



International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)

(Open Access, Double Blind Peer-reviewed, Refereed and Indexed Journal)

www.iasir.net

NB Classifier Analysis among the Meteorological and Solar Activity Parameters

R. K. Tiwari¹, Shailendra Singh² and K. P. Tiwari³

^{1,2}Govt. M.S. Golvalkar College, Rewa, Madhya Pradesh, India

³Govt. Girls College, Bilaspur CG, India

Abstract: Climatologically approach is used to investigate the links between Solar and climate parameters between 1971-2000. We have plotted the cost benefit analysis curve between rainfall, temperature, sunspot number and geomagnetic indices (ap) values for the period covering almost three solar cycles-21, 22, 23. Our finding supports the expected outcome of the correlation for proposed data sets. The results show an important step in identifying reliable connections between meteorological parameters and solar activity parameters in predicting earth's climate.

Keywords: Earth Climate, NB Classifier, Solar and geo-activity parameters, Meteorological parameters

I. Introduction

The Naive Bayesian classifier deals with discrete and continuous attributes and has exceptional performance in real world problems. It explains its decisions as the sum of informational gains. In the present analysis, we have used the Naïve Bayes Classifier. Naïve Bayes is easy and fast to predict class of test data set. It also performs well in multi class prediction. When assumption of independence holds, a naïve bayes classifier performs better compare to other models like logistic regression [1,2]. This algorithm has many advantages. It is fast to train (single scan) and classify. It handles real, discrete and streaming data very well. Naive Bayes classifier and collaborative filtering together builds a recommendation system that uses machine learning and data mining techniques to filter unseen information and predict whether a user would like a given resource or not.

II. Methodology

Naïve Bayes classifier algorithm is the best technique of data mining to make prediction in real time. It is also able to predict the probability of multiple classes of target attribute [3]. In the present work we have selected solar parameters (velocity, sunspot number, ap values) and climate parameters (temperature and rainfall) of three solar cycles, i.e. from year 1971-2000. In this experiment v, ap, SSN and temperature are base class attribute and rainfall is a target attribute class. Naïve Bayes algorithm classify and analyze all the data set in the following steps :

- Collect all relevant data required for study and normalized data set into comma separated value (CSV) format.
- Divide normalized data into training and testing sets as the ratio 3:1 respectively.
- Import training data set into Naïve Bayes Model and generate attribute relation file format.
- The algorithm used in the model convert data set into frequency table and likelihood table by finding probabilities.
- Naïve Bayes algorithm calculates the posterior probability for each class. The class with the highest posterior probability is the outcome of prediction.
- NB algorithm generate summary for each base classes and target class. Finally confusion matrix is generated to predict the result.
- For the testing set, we repeat the above steps. We now compare and analyze the result of both training set and testing set.

III. Result and Discussion:

In the present work, we have generated confusion matrix and evaluation report of training and testing data set with the help of naïve bayes classifier. Fig.-1,2 shows cost/benefit analysis of target class. Classification accuracy observed by threshold curve is 87% and 81% respectively for training and testing set. According to evaluation summary 83% instances are correctly classified for training set and 76% for testing set.

In predictive analysis, a table of confusion with two rows and two columns reports the number of false positives, false negatives, true positives, and true negatives. The final table of confusion contains the average values for all the classes. According to confusion matrix, true positive value shows all the base class attributes (i.e. velocity, sunspot number, ap values, temp.) that affect the rainfall. The percentage of classified instances in

true positive is 83%(57) and 76%(55) for training and testing data sets respectively. Similar study is performed by many researchers in the Implementation of Data Mining Techniques for Meteorological Data Analysis [4-6].

