Project Brief

PROJECT NAME: Diabetes Prediction Project – Classification

BY: Mukesh Kumar
DATE: 24 June 2024

PLAN	• Import data into the workspace. • Open the data and analyse the dependent and independent variables. Here the dependent variable is **labeled** and **categorical.** • Select the correct model to solve the problem. In this case we use classification type model to predict the outcome. • The model should have minimum false negative and false positive prediction outcomes. False positive could lead to unnecessary medical interventions and stress on the other hand false negative is more concerning, this can lead to undiagnosed cases and delayed or missed treatments, potentially causing serious health consequences.
ANALYSE	• This is a diabetes dataset, there are 9 variables in this dataset with 8 independent and one dependent variables. The dependent variable is categorical where 0 indicates the person does not have diabetes and 1 indicates the has diabetes. • The data is slightly **imbalance** with 65% – 0 values and 35% – 1 values • Address the irregularities like null values, zeros, duplicates, outliers. • Understanding the relationships between variables in our dataset. • Standardise and encode the continuous and categorical variables. • Make ready the data to fit in the model.
CONSTRUCT	• Choose the algorithm. We are using **Logistic regression, Support vector machine classifier, Random Forest Classifier and Decision Tree Classifier** algorithms and then select the most accurate one. • Train our model on the data. • Test the model on the test subset • Evaluate the prediction algorithm. Check the accuracy score, cross validation score and confusion matrix.
EXECUTE	• After evaluating all the algorithms we found that Logistic Regression, Support Vector Machine Classifier and Random Forest Classifier have **accuracy 80%.** But **Random Forest Classifier have less false negatives and false positives** than the other two. • Now we can share the prediction model with the stakeholder to test the model and receive the feedback.