iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** VI **Month of publication:** June 2017

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

A Hybrid Technique for creating classification model using Random Committee and Voted Perceptron Classifier

Prashant Sahu[1], Rohit Miri[2]
[1,2] Dr. C.V. Raman University, Bilaspur, Chhattisgarh, India

Abstract*: Various researchers were experimenting with various manual or automated data sets, sometimes we found that due to high error rate various classification tasks of data mining lands into failure zone. This paper tries to give an hybrid approach for reducing error rate during classification tasks with the help of combination of classifiers. All these algorithms are experimented in weka tool. Each algorithm is performed on same data set. It is illustrated that combination of classifiers work better than individual classifiers on some performance measures.
Keyword:* random committee, voted perceptron, weka, classification, MAE, RMSE, RRSE.

I. INTRODUCTION

Classification is an necessary task in data mining which is important for converting critical information to knowledge and also helpful for taking various decisions for complex problems[7]. Random committee is also a classifier which comes under the meta classifier category, it is implemented in weka and proceesed using java classes. It is used for build the platform for randomizable base classifier. Its final prediction is average of the prediction which is generated by base classification algorithms[8], while voted perceptron classifier is based on the perceptron algorithm which is given by Rosenblatt and Frank. It takes data with large margins, it can also be used in high dimensional spaces[9].

II. RELATED WORK

According to [1] Sahilpreet Singh et al (2013) focused on to find the best neural network classification algorithm for intrusion detection. It uses four neural network classifiers namely multilayer perceptron, RBF network, logistic regression and voted perceptron for experiments in weka. This work uses NSL KDD dataset which contains 25192 instances and 41 attributes. In their experiment multilayer perceptron emerges as winner by outperforms other three algorithms on the basis of kappa statistic, RMSE, RAE, RRSE, computation time etc.

According to [2] Sonali Agrawal et al(2012) focused on to illustrate the importance of data mining in education sector for improving the student performance in the available resources. This work uses evaluation approach of different classifiers like vote perceptron,logistic, MLP, SMO, winnow, LIBSVM, RBF network in weka and also analysed the libsvm classification accuracy. In their experiment, libsvm emerged as best choice on the basis of less RRSE, and highest classification accuracy in radial basis.

According to [3] Yugal Kumar et al (2013) focused on to provide the best algorithm for bank data set by evaluating six different classifiers namely bayes net, naive bayes, naive bayes updateable, multilayer perceptron, voted perceptron and J48. The experiment uses 2800 instances and in result, J48 outperforms other algorithms in terms of MAE, RMSE, time taken etc.

According to [4] Nathon Aston et al(2014) focused to evaluate methods for sentiment determination and also analysing tweets which contains top features. It uses sander corpus data set for evolution. After experiment, researchers found that voted perceptron is less effective in terms of best learning rule and gram size but when it is combined with perceptron it becomes effective on both measures.

According to [5] Abdelhmid Salih mohmed salih et al(2014) focued on to produce an decision support system for health care system using multiple meta classifier like adaboost m1, bagging, logitboost, random committe, stacking etc. This work evaluates ensemble design and giving hybrid algorithm to develop decision support system, it based on various parameters like recall, precision etc.

III. METHODOLOGY

Methodology used in this work consist of following steps:

Volume 5 Issue VI, June 2017
ISSN: 2321-9653

A. Data Selection and Preparation

In this work, data set is selected from an online data repository[6] which is in arff file format. This dataset includes 435 instances and 15 attributes.

B. Tool Selection

Weka tool used in this work which supports various data mining tasks and it also includes collection of different classifiers.

C. Preprocessing Data

Now the data set is loaded into weka using browse button and then a supervised filter which is attribute based is applied on the data set which remove unnecessary attributes and therefore 5 attributes were selected.

D. Classification Iteration 1

Then we go to classify tab and set percentage split=66% and select random committee algorithm and select run, output appears on screen.

E. Classification Iteration 2

Then we go to classify tab and set percentage split=66% and select voted perceptron algorithm and select run, output appears on screen.

F. Classification using prposed approach

Then go to vote classifier which is lies under meta classifier section choose vote and then click on vote and under base classifier click and remove Zero R and add random committe and voted perceptron then click on run.

IV. RESULTS AND ANALYSIS

TABLE I
List of Selected Attributes after Preprocessing Data

adoption-of –the-budget-resoultion
Physician-free-freeze
immigration
synfuel-corporation-cutback
Class

TABLE II
Result after classification iteration 1 (Using Random Committee)

Kappa Statistic	MAE	RMSE	RAE	RRSE
0.9447	0.0459	0.1394	9.5896	28.1166

TABLE III
Result after classification iteration 2 (Using Voted Perceptron)

Kappa Statistic	MAE	RMSE	RAE	RRSE
0.9585	0.0203	0.1424	4.2722	28.7173

Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

TABLE IV

Result after Classification using our proposed approach(Random Committe+ Voted Perceptron)

Kappa Statistic	MAE	RMSE	RAE	RRSE
0.9585	0.042	0.1331	8.7777	26.8394

Table V

Comparision Analysis with Previous Works

Performance Measure	According to [1] for Voted Pereptron	Our Approach (Random Committee + Voted Perceptron)
MAE	0.11	0.042
RMSE	0.34	0.1331
RAE	23.71	8.7777
RRSE	68.87	26.8394

