



ANEXO 1
FORMATO PARA LA PRESENTACIÓN DE PROYECTOS DE INVESTIGACIÓN
CON EL FINANCIAMIENTO DEL FEDU

1. Título del proyecto

Implementación de modelos predictivos con base de sistemas inteligentes para la deserción estudiantil en el ámbito educativo universitario

2. Área de Investigación

Área de investigación	Línea de Investigación	Disciplina OCDE
Ciencias de la Ingeniería	Ingeniería de Software, Bases de Datos e Inteligencia de Negocios	Ingeniería y Tecnología

3. Duración del proyecto (meses)

12 Meses

4. Tipo de proyecto

<u>Individual</u>	
<u>Multidisciplinario</u>	X
<u>Director de tesis pregrado</u>	

4. Datos de los integrantes del proyecto

Apellidos y Nombres	TICONA YANQUI FIDEL ERNESTO (AUTOR)
Escuela Profesional	INGENIERIA DE SISTEMAS
Celular	951765555
Correo Electrónico	fticona@unap.edu.pe

Apellidos y Nombres	JIMENEZ CHURA ADOLFO (CO AUTOR)
Escuela Profesional	INGENIERIA DE SISTEMAS
Celular	994125511
Correo Electrónico	ajimenez@unap.edu.pe

I. Título

Implementación de modelos predictivos con base de sistemas inteligentes para la deserción estudiantil en el ámbito educativo universitario

II. Resumen del Proyecto de Tesis

En la actualidad existen muchos factores que contribuyen a la deserción estudiantil en los estudiantes Universitarios, por lo que se hace importante el desarrollo de modelos predictivos como arboles de decisión, minería de datos y redes bayesianas técnicas que son parte fundamental de los sistemas inteligentes que contribuyen a la adecuada toma de decisiones a los directivos universitarios, entre el objetivo principal se tiene la implementación de las técnicas de predicción para determinar la deserción de los estudiantes universitarios en la escuela profesional de Ingeniería de sistemas de la universidad Nacional del Altiplano y la Universidad Andina Néstor Cáceres Velásquez de Juliaca, el tipo de investigación es aplicada cuasi experimental debido al uso de técnicas de



predicción en la administración de información que se recopilara de los sistemas de gestión de información que tienen las escuelas profesionales en estudio, además mediante la implementación de los modelos predictivos se pretende identificar los factores claves de deserción estudiantil para que los directivos universitarios puedan tomar las adecuadas decisiones.

III. Palabras claves (Keywords)

Modelos predictivos, sistemas inteligentes, arboles de decisión, redes bayesianas, minería de datos.

IV. Justificación del proyecto

Son muchos los factores que impulsan a los estudiantes a desertar de una carrera profesional, por lo que se hace importante el análisis de este problema que cada día va en aumento debido a diferentes factores o indicadores como son los económicos, sociales, académicos y entre otros.

El trabajo toma relevancia debido a que se realizara un análisis de los indicadores de deserción mediante el uso de las técnicas de predicción de los sistemas inteligentes mediante el cual se determinara los factores claves de deserción estudiantil en las escuelas profesionales de ingeniería de sistemas de la Universidad Nacional del Altiplano y la Universidad Andina Néstor Cáceres Velásquez.

También es importante mencionar que en el trabajo de investigación se pretende probar diferentes técnicas de predicción de los sistemas inteligentes los cuales contribuirán a seleccionar los indicadores que influyen de gran manera en la deserción estudiantil y de esta forma seleccionar la mejor técnica de predicción.

El trabajo de investigación servirá a los directivos universitarios para que estos puedan tomar las mejores decisiones respecto a la deserción estudiantil y la información también puede ser tomado como referente para los padres de familia, docentes con la finalidad de orientar a los jóvenes a que puedan tomar las mejores decisiones y de esta forma evitar la deserción estudiantil.

V. Antecedentes del proyecto

(Condori y Pablo 2018) El presente trabajo de investigación tiene como objetivo efectuar un análisis predictivo de la deserción estudiantil utilizando Data Warehouse y minería de datos en la Universidad Nacional Jorge Basadre Grohmann 2012-2018. Para ello, el autor ha construido un Data Warehouse para el análisis de datos de la deserción estudiantil de la Universidad Nacional Jorge Basadre Grohmann bajo la metodología de Ralph Kimball. Como conclusión general, el autor efectuó un análisis predictivo de la deserción estudiantil utilizando Data Warehouse y minería de datos bajo metodologías permitiendo entender la información histórica e interpretación de sus resultados predichos.

