ISSN 2320 -2602

International Journal of Advances in Computer Science and Technology (IJACST), Vol.3 , No.11, Pages : 06-11 (2014)
Special Issue of ICCAAC 2014 - Held on December 08,2014,Hyderbad, India

Effectiveness Analysis of ZeroR, RIDOR and PART Classifiers for Credit Risk Appraisal

Lakshmi Devasena C
Dept. of Operations & Systems, ISB Hyderabad, IFHE University,
Hyderabad, India.
devaradhe2007@gmail.com

Abstract: Credit Risk Appraisal is a hazardous task in Financial Industries like Banks. Identifying the defaulter before giving loan is a remarkable and troublesome task of the Bankers. Classification algorithms are the superior choice for predictive analysis like finding the pretender, whether he/she is a unpretentious customer or a fraud. Finding the outstanding classifier is a hard-hitting assignment for any industrialist like a banker. This gives authority to computer science researchers to drill down well-organized research works through evaluating different classifiers and identifying the finest classifier for such predictive problems. This research work inspects the effectiveness of different Rule Based Classifiers (RIDOR, ZeroR and PART Classifiers) for the credit risk prediction and evaluates their strength through various measures. German credit dataset has been taken and used to foresee the credit risk using open source machine learning tool.

Keywords: Credit Risk Appraisal, PART Classifier, Proficiency Comparison, RIDOR Classifier, ZeroR Classifier.

INTRODUCTION

The gigantic volume of business transactions enforced information processing automation an revitalizing factor for high quality standards, cost diminution, with high speed results. Automated data analysis and result of the relevant successes formed by state-of-the art computer algorithms have modified the opinions of many misanthropists. Earlier, people thought that financial market analysis necessitates intuition, knowledge and experience and wondered how this job could be automated. On the contrary, growth of scientific and technological ability, achieved the automation of financial market analysis. In modern days, credit debtor prediction and credit risk appraisal have enthralled great deal of interests from regulators, practitioners, and theorists, in the financial industry. Since, the credit risk of an applicant could be forecasted from the past giant database and the demographic data, it needs automation. Computerization of credit risk prediction can be attained using classification techniques. Identifying the classifier, which predicts credit risk in an proficient manner, is an crucial and decisive task. Rule based classifiers are human understandable and easy to interpret, it easily handle missing values and numeric attributes, so, in this research work three rule based classifiers are randomly selected and compared. This research work judges the credit risk performance of three rule based classifiers, namely, RIDOR, ZeroR and PART Classifier and compares their accuracy in credit risk prediction.

LITERATURE REVIEW

There are many research works proposed to predict credit risk using extensive computing techniques. In [1], a neural network based algorithm for automatic provisioning to credit risk scrutiny in a real world problem is presented. An assimilated back propagation neural network (BPNN) with the customary discriminant analysis approach used to discover the performance of credit scoring is given in [2]. A comparative study of corporate credit rating analysis using back propagation neural network (BPNN) and support vector machines (SVM) is described in [3]. An uncorrelated maximization algorithm within a triple-phase neural network ensemble technique for credit risk evaluation to differentiate good creditors from bad ones are elucidated in [4]. An application of artificial neural network to credit risk assessment using two altered architectures are deliberated in [5]. Credit risk investigation using diverse Data Mining models like C4.5, NN, BP, RIPPER, LR and SMO is likened in [6]. The credit risk of a Tunisian bank through modeling the non-payment risk of its commercial loans is analyzed in [7]. Credit risk valuation using six stage neural network ensemble learning approach is argued in [8]. A modeling framework for credit calculation models is erected using different modeling procedures is explained and its performance is analyzed in [9]. Hybrid method for assessing credit risk using Kolmogorove-Smirnov test, Fuzzy Expert system and DEMATEL method is enlightened in [10]. An Artificial Neural Network centered methodology for Credit Risk supervision is proposed in [11]. Artificial neural networks using Feed-forward back propagation neural network and business rules to correctly determine credit defaulter is proposed in [12]. The performance comparison of Memory based classifiers for credit risk investigation is experimented and précised in [13]. The performance comparison between Instance Based and K Star Classifiers for Credit Risk Inspection is accomplished and pronounced in [14]. The performance comparison among Sequential Minimal Optimization and Logistic Classifiers for Credit Risk Calculation is specified in [15]. The performance comparison between Multilayer Perceptron and SMO Classifier for Credit Risk appraisal is described in [16]. The performance comparison between JRip and PART Classifier for Credit Risk Estimation is explored in [17]. Proficiency comparison between Partial Decision Tree Classifier and Logistic Classifier for Credit Risk Prediction is explored in [18]. This research work scrutinizes the efficiency of different Tree Based Classifiers (RIDOR, ZeroR and PART

