**PROJECT REPORT**

1. PROBLEM STATEMENT

Diabetes Prediction : using KNN Classification

1. APPROACH/ALGORITHM APPLIED
2. Exploring Data

Head, Tail, Shape, Value Counts, Null-Values, info(),

Describe(), Age bucket to show people having diabetes

in different age groups, glucose bucket to show how

many people have diabetes in particular level.

1. Invalid Cases : in this we will explore that if there is any

0 value of glucose level, BMI etc.

1. Data Visualisation

Here we have used different plots ( like bar plot,

Line graph, histogram, scatter plot) to show people

having diabetes in different age groups and many

more.

1. Train-Test Data using KNN algorithm

Get accuracy. Note: In case of classification algorithms score method represents accuracy.

1. Cross Validation

Cross-validation is a technique to evaluate predictive models by partitioning the original sample into a training set to train the model, and a test set to evaluate it.

1. CONCLUSION

Using KNN , we conclude that : (7 neighbours)

Accuracy of K-NN classifier on training set is 0.79

Accuracy of K-NN classifier on test set is 0.73

1. INFERENCE

Thus a knn classifier with number of neighbors as 14 achieves the best score/accuracy of 0.7578 i.e about 76% using cross validation.