



**UNIVERSIDAD TECNOLÓGICA DE PANAMÁ
FACULTAD DE INGENIERÍAS DE SISTEMAS
COMPUTACIONALES**



Licenciatura en Ingeniería de Sistemas y Computación

Probabilidad Aplicada a la TIC

**Relación entre Los Hábitos Alimenticios, la Actividad Física y Los Niveles De
Obesidad: Un Análisis Predictivo Con Machine Learning**

Josip De Lora 8-1028-836

Mateo Del Giudice E-229-194

Rene Rodriguez 8-1034-1629

Ian Nieto 8-1033-124

Dhruv Sakhrani 8-1023-30

Profesor Juan Marcos Castillo, PhD

Grupo 1IL124

Año 2025

Nombre de la investigación:

Relación entre los hábitos alimenticios, la actividad física y los niveles de obesidad: un análisis predictivo con machine learning

Descripción de la base de datos a utilizar:

La base de datos contiene registros de personas con distintas características relacionadas con la salud y el estilo de vida. Incluye variables como:

Nombre del campo	Tipo de dato	Descripción en español
Gender	Categorico	Género
Age	Continuo	Edad
Height	Continuo	Estatura
Weight	Continuo	Peso
family_history_with_overweight	Binario	¿Algún familiar sufre o ha sufrido de sobrepeso?
FAVC	Binario	¿Consumes alimentos altos en calorías con frecuencia?
FCVC	Entero	¿Sueles incluir vegetales en tus comidas?
NCP	Continuo	¿Cuántas comidas principales realizas al día?
CAEC	Categorico	¿Consumes alimentos entre comidas?
SMOKE	Binario	¿Fumas?
CH2O	Continuo	¿Cuánta agua bebes diariamente?
SCC	Binario	¿Controlas las calorías que consumes diariamente?
FAF	Continuo	¿Con qué frecuencia realizas actividad física?
TUE	Entero	¿Cuánto tiempo usas dispositivos tecnológicos como celulares, videojuegos, TV, etc.?
CALC	Categorico	¿Con qué frecuencia consumes alcohol?

MTRANS

Categorico ¿Qué medio de transporte utilizas con mayor frecuencia?

NObeyesdad

Categorico Nivel de obesidad

Esta base es ampliamente utilizada para modelos de clasificación y regresión en temas de salud pública.

Fuente:

<https://archive.ics.uci.edu/dataset/544/estimation+of+obesity+levels+based+on+eating+habits+and+physical+condition>

Introducción:

El análisis de datos se ha convertido en una herramienta clave para entender el comportamiento de distintos fenómenos en diversas áreas. En este caso, se trabaja con una base de datos que incluye información sobre características personales, hábitos alimenticios y actividad física de diferentes individuos. El objetivo del análisis es explorar cómo se relacionan estas variables entre sí, utilizando herramientas como tablas dinámicas, gráficas y medidas estadísticas.

A través del análisis descriptivo se busca resumir y visualizar la información de forma clara, mientras que mediante modelos de regresión se pretende identificar patrones que puedan ser útiles para realizar predicciones. Este enfoque permite transformar los datos en conocimiento que puede apoyar la toma de decisiones o estudios más complejos en el futuro.

Justificación:

La obesidad se ha convertido en uno de los principales problemas de salud pública a nivel mundial, afectando tanto a adultos como a jóvenes. La obesidad mundial en adultos se ha más que duplicado desde 1990, y la obesidad en adolescentes se ha cuadruplicado. Sus consecuencias no se limitan únicamente al aumento de peso, sino que también incrementan el riesgo de desarrollar enfermedades crónicas como diabetes tipo 2, hipertensión, problemas cardiovasculares y ciertos tipos de cáncer. Por esta razón, resulta fundamental contar con herramientas que permitan identificar oportunamente los factores que contribuyen a su aparición.

Este proyecto se justifica en la necesidad de analizar cómo los hábitos alimenticios y los comportamientos relacionados con la actividad física influyen directamente en el desarrollo de la obesidad. Al aplicar modelos estocásticos, se busca generar un enfoque predictivo que facilite la prevención de la obesidad.

De esta manera, el estudio aportará información valiosa para disminuir la prevalencia de obesidad y mejorar la calidad de vida de la población.

