



Diabetes Prediction: Identifying Key Factors

This presentation explores the key factors influencing diabetes risk, using a dataset of medical attributes. We will delve into the dataset, analyze its features, and apply machine learning to predict the likelihood of diabetes.

NS by Nazeema Shaheen Sultana

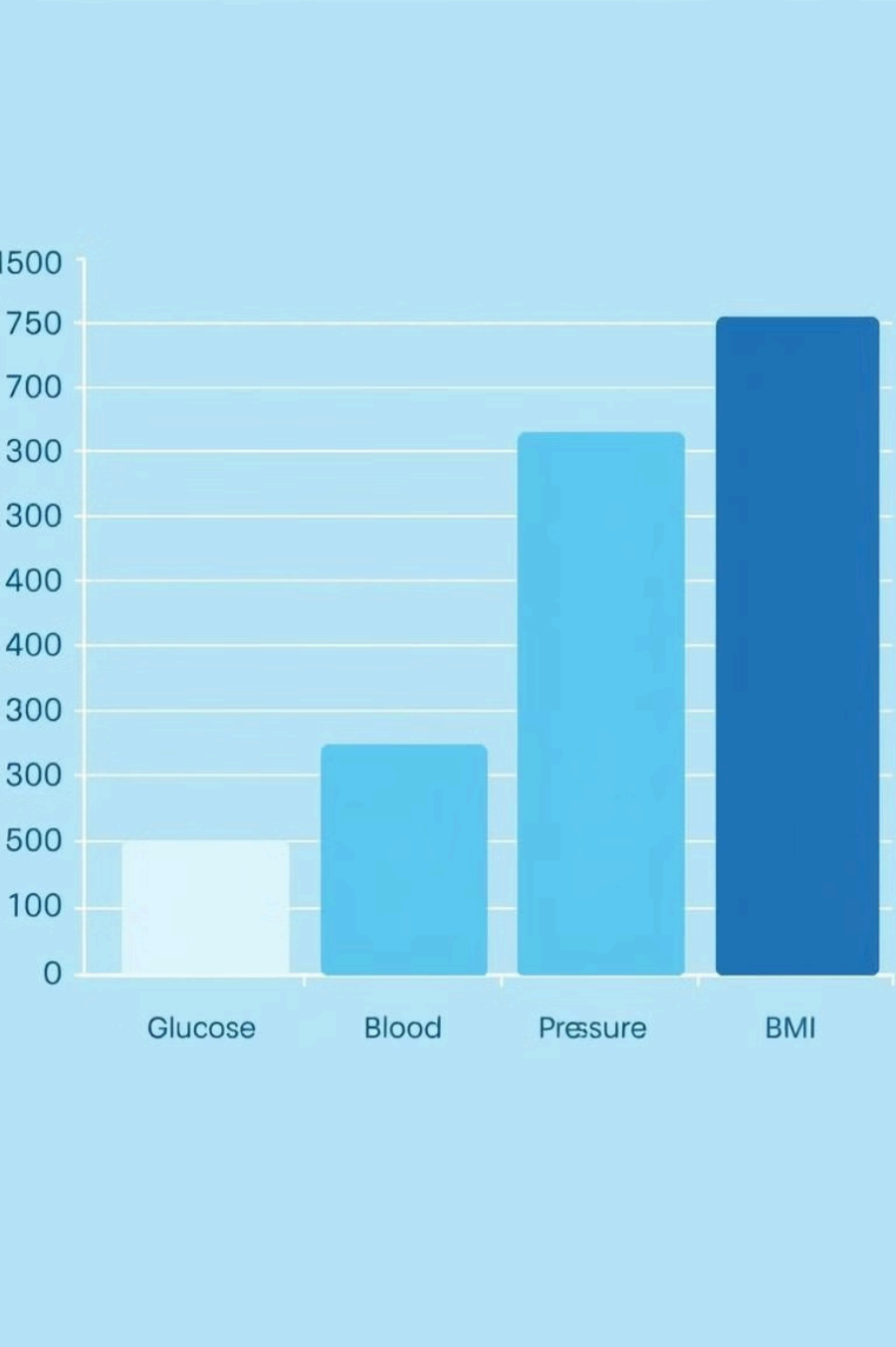
Understanding the Dataset

Data Overview

The dataset contains 768 instances, each representing a patient. Each instance has 9 attributes.

Attributes

These attributes include demographic and medical factors, like age, glucose levels, and family history.



Feature Exploration: Numeric Attributes

Pregnancies

The number of times a woman has been pregnant.

