EXPLORATORY DATA ANALYSIS

DESCRIPTIVE STATISTICAL:-

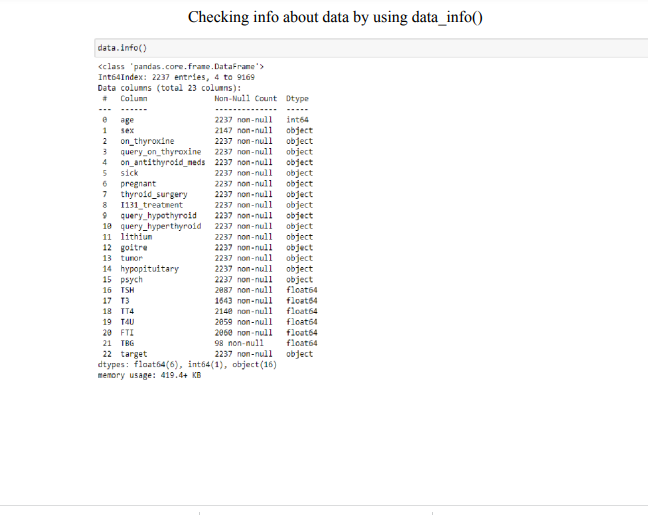
3.1 DATA COLLECTION:- Thyroid data set is collected from UCI repository. The dataset contains 3 classes and 215 samples. These classes are assigned to the values that correspond to the hyper-, hypo-, and normal function of the thyroid gland. The data set content 215 samples and each sample has 5 features. 1. T3-resin uptake test (A percentage). 2. Total serum thyroxin as measured by the isotopic displacement method. 3. Total serum triiodothyronine as measured by radioimmuno assay. 4. Basal thyroid-stimulating hormone (TSH) as measured by radioimmuno assay. 5. Maximal absolute difference of TSH value after injection of 200 μg of thyrotropin-releasing hormone as compared to the basal value. The 150 samples of 215 belong to hyper-function class namely class-1. The 35 samples of 215 belong to hypo function class namely class-2. The 30 samples of 215 belong to normal-function class namely class-3 [4].

3.2 Pre processing All the null and duplicate values has been removed from dataset. e content of data set

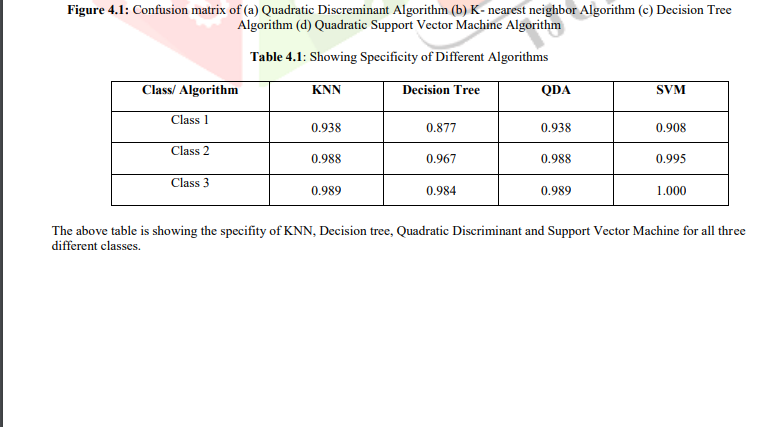
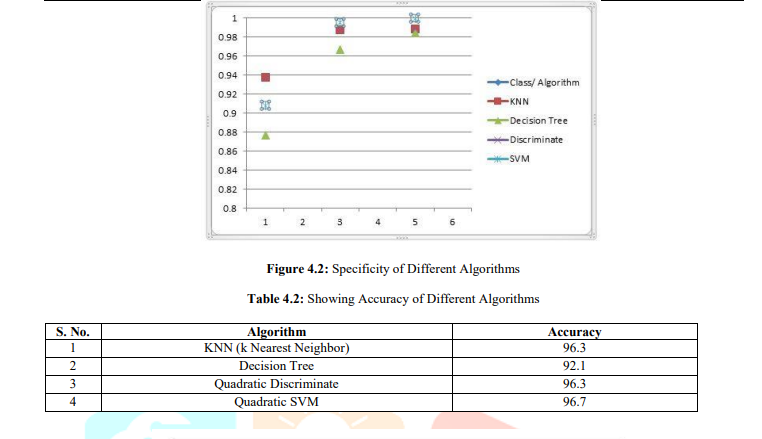
3.3 Dataset Partition In data partition stage input data set has to be divided into two sets i.e. training set and testing set, Data partition stage generate two mutually exclusive data set shares no data among each other and both the set having unique content of data set.

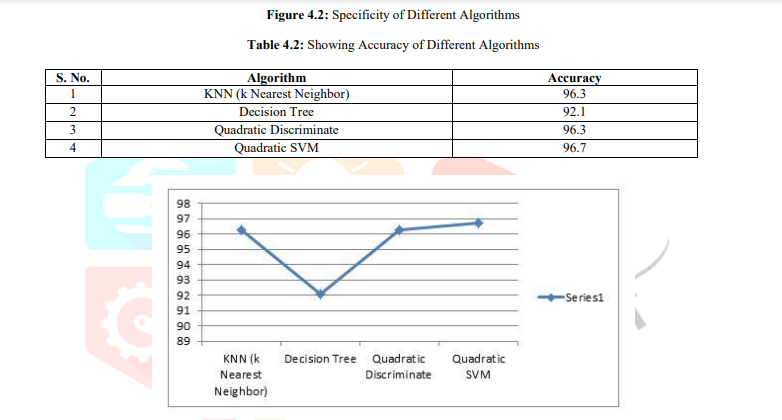
3.4 Classifiers In data partition stage input data set has to be divided into two sets i.e. training set and testing set, Data partition stage generate two mutually exclusive data set shares no data among each other and both the set having unique content of data set

1. retained as the validation data for testing the model, and the remaining k - 1 sub samples are used as training data. The cross-validation process is then repeated k times (the 'folds'), with each of the k subsamples used exactly once as the validation data. The average of the k results gives the validation accuracy of the algorithm [26]. The advantages of k-fold cross validation are that the impact of data dependency is minimized and the reliability of the results can be improved [27].Below is the figure of confusion matrix for each classifiers



VISUAL ANALYSIS





V. CONCLUSION Different Researchers have proposed different techniques to predict the thyroid disorder and different kinds of accuracy level as per used techniques but on comparing this technique with all previous methods we can observe that accuracy of quadratic support vector machine is 96.7% which is higher than all other classifiers shown in table 4.1 and figure 4.3