Antecedentes:

Numerosos estudios han demostrado que los hábitos de alimentación y los comportamientos sedentarios están correlacionados con un incremento en el índice de masa corporal (IMC), especialmente en poblaciones jóvenes y adolescentes. Estas relaciones han sido modeladas mediante técnicas estadísticas y de machine learning para identificar patrones de riesgo y clasificar niveles de obesidad según variables como actividad física y hábitos alimenticios.

Además, investigaciones que usan modelos estocásticos han mostrado que combinaciones de cadenas de Markov permiten modelar con precisión las transiciones entre categorías de peso (normal, sobrepeso y obesidad) en el tiempo, ofreciendo una visión robusta de la evolución de la obesidad poblacional. En el campo de técnicas predictivas, se han aplicado

algoritmos como regresión LASSO, Random Forest, Gradient Boosting y máquinas de gradiente estocástico alcanzando niveles de precisión aceptables para clasificar el estado de obesidad según variables sociodemográficas, hábitos alimenticios y actividad.

En resumen, existe una base sólida de investigaciones que respaldan la modelación de la obesidad mediante enfoques vinculados a los hábitos de alimentación y actividad física, utilizando tanto modelos descriptivos como estocásticos y predictivos basados en datos reales.

Definición del Problema:

La obesidad es uno de los problemas de salud pública más importantes en el mundo. Factores como la alimentación inadecuada, el sedentarismo, y malos hábitos de vida son determinantes clave en su desarrollo. Entender cómo estos factores interactúan y predicen el nivel de obesidad permite generar mejores políticas de prevención y tratamiento. Esta investigación busca explorar dichos factores usando análisis de datos y modelos predictivos.

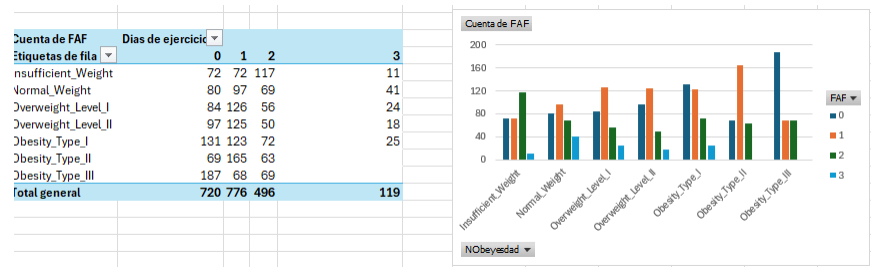
Avance de Estadísticos:

Edades		Estatura		Peso	
Mean	24.315964	Mean	1.701677353	Mean	86.58605813
Standard Error	0.138360916	Standard Error	0.002030766	Standard Error	0.570047193
Median	23	Median	1.700499	Median	83
Mode	21	Mode	1.7	Mode	80
Standard Deviation	6.35707808	Standard Deviation	0.09330482	Standard Deviation	26.19117175
Sample Variance	40.41244171	Sample Variance	0.008705789	Sample Variance	685.9774774
Kurtosis	2.798581823	Kurtosis	-0.562948893	Kurtosis	-0.699898158
Skewness	1.521326133	Skewness	-0.012854646	Skewness	0.255410495
Range	47	Range	0.53	Range	134
Minimum	14	Minimum	1.45	Minimum	39
Maximum	61	Maximum	1.98	Maximum	173
Sum	51331	Sum	3592.240893	Sum	182783.1687
Count	2111	Count	2111	Count	2111

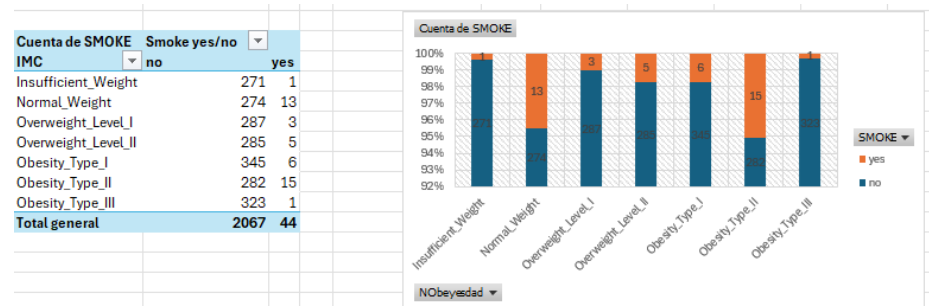
Análisis descriptivo

Es una forma de resumir y entender los datos que tienes. Sirve para ver cómo se comportan las variables, qué valores se repiten más, cuáles son los promedios, y si hay alguna relación entre ellas. También se usan gráficas, tablas y porcentajes para mostrar la información de manera clara.

