KNOWLEDGE STRUCTURE FOR FRAUDULENT FIRM CLASSIFICATION APPROACHES

Dr.G.Ayyappan[1], Dr.A.Kumaravel[2]
Associate Professor[1], Professor[2]
Machine Learning Group[1,2]
Department of Information Technology[1,2],
Bharath Institute of Higher Education and Research[1,2] Chennai

Abstract: Machine learning gives most impact in developing the qualities of an audit work in the future due to enormous development of financial fraud. Here, seven hundred and seventy seven firm's data in fourteen different sectors are collected in this research study. This work proposed number of classification models are estimated in terms of their accuracies and time taken to build the models. The results of Trees and Rules are demonstrating an accuracy of high level for suspicious firm classification. This paper focuses improving the worth of an audit work by implementing a machine learning algorithms.

Keywords: J48, Fraudulent Firm, Audit data, Machine Learning

I. Introduction

Now a day's Fraudulent is a critical issue worldwide. Firms that resort to the unfair practices without the fear of legal repercussion have a grievous consequence on the economy and individuals in the society. Auditing practices are responsible for fraud detection. Audit is defined as the process of examining the financial records of any business to corroborate that their financial statements are in compliance with the standard accounting laws and principles[2]. On successful completion of an audit process, auditors deliver an audit and inspection summary report called audit paras to the company comprising of the details of all the findings from the audit. This may include discrepancies, noncompliance of accounting rules, leakage of revenue, inaccurate calculations, etc. The whole audit process flow of proposed system depicted in Figure 1.

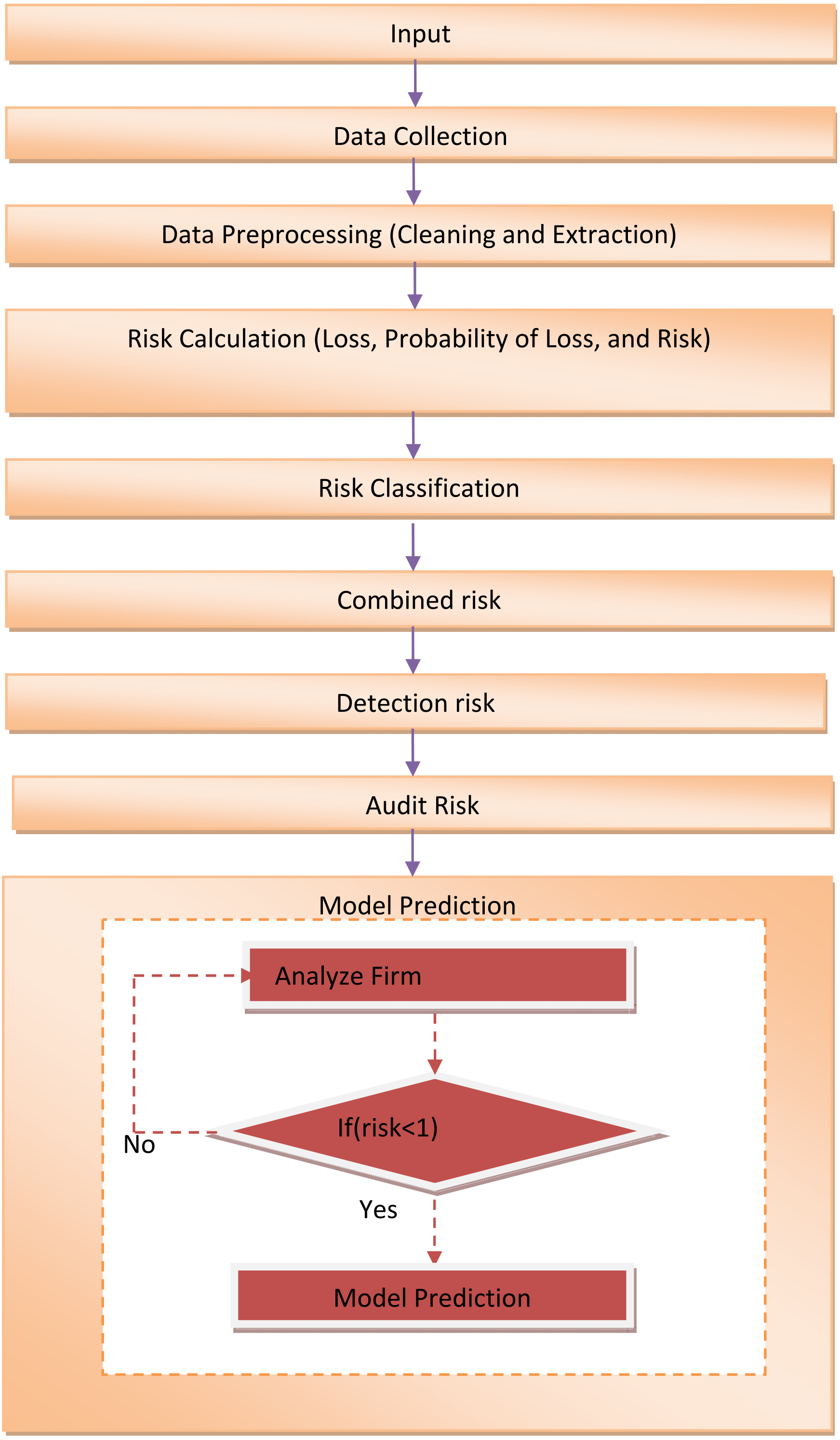

Fig 1.Proposed System

The rest of the paper is organized as follows: In Section 2 Consists the audit dataset, considered features, risk assessment procedure, and the complete methodology. In Section three presents the Experiments and discussed for results. Finally, the conclusion and future scope are discussed in Section 4.

II. Materials and Methods

This section focuses on materials and methods of this research work. The dataset borrowed from UCI repository. Exhaustive one year non-confidential data in the year 2015 to 2016 of firms is collected from the Auditor Office of India to build a predictor for classifying suspicious firms.

Data Set Information:

The goal of the research is to help the auditors by building a classification model that can predict the fraudulent firm on the basis the present and historical risk factors.

S.No	Sector	Firm
1	Irrigation	114
2	Public Health	77
3	Buildings and Roads	82
4	Forest	70
5	Corporate	47
6	Animal Husbandry	95
7	Communication	1
8	Electrical	4
9	Land	5
10	Science and Technology	3
11	Tourism	1
12	Fisheries	41
13	Industries	37
14	Agriculture	200

Attribute Information:

Numerous risk factors are observed from a variety of areas like

- past records of audit office,
- profit-value records,
- firm reputation summary,
- loss-value records,
- environmental conditions reports,
- audit-paras,
- on-going issues report,
- follow-up reports etc.

After in-depth interview with the auditors, important risk factors are evaluated and their probability of existence is calculated from the present and past records.

Methods:

- BayesNet-Bayes Network learning using various search algorithms and quality measures.
- NaiveBayes-Class for a Naive Bayes classifier using estimator classes.
- NaiveBayesMultinomialText-Multinomial naive bayes for text data.
- NaiveBayesUpdateable -Class for a Naive Bayes classifier using estimator classes.
- DecisionTable-Class for building and using a simple decision table majority classifier.
- JRip-This class implements a propositional rule learner, Repeated Incremental Pruning to Produce Error Reduction (RIPPER), which was proposed by William W.
- OneR-Class for building and using a 1R classifier; in other words, uses the minimum-error attribute for prediction, discretizing numeric attributes.
- PART-Class for generating a PART decision list.
- DecisionStump- Class for building and using a decision stump.
- J48-Class for generating a pruned or unpruned C4.
- RandomForest-Class for constructing a forest of random trees.
- RandomTree-Class for constructing a tree that considers K randomly chosen attributes at each node.

