ANEXO 1

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

1. Título del proyecto

Análisis del comportamiento de postulación al Programa de Estudios de Medicina Humana a la Universidad Nacional del Altiplano, periodo 2016 - 2020

2. Área de Investigación

Área de investigación	Línea de Investigación	**Disciplina OCDE**
SISTEMA, COMPUTACIÓN E INFORMÁTICA	INGENIERIA COMPUTACIONAL Y SISTEMAS	

3. Duración del proyecto (meses)

12

4. Tipo de proyecto

Individual	○
Multidisciplinario	X
Director de tesis pregrado	○

4. Datos de los integrantes del proyecto

Apellidos y Nombres	**ADOLFO CARLOS JIMENEZ CHURA**
Escuela Profesional	**INGENIERÍA DE SISTEMAS**
Celular	**994125511**
Correo Electrónico	**ajimenez@unap.edu.pe**

Apellidos y Nombres	**VILLALTA PACORI JULIO CESAR**
Escuela Profesional	**CIENCIAS FÍSICO MATEMÁTICAS**
Celular	**958634975**
Correo Electrónico	**juvipa11@unap.edu.pe**

Apellidos y Nombres	**ZANABRIA ORTEGA MILDER**
Escuela Profesional	**INGENIERÍA DE SISTEMAS**
Celular	**951522828**
Correo Electrónico	**mzanabria@unap.edu.pe**

Apellidos y Nombres	**FERNANDEZ CHAMBI MAYENKA**
Escuela Profesional	**INGENIERÍA DE SISTEMAS**
Celular	**951351766**
Correo Electrónico	**mfernandez@unap.edu.pe**

I. Título (El proyecto de tesis debe llevar un título que exprese en forma sintética su contenido, haciendo referencia en lo posible, al resultado final que se pretende lograr. Máx. palabras 25)

Análisis del comportamiento de postulación al Programa de Estudios de Medicina Humana a la Universidad Nacional del Altiplano, periodo 2016 - 2020

II. Resumen del Proyecto de Tesis (Debe ser suficientemente informativo, presentando -igual que un trabajo científico- una descripción de los principales puntos que se abordarán, objetivos, metodología y resultados que se esperan**)**

El presente trabajo de investigación tiene como objetivo realizar un análisis del comportamiento de los postulantes al Programa de Estudios de Medicina Humana. La muestra a tomar será de los exámenes generales comprendidos entre los años 2016 - 2020 de la Universidad Nacional del Altiplano. Se usará técnicas estadísticas y el lenguaje de programación Python para determinar el comportamiento de los postulantes e ingresantes al programa de estudios en función a variables de estudio como: el género, la edad, número de postulaciones, tipo de colegio, entre otras. Los resultados esperados permitirán tomar las medidas adecuadas para el bien de los estudiantes del programa de estudios.

III. Palabras claves (Keywords) (Colocadas en orden de importancia. Máx. palabras: cinco**)**

Análisis estadístico, machine learning, modelo, predicción, Python, universidad.

IV. Justificación del proyecto (Describa el problema y su relevancia como objeto de investigación. Es importante una clara definición y delimitación del problema que abordará la investigación, ya que temas cuya definición es difusa o amplísima son difíciles de evaluar y desarrollar**)**

El análisis de datos estadísticos es el proceso que permite realizar la interpretación de los datos, con el objetivo de tomar las decisiones más eficaces en toda organización gracias a que existen muchas técnicas de análisis estadístico que van desde una media de datos hasta la aplicación de modelos estadísticos para inferir sobre las relaciones dentro de los datos. Por otro lado, machine learning involucra y utiliza datos, los cuales son descritos y usados en un marco estadístico. Muchos modelos estadísticos pueden hacer predicciones, pero no todos lo hacen con precisión y para llegar a la precisión se necesita crear o usar un algoritmo que haga el trabajo, por ejemplo, predecir el precio de una vivienda en función a muchas variables, entonces el enfoque sería hacía el machine learning. El estudio del comportamiento de postulación al Programa de Estudios de Medicina Humana involucra aplicar una serie de modelos estadísticas o machine learning a poner en práctica en las variables de estudio como: el género, edad, número de postulaciones, tipo de colegio, entre otras.

