**MAJOR PROJECT**

**DIABETES PREDICTION USING ML**

**PROBLEM STATEMENT:**

Diabetes is one of the most frequent diseases worldwide and the number of diabetic patients is growing over the years. The main cause of diabetes remains unknown, yet scientists believe that both genetic factors and environmental lifestyle play a major role in diabetes.

Individuals with diabetes face a risk of developing some secondary health issues such as heart diseases and nerve damage. Thus, early detection and treatment of diabetes can prevent complications and assist in reducing the risk of severe health problems.

**OBJECTIVE:**

To build a model to predict whether an individual is at risk of diabetes or not.

**CONTENTS**

1. **TOOLS I HAVE USED** - Jupyter notebook for implementation of codes.
2. **DATA PREPROCESSESING**

• Include Libraries: Import Libraries such as pandas, numpy, matplotlib, seaborn and some packages from scikit-learn.

• Import Dataset

• Handle the Missing Values: Check whether there are any missing values in the dataset.

1. **DATA VISUALIZATION:** Pairplot, Boxplot, Histogram, Correlation or Confusion Matrix.

**RESULTS AND OUTPUT:**

By comparing the accuracy\_score of the implemented algorithms, the algorithms which gave accuracy are:

Random Forest → 81.81%

Descision Tree → 77.27%

K-Nearest Neighbours → 75.32%

Naïve Bayes → 79.22%

Support Vector Machine → 79.22%

**Logestic Regression → 82.46% - *highest accuracy***

**CONCLUSION:**

I have implemented five popular machine learning algorithms for predicting diabetes. These algorithms include KNN, LR, DT, RF, SVM and NB. From all of them Logistic regression and Random Forest gives highest accuracy for predicting diabetes.

**ML-MAJOR-JUNE**

**K. UMA PRANAVI**