

13 DICAS PARA PREVENIR DIABETES

Machine learning

Diabetes prediction model

Presented by Suljić Anisa

Mentor : prof. Dr. Mirsad Hadžikadić

Academic year 2021-22



/01

/02

Presentation Outline

Topic	03
Workflow	04
Algorithm	05
The predictive data analytics project lifecycle	06
Demo	07



Diabetes

- Diabetes is a chronic health condition that affects how your body turns food into energy. Most of the food you eat is broken down into sugar and released into your bloodstream. When your blood sugar goes up, it signals your pancreas to release insulin.
- Diabetes is a disease that occurs when your blood glucose, also called blood sugar, is too high.



/04

01

Diabetes data

- Downloaded from Kaggle site ([here](#))
- Data and respective labels

03

Data separation

- Splitting data into training and testing data
- Training data to train model
- Testing data to find the accuracy score

05

Trained support vector machine classifier

- Result of training process using algorithm we mentioned in step 04

02

Data preprocessing

- Standardize unsuitable data
- Bring them to the same branch

04

Support vector machine classifier

- Choosed algorithm - ***Support vector machine classifier***
- Feeding data into model

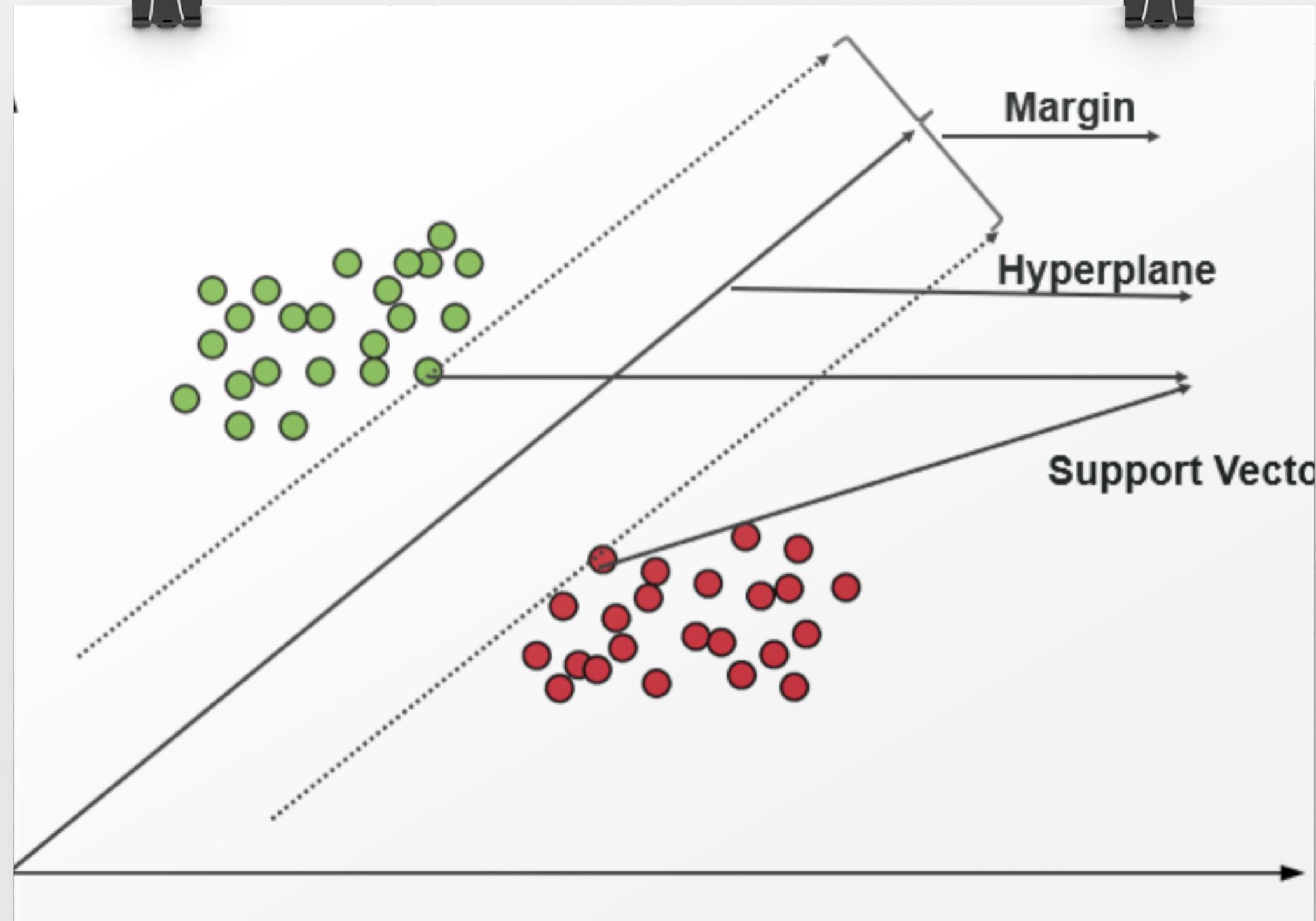
07

Prediction

- Putting new data into trained model and model should be able to predict if person is diabetic or non-diabetic

Support vector machine classifier algorithm

- One of *supervised machine* learning algorithms
- Every "spot" has features which are coordinates
- SVMS perform the classification test by drawing hyperplane that is a line in 2D and a plane in 3D
- *Hyperplane* separates data on two sides
- *Margin* is the maximum distance between hyperplane and hyperplane nearest points
- Hyperplane nearest points are *support vectors*





Business Understanding

- The lifestyle we lead says a lot about our health
- There are many factors that negatively affect on human health
- We are talking about diabetes and the factors that influence the development of this disease
- Looking at the previous conditions of patients who were diabetic and those who were not, we can make a model that will predict whether a person will be diabetic or not
- This model will help in the early detection of diabetes



Data Understanding

- Dataset has 769 records and 9 columns (8 descriptive features and 1 target feature)
- We had to define the ***prediction subject*** for the ABT, ***descriptive features*** and the ***target feature***.
- ***Prediction subject*** : The goal was to develop a model that would predict whether a human would be diabetic. This meant that the prediction subject in this case was a human, so the ABT would need to be built to contain ***one row per human***
- ***Target feature*** : "Outcome" - diabetic / non-diabetic
- ***Descriptive features*** : Pregnancies, Glucose, Blood pressure, Skin thickness, Insulin, BMI, Diabetes pedigree function and Age



Data Preparation

- First I made **full data quality report**
- Then I made some **histograms** for visual presentation of report
- Next step was to **standardize data**
- And before modeling I split data into two sections, **training set** and **test set**

/09



Modeling

- As I said in slides before, I used Support Vector Machine Classifier for Diabetes machine learning model
- 1. First we will use sklearn library and its method SVC(kernel='linear') to create an object
- 2. fit(X_train,Y_train) method help us to train model on training data
After these 2 simple steps we are finished with building our model and we can continue to the evaluation phase

/10



Evaluation

- Before deployment, it is important to check accuracy of training set and test set. The goal is accuracy be as close to one as possible.
- In my case accuracy of training set is 0.786 and accuracy of testing set is 0.772 which is pretty good.
- Also I had to check my model with some test data and after those data passed through the model i checked if prediction result was zero or one.

```
if(prediction==[0]):  
    print("Person is non-diabetic")  
else:  
    print("Person is diabetic")
```



Deployment

- Last step would be model deployment but as medicine evolves and advances every day, this model is nothing new to that science.
- I worked on a small set of data and a small number of features, while systems in medicine work on huge sets and with many more features to make the prediction accuracy as high as possible.



14. November | World Diabetes Day

“The first wealth is health.” – Ralph Waldo Emerson

D I A B E T E S