



UNIVERSIDAD PRIVADA DE TACNA

FACULTAD DE INGENIERIA

Escuela Profesional de Ingeniería de Sistemas

**Proyecto: Dashboard de Detección Temprana de
Abandono Estudiantil para el área de tutoría de
EPIS UPT**

Curso: Inteligencia De Negocios

Docente: Ing. Cuadros Quiroga, Patrick Jose

Integrantes:

<i>Lopez Catunta , Brayar Christian</i>	<i>(2020068946)</i>
<i>Melendez Huarachi, Gabriel Fari</i>	<i>(2021070311)</i>
<i>Cuadros Garcia, Mirian</i>	<i>(2021071083)</i>
<i>Hurtado Ortiz, Leandro</i>	<i>(2015052384)</i>
<i>Briceño Diaz, Jorge Luis</i>	<i>(2017059611)</i>
<i>Chino Rivera, Angel Alessandro</i>	<i>(2021069830)</i>

**Tacna – Perú
2024**

Dashboard de Detección Temprana de Abandono Estudiantil para el área de tutoría de EPIS UPT

Documento de Visión

Versión {1.0}

CONTROL DE VERSIONES					
Versión	Hecha por	Revisada por	Aprobada por	Fecha	Motivo
1.0	JBD/LHO	JLBCG	JLBCG	31/08/2024	Versión Original
1.1	JBD/LHO	JLBCG	JLBCG	21/09/2024	Versión Original

INDICE GENERAL

1.	Introducción	1
1.1	Propósito	1
1.2	Alcance	1
1.3	Definiciones, Siglas y Abreviaturas	1
1.4	Referencias	1
1.5	Visión General	1
2.	Posicionamiento	1
2.1	Oportunidad de negocio	1
2.2	Definición del problema	2
3.	Descripción de los interesados y usuarios	3
3.1	Resumen de los interesados	3
3.2	Resumen de los usuarios	3
3.3	Entorno de usuario	4
3.4	Perfiles de los interesados	4
3.5	Perfiles de los Usuarios	4
3.6	Necesidades de los interesados y usuarios	6
4.	Vista General del Producto	7
4.1	Perspectiva del producto	7
4.2	Resumen de capacidades	8
4.3	Suposiciones y dependencias	8
4.4	Costos y precios	9
		3

4.5	Licenciamiento e instalación	9
5.	Características del producto	9
6.	Restricciones	10
7.	Rangos de calidad	10
8.	Precedencia y Prioridad	10
9.	Otros requerimientos del producto	un10
	aError! Bookmark not defined.	32
	bError! Bookmark not defined.	37
	cError! Bookmark not defined.	42
	dError! Bookmark not defined.	42
	CONCLUSIONES	46
	RECOMENDACIONES	46
	BIBLIOGRAFÍA	46
	WEBGRAFÍA	46

1. Introducción

En el panorama educativo actual, la retención de estudiantes se ha convertido en un desafío clave para las instituciones académicas. Las universidades buscan no solo atraer estudiantes, sino también garantizar que completen sus estudios con éxito. Uno de los principales obstáculos en este proceso es el abandono estudiantil, que puede deberse a factores académicos, socioeconómicos o personales.

El Dashboard de Detección Temprana de Abandono Estudiantil para la Escuela Profesional de Ingeniería de Sistemas de la Universidad Privada de Tacna (UPT) surge como una solución innovadora para identificar estudiantes en riesgo de abandonar sus estudios. Este sistema combina el análisis de datos y técnicas de aprendizaje automático para ofrecer una predicción temprana del riesgo de abandono, permitiendo a los tutores y administradores tomar medidas preventivas adecuadas.

En este informe, se explora cómo esta tecnología avanzada mejorará la capacidad de la universidad para retener a los estudiantes, proporcionando una herramienta que permite tomar decisiones basadas en datos. Desde su concepción hasta su implementación, el sistema busca transformar el proceso de tutoría, contribuyendo a mejorar la retención y el éxito académico en la universidad.

1.1 Propósito

El propósito de este proyecto es desarrollar un sistema de predicción de abandono estudiantil que permita a la universidad identificar de manera temprana a los estudiantes en riesgo. El sistema mejorará los procesos de tutoría y permitirá la optimización de recursos institucionales mediante la visualización de datos y predicciones.