ISSN 2320 -2602

International Journal of Advances in Computer Science and Technology (IJACST), Vol.3 , No.11, Pages : 06-11 (2014)
Special Issue of ICCAAC 2014 - Held on December 08,2014,Hyderbad, India

Classifiers) for the credit risk prediction.

DATASET USED

The German credit data [19] is used to assess the performance of RIDOR, ZeroR and PART Classifiers for credit risk forecast. This data set includes 20 attributes, namely, Duration, Credit History, Checking Status, Purpose, Credit Amount, Employment, Installment Commitment, Saving Status, Personal Status, Other parties, Property magnitude, Age, resident since, Other payment plans, existing credits, job, Housing, No. of dependents, Foreign worker and Own Phone. The data set has 1000 instances of customer credit data with appropriate class. It categorizes the records into two classes, namely, good and bad.

METHODOLOGY USED

In this research work, different Rule Based Classifiers (RIDOR, ZeroR and PART Classifiers) are compared for ability assessment of credit risk evaluation.

ZeroR Classifier

ZeroR is the simplest of the rule based classifiers which relies on the target and ignores all predictors. It simply predicts the majority class. It is based on Frequency Table. The ZeroR classifier takes a look at the target attribute and its possible values. It constructs the frequency table and select its most frequent value. It will ever output the value that is most frequently found for the target attribute in the given dataset. ZeroR as its names suggests; it does not include any rule that works on the non target attributes. So more specifically it predicts the mean (for a numeric type target attribute) or the mode (for a nominal type attribute).

RIDOR Classifier

Ripple Down Rule learner (RIDOR) is also a direct classification method. RIDOR learns rules with exceptions by generating the default rule, using incremental reduced-error pruning to find exceptions with the smallest error rate, finding the best exceptions for each exception, and iterating. It generates a default rule first and then the exceptions for the default rule with the least (weighted) error rate. Then it generates the "best" exceptions for each exception and iterates until pure. Thus it performs a tree-like expansion of exceptions. The exceptions are a set of rules that predict classes other than the default. IREP is used to generate the exceptions. Incremental Reduced Error Pruning IREP is used to create the exceptions. [20] [21] [22].

RIpple-DOwn Rule learner first generates the default rule.The exceptions are generated for the default rule with the lowest (weighted) error rate. Then it generates the "best" exceptions for each exception. Thus it carries out a tree-like expansion of exceptions and its leaf has only default rule without exceptions.
Five inner classes are defined in this class. RIDOR_node class, which implements one node in the RIDOR tree. It's basically built up of a default class and its exception rules. RIDORRule class, which implements a single exception rule using REP.
The rest of the three classes are only used in RIDORRule namely Antd, NumericAntd and NominalAntd. The abstract class Antd class has two subclasses, NumericAntd and NominalAntd, to implement the corresponding abstract functions. These two subclasses implement the functions related to an antecedent with a nominal attribute and a numeric attribute respectively.

PART Classifier

This is a class for generating a PART decision list. It uses separate-and-conquer approach and builds a partial C4.5 decision tree in each iteration and makes the "best" leaf into a rule [21].

PART Classifier Algorithm steps:
1. Build a partial decision tree on the current set of instances
2. Create a rule from the decision tree
– The leaf with the largest coverage is made into a rule
3. Discarded the decision tree
4. Remove the instances covered by the rule
5. Go to step one

PERFORMANCE MEASURES USED

Various scales are used to gauge the performance of the classifiers.

