



Universidad Nacional de
Ingeniería

Maestría en Data
Science:

Tesis 4: Redacción del
proyecto de tesis

Redacción del proyecto de tesis III

Mg. Ing. Luis Alberto Sánchez Alvarado

HIPÓTESIS

Hipótesis

Son explicaciones tentativas del fenómeno o problema investigado formuladas como proposiciones o afirmaciones. La hipótesis se deriva directamente de la pregunta de investigación y debe ser una afirmación específica y comprobable, apoyado en el estado del arte.



Ejemplo:

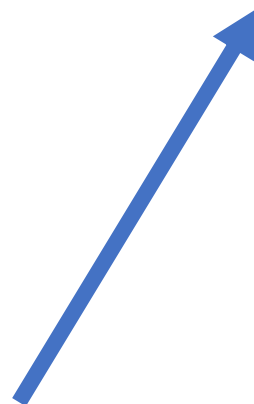
Sobre el estudio de Factores
Determinantes de viralidad en
TikTok en Perú



REDACCIÓN:

PREGUNTAS DE INVESTIGACIÓN

¿Cuáles son los factores determinantes de la viralidad en TikTok en Perú y cómo pueden ser identificados y analizados mediante un modelo predictivo construido a partir del web scraping de videos del 2023, con el fin de proporcionar información útil a los creadores de contenido, marcas y profesionales del marketing para optimizar sus estrategias de contenido y maximizar su impacto en la plataforma?



HIPÓTESIS DE INVESTIGACIÓN

Los factores determinantes de la viralidad en TikTok en Perú, identificados y analizados mediante un modelo predictivo basado en datos recopilados a través de web scraping de videos del 2023, **incluyen el uso de música popular, la frecuencia de publicación, la participación en desafíos virales, y el uso de hashtags específicos**, que ayudan a optimizar las estrategias de marketing en la plataforma.

Apoyado en el estado del arte

Otro ejemplo:

PREGUNTAS DE INVESTIGACIÓN

¿El algoritmo de Random Forest proporciona mejores indicadores de rendimiento (como precisión, sensibilidad y especificidad) cuando se entrena con datos balanceados en comparación con datos no balanceados para la predicción de aceptación de leads en una campaña de tarjeta de crédito?



HIPÓTESIS DE INVESTIGACIÓN

El algoritmo de Random Forest proporciona mejores indicadores de rendimiento, tales como precisión, sensibilidad y especificidad, cuando se entrena con datos balanceados en comparación con datos no balanceados para la predicción de aceptación de leads en una campaña de tarjeta de crédito en una entidad financiera

Otro ejemplo:

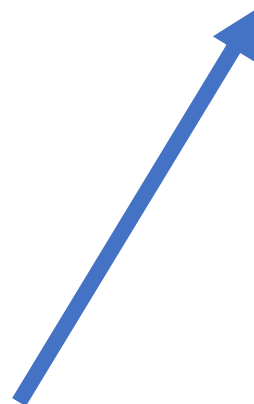
PREGUNTAS DE INVESTIGACIÓN

¿Qué tan eficaces son los algoritmos de clasificación de aprendizaje automático en la predicción de la curación de malformaciones arteriovenosas cerebrales en pacientes tratados con radiocirugía, y cuáles son los factores clave que influyen en la precisión de estas predicciones?



HIPÓTESIS DE INVESTIGACIÓN

Los algoritmos de clasificación de aprendizaje automático pueden predecir con alta precisión la curación de malformaciones arteriovenosas cerebrales en pacientes tratados con radiocirugía, y factores clínicos **específicos como el tamaño de la malformación, la dosis de radiación y las características del paciente (edad, comorbilidades)** influyen significativamente en la precisión de estas predicciones.



Apoyado en el estado del arte

Otro ejemplo:

PREGUNTAS DE INVESTIGACIÓN

¿Cómo pronosticar la cantidad de alumnos aprobados y desaprobados en los cursos del Programa de Estudios Básicos de la Universidad mediante el uso de algoritmos de Machine Learning.?