1.2 Alcance

El alcance de este proyecto incluye el desarrollo, implementación y despliegue de un dashboard interactivo que integre un modelo predictivo basado en datos históricos y actuales de los estudiantes, como rendimiento académico, asistencia y factores socioeconómicos. El dashboard proporcionará a los administradores y tutores una plataforma para visualizar los niveles de riesgo, analizar tendencias y tomar decisiones informadas que mejoren la retención estudiantil.

1.3 Definiciones, Siglas y Abreviaturas

- UI (User Interface): Interfaz de Usuario, el espacio donde interactúan el usuario y el sistema.
- ML (Machine Learning): Aprendizaje Automático, rama de la inteligencia artificial que permite que los sistemas aprendan y mejoren automáticamente a partir de la experiencia sin ser programados explícitamente para ello.
- API (Application Programming Interface): Interfaz de Programación de Aplicaciones, un conjunto de reglas y herramientas que permite que diferentes sistemas de software se comuniquen entre sí.
- DB (Database): Base de Datos, sistema estructurado para almacenar, gestionar y recuperar información.

1.4 Referencias

Documentos Adicionales:

- FD01. (2024). *Informe de Factibilidad*. Documento interno de la Universidad Nacional de Educación Enrique Guzmán y Valle.
 - Descripción: Documento que presenta un análisis detallado de la viabilidad del proyecto, incluyendo costos, beneficios y evaluación de riesgos.

1.5 Visión General

Este documento está estructurado para proporcionar una visión integral del proyecto, comenzando con la definición del propósito y alcance, seguido por una descripción detallada de los interesados y usuarios. A continuación, se presenta una vista general del producto, incluyendo sus características principales, restricciones y criterios de calidad. Finalmente, se abordan aspectos relacionados con la precedencia, prioridades, y otros requerimientos específicos del producto, concluyendo con recomendaciones y referencias adicionales.

2. Posicionamiento

2.1 Oportunidad de negocio

La implementación del Dashboard de Detección Temprana de Abandono Estudiantil para la Universidad Privada de Tacna representa una importante oportunidad tanto para la universidad como para los proveedores de tecnología y servicios asociados. La universidad se beneficiará al optimizar su sistema de tutoría, identificando a los estudiantes en riesgo de abandono de manera temprana, lo que permitirá mejorar las tasas

de retención y el éxito académico. Esto contribuirá a fortalecer su reputación institucional, atrayendo a más estudiantes y generando una mayor demanda de sus programas académicos.

Desde el punto de vista de los proveedores tecnológicos, el desarrollo e implementación de este sistema abre oportunidades para ofrecer soluciones avanzadas en análisis de datos y aprendizaje automático (machine learning). Además, el éxito del sistema en la universidad podría generar casos de estudio replicables en otras instituciones educativas, ampliando el mercado de estas tecnologías y su influencia en el sector educativo.

2.2 Definición del problema

El proceso actual de seguimiento y apoyo a los estudiantes en la Universidad Privada de Tacna enfrenta varios desafíos críticos relacionados con el abandono estudiantil:

- **Crecimiento en el número de estudiantes en riesgo:** A medida que aumenta la población estudiantil, se vuelve más difícil identificar y dar seguimiento a aquellos estudiantes que presentan un mayor riesgo de abandonar sus estudios.
- **Evaluación manual:** El proceso manual para identificar a los estudiantes en riesgo es propenso a errores, consume tiempo y es ineficiente, lo que retrasa las intervenciones oportunas por parte de tutores y administradores.
- **Criterios limitados:** Los criterios actuales para evaluar el riesgo de abandono no abarcan todos los factores que influyen en el éxito académico, como aspectos socioeconómicos o patrones de comportamiento.
- **Falta de herramientas tecnológicas:** La ausencia de herramientas avanzadas de análisis de datos y predicción limita la capacidad de la universidad para anticipar el abandono estudiantil y optimizar los recursos de apoyo.

El **Dashboard de Detección Temprana de Abandono Estudiantil** aborda estos problemas mediante el uso de técnicas de análisis de datos y aprendizaje automático, proporcionando una evaluación más precisa y eficiente del riesgo de abandono. Esto permite intervenciones tempranas y mejora las tasas de retención estudiantil.

3. Descripción de los interesados y usuarios

3.1 Resumen de los interesados

Los principales interesados en el Dashboard de Detección Temprana de Abandono Estudiantil son:

- **Tutora de EPIS:** Responsable de apoyar a los estudiantes en riesgo y coordinar intervenciones.
- **Administración de la universidad:** Interesada en mejorar la retención y el rendimiento académico.
- **Estudiantes:** Beneficiarios directos del sistema, buscando una experiencia educativa más enriquecedora.

3.2 Resumen de los usuarios

Los usuarios del sistema incluirán:

- **Tutora:** Utilizará el dashboard para identificar y apoyar a estudiantes en riesgo, facilitando intervenciones oportunas.
- **Administradores académicos:** Monitorearán las métricas de retención y evaluarán la efectividad del sistema.

3.3 Entorno de usuario

El entorno de usuario incluye:

- **Aulas y oficinas administrativas:** Espacios donde se realizan las interacciones con el sistema.
- **Plataformas de gestión académica:** Integración con sistemas existentes de la universidad.
- **Dispositivos electrónicos:** Computadoras y tablets utilizadas por la tutora y administradores.

3.4 Perfiles de los interesados

- **Tutora de EPIS:**
 - **Rol:** Proporcionar apoyo académico y emocional a los estudiantes.
 - **Intereses:** Mejorar la retención y el éxito académico de los estudiantes.
- **Administración:**
 - **Rol:** Tomar decisiones estratégicas basadas en datos.
 - **Intereses:** Incrementar la reputación institucional y optimizar recursos.
- **Estudiantes:**
 - **Rol:** Completar sus estudios con éxito.
 - **Intereses:** Obtener apoyo académico y evitar el abandono.

3.5 Perfiles de los usuarios

- **Tutora:**
 - **Responsabilidades:** Monitorear el rendimiento y bienestar de los estudiantes, intervenir de manera efectiva y coordinar apoyos.
 - **Habilidades:** Uso de tecnología, análisis de datos, y habilidades de comunicación.

- **Administradores:**

- **Responsabilidades:** Analizar datos de retención y proponer mejoras en el sistema académico.
- **Habilidades:** Gestión de información y toma de decisiones basadas en datos.

3.6 Necesidades de los interesados y usuarios

- **Tutora:** Necesita herramientas para identificar rápidamente a estudiantes en riesgo y métodos eficaces de intervención.
- **Administración:** Requiere datos precisos y en tiempo real sobre la retención estudiantil para fundamentar decisiones informadas.

4. Vista General del Producto

4.1 Perspectiva del producto

El Dashboard de Detección Temprana de Abandono Estudiantil es una herramienta desarrollada para ayudar a la tutora y a la administración de la universidad a identificar y apoyar a estudiantes en riesgo de abandono. Utilizando Power BI, el sistema integrará diversas fuentes de datos académicos y de comportamiento, permitiendo un análisis visual y dinámico que facilitará la toma de decisiones informadas.

4.2 Resumen de capacidades

- **Visualización de datos en tiempo real:** Dashboards interactivos que permiten monitorear el rendimiento estudiantil de forma continua.
- **Alertas tempranas:** Notificaciones automáticas para la tutora cuando un estudiante presenta indicadores de riesgo.

- **Análisis histórico:** Capacidad de revisar tendencias en el rendimiento y la retención de estudiantes a lo largo del tiempo.
- **Interactividad:** Herramientas para que la tutora explore los datos y realice un análisis ad hoc.
- **Informes personalizados:** Generación de informes que pueden ser compartidos con la administración y otros interesados.

4.3 Suposiciones y dependencias

- **Suposiciones:**
 - Los datos necesarios para el análisis estarán disponibles y en formatos compatibles con Power BI.
 - La tutora y los administradores tendrán el entrenamiento adecuado para utilizar el sistema eficientemente.
 - La implementación del dashboard se realizará sin interrupciones significativas en las operaciones académicas.
- **Dependencias:**
 - Integración con sistemas existentes de gestión académica para la recopilación de datos.
 - Requerimientos de infraestructura tecnológica para el uso de Power BI.
 - Soporte continuo de los proveedores de tecnología durante la implementación.

