ML – Assignment 2

Team Members:

1. Keval Morabia (2015A7PS0143H)
2. Poojan Patel (2015A3PS0343H)
3. Abhishek Savani (2015A7PS0087H)

Language Used: Java

Logic:

- Decision tree was generated using the data provided and the ID3 algorithm mentioned in Tom. M. Mitchell.
- Missing values were filled using the value which appeared most frequently in the particular attribute column.
- Continuous values were handled as mentioned in section 3.7.2 of Tom M. Mitchell. First the values were sorted in ascending order, then at the points where value was changing, gain was calculated and finally the column was splited at the point where maximum gain was obtained.
- Reduced Error Pruning was performed by removing a node (one by one) and then checking the accuracy. If accuracy was increased than the node was removed else we move on to check the next node.
- Random forests were generated using 50% attributes and 33% data randomly. 10 forests were generated and accuracy increased compared to the original ID3 algorithm.

Output:

Generating Decision Tree using ID3 Algorithm

Training Time=1.979secs

Accuracy=0.807874209200909

Precision=0.8762364294330519

Recall=0.8727272727272727

F-Score=0.874478330658106

No of nodes in tree = 33223

Applying Reduced Error Pruning on the decision tree generated

Training Time=10.7secs

Accuracy=0.8404889134574043

Precision=0.9467631684760756

Recall=0.8588415523781733

F-Score=0.9006617450177867

No of nodes in tree = 2640

Initializing Random Forest with 10 trees, 0.5 fraction of attributes and 0.33 fraction of training instances in each tree

Training Time=1.618secs

Accuracy=0.8313371414532277

Precision=0.944270205066345

Recall=0.8511779630300834

F-Score=0.8953107129241327

Conclusion:

- Hence decision tree was successfully generated and overfitting was removed using Reduced Error Pruning and Random Forests.