No □ Parcial	
Cifrado en reposo	Datos sensibles cifrados en BD	Media	□ Sí □ No □ Parcial	
Gestión de contraseñas	Hashing con bcrypt/argon2	Alta	□ Sí □ No □ Parcial	
Protección SQL Injection	ORM y prepared statements	Alta	□ Sí □ No □ Parcial	

Protección XSS	Sanitización de inputs	Alta	□ Sí □ No □ Parcial	
Protección CSRF	Tokens CSRF	Media	□ Sí □ No □ Parcial	
Auditoría de accesos	Logs de acceso a datos sensibles	Media	□ Sí □ No □ Parcial	
Backup de datos	Backups automáticos diarios	Alta	□ Sí □ No □ Parcial	

Nivel general de seguridad: □ Alto □ Medio □ Bajo

Aspectos críticos pendientes: _____

G. RESUMEN EJECUTIVO DE VIABILIDAD TÉCNICA

G.1. EVALUACIÓN POR CATEGORÍA

Categoría	Cumplimiento	Criticidad de Brechas	Inversión Requerida
Hardware	□ Alta □ Media □ Baja	□ Alta □ Media □ Baja	Gs._____
Software	□ Alta □ Media □ Baja	□ Alta □ Media □ Baja	Gs._____
RRHH Técnicos	□ Alta □ Media □ Baja	□ Alta □ Media □ Baja	Gs._____
Integración	□ Alta □ Media □ Baja	□ Alta □ Media □ Baja	Gs._____
Seguridad	□ Alta □ Media □ Baja	□ Alta □ Media □ Baja	Gs._____

INVERSIÓN TOTAL ESTIMADA: Gs._____

G.2. CONCLUSIÓN DE VIABILIDAD TÉCNICA **El proyecto es técnicamente viable:**

□ Sí, con infraestructura actual

□ Sí, con inversiones menores (< Gs. 30.000.000)

□ Sí, con inversiones moderadas (Gs. 30.000.000- Gs. 100.000.000)

□ Sí, con inversiones significativas (> Gs. 100.000.000)

□ No es viable en condiciones actuales

Evaluador: _____ Firma: _____ Fecha: _____

6.2. ANEXO 2: Matriz de Perfiles de Usuario

UNIVERSIDAD NACIONAL DE CAAGUAZÚ

FACULTAD DE CIENCIAS Y TECNOLOGÍAS

Proyecto: Dashboard Web para Predicción de Deserción Estudiantil

MATRIZ DE PERFILES DE USUARIO DEL SISTEMA

PERFIL 1: SOPORTE TÉCNICO

Descripción del Rol:

Personal técnico-administrativo responsable de la operación, mantenimiento y soporte del sistema de predicción de deserción.

A. CARACTERÍSTICAS DEL PERFIL

Dimensión	Descripción
Formación académica requerida	Licenciatura en Informática, Sistemas, Ingeniería en Informática o afines
Experiencia mínima	2 años en soporte técnico o administración de sistemas
Conocimientos técnicos requeridos	- Sistemas operativos (Linux/Windows Server) - Bases de datos (SQL básico) - Redes y seguridad informática - Troubleshooting de aplicaciones web
Competencias digitales	Nivel avanzado (8-10/10)
Dedicación estimada	50% de jornada (20 horas/semana)

B. RESPONSABILIDADES EN EL SISTEMA

#	Responsabilidad	Frecuencia
1	Monitorear estado y disponibilidad del sistema	Diaria
2	Gestionar usuarios y permisos de acceso	Semanal
3	Cargar/actualizar datos estudiantiles en el sistema	Semestral
4	Realizar backups de la base de datos	Diaria (automatizada)
5	Atender consultas y problemas técnicos de usuarios	Según demanda
6	Generar reportes técnicos de uso del sistema	Mensual
7	Coordinar actualizaciones y mantenimientos	Según necesidad
8	Documentar incidencias y soluciones	Según demanda

C. FUNCIONALIDADES DEL SISTEMA QUE UTILIZA

Funcionalidad	Nivel de Uso	Criticidad
Panel de administración de usuarios	Alto	Alta
Gestión de roles y permisos	Medio	Alta
Módulo de carga de datos	Alto	Alta
Dashboard de monitoreo del sistema	Alto	Media
Generación de reportes técnicos	Medio	Media
Visualización de logs de sistema	Medio	Media
Configuración de parámetros del sistema	Bajo	Alta