Classification Accuracy

Any classifier could have an error rate and it may fail to categorize correctly. Classification accuracy is calculated as Correctly classified instances divided by Total number of instances multiplied by 100.

Mean Absolute Error

Mean absolute error is the average of the variance between predicted and actual value in all test cases. It is a good measure to gauge the performance.

Root Mean Square Error

Root mean squared error is used to scale dissimilarities between values actually perceived and the values predicted by the model. It is determined by taking the square root of the mean square error.

Confusion Matrix

A confusion matrix encompasses information about actual and predicted groupings done by a classification system.

RESULTS AND DISCUSSION

Open source machine learning tool is used to experiment the performance of different Rule based Classifiers (RIDOR, ZeroR and PART). The performance is tested out using the Training set as well as using different Cross Validation methods. The class is arrived by considering all 20 attributes of the dataset.

International Journal of Advances in Computer Science and Technology (IJACST), Vol.3 , No.11, Pages : 06-11 (2014)
Special Issue of ICCAAC 2014 - Held on December 08,2014,Hyderbad, India

Performance of ZeroR Classifier

The overall assessment summary of ZeroR Classifier using training set and different cross validation methods is given in Table I. The performance of ZeroR Classifier in terms of Correctly Classified Instances and Classification Accuracy is shown in Fig. 1 and Fig. 2. The confusion matrix for different test mode is given in Table II to Table VI. ZeroR Classifier gives 70% accuracy for the training data set. Various cross validation methods are used to check its actual performance. On an average, it gives around 70% of accuracy for credit risk estimation.

Table 1: ZeroR Classifier Complete Evaluation Summary

Test Mode	Correctly Classified Instances	Incorrectly Classified Instances	Accuracy	Mean absolute error	Root Mean Squared Error	Time Taken to Build Model (Sec)
Training Set	700	300	70%	0.4202	0.4583	0
5 Fold CV	700	300	70%	0.4202	0.4583	0
10 Fold CV	700	300	70%	0.4202	0.4583	0
15 Fold CV	700	300	70%	0.4202	0.4583	0
20 Fold CV	700	300	70%	0.4202	0.4583	0

Table 2: Confusion Matrix – ZeroR Classifier (On Training Dataset)

	Good	Bad	Actual (Total)
Good	700	0	700
Bad	300	0	300
Predicted (Total)	1000	0	1000

Table 3: Confusion Matrix – ZeroR Classifier (5 Fold Cross Validation)

	Good	Bad	Actual (Total)
Good	700	0	700
Bad	300	0	300
Predicted (Total)	1000	0	1000

Table 4: Confusion Matrix – ZeroR Classifier (10 Fold Cross Validation)

	Good	Bad	Actual (Total)
Good	700	0	700
Bad	300	0	300
Predicted (Total)	1000	0	1000

Table 5: Confusion Matrix – ZeroR Classifier (15 Fold Cross Validation)

	Good	Bad	Actual (Total)
Good	700	0	700
Bad	300	0	300
Predicted (Total)	1000	0	1000

Table 6: Confusion Matrix – ZeroR Classifier (20 Fold Cross Validation)

	Good	Bad	Actual (Total)
Good	700	0	700
Bad	300	0	300
Predicted (Total)	1000	0	1000