(Perez-Gutierrez 2020) El artículo presenta los resultados de un estudio de comparación de técnicas para apoyar la identificación de deserción estudiantil a partir del registro académico de los estudiantes de una Universidad en Colombia para el programa de Ingeniería de Sistemas. El registro académico se estableció para un periodo de 7 años. Árboles de decisión, regresión logística y Naive Bayes, fueron comparados para lograr establecer la mejor técnica de detección de desertores. Adicionalmente, la herramienta Watson Analytics de IBM fue utilizada para comparar su usabilidad y precisión para un usuario no experto. Nuestra experimentación demostró que el uso de algoritmos simples es suficiente para alcanzar niveles ideales de precisión. Estos resultados son presentados a la comunidad académica para ayudar en la disminución de la deserción estudiantil

(Vegega, Pytel, y Pollo-Cattaneo 2019) La toma de decisiones es un proceso esencial en la vida de las organizaciones siendo particularmente importante para los roles gerenciales que deben tomar decisiones sobre cómo asignar sus recursos. Estas decisiones deben basarse



en predicciones sobre el tiempo, el esfuerzo y/o los riesgos de llevar a cabo sus actividades. Actualmente esta situación se encuentra acentuada por el entorno complejo que rodea a las organizaciones lo cual provoca que deban actuar más allá de los sistemas tradicionales de gestión incorporando nuevos mecanismos, como los provistos por la Inteligencia Artificial, por lo que se estaría desarrollando un Modelo Predictivo Inteligente. En este contexto, el presente trabajo tiene como objetivo proponer la aplicación de un proceso que asista al Ingeniero en Sistemas de Información en la difícil tarea de recolectar, entender, identificar y documentar la información necesaria para implementar un Modelo Predictivo basado en Sistemas Inteligentes

(García Franco 2019) La deserción universitaria se ha convertido en un problema prioritario a ser investigado y tratado. El porcentaje de deserción ha llegado a constituir uno de los principales indicadores de eficiencia interna dentro de cualquier institución de educación superior. El objetivo del presente proyecto consiste en implementar un modelo computacional que permita identificar en forma automática a los estudiantes con mayor riesgo de deserción en la Universidad Peruana Unión Filial Juliaca. Para la implementación de este proyecto se adoptó la metodología CRISP-DM. Se aplicó el modelo de clasificación de Machine Learning, para analizar el comportamiento de los estudiantes, evaluando factores como cantidad de cursos matriculados, cantidad de cursos aprobados, si es independiente o dependiente con respecto al pago de sus estudios, si tiene o no sanción disciplinaria por parte de Bienestar Universitario, cantidad de cursos desaprobados durante el semestre, cantidad de cursos desaprobados dos veces, cantidad de cursos desaprobados de tres veces a más, cantidad de créditos aprobados, cantidad créditos desaprobados, ponderado final del semestre, si la situación del alumnos es regular o irregular, si tiene un saldo a favor o en contra.

(Sandoval Torres 2013) Este artículo presenta la comparación de algoritmos de clasificación para la predicción de la deserción escolar del Centro Universitario UAEM Valle de México en la carrera de Ingeniería en Sistemas y Comunicaciones. El objetivo fue determinar el mejor algoritmo basándose en la precisión de la clasificación, así como en la utilidad en la información provista al tutor. De los experimentos concluimos que los mejores algoritmos son Naïve Bayes Tree, que tiene el menor error, y J48 que provee información útil para el tutor al indicar las materias que debe cuidar, permitiéndole crear estrategias que apoyen el desempeño académico del alumno en las mismas.

(Pérez et al. 2018) El presente estudio tiene como objetivo modelar un sistema predictivo de retención de los estudiantes de primer año de la Universidad Bernardo O'Higgins - UBO (Santiago de Chile), determinando cuáles de las variables de entrada a la educación superior, ya sean estas académicas, sociales o familiares, se revelan significativas para este análisis. La construcción del modelo de investigación se sustentó en una exhaustiva revisión bibliográfica que permitió identificar variables explicativas de la deserción universitaria en el contexto nacional. Consecuentemente, se analizó la relación de cada una de las variables explicativas del estudio con la variable control deserción estudiantil. El modelo multivariado predijo en un 86,4 % la conducta de abandono señalando siete variables categóricas independientes que, finalmente, se develan como factores relevantes del modelo predicción. Las variadas y sustentadas interpretaciones entregadas en los resultados del modelo, así como las sugerencias propuestas para mejorar el índice de retención universitaria, entregan un valor directo al estudio dirigido a optimizar uno de los indicadores más importantes vinculados a una gestión de calidad en las universidades, como es, la retención estudiantil.

