****

**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:

***Lopez Catunta , Brayar Christian (2020068946)***

***Melendez Huarachi, Gabriel Fari (2021070311)***

***Cuadros Garcia, Mirian (2021071083)***

***Hurtado Ortiz, Leandro (2015052384)***

***Briceño Diaz, Jorge Luis (2017059611)***

**Tacna – Perú**

***2024***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| CONTROL DE VERSIONES | | | | | |
| Versión | Hecha por | Revisada por | Aprobada por | Fecha | Motivo |
| 1.0 | JLBD | JLBCG | JLBCG | 31/08/2024 | Versión Original |

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 |

**INDICE GENERAL**

[**1.** **Introducción** 4](#_Toc177202751)

[1.1 Propósito 5](#_Toc177202752)

[1.2 Alcance 5](#_Toc177202753)

[1.3 Definiciones, Siglas y Abreviaturas 6](#_Toc177202754)

[1.4 Referencias 6](#_Toc177202755)

[1.5 Visión General 6](#_Toc177202756)

[**2. Posicionamiento** 7](#_Toc177202757)

[2.1 Oportunidad de negocio 7](#_Toc177202758)

[2.2 Definición del problema 7](#_Toc177202759)

[**3. Descripción de los interesados y usuarios** 8](#_Toc177202760)

[3.1 Resumen de los interesados 8](#_Toc177202761)

[3.2 Resumen de los usuarios 9](#_Toc177202762)

[3.3 Entorno de usuario 9](#_Toc177202763)

[3.4 Perfiles de los interesados 9](#_Toc177202764)

[3.5 Perfiles de los usuarios 10](#_Toc177202765)

[3.6 Necesidades de los interesados y usuarios 10](#_Toc177202766)

[**4. Vista General del Producto** 10](#_Toc177202767)

[4.1 Perspectiva del producto 10](#_Toc177202768)

[4.2 Resumen de capacidades 11](#_Toc177202769)

[4.3 Suposiciones y dependencias 11](#_Toc177202770)

[4.4 Costos y precios 12](#_Toc177202771)

[4.5 Licenciamiento e instalación 12](#_Toc177202772)

[**5.** **Características del producto** 12](#_Toc177202773)

[**6.** **Restricciones** 13](#_Toc177202774)

[**7.** **Rangos de calidad** 13](#_Toc177202775)

[**8.** **Precedencia y Prioridad** 15](#_Toc177202776)

[**9.** **Otros requerimientos del producto** 16](#_Toc177202777)

[a) Estándares legales 16](#_Toc177202778)

[b) Estándares de comunicación 16](#_Toc177202779)

[c) Estándares de cumplimiento de la plataforma 17](#_Toc177202780)

[d) Estándares de calidad y seguridad 17](#_Toc177202781)

[**CONCLUSIONES** 18](#_Toc177202782)

[**RECOMENDACIONES** 18](#_Toc177202783)

[**BIBLIOGRAFIA** 21](#_Toc177202784)

[**WEBGRAFIA** 21](#_Toc177202785)

# **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 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 **Sistema de Predicción de Postulantes** para la Universidad Nacional de Educación Enrique Guzmán y Valle representa una significativa oportunidad de negocio, tanto para la universidad como para los proveedores de tecnología y servicios asociados. La universidad se beneficiará al optimizar su proceso de admisión, mejorando la calidad de los estudiantes admitidos y, por ende, la tasa de éxito académico y la reputación institucional. Esto puede llevar a una mayor atracción de futuros postulantes y a un incremento en la demanda de sus programas académicos.

Desde el punto de vista de proveedores tecnológicos, el desarrollo e implementación de este sistema abre oportunidades para ofrecer soluciones avanzadas en análisis de datos y machine learning. Además, al demostrar la eficacia del sistema en un entorno educativo, se pueden generar casos de éxito que potencialmente se pueden replicar en otras instituciones educativas, ampliando el mercado y la influencia de las tecnologías desarrolladas.

## 2.2 Definición del problema

El proceso actual de admisión en la Universidad Nacional de Educación Enrique Guzmán y Valle enfrenta varios desafíos críticos:

1. **Crecimiento en el número de postulantes:** El aumento en el número de aspirantes ha complicado la tarea de seleccionar candidatos con el mayor potencial académico.
2. **Evaluación manual:** El proceso manual de evaluación es propenso a errores y es ineficiente, lo que puede llevar a decisiones subóptimas.
3. **Criterios estandarizados:** Los criterios actuales no reflejan adecuadamente el potencial real de los postulantes para sobresalir académicamente.
4. **Falta de herramientas tecnológicas:** La ausencia de herramientas avanzadas para el análisis y predicción limita la capacidad de la universidad para mejorar el proceso de selección y aumentar la tasa de éxito de los estudiantes admitidos.