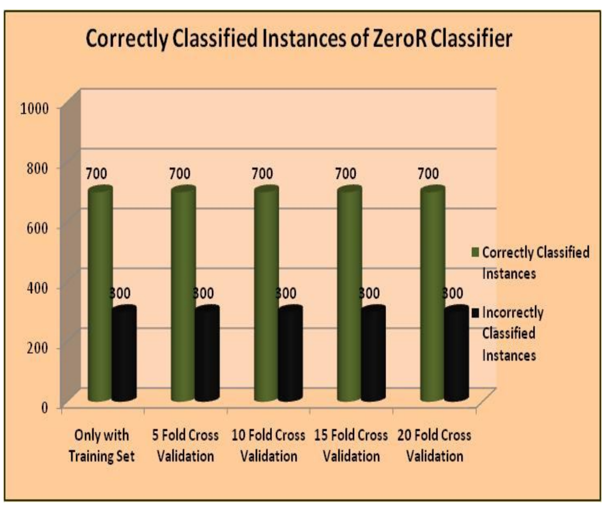

Fig. 1 Correctly Classified instances of ZeroR Classifier

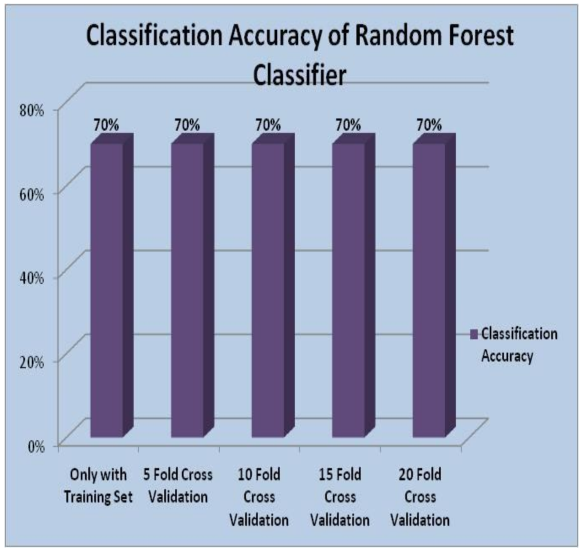

Fig. 2 Classification Accuracy of ZeroR Classifier

Performance of RIDOR Classifier

RIDOR created 5 Rules for credit Risk prediction. The overall assessment summary of RIDOR Classifier using training set and different cross validation methods is given in Table 7. The performance of RIDOR Classifier in terms of Correctly Classified Instances and Classification Accuracy is shown in Fig. 3 and Fig. 4. The confusion matrix for different test mode is given in Table 8 to Table 12. RIDOR Classifier gives 76% accuracy for the training data set. Various cross validation methods are used to check its actual performance. On an average, it gives around 71.4% of accuracy for credit risk estimation.

International Journal of Advances in Computer Science and Technology (IJACST), Vol.3 , No.11, Pages : 06-11 (2014)
Special Issue of ICCAAC 2014 - Held on December 08,2014,Hyderbad, India

Table 7: RIDOR Classifier Overall Evaluation Summary

Test Mode	Correctly Classified Instances	Incorrectly Classified Instances	Accuracy	Mean Absolute Error	Root Mean Squared Error	Time Taken to Build Model (Sec)
Training Set	760	240	76%	0.24	0.4899	0.14
5 Fold CV	714	286	71.4%	0.286	0.5348	0.06
10 Fold CV	719	281	71.9%	0.281	0.5301	0.06
15 Fold CV	708	292	70.8%	0.292	0.5404	0.03
20 Fold CV	717	283	71.7%	0.283	0.532	0.03

Table 8: Confusion Matrix – RIDOR Classifier (On Training Dataset)

	Good	Bad	Actual (Total)
Good	694	6	700
Bad	234	66	300
Predicted (Total)	928	72	1000

Table 9: Confusion Matrix – RIDOR Classifier (5 Fold Cross Validation)

	Good	Bad	Actual (Total)
Good	661	39	700
Bad	247	53	300
Predicted (Total)	908	92	1000

Table 10: Confusion Matrix – RIDOR Classifier (10 Fold Cross Validation)

	Good	Bad	Actual (Total)
Good	650	50	700
Bad	231	69	300
Predicted (Total)	881	119	1000

Table 11: Confusion Matrix – RIDOR Classifier (15 Fold Cross Validation)

	Good	Bad	Actual (Total)
Good	655	45	700
Bad	247	53	300
Predicted (Total)	902	98	1000

Table 12: Confusion Matrix – RIDOR Classifier (20 Fold Cross Validation)

	Good	Bad	Actual (Total)
Good	654	46	700
Bad	237	63	300
Predicted (Total)	891	109	1000

