DIABETES PREDICTIONS USING

Machine Learning(Support Vector Machine)

-Satish Yadav

PROBLEM

- Diabetes is one of the deadliest diseases in the world which is rapidly increases .It is not only diseases but also creator of different kinds of diseases.
- Diabetes should not be ignored if it is untreated then Diabetes may cause some major issues in a person like heart related problems, kidney problem, blood pressure, eye damage, It can also affect other organs of human body etc
- Good thing is diabetes can be controlled if it is predicted earlier.
- To achieve this goal our project will do early prediction of Diabetes in a human body or a patient.

SOLUTION

- The aim of our project is to design and implement Diabetes Prediction using machine learning methods and performance analysis of those methods.
- The techniques we are using are support vector machine.
- Diabetes can be controlled if it is predicted earlier.Machine learning techniques will provide better result for prediction by constructing models from datasets collected from patients.

OUTLINES

- Introduction
- What is Diabetes
- Support vector Machine
- Works Flow
- Analysing Dataset
- Data Splitting
- Conclusion
- References

Introduction

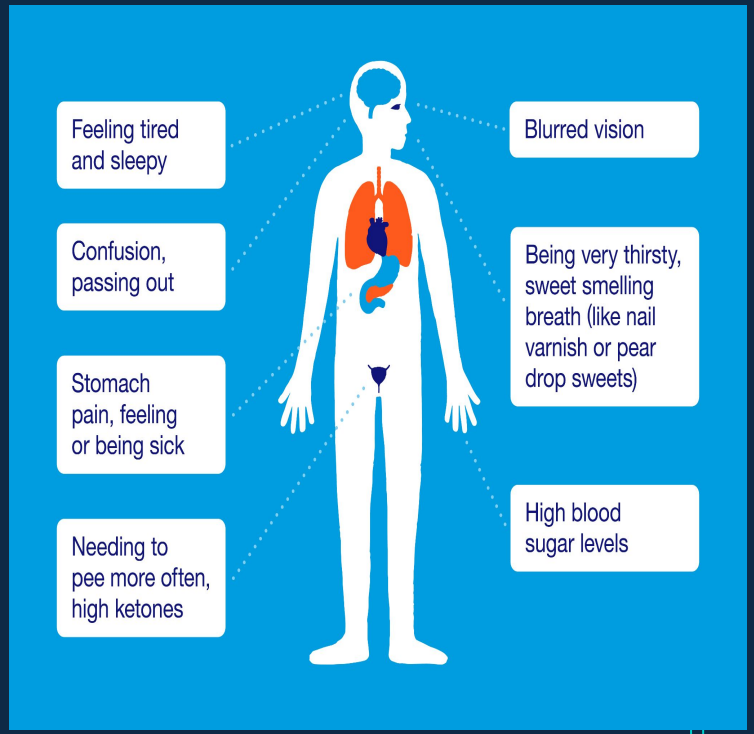
Keywords:

- Diabetes
- Prediction

➢ **Simple image of diabetes testing**

What is Diabetes:

- Diabetes is recognized as one of the most frequent and rapidly growing diseases worldwide .
- It increases the Blood Sugar level Significantly.
- Diabetes enhances the chances of other long term complications like heart attack,cardiovascular disease etc



Support Vector Machine

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane

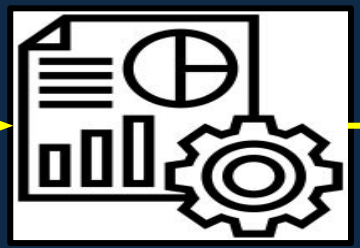
SVM can be Two Type:

- Linear SVM
- Non-Linear SVM

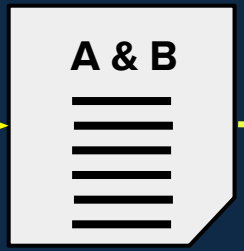
Support Vector Machine Classifier(SVM)

Works Flow

Data Set

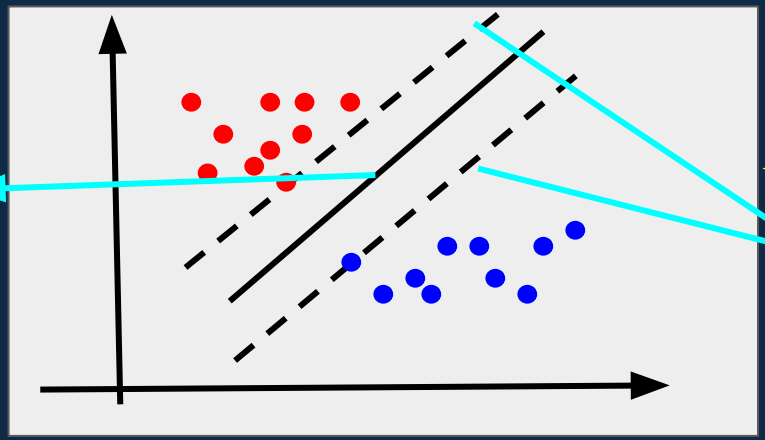


Data Preprocessing



Train Test

Optimal Hyperplane

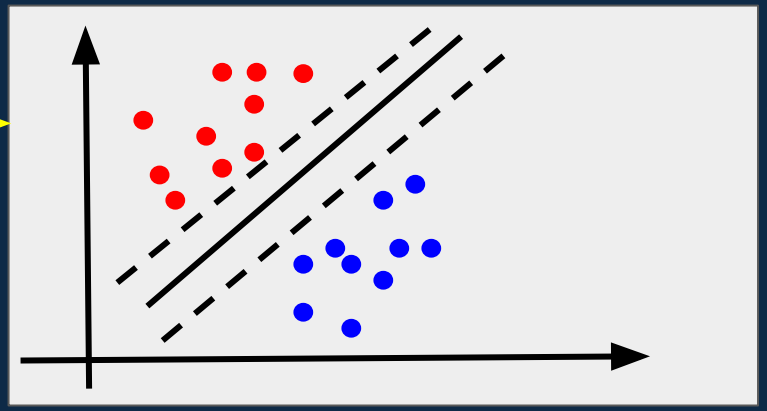


Support Vector

Support Vector Machine Classifier(SVM)

Works Flow

New Data



Diabetes
Or
Non-Diabets
Prediction

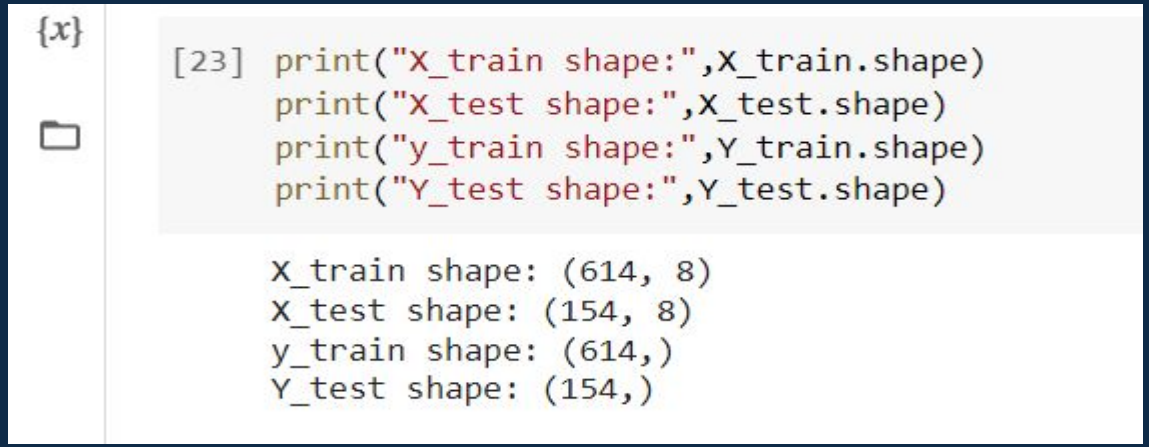
Support Vector Machine Classifier(SVM)

Analysing Dataset

The dataset consists of 768 samples. The dataset has total 9 attributes Pregnancies, Glucose, Blood Pressure, Skin thickness, Insulin, BMI, Diabetes Pedigree Function, Age, Outcome out of which 8 are independent variables and one is the dependent variable i.e. target variable (outcome) which determines whether patient is having diabetes or not.

Data Splitting:

Data is divided into training and testing data into 80:20 ratio. Eighty percent was training data and twenty percent was testing data. Out of 768 records, 614 records were used for trained and 154 records were used for testing.



```
[23] print("X_train shape:",X_train.shape)
     print("X_test shape:",X_test.shape)
     print("y_train shape:",Y_train.shape)
     print("Y_test shape:",Y_test.shape)

     X_train shape: (614, 8)
     X_test shape: (154, 8)
     y_train shape: (614,)
     Y_test shape: (154,)
```

CONCLUSION

As we are doing classification of diabetic and non diabetic patients. SVM works relatively well when there is a clear margin of separation between classes.SVM is more effective in high dimensional spaces. SVM is effective in cases where the number of dimensions is greater than the number of samples and at the last it is relatively memory efficient.

REFERENCES

- https://scikit-learn.org/stable/
- https://scikit-learn.org/stable/modules/svm.html
- https://scikit-learn.org/stable/modules/svm.html
- https://www.python.org/doc/