(Cuji, Gavilanes, y Sanchez 2017) Este artículo muestra la construcción de un modelo predictivo de deserción estudiantil, para pronosticar la probabilidad, que un estudiante abandone su programa académico, mediante técnicas de clasificación, basadas en árboles de decisión. La metodología utilizada, se basa en Knowledge Discovery in Database (KDD), con cinco etapas: selección, procesamiento, transformación, minería de datos y evaluación. Aplicando el algoritmo, Classification and Regression Tree (CART) de la herramienta R, se construyó un árbol con cuatro niveles de profundidad y mismo número reglas, que evalúan a los posibles desertores. Llevando a concluir que las variables nivel y notas tienen mayor influencia en la deserción.

(Ramírez et al. 2018) El objetivo de este estudio es presentar una clasificación basada en árboles de decisión (CBAD) con parámetros optimizados para predecir la deserción de los estudiantes universitarios. El estudio analiza 5288 casos de estudiantes pertenecientes a una universidad pública chilena. Para la técnica CBAD se optimizaron los parámetros para mejorar la predicción usando el software RapidMiner. El resultado de la aplicación de esta técnica con parámetros optimizados logro una razón de precisión de un 87.27%. Se concluye que el uso de técnicas de CBAD con optimización de parámetros resulta en una mejor precisión en comparación a otras investigaciones con un número similar de datos.

(Martínez y Mateus 2020) La deserción estudiantil es una realidad en el país y es un fenómeno complejo. En este artículo, se propone un modelo predictivo que sirva como apoyo a las Universidades colombianas para el análisis de la deserción en estudiantes, principalmente, en programas de pregrado en modalidad virtual. Un modelo predictivo, puede ayudar a las organizaciones a generar ganancias y evitar pérdidas futuras, tomando datos históricos y con ellos arrojar resultados esperados para ser analizados y apoyar en las decisiones. Este modelo se desarrolla, tomando eventos históricos con distintas variables de tipo social, académico, personal, laboral, ingresos a las plataformas e-learning, etc. y posteriormente, a estas variables se le aplican algoritmos de aprendizaje profundo. Se espera la predicción de probabilidades de deserción de cada estudiante, posteriormente, con esta información se puede alertar y aplicar medidas preventivas tempranas con la población estudiantil.

(Fernández-Martín et al. 2019) Dentro de esta línea de investigación, el objetivo de este estudio fue desarrollar modelos explicativos y predictivos de la deserción estudiantil en los programas académicos de grado del Instituto Tecnológico de Costa Rica, con base en diferentes variables registradas en el Sistema de Indicadores de Gestión Institucional (SIGI). Entre los hallazgos se destaca que la probabilidad de deserción se relaciona con variables sociodemográficas, programa de estudio, beneficios obtenidos al ingresar, historial académico y rendimiento en el primer semestre de estudio. Además, resultó que el mejor algoritmo predictor de la deserción fue el "random forest", por identificar correctamente al potencial desertor con una probabilidad de 0.83 y por capturar al 34% de deserción real. Estos resultados son el primer paso hacia la construcción de un modelo predictivo más robusto de la deserción estudiantil, que se espera contribuya a la toma de decisiones preventivas en esta universidad.

VI. Hipótesis del trabajo (Es el aporte proyectado de la investigación en la solución del problema)

Los modelos predictivos de los sistemas inteligentes identifican los indicadores más relevantes de la deserción estudiantil universitaria

VII. Objetivo general

Implementar modelos predictivos a base de sistemas inteligentes para la deserción estudiantil en el ámbito educativo universitario

VIII. Objetivos específicos

- Identificar los indicadores de deserción estudiantil en el ámbito universitario
- Implementar modelos predictivos para el análisis de los indicadores de deserción estudiantil
- Evaluar los modelos predictivos de deserción estudiantil universitaria



IX. Metodología de investigación

El trabajo de investigación es de tipo no experimental por que la investigación se concentra en el análisis de las variables de estudio (Hernández Sampieri, Fernández Collado, y Baptista Lucio 2014), también es de tipo transaccional descriptivo porque permite conocer las variables a través de una exploración inicial en un momento específico de las variables (Horna, A. A. V. 2012).