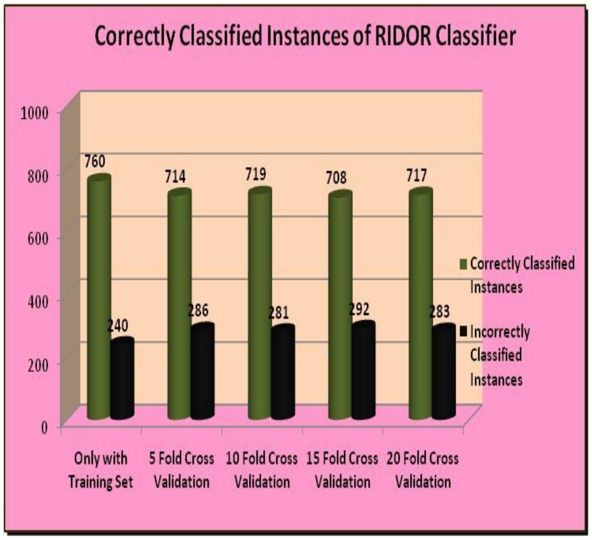

Fig. 3Correctly Classified instances of RIDOR Classifier

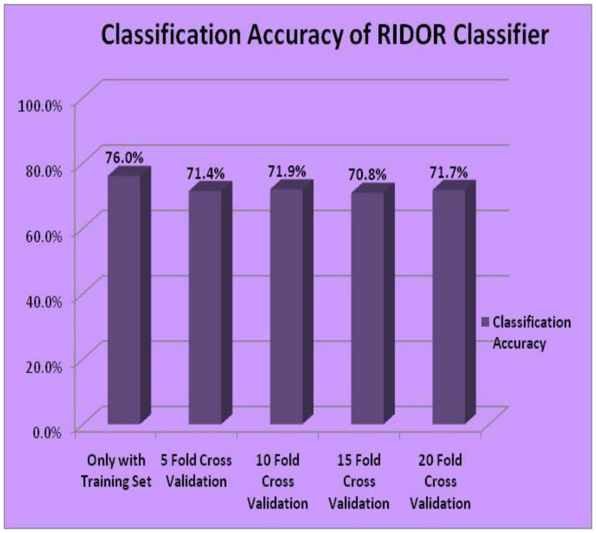

Fig. 4 Classification Accuracy of RIDOR Classifier

Performance of PART Classifier

PART created 78 Rules for credit Risk prediction. The overall assessment summary of PART Classifier using training set and different cross validation methods is given in Table 13. The performance of PART Classifier in terms of Correctly Classified Instances and Classification Accuracy is shown in Fig. 5 and Fig. 6. The confusion matrix for different test mode is given in Table 14 to Table 18. PART Classifier gives 89.7% accuracy for the training data set. Various cross validation methods are used to check its actual performance. On an average, it gives around 70.3% of accuracy for credit risk estimation.

ISSN 2320 -2602

International Journal of Advances in Computer Science and Technology (IJACST), Vol.3 , No.11, Pages : 06-11 (2014)
Special Issue of ICCAAC 2014 - Held on December 08,2014,Hyderbad, India

Table 13: PART Classifier Overall Evaluation Summary

Test Mode	Correctly Classified Instances	Incorrectly Classified Instances	Accuracy	Mean Absolute Error	Root Mean Squared Error	Time Taken to Build Model (Sec)
Training Set	897	103	89.7%	0.1605	0.2833	0.34
5 Fold CV	688	312	68.8%	0.3348	0.5101	0.08
10 Fold CV	702	298	70.2	0.3245	0.4974	0.07
15 Fold CV	726	274	72.6%	0.304	0.4828	0.06
20 Fold CV	696	304	69.6%	0.3253	0.499	0.06

Table 14: Confusion Matrix – PART Classifier (On Training Dataset)

	Good	Bad	Actual (Total)
Good	653	47	700
Bad	56	244	300
Predicted (Total)	709	291	1000

Table 15: Confusion Matrix – PART Classifier (5 Fold Cross Validation)

	Good	Bad	Actual (Total)
Good	548	152	700
Bad	160	140	300
Predicted (Total)	708	292	1000

Table 16: Confusion Matrix – PART Classifier (10 Fold Cross Validation)

	Good	Bad	Actual (Total)
Good	561	139	700
Bad	159	141	300
Predicted (Total)	720	280	1000

Table 17: Confusion Matrix – PART Classifier (15 Fold Cross Validation)

	Good	Bad	Actual (Total)
Good	577	123	700
Bad	151	149	300
Predicted (Total)	728	272	1000

Table 18: Confusion Matrix – PART Classifier (20 Fold Cross Validation)

	Good	Bad	Actual (Total)
Good	562	138	700
Bad	166	134	300
Predicted (Total)	728	272	1000

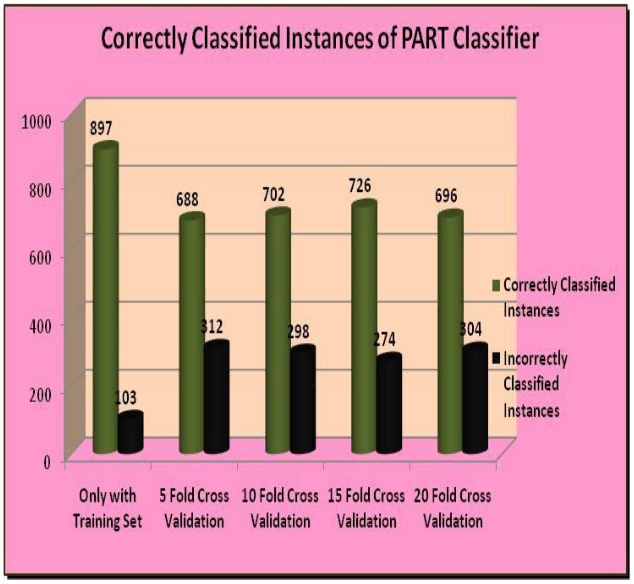

Fig. 5 Correctly Classified instances of PART Classifier

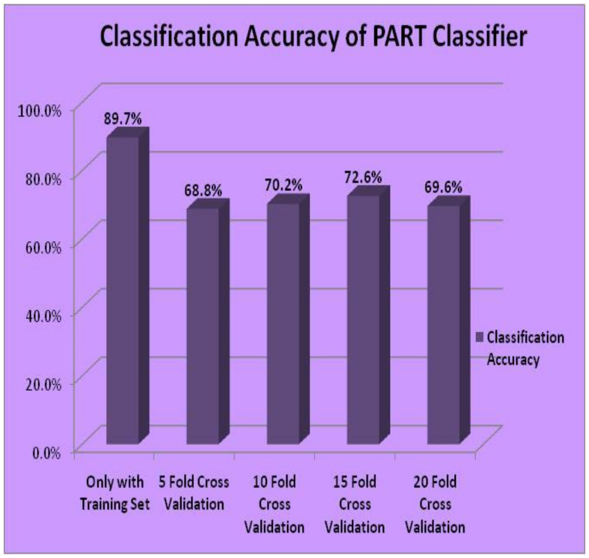

Fig. 6 Classification Accuracy of PART Classifier

6.1. Comparison of RIDOR, ZeroR and PART Classifiers

The comparison of performance between RIDOR, ZeroR and PART Classifiers is depicted in Fig 7, and Fig. 8 in terms of Correctly Classified Instances and Classification Accuracy. The complete ranking is prepared based on correctly classified instances, classification accuracy, MAE and RMSE values and other statistics found using Training Set result and Cross Validation Techniques. Consequently, it is perceived that RIDOR classifier outperforms the other two Classifiers.

