DIABETES DATASET PROJECT REPORT

INTRODUCTION

This project is based on diabetes patient data. The aim of this project is to understand patterns in the dataset such as age, glucose level, BMI, and their relationship with diabetes.

DATA UPLOAD

First, we have to download the dataset. After that, we will write a code to upload the file in Google Colab.

LOAD & VIEW DATA

This command is used in Python to load the data from the file that we have uploaded.

BASIC DATASET

Column types refer to the kind of data stored in each column of a dataset. Missing values are empty or undefined entries in a dataset.

STATISTICAL SUMMARY

The describe() function is used to generate a statistical summary of the dataset. It provides information such as mean, minimum value, maximum value, standard deviation, and quartiles.

MISSING DATA CHECK

The isnull().sum() command is used to check how many missing values are present in each column of the dataset.

DATASET BASIC VISUALIZATION

Age Distribution: A histogram shows how ages are spread across the dataset.

Glucose Levels: A histogram shows the distribution of glucose levels.

BMI vs Glucose: A scatter plot shows the relationship between BMI and glucose.

RESULT

The analysis indicates a strong link between high glucose levels and the likelihood of developing diabetes. Higher BMI values show an increased risk, and older age groups tend to have more diabetes cases.

CONCLUSION

The dataset provides useful insights into the factors influencing diabetes risk. Glucose level, BMI, and age consistently show strong patterns related to diabetes.

Author: Raj Verma

Reg No: 25MIM10213