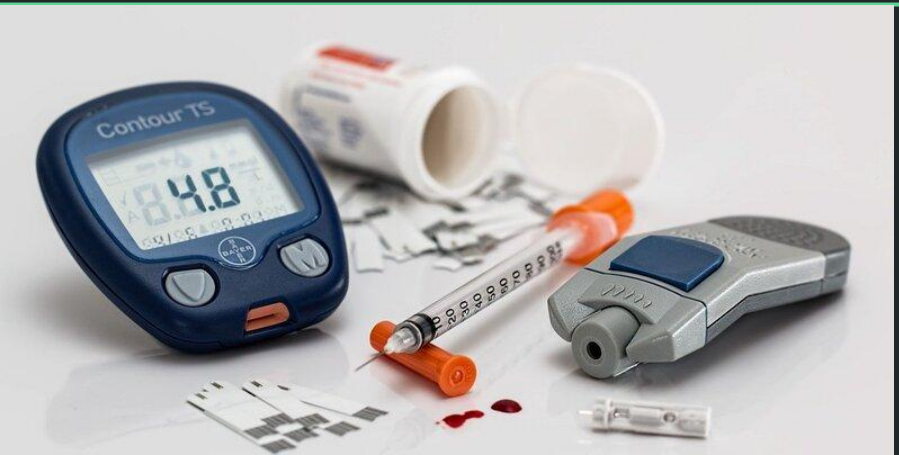
Diabetes Health Indicators Prediction

By - Rohan Sanjay Dange
Data Mining and Machine Learning
A20517920



AGENDA

Overview of Diabetes Health.
Importance of Predicting Diabetes health Indicators.
Project Settings.
Data Set with its Features and Target variables.
EDA
Model and Evaluation Metrics
Demonstration.

Overview of Diabetes Health

- Diabetes is a prevalent chronic disease in the US that impacts millions of Americans.
- It is characterized by the body not making enough insulin or being unable to use it effectively, leading to high blood sugar levels.
- Complications associated with diabetes can cause heart disease, vision loss, amputation, and kidney disease.
- Early diagnosis and lifestyle changes can help mitigate the effects of the disease.
- Type II diabetes is the most common form and its prevalence varies by age, education, income, location, and race, with lower socioeconomic status groups bearing much of the burden.
- The disease places a significant financial burden on the economy, with costs reaching $400 billion annually for diagnosed, undiagnosed, and prediabetes cases

Importance of Predicting Diabetes Health Indicators

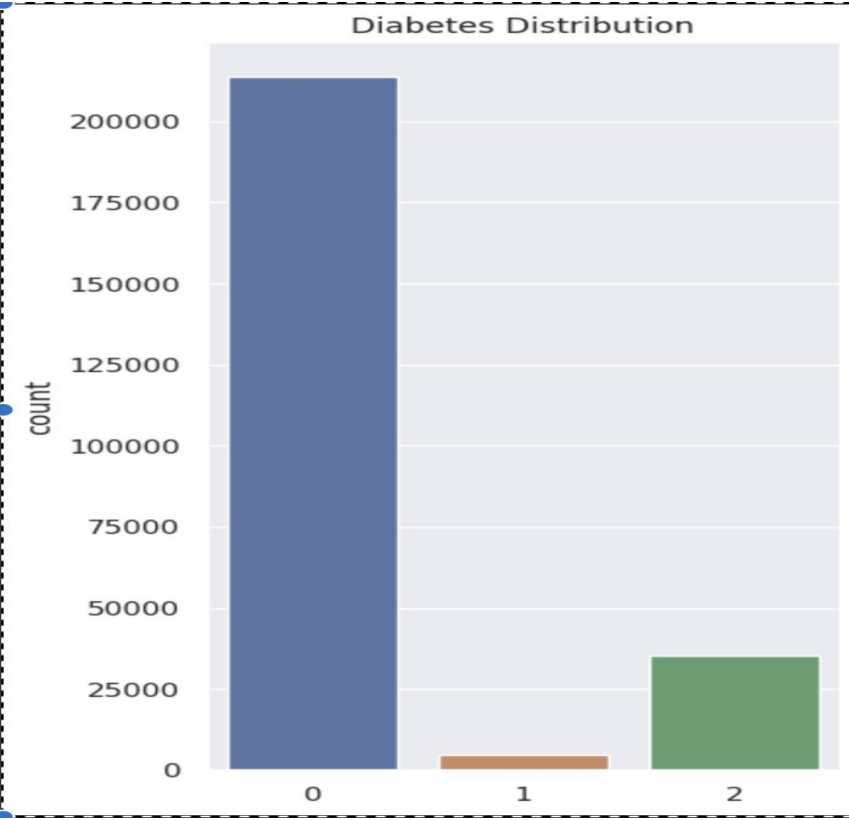
- Early detection of diabetes can lead to timely treatment and better patient outcomes.
- Identifying patients at high risk of complications can allow for targeted interventions to prevent or delay these complications.
- Personalizing treatment plans based on health indicators can optimize treatment and improve patient outcomes.
- Predictive models can help healthcare professionals better understand the risks and outcomes associated with diabetes.
- Predicting health indicators can aid in the development of public health strategies and policies to reduce the prevalence of diabetes and its complications.