Glucose

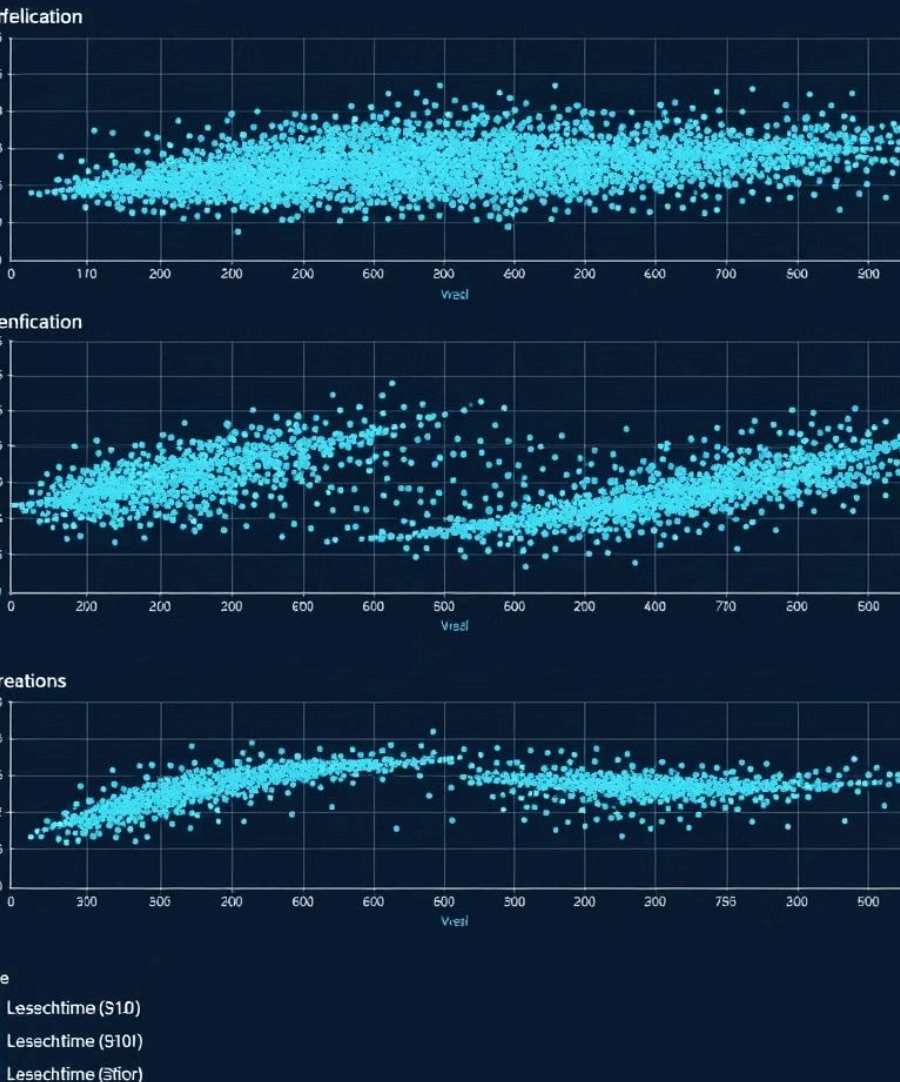
Plasma glucose concentration measured during an oral glucose tolerance test.

Blood Pressure

Diastolic blood pressure measured in millimeters of mercury.

Skin Thickness

Triceps skin fold thickness, a measure of body fat.



Correlations and Relationships



Glucose and Insulin

A strong positive correlation suggests that higher glucose levels are associated with higher insulin levels.



BMI and Insulin

Another positive correlation indicating a link between body mass index and insulin levels.

Preparing the Data for Analysis

- 1 Data cleaning: Addressing missing values, outliers, and inconsistencies in the dataset.
- 2 Feature scaling: Normalizing the range of numeric attributes to improve model performance.
- 3 Data splitting: Dividing the dataset into training and testing sets to evaluate model accuracy.

Data cleans and Timntion

Autopont sociable
projrcce varidells cranired
mer dasis pacing values.



C-Repoodaodzgnrcerscolleates

Reqrer pond aregssing valude,
recomg tittalvction

Acalling cattribute values,
prodon categorical variables



C Reppodaodzgnrcerscclleates



Srade viss categond varidries
acchitite enaniones



Applying Machine Learning Models

1

Logistic Regression

A linear model that estimates the probability of diabetes.

2

Decision Tree

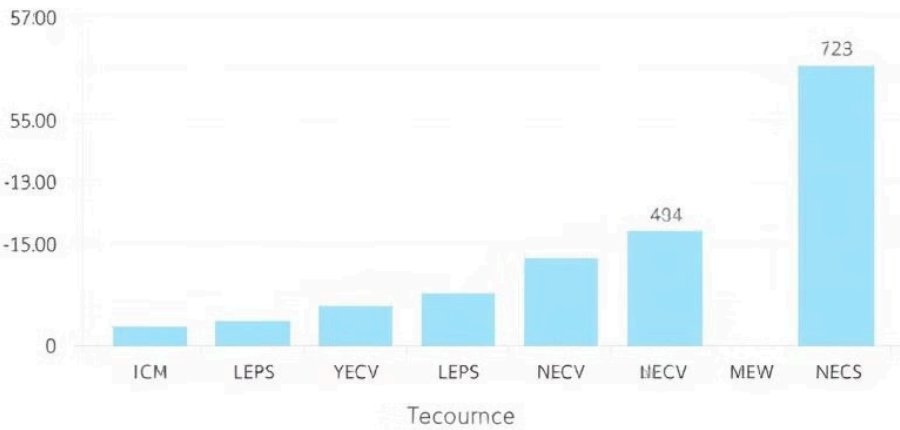
A tree-based model that uses a series of rules to classify instances.

3

Support Vector Machine (SVM)

A powerful algorithm that finds the optimal hyperplane to separate classes.

Total Performance MtTCN



Evaluating Model Performance

0.75

Accuracy

The proportion of correctly predicted instances.

0.80

Precision

The ratio of correctly predicted positive instances to all predicted positive instances.





Insights and Next Steps

The analysis reveals that several factors contribute to diabetes risk, including glucose levels, BMI, and family history. These insights can guide interventions to prevent or manage diabetes.