Predicting probable bankruptcy based on various data attributes.

Hassan M. Mahmudul
ID: 15-28405-1
Serial No: 01
Course: Data Warehousing and Data Mining [A]

Fall 2019-2020

m.mahmudul77@outlook.com

***Abstract*— Data Mining, discovers and extracts useful patterns from large set of data to find observable patterns. This paper been worked with supervised data [1]. All the data are considered for training purpose, and it is used in the five-classification algorithm, this paper presents the analyze and accuracy of probable bankruptcy based on seven different nominal attribute using data mining algorithms for various decision tree approaches using WEKA. Five classification algorithms such as J48, Random Tree (RT), NaiveBayes, REPTree and Random Forest (RF) are used to measure the accuracy. Data mining tool WEKA (Waikato Environment for Knowledge Analysis) has used for performing such five classification algorithms, RandomForest algorithm outperforms other algorithms by yielding an accuracy of 100%.**

***Index Terms*—Data mining, bankruptcy, weka, random tree, J48, RandomForest.**

I. Introduction

Currently Bangladesh has Tk 99,370 crore of defaulted amount, according to data from the Bangladesh Bank [2]. It has grown above the alarming rate. The collected dataset is the actual dataset of banks from India. If we try to collect similar data and use such decisions, then it would be easy to predict any bankruptcy for any the current loan amounts. Weka tool has been used to generate decision trees based on different algorithms. Decision tree generated results and decision tree been added on this paper as well.

II. Related Background

Supervised learning is the machine learning task of learning function that's maps on input to output based on example input to output pairs. The calculations can apply straightforwardly to a dataset. Weka is an open-source information mining apparatus it bolsters information-mining calculations, packing, and boosting. Enlistment is the gaining from class named preparing tuples. In choice tree hubs speak to the information esteems, the edges will highlight all the potential moves, therefore from hub to leaf through the edge it's giving the objective qualities from which we can make order to anticipate. Some supervised learning algorithm required the user to determine certain control parameters.

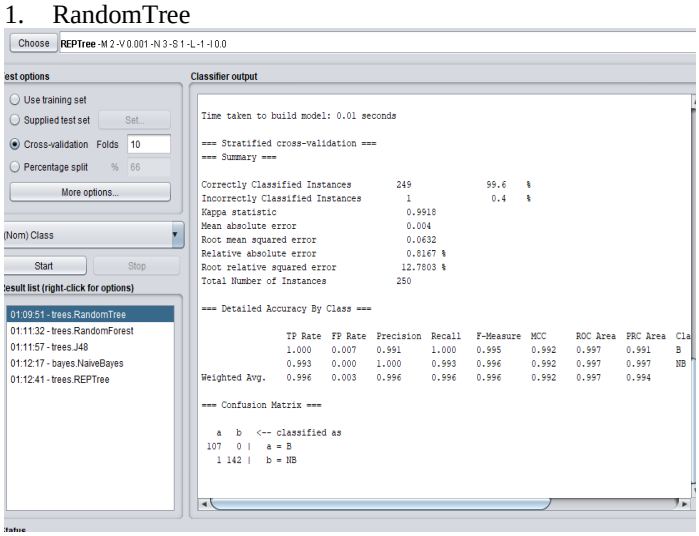
III. Proposed Methodology

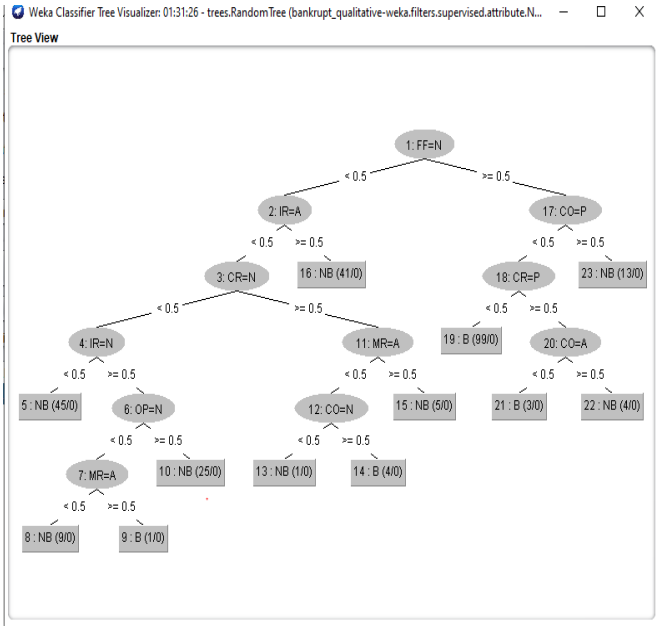
The dataset contains 7 attributes including 250 number of instances containing information on industry risk, management risk, financial Flexibility, credibility, competitiveness, operating Risk. Some sample data is given below, also used data set's url is also attached [3]. Based on the qualitative data parameters from experts, attempt is done to predict possible bankruptcy. Below algorithms are applied to classify correct instances