Python ha ido ganando en seguidores en comunidades de software libre, científica y educacional, por su facilidad de uso y posibilidad de concentrarse en diferentes tipos de problemas. Este lenguaje facilita portabilidad y el acceso a las librerías necesarias para la preparación de los datos, modelado y análisis de datos y finalmente la visualización.

V. Antecedentes del proyecto (Incluya el estado actual del conocimiento en el ámbito nacional e internacional. La revisión bibliográfica debe incluir en lo posible artículos científicos actuales, para evidenciar el conocimiento existente y el aporte de la Tesis propuesta. Esto es importante para el futuro artículo que resultará como producto de este trabajo)

(Evidencia 47, 2020), en su informe se entregan detalles sobre el proceso 2019 (admisión 2020) mediante algunas estadísticas descriptivas y se analiza la distancia entre el postulante y los establecimientos a los cuales postula, entregando un marco de trabajo sugerido para este cálculo. Los resultados indican que la distancia de las postulaciones se ve afectada por variables como orden de preferencia, nivel, ruralidad y dependencia, mientras que no se observan grandes diferencias al considera el sexo del alumno. Al analizar las cartografías se pudo observar que las capitales regionales presentan una mayor demanda por cupos, y que la oferta de estos fue especialmente insuficiente en comunas de la provincia de Santiago.

(Sanchez, Villena 2016) en su investigación presenta los resultados de un estudio cuyo propósito fue conocer y comparar la opinión de un grupo de estudiantes de Ingeniería respecto de las características más representativas de los lenguajes de programación R y Python a la luz de su uso en el análisis estadístico de un caso de estudio real. A los 61 estudiantes participantes, quienes cursaban asignaturas de probabilidad y estadística; se les presentó un caso de estudio real y contextual para su examen. El 98,3% realizó el análisis y documentó sus resultados de acuerdo con los lineamientos establecidos; en el caso de R, el 81,97% optó por usar paquetes externos al núcleo básico para elaborar el informe reproducible, el 52,45% indicó usar R Markdown en detrimento de otra tecnología. En el caso de Python, el 88,52% usó las librerías Scipy, Matplotlib, Numpy y Pandas para el análisis; el 67,21% utilizó Markdown-Python para la redacción del informe. De Python destacaron la facilidad para escribir código; de R distinguieron su potencia para organizar, visualizar y efectuar cálculo estadístico. Se recomienda efectuar un estudio experimental que permita probar métodos pedagógicos que integren prácticas distintas a las predominantes durante las últimas tres décadas en la educación estadística.

(Verjaga-Felgueras, 2018), en su trabajo de fin de grado se ha hecho un estudio (Teórico y práctico) sobre Big Data y tecnologías de tratamiento de datos. La parte teórica se centra en una profunda descripción de Big Data, los algoritmos más utilizados en el tratamiento de datos y se profundiza en la tecnología Spark. En la parte práctica se ha hecho una preparación del entorno de trabajo con todas las instalaciones y configuraciones necesarias de Spark y se ha creado un clúster con varios nodos. Todo ello en dos escenarios: máquinas virtuales en el ordenador portátil como modo experimental y computadores físicos donde se han hecho pruebas

ejecutando un algoritmo escrito en Python que incluye técnicas de Machine Learning. Se va a evaluar la eficacia de Spark haciendo un estudio comparativo con tiempos de ejecución y considerando la distribución de las tareas.

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

El análisis del comportamiento de postulación es irrelevante para el Programa de Estudios de Medicina Humana a la Universidad Nacional del Altiplano, periodo 2016 – 2020

VII. Objetivo general

Realizar un análisis del comportamiento de postulación al Programa de Estudios de Medicina Humana a la Universidad Nacional del Altiplano, periodo 2016 - 2020

VIII. Objetivos específicos

Realizar la limpieza y el preprocesamiento de los datos de los procesos de admisión
Aplicar las técnicas de análisis estadístico a los datos de los procesos de admisión
Visualización de la información de los procesos de admisión

