International Journal of Future Innovative Science and Engineering Research (IJFISER)
Volume - 2, Issue - I ISSN (Online): 2454- 1966

Research Manuscript Title

Analysis of Different Classification Algorithms Applied to Anneal Dataset Using Data Mining Techniques

V.Kamalakkannan[1], Dr.D. Ramyachitra[2],

[1]M Phil Research Scholar, [2]Assistant Professor,

[(1&2)]Computer Science Department

Bharathiar University,

Coimbatore, Tamil Nadu, India

E-Mail: *kamalakkannan.vaiyapuri@gmail.com, jaichitra1@gmail.com*

March – 2016

www.istpublications.com

International Journal of Future Innovative Science and Engineering Research (IJFISER), Volume - 2, Issue – I,
ISSN (Online): 2454- 1966 www.istpublications.com.

Analysis of Different Classification Algorithms Applied to Anneal Dataset Using Data Mining Techniques

V.Kamalakkannan[1], Dr.D. Ramyachitra[2],

[1]M Phil Research Scholar, [2]Assistant Professor,
[(1&2)]Computer Science Department
Bharathiar University,
Coimbatore, Tamil Nadu, India
E-Mail: *kamalakkannan.vaiyapuri@gmail.com, jaichitra1@gmail.com*

ABSTRACT

The classification manner refers to a step by step procedure for a given data into anybody of the specific classes within the desktop finding out. There are countless classification drawback happens and need to be solved. For the period of this paper we are examining the efficiency of the three Rule headquartered classifiers algorithms namely Naïve Bayes, SMO, OneR and J48 algorithms. The Anneal dataset information set is used for the experimental assessment of the classification algorithms. To foretell the exact classification of the algorithms the learning set parameter is used and the results have been tabulated. For experimental assessments we used the efficiency metrics such because the classification accuracy and the error charges to inspecting the results. Sooner or later, we discovered great performance algorithm for the anneal dataset.

Keywords- *Naïve Bayes, SMO, OneR, Classification Algorithms, WEKA, Anneal dataset.*

1. INTRODUCTION

The classification technique is a logical process to construct classification models from an enter knowledge set. For examples incorporate rule-established classifications, determination tree classifications, neural networks, and Naïve Bayes classifications. Each method employs finding out algorithm to establish a mannequin that first-class matches the association among the attributes set and sophistication label of the enter information [1].

The mannequin generated with the aid of a learning algorithm will have to each match the input data best with competently predict the class label of records it has not ever obvious before. Then, a key objective of the educational algorithm is to construct model and pleasant generalization capability. The units appropriately predict the class labels of before unidentified documents. It is quite often worthwhile to calculate the performance of the model on the test set because such calculate provides balanced estimate of its simplification error. The accuracy computed from the test set can also be used to evaluate the relative performance of more than a few classifiers on the equivalent discipline [2].

In this article assessment has been achieved with computer studying algorithms to find which experiment alternative is quality for Different classifier algorithms namely Naïve Bayes, SMO, OneR and J48. Within the testing choice there are 4 forms of parameters corresponding to training set, percentage split, move validation, and provided experiment set. This paper makes use of the Anneal dataset data set for assessment of those algorithms. Part 2 describes the literature evaluate, section 3 describes the methodology for the Anneal dataset data set and part four describes our experiments analysis. And subsequently section 5 gives the conclusion and future work.

2. LITERATURE REVIEW

V. Muralidharan et al., presented the use of Naïve Bayes algorithm and Bayes Net algorithm for fault diagnosis through discrete wavelet features extracted from vibration signals of good and faulty conditions of the components of centrifugal pump. Classification accuracies of unusual discrete wavelet families were calculated and compared to find the best wavelet for the fault diagnosis of the centrifugal pump [3].

R. Sujatha, determined the different classification techniques to produces the excellent prediction for Diabetes, Soy Beans and Wheat seed datasets. They used the various classification techniques are ZeroR, OneR, Decision Table, Naïve Bayes, PART, SMO, J48, Random Tree. Then the results are evaluated based on the accuracy, paired T test and statistical methods.

They decided SMO is the best classification algorithm for those datasets when compared with other algorithms in their research [4].