La población estará compuesta por los estudiantes de la escuela profesional de Ingeniería de sistemas de la Universidad Nacional del Altiplano y la Universidad Andina Néstor Cáceres Velásquez, en donde para las pruebas de los modelos predictivos se utilizará los sistemas de gestión de calidad que tienen implementado estas universidades.

X. Referencias (Listar las citas bibliográficas con el estilo adecuado a su especialidad)

Condori, Mollo, y Nelson Abrahan Pablo. 2018. «Análisis predictivo de la deserción estudiantil utilizando data warehouse y minería de datos en la Universidad Nacional Jorge Basadre Grohmann – Tacna, 2012-2018». *Universidad Nacional Jorge Basadre Grohmann*.

Cuji, Blanca, Wilma Gavilanes, y Rina Sanchez. s. f. «Modelo predictivo de deserción estudiantil basado en arboles de decisión». 9.

Fernández-Martín, Tatiana, Martín Solís-Salazar, María Teresa Hernández-Jiménez, Tania Elena Moreira-Mora, Tatiana Fernández-Martín, Martín Solís-Salazar, María Teresa Hernández-Jiménez, y Tania Elena Moreira-Mora. 2019. «Un análisis multinomial y predictivo de los factores asociados a la deserción universitaria». *Revista Electrónica Educare* 23(1):73-97. doi: 10.15359/ree.23-1.5

García Franco, Jacobo. 2019. «Implementación de un Modelo Computacional basado en Reglas de Clasificación Supervisadas para la Predicción de la Deserción Estudiantil en la Universidad Peruana Unión Filial Juliaca». *Universidad Peruana Unión*.

Hernández Sampieri, Roberto, Carlos Fernández Collado, y Pilar Baptista Lucio. 2014. *Metodología de la investigación*.

Horna, A. A. V. (2012). Desde la idea hasta la sustentación: 7 pasos para una tesis exitosa. *Instituto de Investigación de la Facultad de Ciencias Administrativas y Recursos Humanos. Universidad de San Martín de Porres. Lima*.

Martínez, Julio César, y Sandra Patricia Mateus. 2020. «Propuesta de un Modelo Predictivo utilizando Aprendizaje Profundo para el análisis de deserción estudiantil en Universidades Colombianas Virtuales». *Revista Innovación y Desarrollo Sostenible* 1(1):51-57. doi: 10.47185/27113760.v1n1.8.

Pérez, Alexis Matheu, Claudio Ruff Escobar, Marcelo Ruiz Toledo, Luis Benites Gutierrez, y Germán Morong Reyes. 2018. «Modelo de predicción de la deserción estudiantil de primer año en la Universidad Bernardo O'Higgins». *Educação e Pesquisa: Revista da Faculdade de Educação da Universidade de São Paulo* 44(0):86.

Perez-Gutierrez, Boris Rainiero. 2020. «Comparación de técnicas de minería de datos para identificar indicios de deserción estudiantil, a partir del desempeño académico». *Revista UIS Ingenierías* 19(1):193-204. doi: 10.18273/revuin.v19n1-2020018.

Ramírez, Patricio E., Elizabeth E. Grandón, Patricio E. Ramírez, y Elizabeth E. Grandón. 2018. «Predicción de la Deserción Académica en una Universidad Pública Chilena a través de la Clasificación basada en Árboles de Decisión con Parámetros Optimizados». *Formación universitaria* 11(3):3-10. doi: 10.4067/S0718-50062018000300003.

Sandoval Torres, Sadath. 2013. *Modeling Chemical Reactions and Heat and Mass Transport in a Hydrating Cement Particle Using Comsol Multiphysics 3.5a*.



Vegega, Cinthia, Pablo Pytel, y María Florencia Pollo-Cattaneo. 2019. «Elicitación de Requerimientos para la Construcción de Modelos Predictivos basados en Sistemas Inteligentes dentro del Ámbito Educativo». 10.