TRAINING SET							TESTING SET									
=== Detailed Accuracy By Class ===							=== Detailed Accuracy By Class ===									
TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area			
PRC Area Cls							PRC Area Cls									
0.864	0.235	0.877	0.864	0.870	0.624	0.889	0.809	0.344	0.833	0.809	0.821	0.458	0.833			
0.931	c0						0.906	c0								
0.765	0.136	0.743	0.765	0.754	0.624	0.889	0.656	0.191	0.618	0.656	0.636	0.458	0.834			
0.799	c1						0.720	c1								
Wgd.Avg.							Wgd.Avg.									
0.830	0.202	0.831	0.830	0.831	0.624	0.889	0.760	0.295	0.764	0.760	0.762	0.458	0.833			
0.886							0.846									
=== Confusion Matrix ===							=== Confusion Matrix ===									
Class	True Positive			False Positive			Class	True Positive			False Positive					
a = c0	(57)			(9)			a = c0	(55)			(13)					
b = c1	False Negative			True Negative			b = c1	False Negative			True Negative					
(8)							(11)							(21)		
=== Evaluation on training and testing set ===																
=== Summary ===							Training Period			Testing Period						
Correctly Classified Instances							83 (83%)			76 (76%)						
Incorrectly Classified Instances							17 (17%)			24 (24%)						
Kappa statistic							0.6239			0.4575						
Mean absolute error							0.2751			0.2987						
Root mean squared error							0.3536			0.3867						
Relative absolute error							61.1 (66%)			67.34 (56%)						
Root relative squared error							74.65 (16%)			82.78 (71%)						
Coverage of cases (0.95 level)							99%			99%						
Mean rel. region size (0.95 level)							91%			90.05%						
Total Number of Instances							100			100						

Output of Naïve Bayes Classifier:

Class	Training Set
C0 _{min}	
C0 _{ma} x	

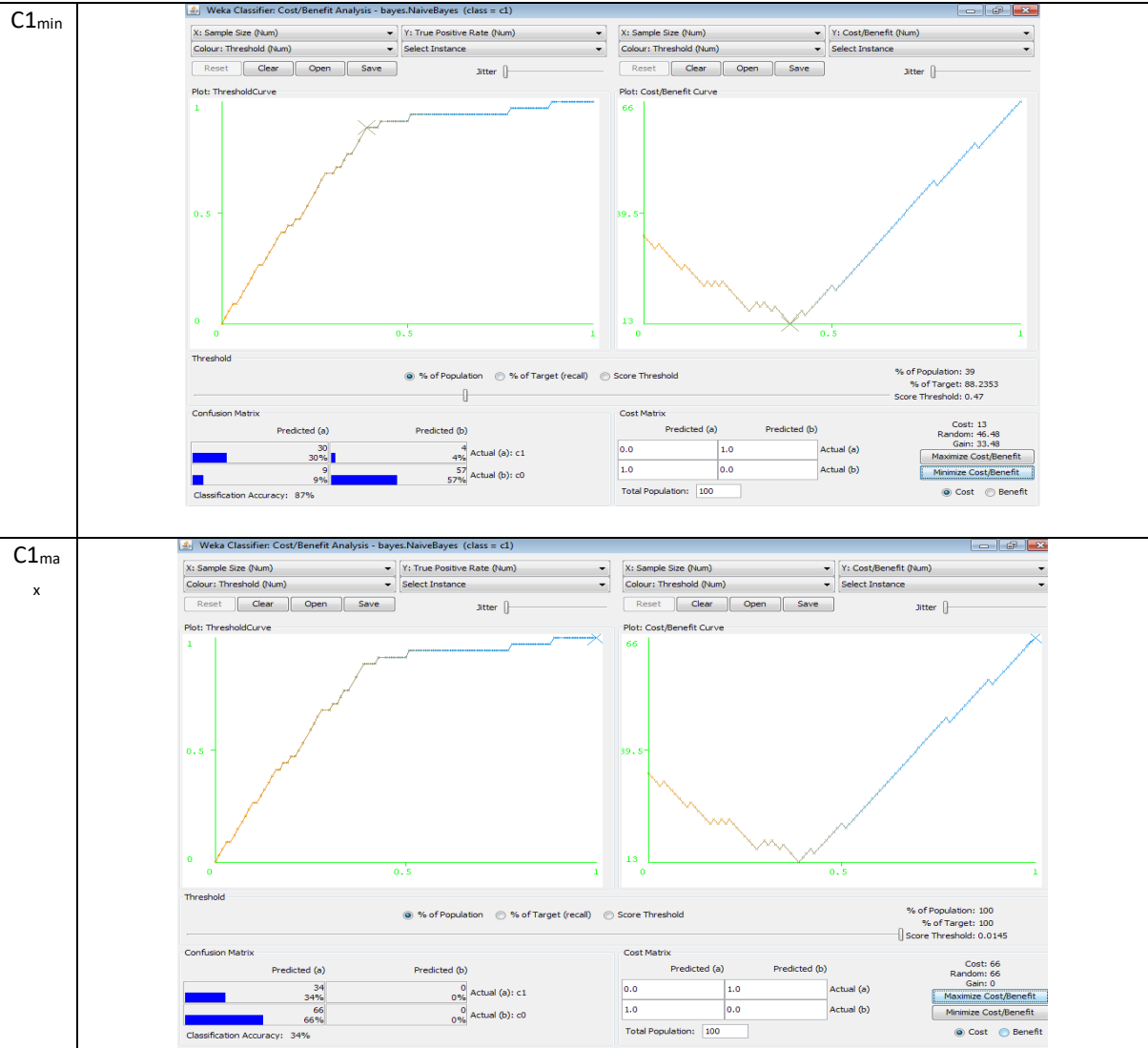
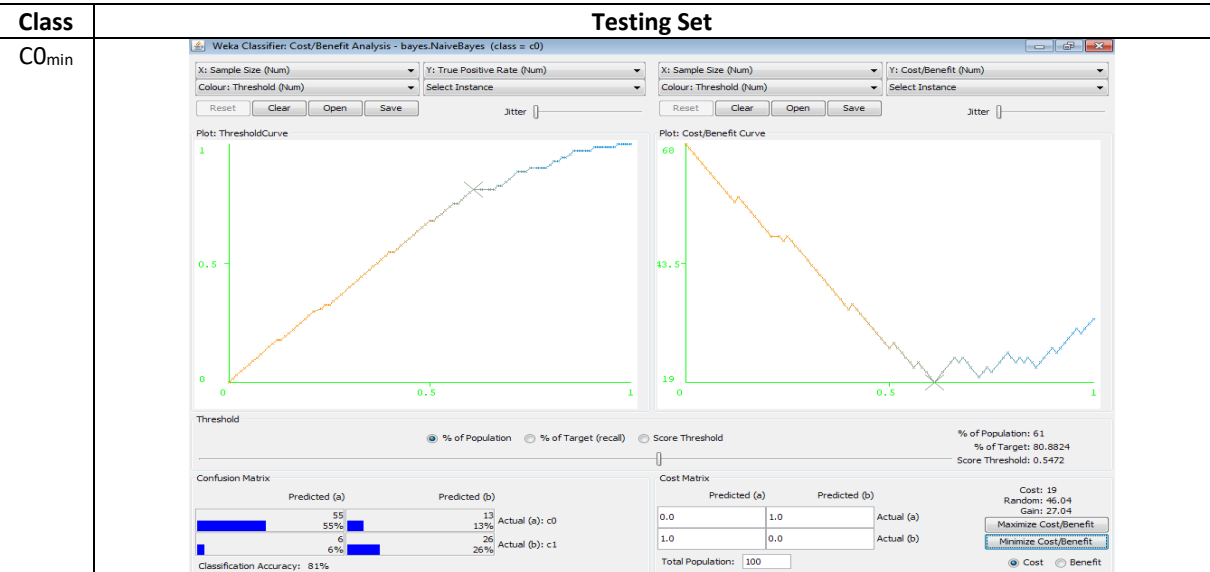


Fig. 1: Cost/Benefit analysis of target class of training set.