S. R. Kalmegh et al., discussed classification of Indigenous news the use of decision table and OneR algorithm. Their experiment end result it's miles determined that OneR algorithm perform well in categorizing within the all the information. The OneR algorithm plays nicely in categorizing in the news associated with medical, education and sports activities. as a result the OneR classification set of rules as better overall performance as compared with choice desk algorithms [5].

P.Yasodha et al., proposed the J48, J48Graft, Random Tree, REP and LAD machine learning classification algorithms using WEKA Tool. The diabetic dataset has been predicted using classification algorithms. Thus the experimental analyze results had been discussed with performance metrics, Accuracy and time estimation using with diabetic dataset. At last found as J48 as high performance when compared with other classification algorithms. Then, the J48Graft founded as low performance [6].

3. METHDOLOGY

The classification algorithms are generally used in different applications. This paper analyzing different classification algorithms such as Naïve Bayes, SMO, OneR and J48 to find which is more effectively for classification of a anneal cancer.

Naïve Bayes

The Naïve Bayes is a easy probabilistic classifier. Naïve Bayes is based on the idea of mutual independency of attributes. This classifier is especially scalable, requiring a no of parameters. The algorithm works on the belief, that variables provided to the classifier are unbiased. The Naive Bayesian classifier is based totally on Bayes' theorem with independence assumptions among predictors [7]. A Naive Bayesian model is easy to construct and not using a complicated iterative parameter estimation which makes it mainly useful for very large datasets. No matter its simplicity, the Naive Bayesian classifier frequently does notably properly and is broadly used because it often outperforms greater state-of-the-art type methods [8].

Sequential Minimal Optimization (SMO)

The SVM wants the answer of very enormous Quadratic Programming (QP).SMO breaks gigantic QP situation into a series of smallest feasible QP disorders. These small QP problems are solved analytically, which avoids using a time-drinking numerical QP optimization as an internal loop. The variety of reminiscence required for SMO is linear in the coaching set dimension, which allows SMO to handle very gigantic coaching units. Sequential Minimal Optimization (SMO) is a simple algorithm that can quickly solve the SVM QP quandary with none extra matrix storage and without utilising numerical QP optimization steps in any recognize. SMO decomposes the whole QP obstacle into QP sub- issues, utilizing Osuna's theorem to make exact convergence. No longer just like the prior approaches, SMO chooses to resolve the smallest viable optimization quandary at each and every step. For the typical SVM QP concern, the smallest feasible optimization dilemma entails two Lagrange multipliers, considering the Lagrange multipliers ought to obey a linear equality constraint. At every step, SMO chooses two Lagrange multipliers to collectively optimize, finds the most beneficial values for these multipliers, and updates the SVM to mirror the new most useful values [9].

OneR Classifier

The OneR or "One Rule" is a simple Classification algorithm. The OneR algorithm creates one rule for each attribute of training data and then chooses the rule with the least error rate. To make a rule for an attribute, the most often class for each attribute worth must be determined. The most often class is just the class that appears most frequent for that attribute value. A rule is just a set of attribute value bound to their most class. The OneR chooses the rule by the least error rate. If two or more rules have the similar error rate, then the rule is selected at random [10].

J48

The J48 is an implementation of C4.5 algorithm. C4.five algorithm produce decision tree classification for a given dataset through recursive department of the records and the decision tree is grown the use of intensity-first strategy [11]. On statistics checking out this algorithm will emphasised on splitting dataset and through choosing a check in order to provide quality bring about information benefit. In discrete attributes as well, these algorithms recollect a take a look at with a end result of many because the quantity of various values and check binary characteristic for every characteristic will keep growing in one-of-a-kind values each attribute could be considered. so one can gather the entropy advantage of these types of binary checks efficiently, the education data set belonging to the node in attention is taken care of for the values of the non-stop characteristic and the entropy profits of the binary reduce based on each wonderful values are calculated in a single scan of the sorted records. This process is repeated for each continuous attributes [12].

4.EXPERIMENTS ANALYSIS

In this experimental analysis the Anneal dataset data set is taken as input for classification in Weka tool with accuracy and high efficiency. Weka is a collection of machine learning algorithm for data mining tasks. This research used the Weka open source data mining tools for modelling of anneal dataset data. To determine the best technique for the anneal dataset from the classification by using the following the performance factors such as highest accuracy, and least error rate.