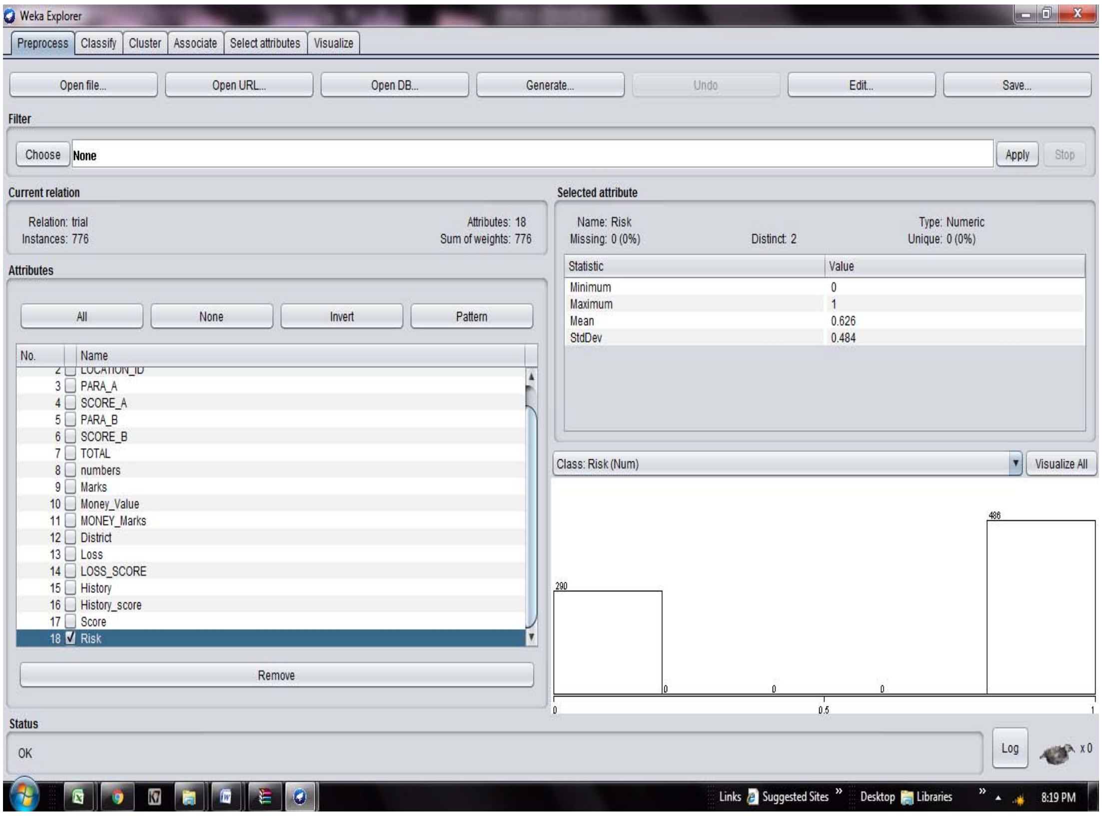

Fig 2: List of Attributes in Weka in Numeric

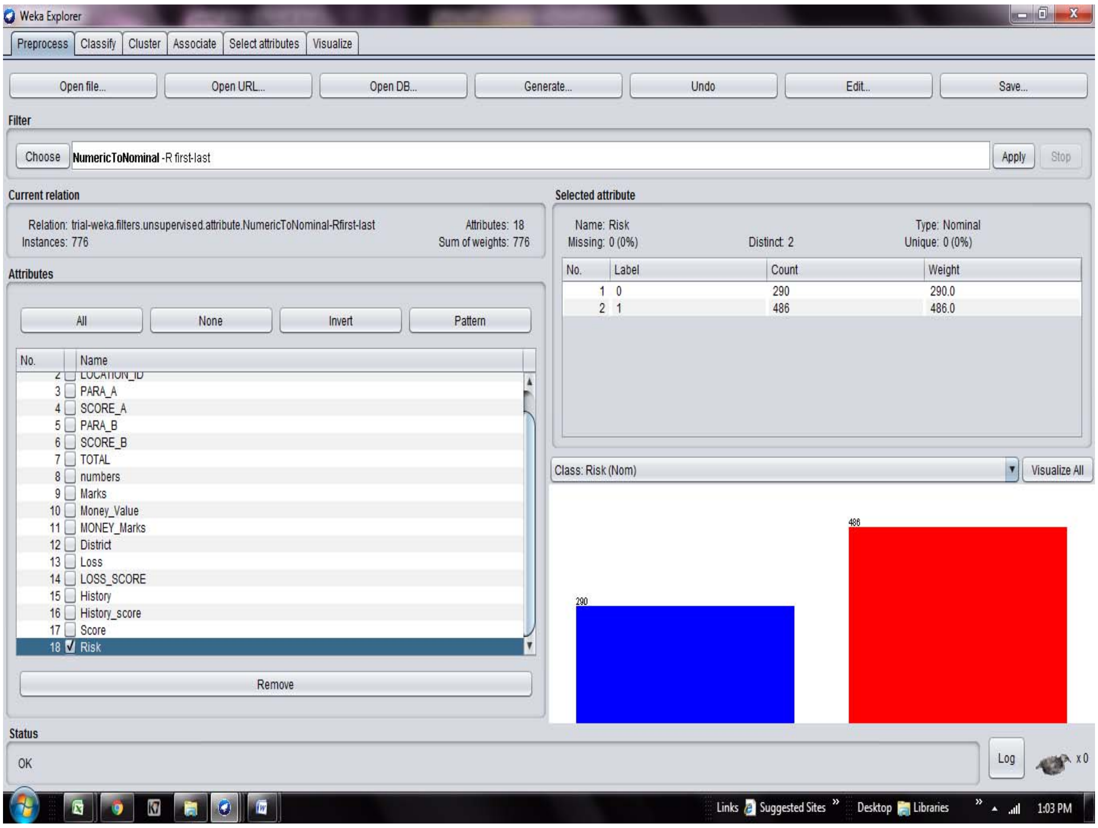

Fig 3: List of Attributes in Weka in Nominal

III. Results and Discussions

In this section presents the experiments and interpretations of this research work for applying various classifiers in machine learning algorithms.

Table 1: Baye Classifiers

S.No	Category	Classifier	Accuracy	Time Taken to Build the Model
1	Bayes	BayesNet	99.61%	0.06
2		NaiveBayes	99.23%	0.02
3		NaiveBayesMultinomialText	62.63%	0.01
4		NaiveBayesUpdateable	99.23%	0

The above table depicts on bayes classifiers applied in this research work. They are BayeNet classifier has 99.61% level of accuracy , NaiveBayes classifier has 99.23% level of accuracy, NaiveBayesMultinomialText classifier has 62.63% level of accuracy and NaiveBayesUpdateable classifier has 99.23% level of accuracy.

