**PROJECT REPORT**

**Description**: Our task 1 project report is all about “classification algorithms”. In this we are focusing on two algorithms mainly “Decision Tree Algorithm“ and “Naïve based Algorithm”.

1. **Selection of Datasets**:

We have chosen two different data sets for two algorithms.

**Decision Tree:**

We have used the data set “**Database for fitting contact lenses**” which has four nominal attributes along with 24 instances for decision tree. We have chosen the class label as “target” in this dataset. This dataset was observed to be complete and noise free and it highly simplified the problem. The attributes are clear and complete (all possible combinations of attribute-value pairs are represented). Each instance is complete and correct. 9 rules cover the training set.

**Naïve Based Algorithm:** In this we have considered “Space shuttle O-Ring dataset”. This data set relates to explosion of USA Space shuttle Challenge. Finding out the probability on this would give me the estimation of the explosion.

**# Packages used for Classification Algorithms**

**“Decision Tree: “**

Package (rpart): Recursive partitioning for classification, regression and survival trees.

Package (caret): The caret package (short for Classification and Regression *T*raining) is a set of functions that attempt to streamline the process for creating predictive models.

Package (ROCR): Visualizing the performance of scoring classifiers.

Package (party): A computational toolbox for recursive partitioning.

**Naïve Based:**

Package(klaR): Miscellaneous functions for classification and visualization.

Package (ROCR): Visualizing the performance of scoring classifiers

Package(E1071): Functions for latent class analysis, short time Fourier transform, fuzzy clustering, support vector machines, shortest path computation, bagged clustering, naive Bayes classifier.

# Accuracy, Recall & Precision description for classification trees

“For Decision tree”

Overall Accuracy: “66.667%”. (same for both training & testing data).

**Precision & Recall for training data:**

|  |  |  |
| --- | --- | --- |
| **Class label** | **Recall** | **Precision** |
| Class 1 | 66.667% | 66.667% |
| Class 2 | 50% | 50% |
| Class 3 | 72.727% | 72.727% |

The above table will explain the efficiency and effective ness between training and testing data.

**“Naïve Bayes Classification”**

|  |  |  |  |
| --- | --- | --- | --- |
| **Class labels** | **Accuracy** | **Recall** | **Precision** |
| **Class 0** | **83.33%** | **100%** | **83.33%** |
| **Class 1** | **83.33%** | **0%** | **-** |

By Comparing the above two “classification trees”, Decision tree and Naïve Bayes alogorithm show a very efficient output. By observing the accuracy of two algorithms we can confirm that **Accuracy of decision tree is “66.667%” where as for Naïve Bayes it is “83.33%”. By this understanding we can say that “Naïve Bayes ” can be considered as much efficient classification tree.**