TABLE I: ACCURACY MEASURE BY CLASS FOR RULE BASED CLASSIFIER

Algorithms	Naïve Bayes	SMO	OneR	J48
Correctly Classified	97.66	97.01	83.65	95.88
Incorrectly Classified	2.34	2.99	16.35	4.12

By using the training set option to predict the potential relationship between the correctly classified instances and incorrectly classified instances for different classification algorithms. The accuracy measure by class for the Rule based classifier as depicted in Table 1. Therefore the Naïve Bayes classification algorithm has high accuracy to compare with other algorithms, which has been shown figure 1.

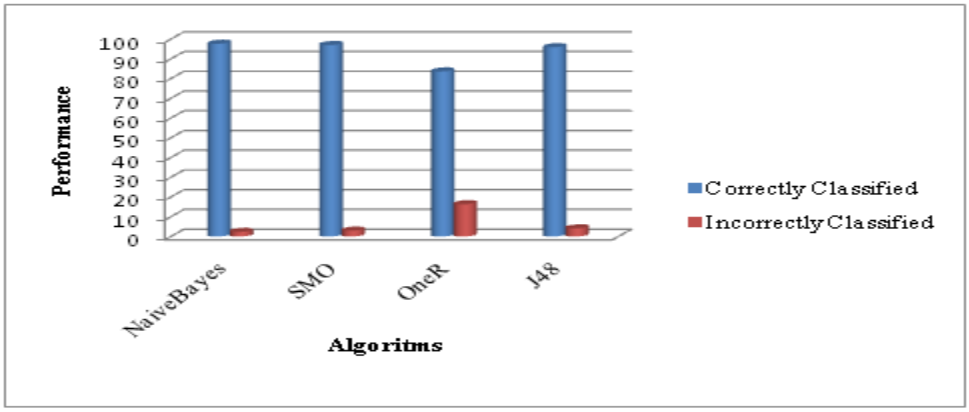

Figure 1: Accuracy Measure for Rule Based Classifier

TABLE II: ACCURACY MEASURE FOR IN RULE BASED CLASSIFIERS

Algorithms	TP Rate	FP Rate	Precision	F-Measure	ROC Area	Kappa Statistics
Naïve Bayes	0.97	0.03	0.97	0.97	0.99	0.94
SMO	0.97	0.03	0.97	0.97	0.97	0.94
OneR	0.83	0.52	0.7	0.76	0.66	0.44
J48	0.96	0.06	0.96	0.96	0.96	0.9

For Naïve Bayes algorithm it is inferred that for training set, TP Rate, Precision, F-Measure, ROC area and Kappa Statistics values fluctuates when the dataset is increased. The Naïve Bayes and SMO performs high TP rate, Precision value, F-Measure and Kappa Statistics. But the ROC Value high when compared with other algorithms. Then the Naïve Bayes had been highest accuracy when compared with other algorithms that has given table 2. From the given graph (Figure 2) the Naïve Bayes classification algorithm performs better, because it contains highest accuracy when compared with SMO, OneR and J48 algorithms.

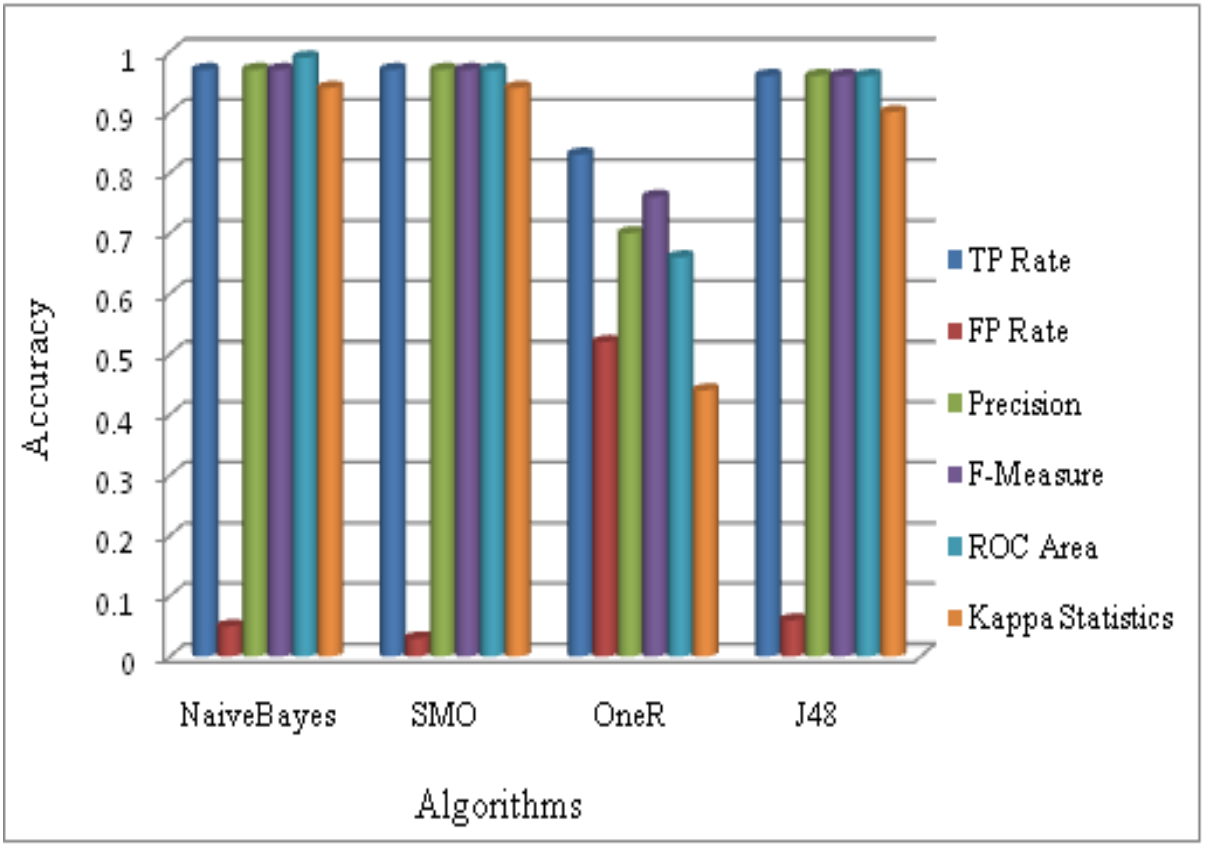

Figure 2: Accuracy Parameters for Rule based classification

TABLE III: ACCURACY ERROR RATE MEASURE FOR RULE BASED CLASSIFICATION

Algorithms	MAE	RMSE	RAE	RRSE
Naïve Bayes	0.01	0.08	13.46	33.67
SMO	0.22	0.31	165.32	120.06
OneR	0.05	0.23	40.5	90
J48	0.02	0.09	15.44	36.08

The SMO, OneR and J48 classification algorithms reach the highest error rate. Then the Naïve Bayes classification algorithm performs better because it carry lowest error rate when compared to SMO, OneR and J48 algorithms. Thus the error rate measure by class for the different classifiers as depicted in Table 3.