IX. Metodología de investigación (Describir el(los) método(s) científico(s) que se empleará(n) para alcanzar los objetivos específicos, en forma coherente a la hipótesis de la investigación. Sustentar, con base bibliográfica, la pertinencia del(los) método(s) en términos de la representatividad de la muestra y de los resultados que se esperan alcanzar. Incluir los análisis estadísticos a utilizar**)**

Para lograr los objetivos planteados, se tomará en cuenta una muestra significativa de los procesos de admisión ordinario general comprendidos entre los períodos 2016 – 2020 de nuestra universidad.
De acuerdo a (Hernández Sampieri 2016) la presente investigación es de tipo aplicada pues tiene como finalidad la búsqueda y consolidación del marco teórico para aplicarlos en el análisis y evaluación del comportamiento de las variables.
De nivel correlacional ya que este estudio tiene como propósito conocer la relación o grado de asociación que existe entre las variables en una muestra o contexto en particular.

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

Guerra, L. (2012). Modelo para el análisis del comportamiento de los estudiantes durante actividades colaborativas a través de Internet. https://www.tesisenred.net/bitstream/handle/10803/113562/TLGT1de1.pdf?sequence=1
Centro de Estudios Mineduc. (2020). Santiago de Chile, Chile. Análisis de postulaciones Sistema de Admisión Escolar. Recuperado de https://www.researchgate.net/profile/Fabian-Ramirez-11/publication/349605104_Analisis_de_Postulaciones_Sistema_de_Admision_Esc

olar/links/6037cdb9a6fdcc37a8514079/Analisis-de-Postulaciones-Sistema-de-Admision-Escolar.pdf
Sánchez-Villena, A. (2019). Uso de programas estadísticos libres para el análisis de datos: Jamovi, Jasp y R. *Revista perspectiva*, *20*(1), 112-114.
Acosta, J. H. C., Solís, M. C., & Contreras, I. S. (2019). Programación computacional y análisis de datos en educación estadística. *Areté: Revista Digital del Doctorado en Educación de la Universidad Central de Venezuela*, *5*(9), 11-27.
Verjaga-Felgueras, M. E. (2018). Análisis de datos y extracción de conocimiento utilizando Big Data.

XI. Uso de los resultados y contribuciones del proyecto (Señalar el posible uso de los resultados y la contribución de los mismos)

Se acercará a la UNA – Puno con fines de estadísticos e históricos de las postulaciones realizadas para obtener una vacante en el Programa de Estudios de Medicina humana y servirá como base para realizar otras investigaciones.

XII. Impactos esperados

i. Impactos en Ciencia y Tecnología

Uso del lenguaje de programación Python orientado a objetos y con posibilidad de desarrollar proyectos de investigación.

ii. Impactos económicos

Al convertirse a la UNA - Puno el resultado del análisis será más productivo, lo que redundaría en mayores ingresos económicos.

iii. Impactos sociales

La toma de decisiones en las empresas e instituciones se realizará en el menor tiempo posible con el uso de los resultados de la investigación.

iv. Impactos ambientales

Ahorro de recursos de lectura en disco, lo que genera un consumo de energía bajo ya que los datos se almacenan en memoria.

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: Sólo nuestras personas son el recurso humano para la elaboración del presente proyecto.

Recursos Materiales:

Hardware:
Computadora Personal Intel Core i5.
Impresora EPSON STYLUS 355.
Memoria USB 32.Gb

Software:
Windows 10
Microsoft Office 2016
Microsoft Project 2016
Mozilla Firefox
Lenguaje de programación python
Gestor de base de datos MySQL

Servicios:
Conexión a Internet.

XIV. Localización del proyecto (indicar donde se llevará a cabo el proyecto)

Puno - ciudad

XV. Cronograma de actividades

Actividad	Trimestres											
	E	F	M	A	M	J	J	A	S	O	N	D
Estudio preliminar y determinación de requerimientos												
Diseño y desarrollo del proyecto												
Prueba e implantación del Proyecto												
Evaluación del Proyecto y elaboración del artículo final												

XVI. Presupuesto

Descripción	Unidad de medida	Costo Unitario (S/.)	Cantidad	Costo total (S/.)
Responsable del proyecto	persona	2400.00	03	7 200.00
Materiales	unidades	30.00	30	900.00
Servicio de internet	meses	90.00	12	1 080.00
Total				9 180.00