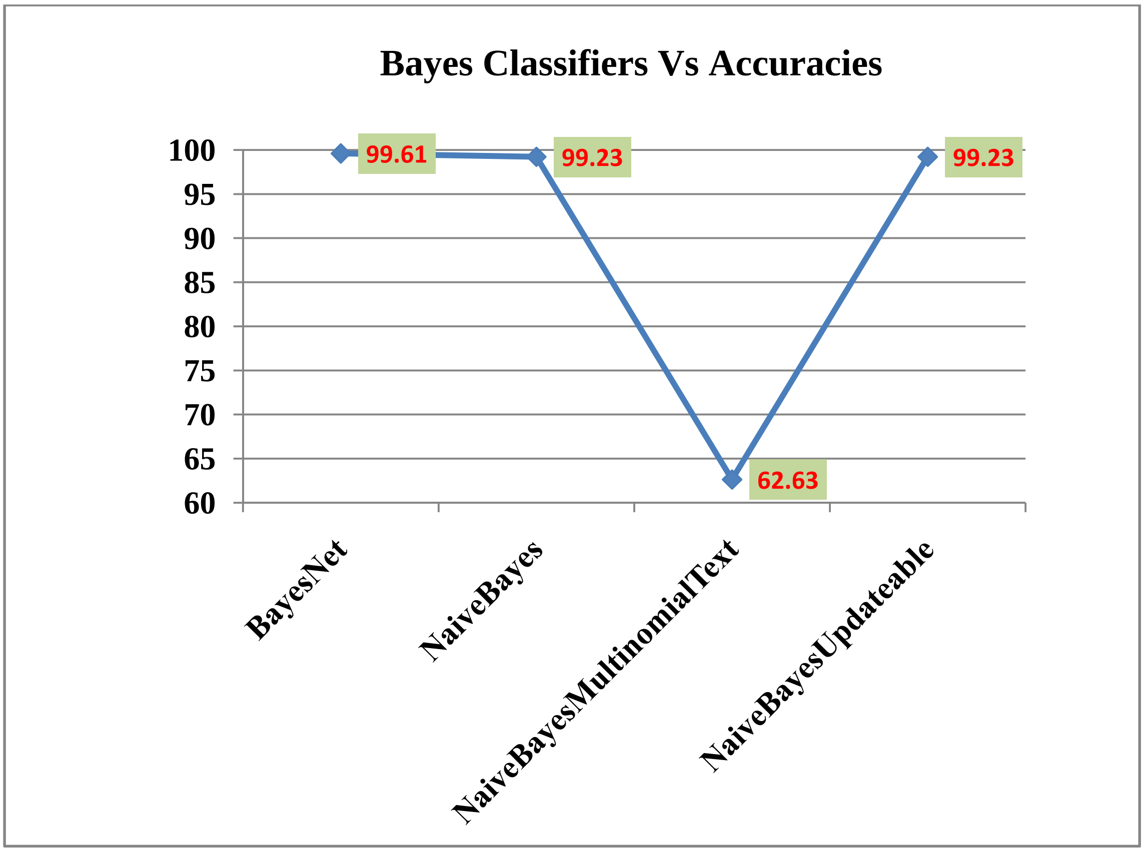

Fig 4: Graphical Representation of Bayes Classifiers Vs Accuracies

The above diagram clearly depicts on NaiveBayesMultinomialText classifier has 62.63% has lowest accuracy compare with other bayes classifier models.

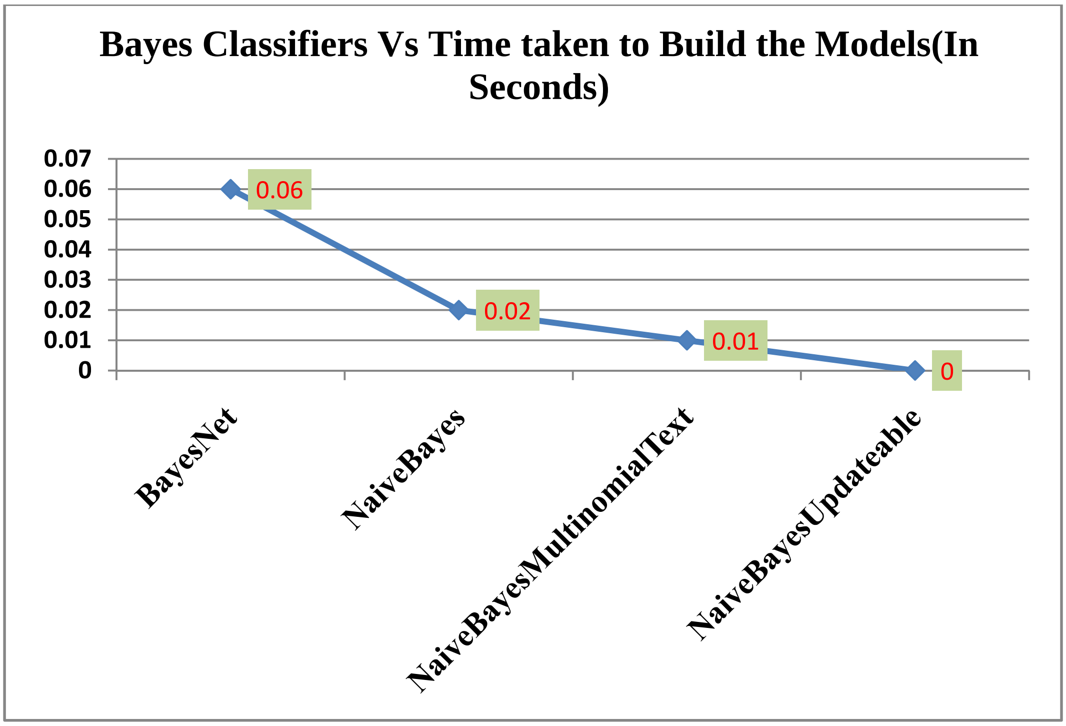

Fig 5: Graphical Representation of Bayes Classifiers Vs Time to build the model

The above diagram has represented BayesNet has 0.06 seconds for taken to build the model. Other models are having low time consumption to build the models.

Table 2: Rules Classifiers

S.No	Category	Classifier	Accuracy	Time Taken to Build the Model
1	Rules	DecisionTable	99.74%	0.21
2		JRip	100%	0.06
3		OneR	100%	0.01
4		PART	100%	0.03

The above table depicts on Rules classifiers applied in this research work. They are DecisionTable classifier has 99.74% level of accuracy , JRip classifier has 100% level of accuracy, OneR classifier has 100% level of accuracy and PART classifier has 100% level of accuracy.

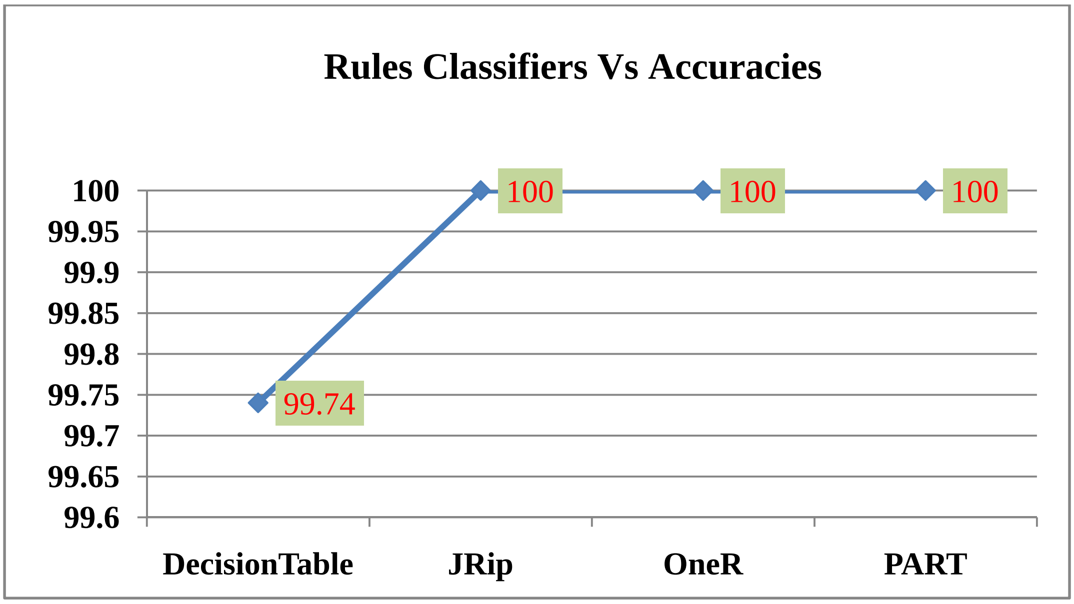

Fig 6: Graphical Representation of Rules Classifiers Vs Accuracies

The above diagram depicts on JRip ,OneR and PART classifiers are having 100% accuracy levels and equal value., But DecisionTable classifier has 99.74% level of accuracy.

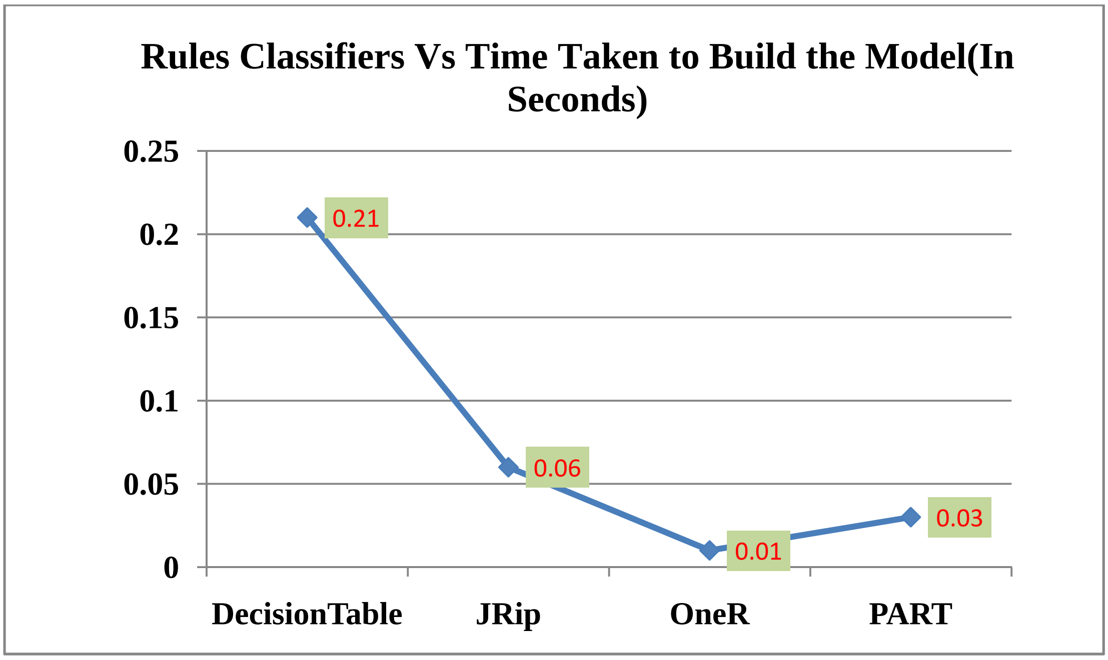

Fig 7: Graphical Representation of Rules Classifiers Vs Time to build the model

The above diagram depicts OneR classifier is having low time consumption to build the model compare with other models.

Table 3:Trees Classifiers

S.No	Category	Classifier	Accuracy	Time Taken to Build the Model
1	Trees	DecisionStump	100%	0.01
2		J48	99.74%	0.09
3		RandomForest	99.48%	0.56
4		RandomTree	87.24%	0.01

The above table depicts on Trees classifiers applied in this research work. They are DecisionStump classifier has 100% level of accuracy , J48 classifier has 99.74% level of accuracy, RandomForest classifier has 99.48% level of accuracy and RandomTree classifier has 87.24% level of accuracy.

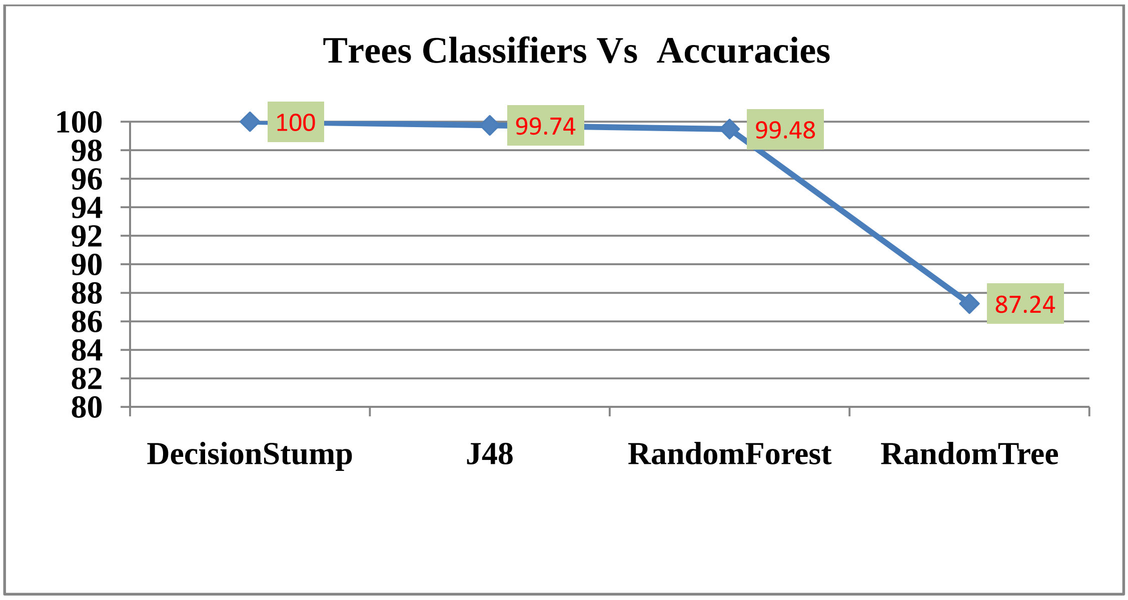

Fig 8: Graphical Representation of Trees Classifiers Vs Accuracies

The above diagram represents the DecisionStump classifier has high accuracy level100%. Other classifiers are having lowest accuracies compare with DecisionStump.

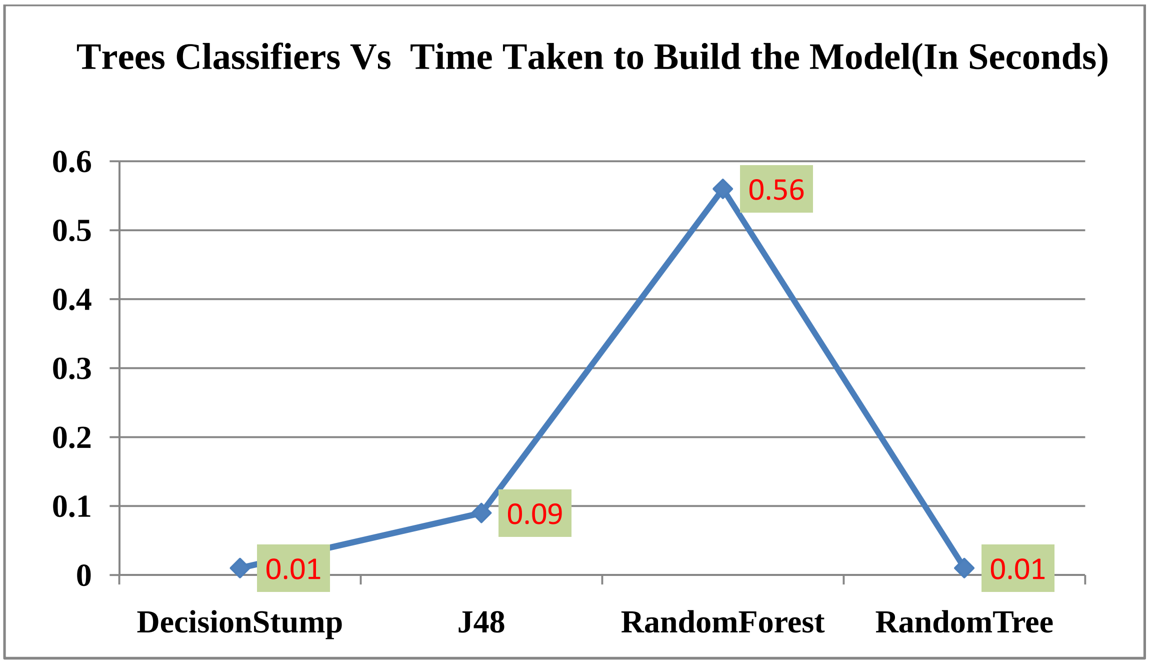

Fig 9: Graphical Representation of Trees Classifiers Vs Time to build the model

The above diagram represents the DecisionStump and RandomTree models are having 0.01 time consumption to build the model. Others are J48and RandomForest are having high time consumption to build the model compare with DesicionStump and RandomTree.

IV. Conclusions

In this research work concludes that after collecting the data of seven hundred and seventy seven firms from fourteen different sectors, it is cleaned, transformed, and useful risk factors are examined with the help of an in-depth interview with the auditors. Twelve state-of-the-art classifiers like BayesNet, NaiveBayes, NaiveBayesMultinomialText, NaiveBayesUpdateable, DecisionTable, JRip, OneR, PART, DecisionStump, J48, RandomForest, RandomTree, etc. are applied in this research work. For assessment of all the classifiers 12 different evaluation criteria using subjective ranking of criteria by audit experts are considered. JRip, OneR, PART, and DecisionStump also give stable results on K-fold validation testing, helping as a proof of eligibility of classifiers to perform an efficient risk assessment of the doubtful firms in the audit field work decision-making process.

References

[1] Nishtha Hooda, Seema Bawa & Prashant Singh Rana (2018) Fraudulent Firm Classification: A Case Study of an External Audit, Applied Artificial Intelligence, 32:1, 48-64, DOI: 10.1080/08839514.2018.1451032
[2] Couceiro, M. 2016. Particle swarm optimization. In Fractional order darwinian particle swarm optimization: Applications and evaluation of an evolutionary algorithm, 1–10. Boston, MA: Springer.
[3] Staff, A. 2014. Reimagining auditing in a wired world1. Technical report, University of Zurich, Department of Informatics. Zurich: Citeseer.
[4] Triantaphyllou, E. 2013. Multi-criteria decision making methods: A comparative study, Vol. 44 Boston, MA: Springer.
[5] Sharma, A. 2013. A review of financial accounting fraud detection based on data mining techniques. International Journal of Computer Applications.
[6] Smith-Miles, K. A. 2009. Cross-disciplinary perspectives on meta-learning for algorithm selection. ACM Computing Surveys (CSUR) 41 (1):16–19.
[7] Houston, R. W., M. F. Peters, and J. H. Pratt. 1999. The audit risk model, business risk and audit-planning decisions. The Accounting Review 74 (3):281–98. doi:10.2308/ accr.1999.74.3.281.
[8] Cosserat, G. 2009. Accepting the engagement and planning the audit. In Modern auditing, ed. G. Cosserat and N. Rodda, 3rd ed., 734–36. John Wiley & Sons.
[9] Buntine, W. 2016. Learning classification rules using bayes. Proceedings of the sixth international workshop on Machine learning, Sydney, Australia, ACM,94–98.
[10] Tysiac, K. 2015. Data analytics helps auditors gain deep insight. Journal of Accountancy. Accessed, Auguest 11, 2019.