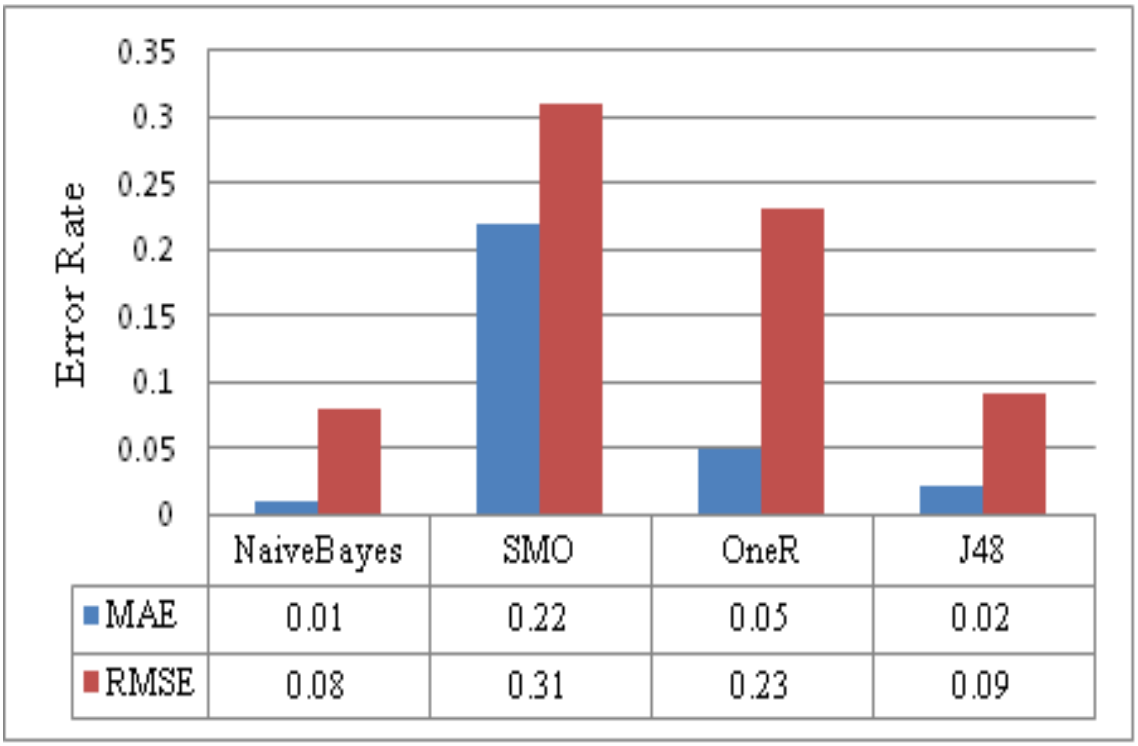

Figure 3: Accuracy Error Rate of MAE & RMSE for

The Mean absolute error and Root mean squared error to find the least error for different classification algorithm. Then the Naïve Bayes algorithm had been performed better when compared with SMO, OneR and J48 algorithms. Relative absolute and Root relative squared errors are used to determine the better performance when compared with above mentioned different classification algorithms. The accuracy error rate measure for Rule based classifier can be exposed in Figure 3 & 4.