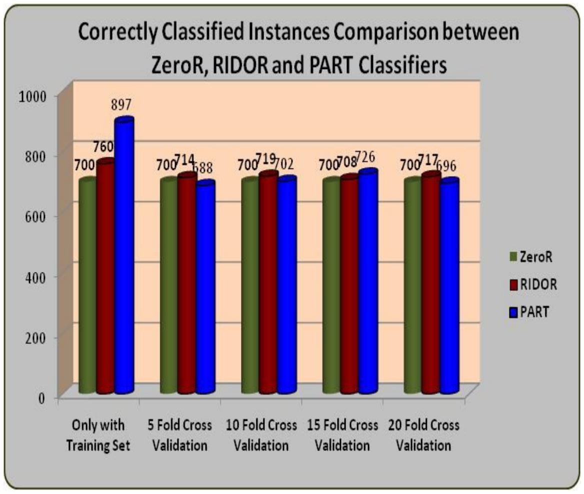

Fig. 7 Correctly Classified Instances Comparison between RIDOR, ZeroR and PART Classifiers

International Journal of Advances in Computer Science and Technology (IJACST), Vol.3 , No.11, Pages : 06-11 (2014)
Special Issue of ICCAAC 2014 - Held on December 08,2014,Hyderbad, India

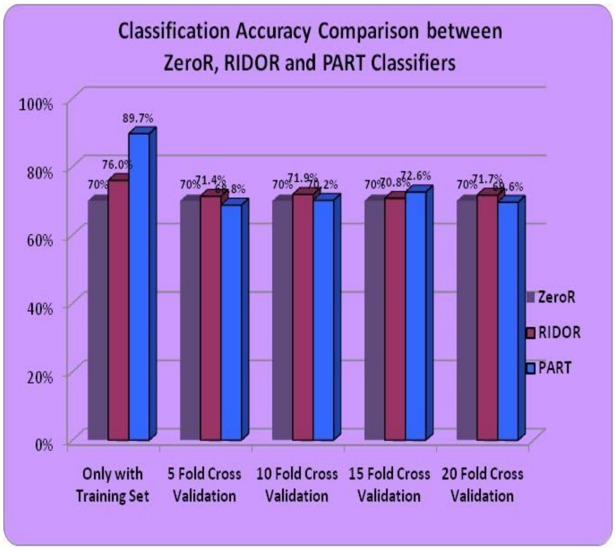

Fig. 8 Classification Accuracy Comparison between RIDOR, ZeroR and PART Classifiers

CONCLUSION

This work investigated the efficiency of three different classifiers namely, RIDOR, ZeroR and PART Classifiers for credit risk prediction. Testing is accomplished using the open source machine learning tool. Also, effectiveness comparison of both the classifiers has been done in view of different scales of performance evaluation. At last, it is observed that RIDOR Classifier performs best, followed by PART Classifier and then by ZeroR Classifier for credit risk prediction by taking various measures including Classification accuracy, Mean Absolute Error and Time taken to build the model.

ACKNOWLEDGMENT

The author expresses her deep gratitude to the Management of IBS Hyderabad, IFHE University and Operations & IT Department of IBS Hyderabad for constant support and motivation.

REFERENCES

[1] Germano C. Vasconcelos, Paulo J. L. Adeodato and Domingos S. M. P. Monteiro, “A Neural Network Based Solution for the Credit Risk Assessment Problem,” Proceedings of the IV Brazilian Conference on Neural Networks - IV Congresso Brasileiro de Redes Neurais pp. 269-274, July 20-22, 1999.

[2] Tian-Shyug Lee, Chih-Chou Chiu, Chi-Jie Lu and I-Fei Chen, “Credit scoring using the hybrid neural discriminant technique,” Expert Systems with Applications (Elsevier) 23, pp. 245–254, 2002.

[3] Zan Huang, Hsinchun Chena, Chia-Jung Hsu, Wun-Hwa Chen and Soushan Wu, “Credit rating analysis with support vector machines and neural networks: a market comparative study,” Decision Support Systems (Elsevier) 37, pp. 543– 558, 2004.

[4] Kin Keung Lai, Lean Yu, Shouyang Wang, and Ligang Zhou, “Credit Risk Analysis Using a Reliability-Based Neural Network Ensemble Model,” S. Kollias et al. (Eds.): ICANN 2006, Part II, Springer LNCS 4132, pp. 682 – 690, 2006.