4.4 Costos y precios

Los costos asociados con el desarrollo e implementación del Dashboard de Detección Temprana de Abandono Estudiantil están detallados en el documento **FD01-Informe de Factibilidad**.

4.5 Licenciamiento e instalación

El uso de Power BI requiere la adquisición de licencias adecuadas que se alineen con el número de usuarios y las características del sistema. El proceso de instalación incluirá:

- Configuración del entorno de Power BI.
- Integración de las fuentes de datos.
- Capacitación para la tutora y los administradores sobre el uso del dashboard.
- Soporte técnico inicial durante la fase de implementación para asegurar una transición fluida.

5. Características del producto

- **Modelo Predictivo:** Algoritmo de machine learning para estimar el riesgo de abandono estudiantil basado en datos históricos y actuales.
- **Dashboard Interactivo:** Interfaz gráfica para visualizar el riesgo de abandono, tendencias académicas y estadísticas en tiempo real.
- **Visualización de Datos:** Gráficas y tablas para analizar factores de riesgo como rendimiento académico, asistencia y datos socioeconómicos.
- **Filtrado y Segmentación:** Opciones para filtrar datos por carrera, nivel académico, y otros criterios relevantes.
- **Generación de Reportes:** Herramienta para crear y exportar reportes sobre estudiantes en riesgo en formatos como PDF y Excel.
- **Alertas y Notificaciones:** Sistema para alertar a tutores y administradores sobre estudiantes que superen umbrales de riesgo críticos.

6. Restricciones

- Limitaciones Técnicas: Dependencia de la calidad de los datos y la capacidad del algoritmo de machine learning para adaptarse a nuevas tendencias.
- Tiempo: Plazos para el desarrollo y la implementación del sistema pueden verse afectados por la disponibilidad de recursos y la complejidad del modelo predictivo.
- Recursos: Disponibilidad de personal especializado en machine learning y análisis de datos, así como de infraestructura tecnológica adecuada.

7. Rangos de calidad

Desarrollo del Modelo Predictivo:

- **Rango de calidad alto:** Algoritmo de machine learning con alta precisión y fiabilidad en la detección del riesgo de abandono, sin errores significativos y capaz de adaptarse a nuevas tendencias en el comportamiento de los estudiantes.
- **Rango de calidad medio:** Algoritmo con una precisión aceptable, con algunos errores menores que no afectan gravemente la detección de riesgo y que pueden corregirse con ajustes menores.
- **Rango de calidad bajo:** Algoritmo con problemas frecuentes en la detección de riesgo, errores significativos que afectan la precisión y la capacidad de adaptación a nuevos patrones.

Experiencia del Usuario:

- **Rango de calidad alto:** Dashboard intuitivo y fácil de usar, con navegación fluida, tiempos de respuesta rápidos, y visualización clara de los datos de los estudiantes y sus riesgos.

- **Rango de calidad medio:** Algunas áreas del dashboard pueden ser mejoradas para mayor usabilidad, pero en general, la experiencia del usuario es satisfactoria y cumple con los requisitos básicos.
- **Rango de calidad bajo:** Dashboard confuso o difícil de usar, con navegación lenta y problemas significativos en la visualización de datos que afectan la eficiencia en el análisis del riesgo.

Seguridad de Datos:

- **Rango de calidad alto:** Implementación robusta de medidas de seguridad, con protección adecuada de los datos de los estudiantes y cumplimiento de las normativas vigentes sobre privacidad y seguridad.
- **Rango de calidad medio:** Algunas áreas de mejora en la seguridad de los datos, pero en general, se siguen las mejores prácticas estándar y se realiza un manejo adecuado de la información.
- **Rango de calidad bajo:** Fallos graves en la seguridad de datos, como vulnerabilidades significativas o falta de medidas adecuadas para proteger la información personal de los estudiantes.

Eficiencia Operativa:

- **Rango de calidad alto:** Sistema de detección y dashboard operando de manera óptima, con procesos bien gestionados, tiempos de carga rápidos y capacidad para manejar grandes volúmenes de datos sin problemas.
- **Rango de calidad medio:** Algunas áreas de eficiencia operativa pueden necesitar ajustes, pero el sistema funciona de manera efectiva y cumple con la mayoría de los requisitos operacionales.

- **Rango de calidad bajo:** Problemas significativos en la operación del sistema, como tiempos de respuesta lentos, dificultades para manejar grandes volúmenes de datos o errores frecuentes en el procesamiento.

8. Precedencia y Prioridad

Desarrollo del Modelo Predictivo

- **Prioridad:** Alta
- **Descripción:** Este es el componente central del proyecto. El modelo predictivo debe ser desarrollado y validado primero, ya que proporciona la base para identificar a los estudiantes en riesgo de abandono.

Implementación del Dashboard Interactivo

- **Prioridad:** Alta
- **Descripción:** Una vez que el modelo predictivo esté funcional, se debe construir el **dashboard** para visualizar los resultados y permitir a los usuarios interactuar con los datos de riesgo de manera efectiva.

Generación de Reportes y Alertas

- **Prioridad:** Media
- **Descripción:** Después de que el **dashboard** esté en funcionamiento, se debe desarrollar la capacidad para generar reportes y configurar alertas sobre estudiantes que superen umbrales críticos de riesgo.

Capacitación y Documentación

- **Prioridad:** Baja

- **Descripción:** Finalmente, se deben proporcionar recursos educativos y documentación para que los usuarios puedan utilizar el sistema de manera efectiva. Esto incluye manuales de usuario y formación sobre el uso del sistema y la interpretación de los datos relacionados con el riesgo de abandono.

9. Otros requerimientos del producto

a) Estándares legales

- **Protección de Datos Personales:** Cumplir con la Ley N° 29733, Ley de Protección de Datos Personales, y su reglamento en Perú, que establece cómo se debe manejar y proteger la información personal de los estudiantes.
- **Regulación sobre Análisis Predictivo:** Asegurarse de que el uso de modelos predictivos cumpla con las regulaciones locales aplicables en el ámbito educativo y de datos.
- **Normativas sobre Seguridad Informática:** Seguir las directrices establecidas por el Instituto Nacional de Defensa Civil (INDECI) y el Reglamento de Seguridad de la Información para garantizar la seguridad y protección de la información.

b) Estándares de comunicación

- **Protocolos de Comunicación Segura:** Utilizar HTTPS para la transmisión segura de datos entre el dashboard y el servidor, y garantizar que toda la comunicación de datos esté cifrada, conforme a las mejores prácticas internacionales adaptadas a la normativa peruana.
- **Interoperabilidad de Sistemas:** Implementar API (Application Programming Interface) con estándares comunes, asegurando compatibilidad con los sistemas y plataformas utilizadas por la universidad.

- Documentación de API: Proporcionar documentación completa y actualizada para las interfaces de programación, facilitando la integración con otros sistemas y garantizando una comunicación eficiente.

c) Estándares de cumplimiento de la plataforma

- Compatibilidad con Navegadores: Asegurarse de que el dashboard sea compatible con los principales navegadores web (Chrome, Firefox, Edge, Safari) y versiones recientes, adaptándose a las tecnologías utilizadas por los usuarios en Perú.
- Integración con Sistemas Existentes: El producto debe integrarse sin problemas con los sistemas actuales de la universidad, como bases de datos y plataformas de gestión de estudiantes en el contexto peruano.
- Escalabilidad: Diseñar el sistema para que pueda escalar según las necesidades de la universidad, soportando un aumento en el número de usuarios y datos sin pérdida de rendimiento, teniendo en cuenta las proyecciones de crecimiento local.

d) Estándares de calidad y seguridad

- Precisión del Modelo: Asegurar que el modelo predictivo mantenga un nivel de precisión definido, acorde con los requisitos y expectativas de la universidad, como una tasa de precisión del 90% en las predicciones.
- Usabilidad: El dashboard debe ser intuitivo y fácil de usar, con una experiencia de usuario fluida y sin errores críticos, adaptado a los perfiles y necesidades de los usuarios peruanos.
- Seguridad de Datos: Implementar medidas de seguridad robustas para proteger la información de los postulantes, incluyendo autenticación de usuarios, cifrado de datos en reposo y en tránsito, y auditorías de seguridad regulares, siguiendo las directrices del Ministerio de Justicia y Derechos Humanos.