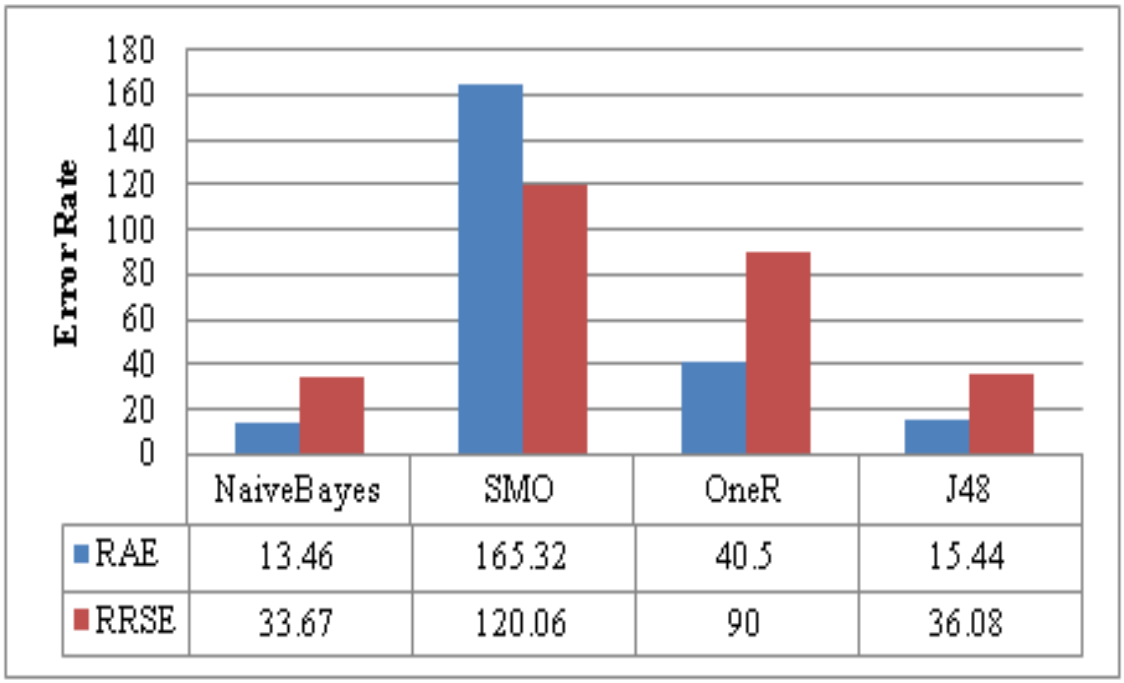

Figure 4: Accuracy Error Rate of RAE & RRSE for Anneal dataset.

5. CONCLUSION

On this paper used classification algorithms to understand the nice classifier from the sections of Naïve Bayes, SMO, OneR and J48 algorithms. Here used the anneal dataset for calculating the performance with the aid of using training set experiment. The different classification algorithms are used experiment comparison with accuracy and error fee via utilising Weka instrument. The algorithms conclude that the Naïve Bayes suits the first-class from Classifier compare to SMO, OneR and J48.

In future the other classification algorithms will also use for experimented with other dataset. Then in future we are able to alternate or merge the Naïve Bayes algorithm to receive more potent outcome. The rule classification algorithms can analyze utilizing a lot of parameters specifically cross validation, furnished test set.

REFERENCES

[1] Ian H. Witten, Eibe Frank, Mark A. Hall,"Data Mining Practical Machine Learning Tools and Techniques",3rd Edition, Morgan Kaufmann Publishers is an imprint of Elsevier 30 Corporate Drive, Suite 400, Burlington, MA 01803, USA

[2] V. Cherkassky and F. Milier, Learning from Data: Concepts, Theory, and Methos. Wiley Interscience, 1998

[3] V. Muralidharan et al., "A Comparative study of Naïve Bayes classifier and Bayes net classifier for fault diagnosis of monoblock centrifugal pump using wavelet analysis", Applied soft computing 12(2012) 2023-2029.

[4] R.Sujatha, D.Ezhilmaran, "Evaluation of Classifiers to Enhance Model Selection", International Journal of Computer Science & Engineering Technology (IJCSET) ISSN: 2229-3345 Vol. 4 No. 01 Jan 2013.

[5] S. R. Kalmegh, S. N. Deshmukh," Effective Evaluation of Classification of Indigenous News Using Decision Table and OneR Algorithm", International Journal of Advanced Information Science and Technology (IJAIST), Vol.26, No.26, ISSN: 2319:2682.

[6] P.Yasodha, N.P. Ananthanarayanan, "Comparative Study of Diabetic Patient Data's Using Classification Algorithm in WEKA tool", International journal of Computer Applications Technology and Research, Volume 3-Issue 9, 2014, ISSN: 2319- 8656.

[7] Nong Y., *the Handbook of Data Mining* (Lawrence Earlbaum Associates, 2003).

[8] Mohit, Rohit Ranjan Verma, Sameeksha Katoch, Ashoka Vanjar, S N Omkar, "Classification of Complex UCI Datasets Using Machine lerning Algorithms Using Hadoop", Internation Journal of Computer Sceince and Software Engineering (IJCSSE), Volume 4, Issue 7, July, ISSN: 2409-4285.

[9] Gauray Taneja, Ashwin Senthi, "Comparison of Classification in Data Mining", IJCSMC, Vol.3, Issue 11, Pg: 102-115, ISSN 2320 – 088X

[10] Y. Zhao and Y. Zhang, "Comparison of Decision Tree Methods for Finding Active Objects," National Astronomical Observatories, Advances of Space Research, 2007.

[11] I. H. Witten, and E. Frank, "Data Mining Practical Machine Learning Tools and Techniques," Second Edition, Morgan Kaufmann Publisher, United States of America, 2005.

[12] Ian H. Witten, Eibe Frank. Data Mining: Practical Machine Learning Tools and Techniques with Java Implementation. Morgan Kaufmann Publishers. ISBN 1-55860-552-5