[5] Eliana Angelini, Giacomo di Tollo, and Andrea Roli “A Neural Network Approach for Credit Risk Evaluation,” Kluwer Academic Publishers, pp. 1 – 22, 2006.

[6] S. Kotsiantis, “Credit risk analysis using a hybrid data mining model,” Int. J. Intelligent Systems Technologies and Applications, Vol. 2, No. 4, pp. 345 – 356, 2007.

[7] Hamadi Matoussi and Aida Krichene, “Credit risk assessment using Multilayer Neural Network Models - Case of a Tunisian bank,” 2007.

[8] Lean Yu, Shouyang Wang, Kin Keung Lai, “Credit risk assessment with a multistage neural network ensemble learning approach”, Expert Systems with Applications (Elsevier) 34, pp.1434–1444, 2008.

[9] Arnar Ingi Einarsson, “Credit Risk Modeling”, Ph.D Thesis, Technical University of Denmark, 2008.

[10] Sanaz Pourdarab, Ahmad Nadali and Hamid Eslami Nosratabadi, “A Hybrid Method for Credit Risk Assessment of Bank Customers,” International Journal of Trade, Economics and Finance, Vol. 2, No. 2, April 2011.

[11] Vincenzo Pacelli and Michele Azzollini, “An Artificial Neural Network Approach for Credit Risk Management”, Journal of Intelligent Learning Systems and Applications, 3, pp. 103-112, 2011.

[12] A.R.Ghatge and P.P.Halkarnikar, “Ensemble Neural Network Strategy for Predicting Credit Default Evaluation” International Journal of Engineering and Innovative Technology (IJEIT) Volume 2, Issue 7, January 2013 pp. 223 – 225.

[13] Lakshmi Devasena, C., “Adeptness Evaluation of Memory Based Classifiers for Credit Risk Analysis,” Proc. of International Conference on Intelligent Computing Applications - ICICA 2014, 978-1-4799-3966-4/14 (IEEE Explore), 6-7 March 2014, pp. 143-147, 2014.

[14] Lakshmi Devasena, C., “Adeptness Comparison between Instance Based and K Star Classifiers for Credit Risk Scrutiny,” International Journal of Innovative Research in Computer and Communication Engineering, Vol.2, Special Issue 1, March 2014.

[15] Lakshmi Devasena, C., “Effectiveness Assessment between Sequential Minimal Optimization and Logistic Classifiers for Credit Risk Prediction,” International Journal of Application or Innovation in Engineering & Management, Volume3, Issue 4, April 2014.

[16] Lakshmi Devasena, C., “Efficiency Comparison of Multilayer Perceptron and SMO Classifier for Credit Risk Prediction,” International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 4, 2014.

[17] Lakshmi Devasena, C. “Competency Assessment between JRip and Partial Decision Tree Classifiers for Credit Risk Estimation”, International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 4 (5), May – 2014, pp. 164-173.

[18] Lakshmi Devasena, C. 2014. Competency Comparison between Logistic Classifier and Partial Decision Tree Classifier for Credit Risk Prediction. Operations Research and Applications: An International Journal (ORAJ), Vol. 1, No.1, August 2014, pp. 31 – 40.

[19] UCI Machine Learning Data Repository – http://archive.ics.uci.edu/ml/datasets.

[20] Ian H. Witten, Eibe Frank, Mark A. Hall. Data Mining Practical Machine Learning Tools and Techniques, Third Edition, Morgan Kaufmann Publishers is an imprint of Elsevier.

[21] C. Lakshmi Devasena, T. Sumathi, V.V. Gomathi and M. Hemalatha. Effectiveness Evaluation of Rule Based Classifiers for the Classification of Iris Data Set. Bonfring International Journal of Man Machine Interface, Vol. 1, Special Issue, December 2011, pp. 5 - 9.

[22] M. Thangaraj, C.R.Vijayalakshmi. Performance Study on Rule-based Classification Techniques across Multiple Database Relations. International Journal of Applied Information Systems (IJAIS), Foundation of Computer Science FCS, New York, USA Volume 5– No.4, March 2013.