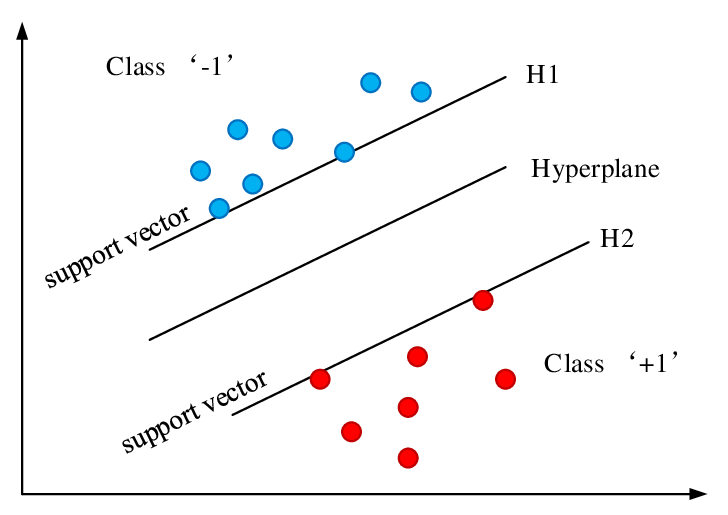
Diabetes Prediction Model Using Support Vector Machine Algorithm

Support Vector Machine: Abbreviated as SVM is a supervised learning algorithm that allows one to solve problems or to make predictions concerning classification or outliers’ detection, its principle is to separate data using a margin in which the purpose is to maximize it to have better results away from noise.

*“****A support vector machine (SVM) is a machine learning algorithm that uses supervised learning models to solve complex classification, regression, and outlier detection problems by performing optimal data transformations that determine boundaries between data points based on predefined classes, labels, or outputs. SVMs are widely adopted across disciplines such as healthcare, natural language processing, signal processing applications, and speech & image recognition fields.”*  According to Spiceworks.**



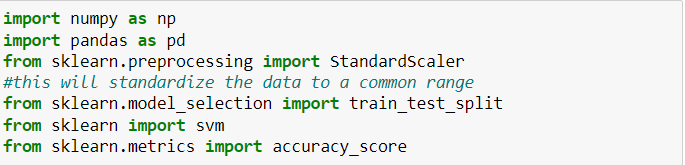
**researchgate.net**

**Libraries/Dependencies/Dataset Used in the Project:**

* **NumPy**
* **Pandas**
* **Scikit-learn>Preprocessing>StandardScaler**
* **Scikit-learn>model\_selection>train\_test\_split**
* **Scikit-learn>SVM**
* **Scikit-learn>metrics>accuracy\_score**
* **PIMA Diabetes Dataset**

**Steps to realize the project:**

1. **Importing dependencies:**

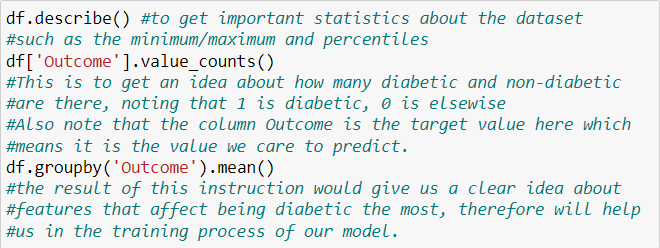


1. **Loading the dataset:**

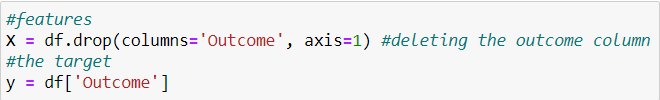


**Please note that the second instruction displays the first 10 elements of the dataset that was named df.**

1. **Analyzing and interpreting your dataset**

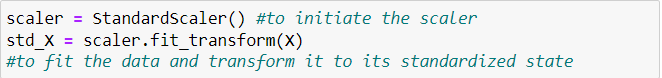


1. Separating features and the target:



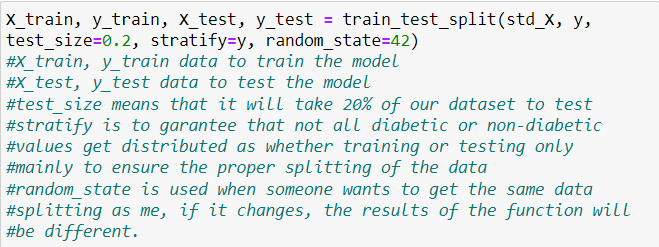
You can always use print to display any part of the results or the dataset whenever you feel like it throughout the project.

1. Standardizing the data

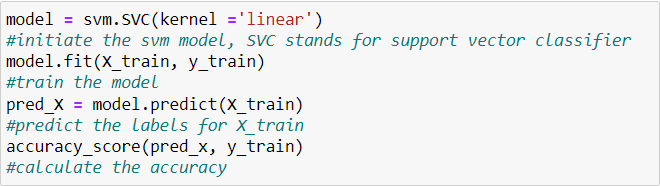
If we look at the ranges of different values of each column, we can notice that the range of the data is not the same when it comes to every feature which would make it very difficult for our model to make the right prediction leading to lower quality results, therefore, we use the standardization to fix this problem.

We are going to be using std\_X as the training data for our model in what comes next.

1. Splitting the dataset into training and testing parts:

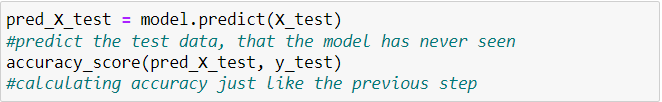


1. Training the SVM model and get the accuracy:



If the accuracy of your training is not very high, don’t be so disappointed since it is normal, the real accuracy to be counted is the one measured on the testing dataset which will be the next step, moreover note that you shouldn’t expect much of the model since the dataset is relatively very small.

1. Accuracy on test data:



Congratulations, don’t hesitate to share your accuracies with me ;)