El **Sistema de Predicción de Postulantes** aborda estos problemas mediante el uso de técnicas de análisis de datos y machine learning para proporcionar una evaluación más precisa y eficiente de los candidatos, optimizando así el proceso de admisión.

# **3. Descripción de los interesados y usuarios**

## 3.1 Resumen de los interesados

* **Administradores Universitarios**: Responsables de la toma de decisiones en el proceso de admisión y evaluación de postulantes. Están interesados en herramientas que les permitan tomar decisiones informadas y optimizar el proceso de selección.
* **Personal de TI**: Encargados de la implementación y mantenimiento del sistema. Necesitan una solución que sea fácil de integrar con las plataformas existentes y que requiera un mantenimiento mínimo.

## 3.2 Resumen de los usuarios

* **Administradores del Sistema**: Utilizarán el dashboard para consultar y analizar datos de postulantes. Necesitan una interfaz intuitiva y herramientas para visualizar las predicciones y datos relevantes.

## 3.3 Entorno de usuario

* **Entorno Físico**: El sistema será utilizado principalmente en entornos de oficina dentro de la universidad, accesible desde estaciones de trabajo con una red de comunicación estable.
* **Entorno Tecnológico**: Se operará sobre servidores con Linux y estaciones de trabajo con Windows o Linux. El dashboard estará disponible a través de navegadores web, y se integrará con la infraestructura de red existente de la universidad.

## 3.4 Perfiles de los interesados

* **Administrador Universitario**:
  + **Funciones**: Toma de decisiones estratégicas y operativas, supervisión del proceso de admisión.
  + **Intereses**: Mejorar la calidad del proceso de admisión, optimizar la selección de candidatos.
  + **Expectativas**: Herramientas que proporcionen datos precisos y fáciles de interpretar.
* **Personal de TI**:
  + **Funciones**: Implementación y mantenimiento del sistema.
  + **Intereses**: Integración sencilla, mantenimiento fácil, y soporte técnico adecuado.
  + **Expectativas**: Documentación clara y soporte continuo.

## 3.5 Perfiles de los usuarios

* **Administrador del Sistema**:
  + **Funciones**: Gestión de la plataforma, generación de informes.
  + **Intereses**: Facilidad de uso y acceso a datos completos.
  + **Expectativas**: Interfaz de administración clara y funcionalidades para gestionar datos de postulantes.

## 3.6 Necesidades de los interesados y usuarios

* **Administradores Universitarios**: Necesitan herramientas que faciliten la toma de decisiones informadas y que proporcionen una visión clara de las predicciones y análisis de los postulantes.
* **Personal de TI**: Requieren una integración sencilla con sistemas existentes, documentación técnica adecuada y soporte para resolver problemas técnicos.

# **4. Vista General del Producto**

## 4.1 Perspectiva del producto

El producto final es un dashboard interactivo dirigido a los administradores de la universidad. Este dashboard permitirá prever el número de aspirantes y sus características, facilitando la toma de decisiones en el proceso de admisión.

## 4.2 Resumen de capacidades

* **Predicción de postulantes**: Estimación del número de postulantes por carrera.
* **Análisis demográfico**: Visualización de la distribución geográfica y perfiles de los postulantes.
* **Visualización de tendencias**: Identificación de patrones en la demanda de admisiones.
* **Generación de reportes**: Exportación de datos en formatos como PDF y Excel.
* **Alertas**: Notificaciones sobre cambios significativos en las predicciones.

## 4.3 Suposiciones y dependencias

* **Suposiciones**:
  + Se asume que se dispondrá de datos históricos precisos y completos de postulantes anteriores.
  + Se asume que los algoritmos de machine learning serán capaces de adaptarse a cambios en las tendencias de postulación.
  + Se asume que los usuarios tendrán acceso regular a Internet para interactuar con el dashboard.
* **Dependencias**:
  + Dependencia en la calidad y disponibilidad de los datos: La precisión de las predicciones depende en gran medida de la calidad de los datos proporcionados.
  + Dependencia de la infraestructura tecnológica: El funcionamiento óptimo del sistema requiere servidores y bases de datos con capacidades suficientes para manejar grandes volúmenes de datos.
  + Dependencia de la colaboración de las áreas de admisión y tecnología de la universidad para la correcta implementación y mantenimiento del sistema.

## 4.4 Costos y precios

Los detalles sobre los costos y precios asociados con el desarrollo e implementación del Sistema de Predicción de Postulantes se encuentran en el documento separado titulado **FD01-Informe de Factibilidad**. Este informe proporciona un análisis detallado de los costos involucrados en cada fase del proyecto.

## 4.5 Licenciamiento e instalación

El producto se licenciará bajo un modelo de suscripción para la universidad. La instalación se realizará en los servidores de la universidad o en la nube, según las necesidades.

# **5. Características del producto**

* Modelo Predictivo: Algoritmo de machine learning para estimar el número de postulantes y sus características basadas en datos históricos y actuales.
* Dashboard Interactivo: Interfaz gráfica para visualizar predicciones, tendencias demográficas y estadísticas en tiempo real.
* Visualización de Datos: Gráficas y tablas para analizar la distribución geográfica y demográfica de los postulantes.
* Filtrado y Segmentación: Opciones para filtrar datos por carrera, modalidad, y otros criterios relevantes.
* Generación de Reportes: Herramienta para crear y exportar reportes en formatos como PDF y Excel.
* Alertas y Notificaciones: Sistema para alertar sobre cambios significativos en las predicciones y umbrales definidos.

# **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 las predicciones, sin errores significativos y capaz de adaptarse a nuevas tendencias con eficacia.
* Rango de calidad medio: Algoritmo con una precisión aceptable, con algunos errores menores que no afectan gravemente las predicciones y que pueden corregirse con ajustes menores.
* Rango de calidad bajo: Algoritmo con problemas frecuentes en las predicciones, errores significativos que afectan la precisión y la capacidad de adaptación a nuevas tendencias.

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 y predicciones.
* Rango de calidad medio: Algunas áreas del dashboard pueden ser mejoradas para una 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.

Seguridad de Datos:

* Rango de calidad alto: Implementación robusta de medidas de seguridad, con protección adecuada de los datos de los postulantes 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 postulantes.

Eficiencia Operativa:

* Rango de calidad alto: Sistema de predicción y dashboard operando de manera óptima, con procesos internos 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 todas las predicciones sobre los postulantes.

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 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 cambios significativos en las predicciones.

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.

# **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 postulantes.
* 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 la cantidad y características de los postulantes, mejorando la planificación y asignación de recursos.
* El dashboard interactivo facilitará a los administradores el análisis de tendencias y la toma de decisiones basadas en datos actualizados y visualizados en tiempo real.

# **RECOMENDACIONES**

**Validación y Verificación de Datos:**

* **Recomendación:** Implementar procesos rigurosos para la validación y verificación de datos históricos utilizados para el entrenamiento del modelo predictivo. Esto incluye la limpieza de datos, la eliminación de errores y la comprobación de la integridad y precisión de los datos.
* **Justificación:** La precisión del modelo predictivo depende en gran medida de la calidad de los datos. Datos erróneos o incompletos pueden llevar a predicciones inexactas y decisiones subóptimas en el proceso de admisión.

**Capacitación Continua del Personal:**

* **Recomendación:** Ofrecer capacitación continua y actualizaciones para los administradores y evaluadores académicos sobre el uso del sistema y sus funcionalidades. Asegurarse de que el personal esté bien capacitado para interpretar las predicciones y utilizar el sistema de manera efectiva.
* **Justificación:** La correcta utilización del sistema y la interpretación precisa de sus resultados son cruciales para maximizar sus beneficios. La capacitación regular ayudará a mantener la competencia y confianza en el uso del sistema.

**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 admisión. Realizar evaluaciones periódicas del rendimiento del modelo y ajustar los parámetros según sea necesario.
* **Justificación:** Las condiciones y las tendencias pueden cambiar con el tiempo, por lo que es importante que el modelo se mantenga relevante y preciso en el contexto actual.

**Integración y Compatibilidad:**

* **Recomendación:** Asegurarse de que el sistema de predicción se integre de manera fluida con las plataformas existentes de la universidad. Realizar pruebas exhaustivas de integración para identificar y resolver problemas potenciales antes de la implementación completa.
* **Justificación:** Una integración exitosa minimizará la resistencia al cambio y garantizará una transición más suave, reduciendo el riesgo de problemas técnicos y operativos.

**Gestión de la Seguridad de Datos:**

* **Recomendación:** Implementar medidas de seguridad robustas para proteger la información sensible de los postulantes. 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 postulantes es crucial 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