D. COMPETENCIAS DIGITALES REQUERIDAS

Competencia	Nivel Requerido	Evaluación (1-10)
Uso de sistemas operativos (Linux/Windows)	Avanzado	8-10
Gestión de bases de datos	Intermedio	6-8
Navegación y uso de aplicaciones web	Avanzado	8-10
Resolución de problemas técnicos	Avanzado	8-10
Interpretación de logs y mensajes de error	Avanzado	7-9
Gestión de usuarios y permisos	Intermedio	6-8
Comprensión de conceptos de seguridad informática	Intermedio	6-8

PERFIL 2: TÉCNICO ACADÉMICO

Descripción del Rol:

Personal académico (coordinadores de carrera, docentes tutores, autoridades) que utiliza el sistema para identificar estudiantes en riesgo y tomar decisiones sobre intervenciones académicas.

A. CARACTERÍSTICAS DEL PERFIL

Dimensión	Descripción
Formación académica requerida	Título de grado en área relevante + experiencia docente
Experiencia mínima	3 años como docente, preferiblemente con experiencia en tutoría o coordinación
Conocimientos técnicos requeridos	- Uso de navegadores web - Manejo de aplicaciones ofimáticas - Interpretación de gráficos y estadísticas básicas
Competencias digitales	Nivel intermedio (5-7/10)
Dedicación estimada	10-15% de jornada (4-6 horas/semana)

B. RESPONSABILIDADES EN EL SISTEMA

#	Responsabilidad	Frecuencia

1	Consultar lista de estudiantes en riesgo de su carrera	Semanal
2	Revisar perfiles individuales de estudiantes en riesgo	Semanal
3	Identificar factores de riesgo específicos por estudiante	Según necesidad
4	Generar reportes académicos de estudiantes en riesgo	Mensual
5	Analizar tendencias de deserción en su carrera	Semestral
6	Comparar indicadores entre períodos académicos	Semestral
7	Exportar datos para análisis adicionales	Según necesidad
8	Planificar intervenciones basadas en información del sistema	Continua

C. FUNCIONALIDADES DEL SISTEMA QUE UTILIZA

Funcionalidad	Nivel de Uso	Criticidad
Dashboard principal con indicadores de su carrera	Alto	Alta
Lista de estudiantes en riesgo (filtrable y ordenable)	Alto	Alta
Vista detallada de estudiante individual	Alto	Alta
Visualización de factores de riesgo por estudiante	Alto	Alta
Gráficos de evolución de rendimiento académico	Medio	Media
Comparativas entre carreras	Bajo	Baja
Generación de reportes académicos	Medio	Media
Exportación de datos (PDF, Excel)	Medio	Media

D. COMPETENCIAS DIGITALES REQUERIDAS

Competencia	Nivel Requerido	Evaluación (1-10)
Navegación en aplicaciones web	Intermedio	5-7
Interpretación de gráficos estadísticos	Intermedio	5-7
Uso de filtros y búsquedas	Básico-Intermedio	4-6

Comprensión de indicadores numéricos	Intermedio	5-7
Exportación y manejo de archivos (PDF, Excel)	Básico-Intermedio	4-6
Comprensión de conceptos de probabilidad	Básico	3-5
Uso de aplicaciones ofimáticas	Intermedio	5-7

COMPARACIÓN DE PERFILES

Dimensión	Soporte Técnico	Técnico Académico
Competencias digitales requeridas	Avanzado (8-10/10)	Intermedio (5-7/10)
Enfoque principal	Operación y mantenimiento del sistema	Uso de información para decisiones académicas
Dedicación estimada	50% jornada (20 hrs/sem)	10-15% jornada (4-6 hrs/sem)
Capacitación requerida	16 horas (énfasis técnico)	13 horas (énfasis funcional)
Criticidad del rol	Alta (sistema no funciona sin este rol)	Alta (sistema no cumple objetivo sin este rol)

6.3. ANEXO 4: Necesidades de Capacitación por Rol

FACULTAD DE CIENCIAS Y TECNOLOGÍAS

Proyecto: Dashboard Web para Predicción de Deserción Estudiantil

CHECKLIST DE IDENTIFICACIÓN DE NECESIDADES DE CAPACITACIÓN

Fecha de evaluación: _____

Evaluador: _____

INSTRUCCIONES

Este instrumento permite identificar las necesidades de capacitación específicas para cada perfil de usuario del sistema. Para cada competencia:

1. Evalúe el **nivel actual** del usuario/equipo (escala 1-5)
2. Compare con el **nivel requerido** por el sistema
3. Identifique la **brecha** (diferencia entre requerido y actual)
4. Marque si requiere capacitación (brecha ≥ 2 puntos = capacitación prioritaria) **Escala de evaluación:**

- **1:** Sin conocimiento/experiencia
- **2:** Conocimiento básico/teórico
- **3:** Conocimiento intermedio/práctica limitada
- **4:** Conocimiento avanzado/práctica regular
- **5:** Experto/dominio completo

PERFIL 1: SOPORTE TÉCNICO **Datos del evaluado:**

- Nombre: _____
- Cargo actual: _____
- Años de experiencia en área técnica: _____

A. COMPETENCIAS TÉCNICAS GENERALES

#	Competencia	Nivel Requerido	Nivel Actual	Brecha	¿Requiere Capacitación?
1	Administración de sistemas operativos Linux	4			□ Sí □ No
2	Administración de sistemas operativos Windows Server	3			□ Sí □ No
3	Gestión de bases de datos relacionales (SQL)	4			□ Sí □ No
4	Administración de PostgreSQL	4			□ Sí □ No
5	Conceptos de redes y protocolos TCP/IP	3			□ Sí □ No
6	Configuración de firewalls y seguridad de red	3			□ Sí □ No
7	Gestión de certificados SSL/TLS	3			□ Sí □ No
8	Uso de Docker y contenedores	3			□ Sí □ No
9	Conceptos de APIs REST	3			□ Sí □ No

10	Interpretación de logs de sistema	4			□ Sí □ No

Subtotal brechas identificadas: _____ competencias requieren capacitación

B. COMPETENCIAS ESPECÍFICAS DEL SISTEMA

#	Competencia	Nivel Requerido	Nivel Actual	Brecha	¿Requiere Capacitación?
11	Comprensión de arquitectura de aplicaciones web (frontend/backend)	4			□ Sí □ No
12	Conocimiento de Python y Flask	3			□ Sí □ No
13	Conocimiento de React.js	2			□ Sí □ No
14	Conceptos básicos de Machine Learning	2			□ Sí □ No
15	Gestión de usuarios y control de acceso (RBAC)	4			□ Sí □ No
16	Procedimientos de backup y recuperación de datos	5			□ Sí □ No
17	Troubleshooting de aplicaciones web	4			□ Sí □ No
18	Monitoreo de desempeño de sistemas	4			□ Sí □ No
19	Documentación técnica de incidencias	4			□ Sí □ No

20	Atención y soporte a usuarios finales	4			□ Sí □ No

Subtotal brechas identificadas: _____ competencias requieren capacitación

---

C. CONOCIMIENTOS ESPECÍFICOS DEL DOMINIO

#	Conocimiento	Nivel Requerido	Nivel Actual	Brecha	¿Requiere Capacitación?
21	Estructura de datos académicos (estudiantes, materias, calificaciones)	3			□ Sí □ No
22	Procesos académicos universitarios	2			□ Sí □ No
23	Normativa de protección de datos personales	3			□ Sí □ No
24	Conceptos de deserción estudiantil	2			□ Sí □ No
25	Interpretación de métricas de predicción (accuracy, recall, etc.)	2			□ Sí □ No

Subtotal brechas identificadas: _____ conocimientos requieren capacitación

---

D. RESUMEN DE NECESIDADES - PERFIL SOPORTE TÉCNICO **Total de brechas identificadas:** _____ de 25 competencias/conocimientos

Prioridad de capacitación:

□ **ALTA:** 10 o más brechas identificadas (≥40% de competencias)

□ **MEDIA:** 5-9 brechas identificadas (20-36% de competencias)

□ **BAJA:** Menos de 5 brechas identificadas (<20% de competencias)

Duración estimada de capacitación requerida: _____ horas

Modalidad recomendada: □ Presencial □ Virtual □ Mixta □ Autoformación con seguimiento

---

PERFIL 2: TÉCNICO ACADÉMICO **Datos del evaluado:**

- Nombre: _____
- Cargo actual: _____
- Carrera/Departamento: _____
- Años de experiencia docente: _____

A. COMPETENCIAS DIGITALES BÁSICAS

#	Competencia	Nivel Requerido	Nivel Actual	Brecha	¿Requiere Capacitación?
1	Uso de navegadores web y aplicaciones online	4			□ Sí □ No
2	Manejo de aplicaciones ofimáticas (Word, Excel, PDF)	3			□ Sí □ No
3	Gestión de archivos y carpetas	3			□ Sí □ No

4	Uso de correo electrónico institucional	3			□ Sí □ No
5	Descarga e instalación de archivos	2			□ Sí □ No

Subtotal brechas identificadas: _____ competencias requieren capacitación

---

B. COMPETENCIAS DE ANÁLISIS DE INFORMACIÓN

#	Competencia	Nivel Requerido	Nivel Actual	Brecha	¿Requiere Capacitación?
6	Interpretación de gráficos estadísticos (barras, líneas, tortas)	4			□ Sí □ No
7	Comprensión de indicadores numéricos (promedios, porcentajes)	4			□ Sí □ No
8	Interpretación de tablas de datos	4			□ Sí □ No
9	Análisis de tendencias temporales	3			□ Sí □ No
10	Comparación de datos entre grupos	3			□ Sí □ No
11	Comprensión de conceptos de probabilidad básica	3			□ Sí □ No

Subtotal brechas identificadas: _____ competencias requieren capacitación

---

C. COMPETENCIAS ESPECÍFICAS DEL SISTEMA

#	Competencia	Nivel Requerido	Nivel Actual	Brecha	¿Requiere Capacitación?
12	Navegación en dashboards interactivos	4			□ Sí □ No
13	Uso de filtros y búsquedas en aplicaciones web	4			□ Sí □ No
14	Interpretación de indicadores de riesgo (semáforos, colores)	4			□ Sí □ No
15	Exportación de reportes (PDF, Excel)	3			□ Sí □ No
16	Comprensión de predicciones probabilísticas	3			□ Sí □ No
17	Identificación de factores de riesgo académico	4			□ Sí □ No
18	Análisis de trayectorias académicas individuales	4			□ Sí □ No

D. CONOCIMIENTOS DEL DOMINIO ACADÉMICO

#	Conocimiento	Nivel Requerido	Nivel Actual	Brecha	¿Requiere Capacitación?
19	Factores asociados a deserción estudiantil	4			□ Sí □ No
20	Estrategias de intervención académica	4			□ Sí □ No

21	Tutoría y acompañamiento estudiantil	4			□ Sí □ No
22	Normativa académica institucional	4			□ Sí □ No
23	Conceptos básicos de inteligencia artificial aplicada a educación	2			□ Sí □ No
24	Limitaciones y alcances de modelos predictivos	3			□ Sí □ No
25	Ética en el uso de datos estudiantiles	4			□ Sí □ No

Subtotal brechas identificadas: _____ conocimientos requieren capacitación

---

E. RESUMEN DE NECESIDADES - PERFIL TÉCNICO ACADÉMICO

Total de brechas identificadas: _____ de 25 competencias/conocimientos

Prioridad de capacitación:

□ **ALTA:** 10 o más brechas identificadas (≥40% de competencias)

□ **MEDIA:** 5-9 brechas identificadas (20-36% de competencias)

□ **BAJA:** Menos de 5 brechas identificadas (<20% de competencias)

Duración estimada de capacitación requerida: _____ horas

Modalidad recomendada: □ Presencial □ Virtual □ Mixta □ Autoformación con seguimiento

6.4. ANEXO 5: Cálculo de retorno de la inversión

DESARROLLO DETALLADO DE CÁLCULOS ECONÓMICOS

A. CÁLCULO DE COSTOS DEL PROYECTO

A.1. Costos de Desarrollo (Año 0)

Tiempo de desarrollo valorizado:

El desarrollo del sistema requirió aproximadamente 6 meses de trabajo de un desarrollador full-stack. Considerando un salario mensual de referencia para un ingeniero en sistemas con experiencia en desarrollo web y machine learning de Gs. 8.000.000:

Costo_desarrollo_personal = 6 meses × 8.000.000 Gs./mes = 48.000.000 Gs.

Hardware y equipamiento:

Todo el hardware necesario (computadora de desarrollo, periféricos) ya estaba disponible en la institución, por lo que no se incurrió en costos adicionales.

Costo_hardware_desarrollo = 0 Gs.

Software y licencias:

Todo el software utilizado (Python, Flask, React, PostgreSQL, herramientas de desarrollo) es de código abierto y gratuito.

Costo_software_desarrollo = 0 Gs.

Total Costos de Desarrollo:

Costos_desarrollo = 48.000.000 + 0 + 0 = 48.000.000 Gs.

A.2. Costos de Implementación (Año 0)

Infraestructura de servidores:

Según el Anexo 1 (Evaluación de Viabilidad Técnica), toda la infraestructura de hardware necesaria está disponible en la institución.

Costo_infraestructura = 0 Gs.

Licencias de software:

Todo el software de sistema y aplicación es open source.

Costo_licencias = 0 Gs.

Instalación y configuración:

Se estima 2 semanas de trabajo (80 horas) para instalación, configuración de servidores, migración de datos y pruebas finales. Considerando una tarifa de Gs. 150.000/hora para trabajo técnico especializado:

Costo_instalación = 80 horas × 150.000 Gs./hora = 12.000.000 Gs.

Capacitación inicial:

Según el Anexo 5 (Planes de Capacitación por Rol), la capacitación será impartida por el equipo de desarrollo sin costo adicional, utilizando instalaciones y equipamiento institucional.

Costo_capacitación = 0 Gs.

Total Costos de Implementación:

Costos_implementación = 0 + 0 + 12.000.000 + 0 = 12.000.000 Gs.

A.3. Costos Operativos Anuales (Años 1-3)

Personal de soporte técnico: Según el Anexo 3, se requiere:

- Administrador de Sistemas: 25% dedicación (10 hrs/semana)
- Soporte Técnico: 50% dedicación (20 hrs/semana)

Considerando salarios mensuales de Gs. 6.000.000 y Gs. 4.000.000 respectivamente:

Costo_admin = 6.000.000 × 0.25 × 12 = 18.000.000 Gs./año

Costo_soporte = 4.000.000 × 0.50 × 12 = 24.000.000 Gs./año

Costo_personal_anual = 18.000.000 + 24.000.000 = 42.000.000 Gs./año

Sin embargo, este personal ya forma parte de la estructura de la FCyT y sus salarios ya están presupuestados. El costo incremental real es solo el tiempo adicional dedicado al sistema, que representa aproximadamente el 20% de su carga laboral total destinada a este proyecto específico:

Costo_personal_incremental = 42.000.000 × 0.20 = 8.400.000 Gs./año

Mantenimiento de infraestructura:

- Electricidad para servidores: Gs. 300.000/mes
- Conectividad (proporción atribuible): Gs. 200.000/mes
- Renovación certificado SSL: Gs. 0 (Let's Encrypt gratuito)

Costo_infraestructura_anual = (300.000 + 200.000) × 12 = 6.000.000 Gs./año

Actualizaciones y mejoras:

Se estima un 10% del costo de desarrollo anual para actualizaciones menores, corrección de bugs y mejoras incrementales:

Costo_actualizaciones = 48.000.000 × 0.10 / 6 = 800.000 Gs./año

Total Costos Operativos Anuales:

Costos_operativos_anuales = 8.400.000 + 600.000 + 800.000 = 9.200.000 Gs./año

Nota: Se utilizó el costo de infraestructura reducido considerando solo el consumo eléctrico adicional de los servidores.

Costos Operativos Totales (3 años):

Costos_operativos_3años = 9.200.000 × 3 = 27.600.000 Gs.

A.4. Costos Totales del Proyecto (3 años)

Costos_totales = Costos_desarrollo + Costos_implementación + Costos_operativos_3años

Costos_totales = 48.000.000 + 12.000.000 + 27.600.000 = 87.600.000 Gs.

B. CÁLCULO DE BENEFICIOS ECONÓMICOS

B.1. Datos Base de Deserción Estudiantil

Según datos históricos de la FCyT-UNCA:

- **Matrícula total promedio:** 1,200 estudiantes
- **Tasa de deserción anual histórica:** 15%
- **Estudiantes que desertan anualmente:** 1,200 × 0.15 = 180 estudiantes/año
- **Arancel anual promedio por estudiante:** Gs. 3.600.000

B.2. Proyección de Reducción de Deserción

Capacidad de detección del sistema:

El modelo de machine learning tiene un Recall (sensibilidad) del 82.5%, lo que significa que identifica correctamente al 82.5% de los estudiantes que realmente van a desertar. Para ser conservadores, se utiliza el 80%:

Estudiantes_detectados = 180 × 0.80 = 144 estudiantes/año

Efectividad de las intervenciones:

Basándose en literatura sobre sistemas de alerta temprana en educación superior (Tinto, 2012; Arnold & Pistilli, 2012), se estima que las intervenciones tempranas (tutoría académica, apoyo psicopedagógico, seguimiento personalizado) logran retener al 50% de los estudiantes identificados:

Estudiantes_retenidos = 144 × 0.50 = 72 estudiantes/año

Reducción neta de deserción:

Reducción_porcentual = (72 / 180) × 100 = 40%

Esto significa que el sistema permite reducir la deserción del 15% actual al 9% (una reducción del 40% relativo).

B.3. Valorización de Beneficios Económicos

Beneficio por estudiante retenido:

Cada estudiante retenido que continúa sus estudios genera ingresos por aranceles. Considerando que un estudiante promedio cursa 4 años, y que al retenerlo en el primer año crítico se asegura su permanencia por los años restantes:

Para simplificar el cálculo y ser conservadores, se considera solo el beneficio del año inmediato:

Beneficio_por_estudiante = 3.600.000 Gs./año

Beneficio anual total:

Beneficio_anual = 72 estudiantes × 3.600.000 = 259.200.000 Gs./año

Sin embargo, para ser más conservadores y considerar que no todos los estudiantes retenidos completan el año completo, aplicamos un factor de ajuste del 90%:

Beneficio_anual_ajustado = 259.200.000 × 0.90 = 233.280.000 Gs./año

Ajuste adicional por efecto gradual:

En el primer año de implementación, el sistema no operará a plena capacidad todo el año. Se estima que el primer año alcanza solo el 60% de efectividad, el segundo año el 100%, y el tercer año el 100%:

Beneficio_año1 = 233.280.000 × 0.60 = 139.968.000 Gs.

Beneficio_año2 = 233.280.000 × 1.00 = 233.280.000 Gs.

Beneficio_año3 = 233.280.000 × 1.00 = 233.280.000 Gs.

Beneficios totales (3 años):

Beneficios_totales = 139.968.000 + 233.280.000 + 233.280.000 = 606.528.000 Gs.

Nota: Para simplificar el análisis y ser más conservadores, utilizaremos un beneficio anual constante de Gs. 144.000.000 (equivalente a 40 estudiantes retenidos × Gs. 3.600.000):

Beneficios_totales_conservador = 144.000.000 × 3 = 432.000.000 Gs.

C. CÁLCULO DE INDICADORES FINANCIEROS

C.1. Beneficio Neto

Beneficio_neto = Beneficios_totales - Costos_totales

Beneficio_neto = 432.000.000 - 87.600.000 = 344.400.000 Gs.

C.2. Retorno de Inversión (ROI)

(Beneficios_totales - Costos_totales)

ROI = ——————————————————————————— × 100

Costos_totales

(432.000.000 - 87.600.000)

ROI = ——————————————————— × 100 87.600.000

344.400.000

ROI = ————————————— × 100 = 3.932 × 100 = 393.2% 87.600.000

Interpretación: Por cada guaraní invertido en el proyecto, se obtienen Gs. 3.93 de beneficio neto, o expresado de otra forma, la inversión se multiplica por 4.93 veces en tres años.

C.3. Período de Recuperación de la Inversión

(Costos_desarrollo + Costos_implementación)

Período_recuperación =

——————————————————————————————————————

(Beneficio_anual - Costos_operativos_anuales)

(48.000.000 + 12.000.000)

Período_recuperación = ——————————————————————————————

(144.000.000 - 9.200.000)

60.000.000

Período_recuperación = ——————————— = 0.445 años 134.800.000

Período_recuperación = 0.445 × 12 = 5.34 meses ≈ 5.4 meses

Interpretación: La inversión inicial (desarrollo + implementación) se recupera completamente en aproximadamente 5 meses y medio de operación del sistema.