What is SVM in data analytics?

SVM or Support Vector Machine is **a linear model for classification and regression problems**. It can solve linear and non-linear problems and work well for many practical problems. The idea of SVM is simple: The algorithm creates a line or a hyperplane which separates the data into classes.

We use a standard machine learning procedure. First, we gathered reliable ECG data. The data was then preprocessed, with selected attributes that will influence the target and outliers removed, and the data was split into test set and train set, with train set used for training algorithms and test set used to test the models.

Data Collection and Processing Data Collection

The dataset was compiled from multiple patient records that included 14 attributes such as restagc, fbs, thal, ca, cp, sex, age, thalach, chol, trestbps, slope, oldpeak, exangand target. The table below contains a description of the attributes used for analysis. The visualization shows that age does not play a significant role in predicting heart disease because the same age groups have an equal number of people with and without heart disease. Outlier detection and data preprocessing For each attribute or feature, we calculate the Z-score of each individual value of that attribute in relation to the column mean and standard deviation. After that, take the magnitude or absolute value of the obtained z score. If the z score is less than a certain threshold, a particular row or record is outlier and is removed.

Data Transformation: (Standardization)

It is a technique used to scale where the values are scaled around the mean with standard deviation as unity.

$$X=X-\mu\sigma \qquad (1)$$

Attribute Selection (Pearson Correlation)

This measures the relationship between two sets of data. The value of the correlation varies from −1to1. Here+1 denotes a strong positive correlation, a strong negative correlation is denoted by −1, and no correlation is zero. If the absolute value of the correlation is less than 0.2, then it means the two datasets are not correlated and can omit during analysis. If an attribute having correlation closer to zero, then that attributes can be eliminated. is the graph showing importance of each feature and it is clear that ‘thal’ is a more important feature and ‘age’ is a less important feature.

Conclusion

The processing and identification of heart health data will aid in the prediction of abnormal heart conditions and the rapid rescue of human lives. To provide a new approach to heart disease, ML techniques were used to process relevant data. In the medical world, heart disease prevention is complicated and important. However, if the condition is detected early on and prevention is implemented as soon as possible, the lethal rate can be managed.