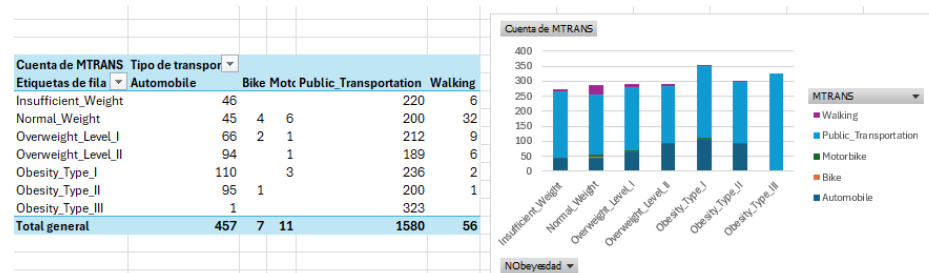
IMC con respecto a cuántos días de ejercicios realizan. Se observa que la mayor cantidad de personas que no realizan ejercicio son las personas con obesidad tipo 3.



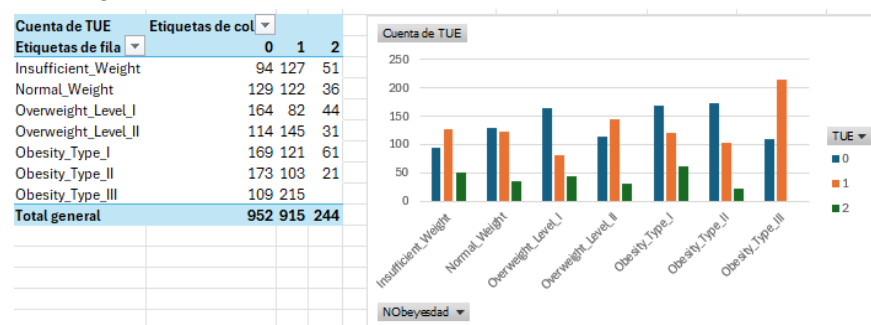
IMC con respecto al hábito de fumar. Se observa que las personas que más fuman en la base de datos son los de grado de obesidad tipo 2.



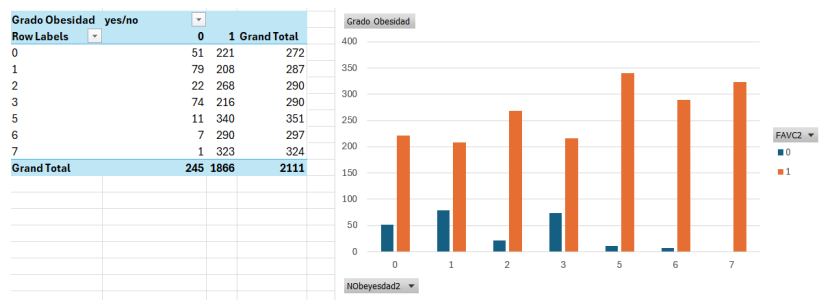
IMC con respecto a de qué manera se transporta la persona: caminando, transporte público, motocicleta, bicicleta, carro. Se observa que casi ninguna persona suele caminar en su día a día para transportarse.



IMC con respecto al tiempo de uso de pantalla. Se observa el mayor índice de uso en el último grado de obesidad.



Se relaciona el grado de obesidad con el campo de “Comes comida con altas calorías”, donde se puede visualizar que la mayoría de personas que consumen comidas con altas calorías son las personas en el grado de obesidad.



Se relaciona el consumo del agua con el nivel de consumo de litro de agua diario. Se observa una tendencia donde en la gran mayoría de respuestas seleccionan 1 litro de agua diario.

