

UNIVERSIDAD PERUANA UNIÓN
FACULTAD DE INGENIERÍA Y ARQUITECTURA
Escuela Profesional de Ingeniería de Sistemas



**Modelo de Machine Learning Basado en los Factores de Riesgo para
Predecir el Sobrepeso y la Obesidad**

Tesis para obtener el Título Profesional de Ingeniero de Sistemas

Autor:

Willian Quispe Calcina
Victor Lee Chuquin Huarachi
Alex Alejandro Azurin Pereyra

Asesor:

Msc. Ferdinand Edgardo Pineda Ancco

Juliaca, febrero de 2025

DECLARACIÓN JURADA DE ORIGINALIDAD DE TESIS

Yo Msc. Ferdinand Edgardo Pineda Ancco, docente de la Facultad de Ingeniería y Arquitectura, Escuela Profesional de Ingeniería de Sistemas, de la Universidad Peruana Unión.

DECLARO:

Que la presente investigación titulada: **“MODELO DE MACHINE LEARNING BASADO EN LOS FACTORES DE RIESGO PARA PREDECIR EL SOBREPESO Y LA OBESIDAD”** de los autores **Willian Quispe Calcina, Victor Lee Chuquin Huarachi y Alex Alejandro Azurin Pereyra**, tiene un índice de similitud de 13% verificable en el informe del programa Turnitin, y fue realizada en la Universidad Peruana Unión bajo mi dirección.

En tal sentido asumo la responsabilidad que corresponde ante cualquier falsedad u omisión de los documentos como de la información aportada, firmo la presente declaración en la ciudad de Juliaca, a los 2 días del mes de mayo del año 2025



Msc. Ferdinand Edgardo Pineda

Ancco

Asesor



ACTA DE SUSTENTACIÓN DE TESIS

En Puno, Arequipa, Villa Chillonquani, a 10 día(s) del mes de febrero del año 2025 siendo las 10:00 horas, se reunieron los miembros del Jurado en la Universidad Peruana Unión Campus Juliaca, bajo la dirección del (de la) presidente(a):

Ma. Bonuzet Branas Herrera Yucra el (el) secretario(a): Mg. Harold Mamani Pora

y los demás miembros Mg. Xenias Suboya Rios y el (la) asesor(a) Max Ferdinand Edgardo Pardo Anco

con el propósito de administrar el acto académico de sustentación de la tesis titulado Modelo de Machine Learning Basado en los Factores de Riesgo para Predecir el Sobrepeso y la Obesidad

- del(los) bachiller(es): a) William Quijpe Galena b) Alex Alejandro Azurin Pereyra c) Victor Lee Shuquin Huarachi

conducente a la obtención del título profesional de Ingeniero de Sistemas

El Presidente inició el acto académico de sustentación invitando al (a la) / a (los) (las) candidato(a)s hacer uso del tiempo determinado para su exposición. Concluida la exposición, el Presidente invitó a los demás miembros del jurado a efectuar las preguntas, y adaptaciones pertinentes, las cuales fueron absueltas por al (a la) / a (los) (las) candidato(a)s. Luego, se produjo un receso para las deliberaciones y la emisión del dictamen del jurado. Posteriormente, el jurado procedió a dejar constancia escrita sobre la evaluación en la presente acta, con el dictamen siguiente:

Bachiller (a): William Quijpe Galena

Table with columns: CALIFICACIÓN, ESCALAS (Vigesimal, Literal, Cualitativa), Mérito

Bachiller (b): Alex Alejandro Azurin Pereyra

Table with columns: CALIFICACIÓN, ESCALAS (Vigesimal, Literal, Cualitativa), Mérito

Bachiller (c): Victor Lee Shuquin Huarachi

Table with columns: CALIFICACIÓN, ESCALAS (Vigesimal, Literal, Cualitativa), Mérito

(*) Ver parte posterior

Finalmente, el Presidente del jurado invitó al (a la) / a (los) (las) candidato(a)s a ponerse de pie, para recibir la evaluación final y concluir el acto académico de sustentación procediéndose a registrar las firmas respectivas.

Signature of President

Signature of Asesor(a)

Signature of Bachiller (a)

Signature of Miembro

Signature of Bachiller (b)

Signature of Secretaria

Signature of Miembro

Signature of Bachiller (c)

Modelo de Machine Learning Basado en los Factores de Riesgo para Predecir el Sobrepeso y la Obesidad

Risk Factor-Based Machine Learning Model to Predict Overweight and Obesity

Willian Quispe, Victor Lee Chuquin, Alex A. Azurin

Resumen

Este estudio presenta un modelo de Machine Learning basado en factores de riesgo para predecir el sobrepeso y la obesidad, destacando su importancia en la prevención y control de esta condición. La metodología incluye limpieza de datos, imputación de valores faltantes, selección de características y evaluación de modelos de regresión como RFR, DTR, SVR y XGB. Los resultados muestran que el GBR alcanzó el mejor desempeño, seguido de cerca por XGB, al predecir el índice de masa corporal (IMC) utilizando 20 características clave, como colesterol LDL, perímetro abdominal, medidas diastólicas, presión arterial, hemoglobina y glucosa. La reducción de características mejoró notablemente el rendimiento de los modelos, especialmente los basados en técnicas de Boosting. Estos hallazgos destacan la precisión de los modelos de Machine Learning en la predicción del IMC y su potencial para la identificación temprana de riesgos relacionados con la obesidad en la población adulta peruana.

Palabras clave: Machine Learning, Obesidad, Sobrepeso, Modelos de Regresión.

Abstract

This study presents a Machine Learning model based on risk factors to predict overweight and obesity, highlighting its importance in the prevention and control of this condition. The methodology includes data cleaning, missing value imputation, feature selection and evaluation of regression models such as RFR, DTR, SVR and XGB. The results show that RBF achieved the best performance, closely followed by XGB, in predicting body mass index (BMI) using 20 key features, such as LDL cholesterol, abdominal circumference, diastolic measurements, blood pressure, hemoglobin and glucose. Feature reduction significantly improved the performance of the models, especially those based on boosting techniques. These findings highlight the accuracy of Machine Learning models in BMI prediction and their potential for early identification of obesity-related risks in the Peruvian adult population.

Keywords: Machine Learning, Obesity, Overweight, Regression Models.