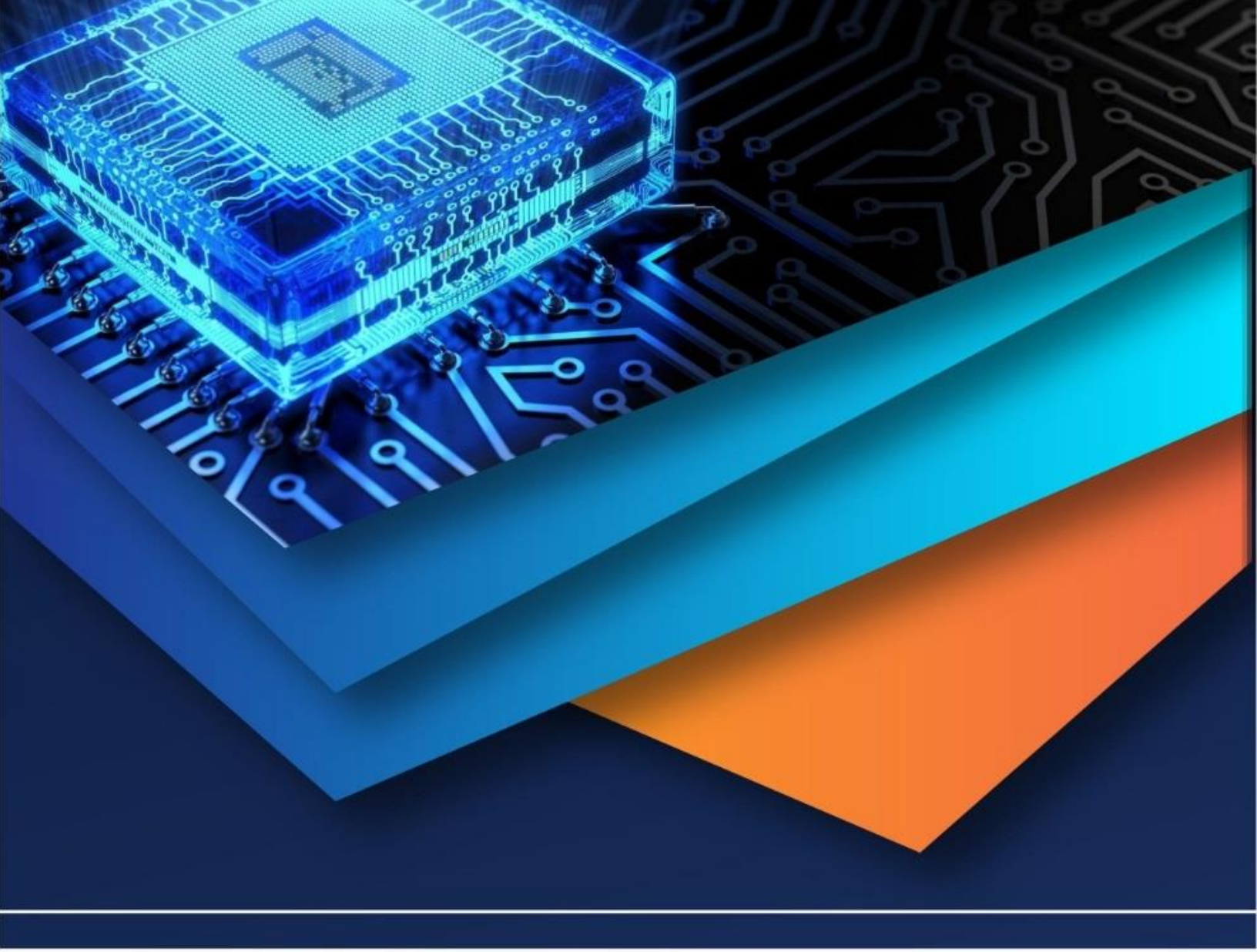
By this comparison we can say that our proposed approach shows minor improvement in MAE, RMSE and major improvement in RAE, RRSE.

V. CONCLUSION

This research work is performed to minimize the error rate so that we can reach the nearest possible to correct classification. For this we have present a hybrid approach by combining two classifiers. After doing this research work it is clear that our approach is successful in reducing error rate. The algorithms are applied to vote dataset after reducing its attributes and implemented using weka tool.

REFERENCES

[1] Sahilpreet Singh, Meenakshi bansal,"Improvement of intrusion detection system in data mining using neural network", International Journal Of Advanced Research in computer science and software engineering(IJARCSSE), Septemeber 2013
[2] Sonali Agrawal, G.N. Pandey, M.D. Tiwari,"Data Mining in education: data classification and decision tree approach", International journal of e-education, e-business, e-business, e-management and e-learning, April 2012
[3] Yugal Kumar, G. Sahoo, "Study of parametric performance evaolution of machine learning and statistical classifier", I.J. Information technology and computer science, 2013
[4] Nathon Aston, Jacob Liddle, WeiHu,"Twitter sentiment in data stream with perceptron", Journal of computer and communications ,2014
[5] Abdelhmid Salih Mohmed salih, Ajith Abraham,"Novel ensemble decision support and health care monitoring system", 2014
[6] http://storm.cis.fordham.edu/~gweiss/data-mining/weka-data/vote.arff
[7] Prashant Sahu, Jageshwer Shriwas,RohitMiri and Asha Miri,"Survey Paper on Various Classification Approaches in Data Mining",CiiT International Journal of Sotware Engineering, Vol , No 1, Jnauary 2017
[8] http://weka.sourceforge.net/doc.dev/weka/classifiers/meta/RandomCommittee.html
[9] http://curtis.ml.cmu.edu/w/courses/index.php/Voted_Perceptron



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY