Assignment - 12

* Aim → On the given data, perform the performance measurements using Simple Naive Bayes algorithm such as accuracy, error rate, precision, recall, TPR, FPR, TNR etc. (using WEKA)

* Theory →

- Naive Bayes Algorithm:

It is a classification technique based on Bayes theorem with an assumption of independence among predictors. In simple terms a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.

Naive Bayes model is easy to build and particularly useful for very large datasets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods.

Bayes theorem provides a way of calculating Posterior probability P(c|x) from P(c), P(x) and P(c|x).

$$P(c|x) = \frac{P(x|c) \cdot P(c)}{P(x)}$$

P(x|c) → Likelihood

P(c) → Class probability

P(x) → Predictor prior probability

- Advantages:

i) It is easy and fast to predict class data set. It also performs well in multi class prediction

ii) When assumption of independence holds a Naive Bayes classifier performs better compare to the other models like logistic regression, and you need less training data.

iii) It performs well in case of categorical input variables compared to numerical variables.

- Disadvantages:

i) If categorical variable has a category which was not observed in training the model will assign zero (0) probability and will be unable to make a prediction

ii) In real life it is almost impossible that we get a set of predictors which are completly ~~enter~~ independent

- Applications :

i) Real - time prediction
ii) Multi - class prediction
iii) Text classification / spam filtering / sentiment analysis
iv) Recommendation system.

* Conclusion → In this assignment we have implemented and understood the concept of the Naive Bayes algorithm.