HIPÓTESIS DE INVESTIGACIÓN

Si se aplican los algoritmos de Machine Learning, entonces, se incrementará la eficiencia en la predicción de la cantidad de alumnos aprobados y desaprobados en los cursos del Programa de Estudios Básicos de la Universidad

Otro ejemplo:

PREGUNTAS DE INVESTIGACIÓN

¿Cuáles son las variables que influyen en las ventas de acuerdo a los servicios que ofrece la empresa Travico Perú S.A.C.?



HIPÓTESIS DE INVESTIGACIÓN

Mediante la aplicación de ciencia de datos y de aprendizaje automático se identificará cuáles son las variables más relevantes que influyen en las ventas de los servicios, de tal manera que esta información permita identificar oportunidades de negocio y priorizar esfuerzos

METODOLOGÍA

Metodología

Conjunto de técnicas, métodos y procedimientos sistemáticos empleados para llevar a cabo toda la investigación de manera estructurada y lógica. Este apartado incluye la descripción detallada del diseño de investigación, las estrategias de recolección y análisis de datos, y la aplicación de algoritmos de Machine Learning o modelos estadísticos. La metodología se encuentra fundamentada en el marco teórico y responde a la pregunta de investigación.




Redacción del proyecto de tesis




Población en estudio

Descripción de los elementos por analizar en el estudio



Descripción de los datos

Mencionar de donde se tomarán los datos y en que forma, describiendo en lo posible las variables a considerar




Análisis para la preparación de los datos

Algoritmos para identificación de atípicos, imputación de datos, etc.



Modelamiento

Describir los modelos que serán puestos a pruebas en el estudio



Evaluación de los modelos

Mencionar las métricas que se usarán

RESUMEN

Resumen

Proporciona una visión general del trabajo que se realizará y permite a los lectores entender el propósito y la importancia de la investigación.

Resumen y palabras clave

(Máximo 200 palabras, 5 palabras clave)



Redacción del proyecto de tesis

Contexto

Problema

Datos a emplear

Anterior al año 2017 la SUNAT, realizaba de manera aleatoria y empírica el control aduanero a los pasajeros de vuelos internacionales que arriban al aeropuerto Internacional Jorge Chávez. La selección para realizar el control aduanero dependía de la experiencia del funcionario y se realizaba antes de la salida del pasajero del recinto aduanero, no siendo un control muy efectivo. Con Resolución Directoral N° 139-2017-MTC/12 en 2018 SUNAT logró acceder a la transmisión de Información Anticipada de Pasajeros (API), esto es el envío de datos electrónicos con información de los pasajeros que reemplaza el manifiesto de pasajeros físico que entregaban las compañías aéreas. Aprovechando estos datos y con el fin de optimizar los recursos de la administración aduanera se busca implementar en los controles aduaneros una herramienta tecnológica para identificar perfiles de riesgo, con el fin de procesar los datos electrónicos del API y mediante técnicas de gestión de riesgo permita identificar perfiles de posibles pasajeros propensos a cometer alguna infracción aduanera. Esto permitirá identificar a quienes revisar antes del arribo del vuelo facilitando el control, basado en un análisis y no en forma subjetiva.

gestión de riesgo, perfil de riesgo, Información Anticipada de Pasajeros (API).

Objetivo general+algoritmos

Importancia / aporte

PLAN DE TRABAJO

ACTIVIDADES	Set-24				Oct-24				Nov-24				Dic-24			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1. Documentación																
2. Redacción de informe																
Planteamiento del Problema			X	X	X											
Métodos de solución						X	X									
Desarrollo de la solución								X	X							
Resultados										X	X					
Conclusiones										X	X					
3. Preparación del documento según normas APA																
a. Edición de contenidos											X	X				
b. Edición de cuadros, gráficos y figuras													X	X		
c. Edición de anexos													X	X		
4. Edición del documento final																