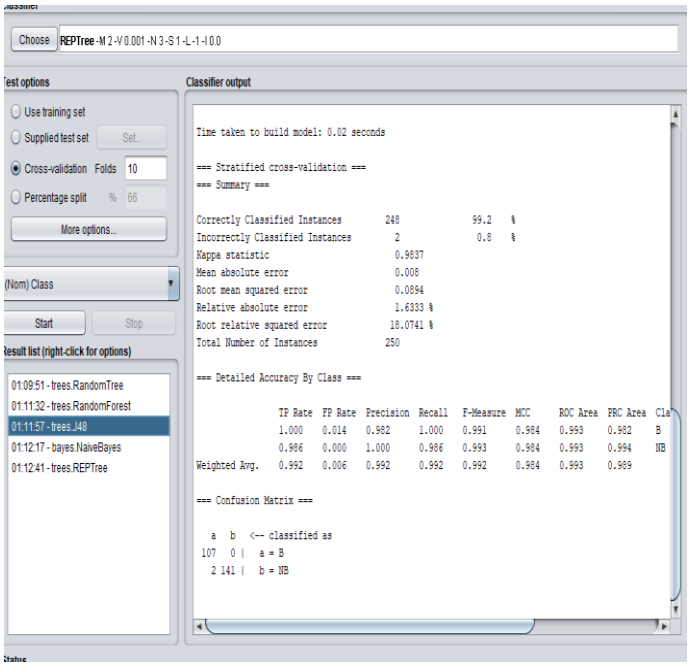
- RandomTree
- RandomForest
- J48
- NaiveBayes
- REPTree

IV. Result and Analysis

1. RandomTree

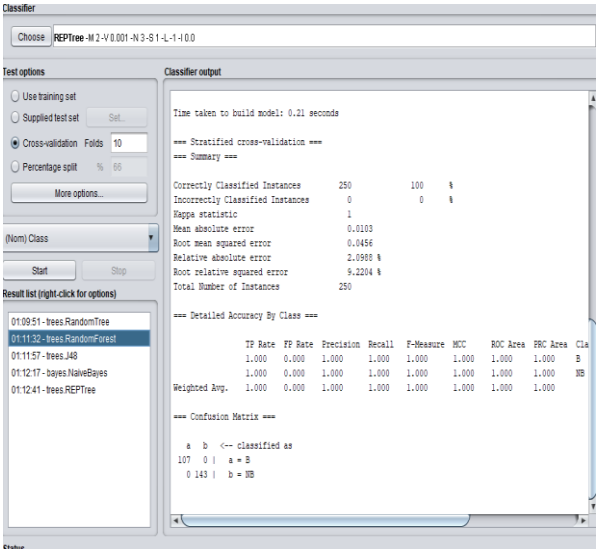






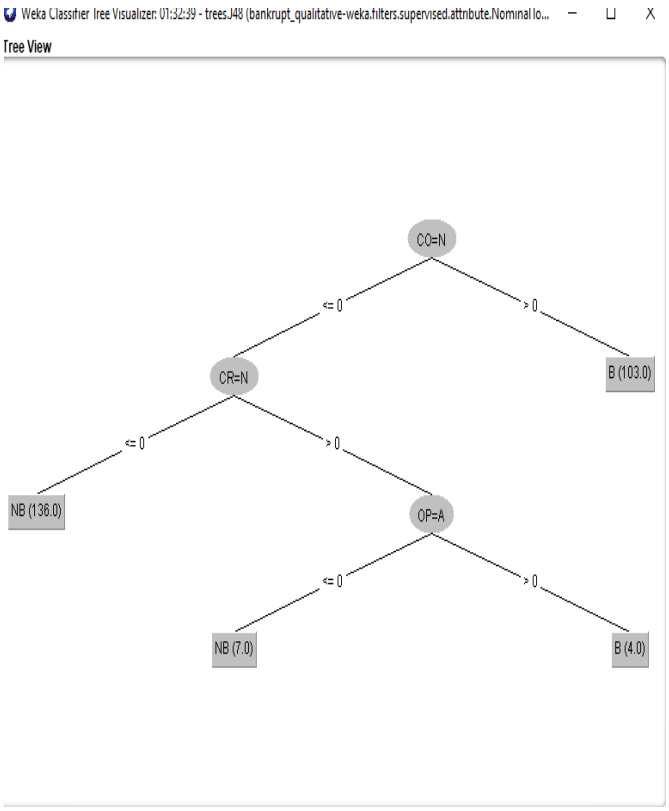
Analysis
Correctly Classified instance = 249, 99.6%
Incorrectly Classified instance = 1, 0.4%