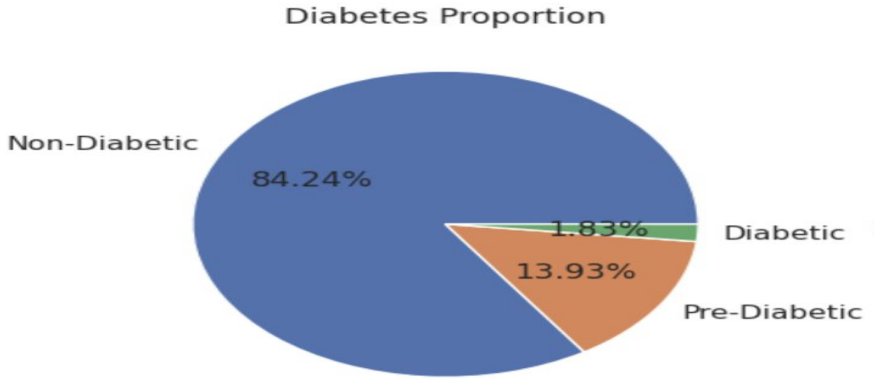
PROJECT SETTINGS

Data Set With its Features and Target Variables

- The dataset consists of 253,680 survey responses with 21 feature variables and a target variable called "Diabetes Level."
- The feature variables include information such as (HighBP, HighChol, CholCheck, BMI, Smoker, Stroke, HeartDiseaseorAttack, PhysActivity, Fruits, Veggies, HvyAlcoholConsump, AnyHealthcare, NoDocbcCost, GenHlth, MentHlth, PhysHlth, DiffWalk, Sex, Age, Education, and Income)
- The target variable, "Diabetes Level," is assigned the values 0 (no diabetes or only during pregnancy), 1 (pre-diabetes), and 2 (diabetes).
- The train-test split performed is 80% of the data for training the model and 20% of the data for testing the model
- Outliers in the response variable were identified and treated to ensure data integrity and reliability.

Exploratory Data Analysis





MODELS AND EVALUATION METRICS

	Accuracy	F-1	Recall	Precision	AUC
Logistic Regression	0.84	0.84	0.84	0.84	0.77
Naive Bayes	0.76	0.76	0.76	0.76	0.73
ANN	0.85	0.85	0.85	0.85	0.77
KNN	0.82	0.82	0.82	0.82	0.61
Decision Tree	0.77	0.77	0.77	0.77	0.56

DEMONSTRATION


THANK YOU!