- Cumplimiento de Estándares de Calidad: Seguir normas de desarrollo de software de calidad, adaptadas a la realidad peruana, para garantizar un producto confiable y eficaz.

CONCLUSIONES

- El sistema permitirá a la universidad prever el riesgo de abandono de los estudiantes, mejorando así la planificación y asignación de recursos para su retención.
- El dashboard interactivo facilitará a los administradores el análisis de tendencias relacionadas con el abandono y la toma de decisiones basadas en datos actualizados y visualizados en tiempo real.

RECOMENDACIONES

1. Validación y Verificación de Datos:

- **Recomendación:** Implementar procesos rigurosos para la validación y verificación de los datos históricos utilizados en el modelo de predicción del abandono estudiantil. Esto incluye la limpieza de datos, la eliminación de errores y la comprobación de la integridad y precisión de la información académica y socioeconómica de los estudiantes.
- **Justificación:** La efectividad del modelo predictivo depende de la calidad de los datos. Datos erróneos o incompletos pueden resultar en predicciones inexactas, afectando negativamente la capacidad de intervención temprana y las decisiones de tutoría.

2. Capacitación Continua del Personal:

- **Recomendación:** Proporcionar capacitación continua y actualizaciones a tutores y administradores sobre el uso del sistema y sus funcionalidades. Asegurarse de que el personal esté bien capacitado para interpretar las predicciones y utilizar el dashboard de manera efectiva.
- **Justificación:** Una correcta interpretación de las predicciones es crucial para maximizar los beneficios del sistema. La capacitación regular ayudará a mantener la competencia y confianza en el uso de la herramienta.

3. Actualización y Mantenimiento del Modelo Predictivo:

- **Recomendación:** Establecer un proceso continuo para la actualización y ajuste del modelo predictivo en función de nuevos datos y cambios en las tendencias de abandono estudiantil. Realizar evaluaciones periódicas del rendimiento del modelo y ajustar los parámetros según sea necesario.
- **Justificación:** Las condiciones académicas y sociales pueden cambiar con el tiempo. Es fundamental que el modelo se mantenga relevante y preciso para abordar las nuevas realidades del entorno educativo.

4. Integración y Compatibilidad:

- **Recomendación:** Asegurarse de que el sistema de predicción se integre de manera fluida con las plataformas y sistemas existentes de la universidad. Realizar pruebas exhaustivas de integración para identificar y resolver problemas potenciales antes de la implementación completa.
- **Justificación:** Una integración exitosa facilitará la adopción del sistema por parte de los usuarios y garantizará una transición más suave, reduciendo el riesgo de problemas técnicos y operativos.

5. Gestión de la Seguridad de Datos:

- **Recomendación:** Implementar medidas de seguridad robustas para proteger la información sensible de los estudiantes. Esto incluye encriptación de datos, autenticación de usuarios y protocolos de seguridad para prevenir accesos no autorizados.
- **Justificación:** La protección de los datos de los estudiantes es esencial para mantener la confianza en el sistema y cumplir con las normativas de privacidad y seguridad.

BIBLIOGRAFIA

Breiman, L. (2001). "Random Forests". *Machine Learning*, 45(1), 5-32.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press.

Python Software Foundation. (2021). *Python Documentation*. Retrieved from <https://docs.python.org/3/>

TensorFlow. (2021). *TensorFlow Documentation*. Retrieved from <https://www.tensorflow.org/learn>

WEBGRAFIA

Scikit-learn Documentation: https://scikit-learn.org/stable/user_guide.html

TensorFlow Documentation: <https://www.tensorflow.org/learn>

Towards Data Science - An Introduction to Machine Learning: <https://towardsdatascience.com/an-introduction-to-machine-learning-14dd2da2ef5d>

KDnuggets - Data Science and Machine Learning News: <https://www.kdnuggets.com/>

Coursera - Machine Learning by Andrew Ng: <https://www.coursera.org/learn/machine-learning>

Kaggle - Learn Data Science: <https://www.kaggle.com/learn/overview>