2. RandomForest



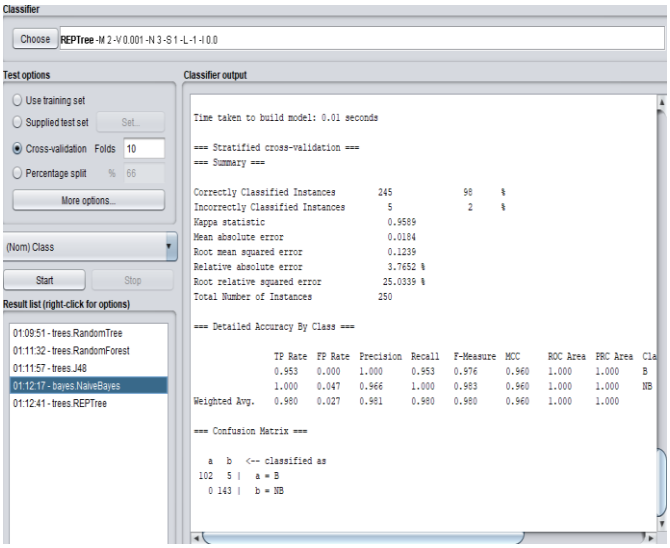
Analysis
Correctly Classified instance = 250, 100%
Incorrectly Classified instance = 0, 0%

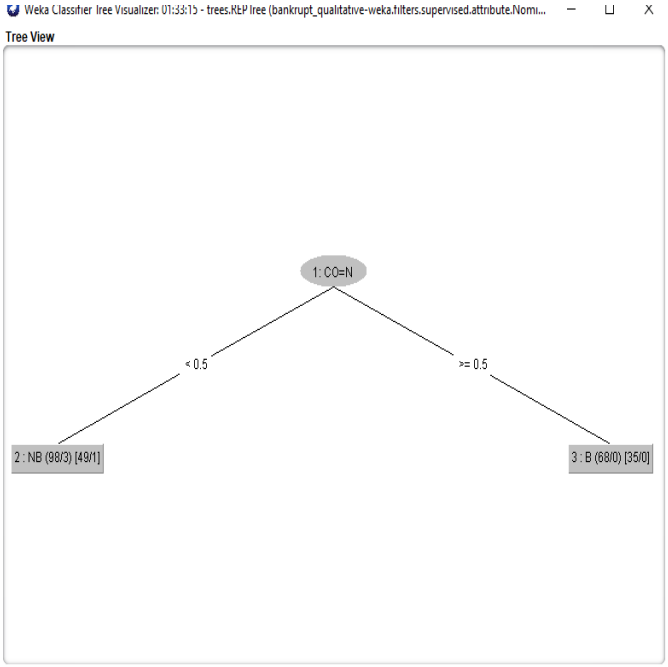
3. J48



Analysis
Correctly Classified instance = 248, 99.2%
Incorrectly Classified instance = 2, 0.8%

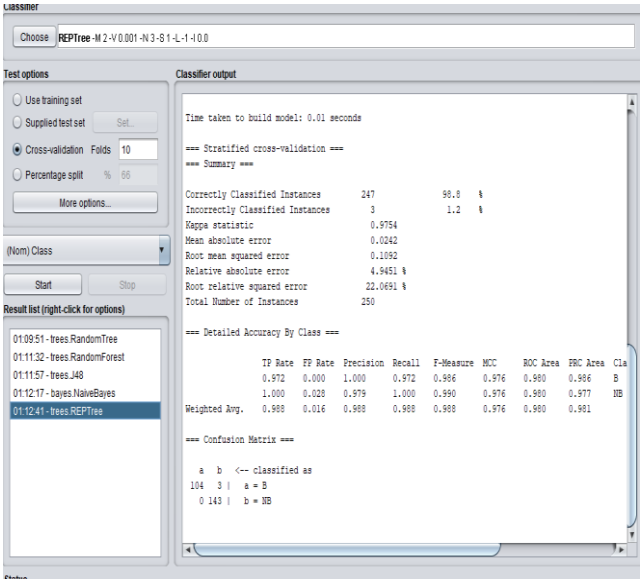
4. NaiveBayes





Analysis
Correctly Classified instance = 245, 98%
Incorrectly Classified instance = 5, 2%

5. REPTree



Analysis
Correctly Classified instance = 247, 98.8%
Incorrectly Classified instance = 3, 1.2%

Result Analysis: (Correctly classified instances)

- RandomTree: 99.6%
- RandomForest: 100%
- J48: 99.2%
- NaiveBayes: 98%
- REPTree: 98.8%

V. CONCLUSION

As we examined the tree calculations, we can reach resolution that for credit informational index RandomForest tree is most appropriate for basic leadership as it is giving 100 percent of accuracy. In future, the proposed technique will be stretched out to other informational indexes from the regions like banking, and financial exchange and so on.

REFERENCES

[1] https://blogs.nvidia.com/blog/2018/08/02/supervised-unsupervised-learning/

[2] https://www.thedailystar.net/business/banking/bangladesh-bank-moves-amend-bankruptcy-act-1997-1698106.

[3] https://archive.ics.uci.edu/ml/datasets/qualitative_bankruptcy.