XI. Uso de los resultados y contribuciones del proyecto

Los usos que se podrán realizar para el presente trabajo de investigación es el de ayudar en la toma de decisiones de los directivos de las universidades en estudio, como también para proponer políticas de trabajo y estrategias con la finalidad de evitar la deserción estudiantil en las universidades.

Brindar información a los docentes, directivos y autoridades universitarias acerca de los principales indicadores de deserción en los que incurren los estudiantes universitarios de las universidades en estudio, además es política del estado evitar la deserción estudiantil en las universidades, para lo cual el presente trabajo de investigación contribuirá con la solución de un problema social que se tiene en nuestra nación.

XII. Impactos esperados

i. Impactos en Ciencia y Tecnología

- Experimentar con los diferentes modelos predictivos de minería de datos, arboles de decisión y redes bayesianas para encontrar un modelo que sea eficiente en el análisis de los indicadores que están relacionados a la deserción estudiantil universitaria.
- Implementar un modelo predictivo de los sistemas inteligentes en la solución del problema planteado.

ii. Impactos económicos

- Aumento del presupuesto de las universidades en lo referente a tutoría, bienestar social del estudiante y otras áreas que se relacionan con la deserción estudiantil.
- Aumento de ingresos en la universidad debido a la cantidad de estudiantes que egresan y continúan sus estudios en forma ininterrumpida.

iii. Impactos sociales

- Implementar políticas para evitar la deserción estudiantil mediante el uso de los modelos predictivos planteados en la investigación permitirá que se aumente la cantidad de profesionales y más hogares puedan estar satisfechos con los servicios que prestan las universidades.
- El trabajo de investigación contribuirá con los profesionales encargados de analizar este problema de la deserción estudiantil que tienen las universidades mediante el uso de los modelos predictivos.

iv. Impactos ambientales

El desarrollo de sistemas inteligente provocara que se reduzca el uso de papeles de escritorio y otros materiales que normalmente son utilizados para realizar análisis de la deserción estudiantil universitario el cual muchas veces es realizado en forma manual.

XIII. Recursos necesarios (Infraestructura, equipos y principales tecnologías en uso relacionadas con la temática del proyecto, señale medios y recursos para realizar el



proyecto)

Recursos Humanos:	
Analista de sistemas	(cantidad 1)
Programador de sistemas	(cantidad 1)
Recursos Físicos:	
Computadora i7 de decima generación	(cantidad 1)
Impresora HP Smart 410	(cantidad 1)
Cartuchos de tinta	(cantidad 1)
Recursos Lógicos:	
Software Sql Server	(cantidad 1)
Software Python	(cantidad 1)
Software Anaconda	(cantidad 1)
Recursos de Escritorio:	
Papeles	(2 millares)
Bolígrafos	(una docena)
Cuadernos	(6 cuadernos)
Recurso económico:	
Inversión Inicial	S/. 6 000 (seis mil nuevos soles)
Inversión total	S/. 11 430 (once mil cuatro cientos treinta nuevos soles)

XIV. Localización del proyecto

Universidad Nacional del Altiplano, Facultad de Ingeniería Mecánica Eléctrica, Electrónica y de Sistemas, Escuela Profesional de Ingeniería de Sistemas.
Universidad Andina Néstor Cáceres Velásquez, Facultad de ingeniería de Sistemas, Escuela Profesional de Ingeniería de Sistemas.

XV. Cronograma de actividades

Actividad	MENSUAL											
	E	F	M	A	M	J	J	A	S	O	N	D
Elaboración de proyecto de investigación	X											
Revisión de literatura	X	X	X									
Recopilación de información				X	X	X						
Implementación de modelos predictivos						X	X	X				
Pruebas de sistemas inteligentes								X	X			
Redacción del borrador de trabajo de investigación										X	X	
Presentación del informe final												X

XVI. Presupuesto

Descripción	Unidad de medida	Costo Unitario (S/.)	Cantidad	Costo total (S/.)
Analista de sistemas	1 persona	400	10 meses	4000
Programador de sistemas	1 persona	300	10 meses	3000
Computadora i7 de decima generación	1 cantidad	3500	1	3500
Impresora HP Smart 410	1 cantidad	600	1	600



Cartuchos de tinta	1 cantidad	100	1	100
Software Sql Server	1 permiso	100	1	100
Software Python	1 permiso	50	1	50
Software Anaconda	1 permiso	50	1	50
Papeles	2 millares	15	2	30