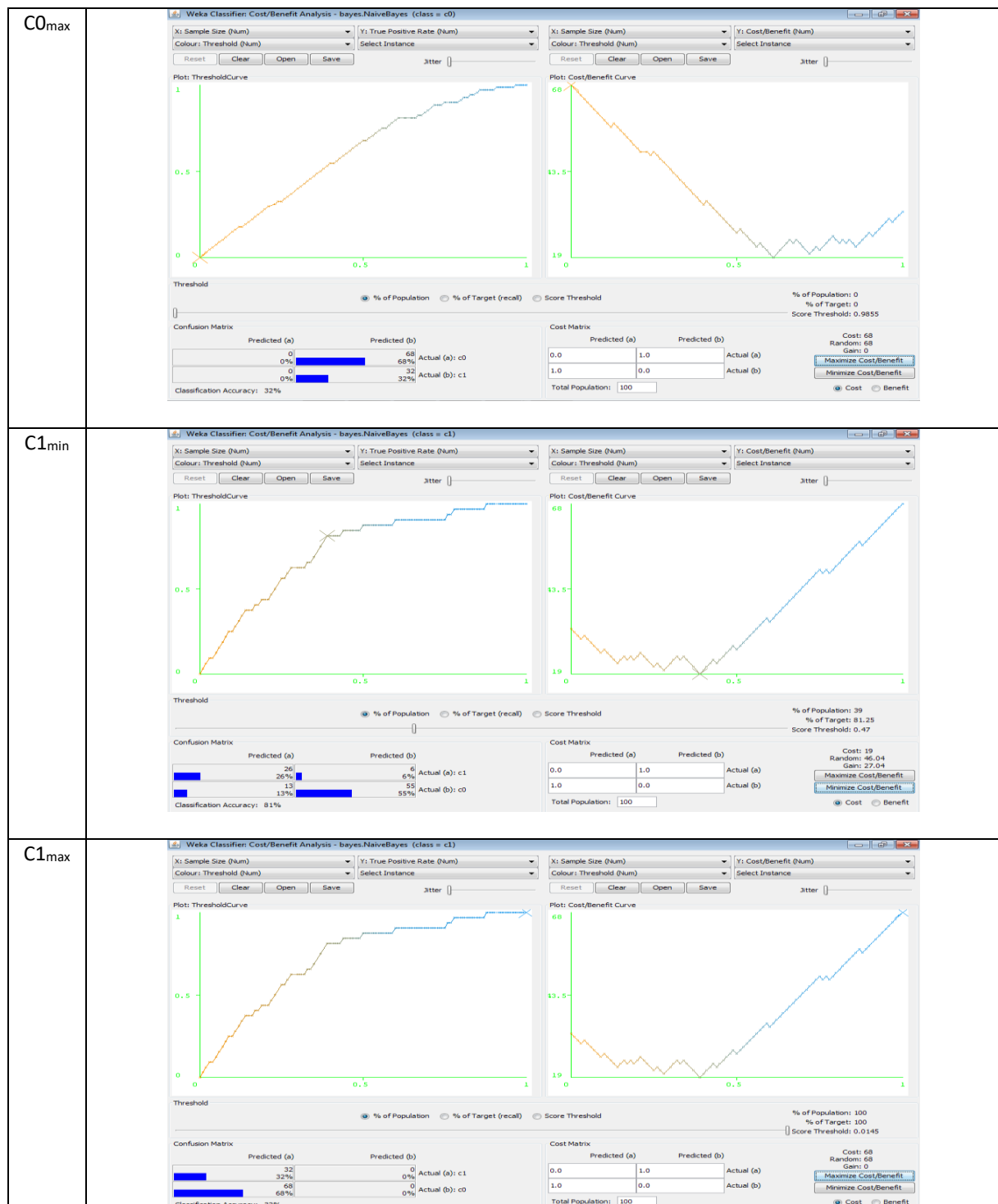


Fig. 2: Cost/Benefit analysis of target class of testing set

Acknowledgement: The authors thank to MPCST, Bhopal for the financial support. The climate data are received from IMD Pune.

References

- [1]. Chai, K.; H. T. Hn, H. L. Chieu; "Bayesian Online Classifiers for Text Classification and Filtering", Proceedings of the 25th annual international ACM SIGIR conference on Research and Development in Information Retrieval, August 2002, pp 97-104.
- [2]. Han, J., Micheline K., "Data Mining Concepts and Techniques, Morgan Kaufman Publishers, 2003.
- [3]. C. C. Aggarwal and P. S. Yu, "A survey of uncertain data algorithms and applications," IEEE Trans. Knowl. Data Eng., vol. 21, no. 5, pp. 609-623, May 2009.
- [4]. Sarah N. Kohail, Alaa M. El-Halees, "Implementation of Data Mining Techniques for Meteorological Data Analysis". July 2011.
- [5]. Olaiya F. and Adeyemo A.B., 2012, "Application of Data Mining Techniques in Weather Prediction and Climate Change Studies".
- [6]. Sangari. R.S and Balamurugan M., "A Survey on Rainfall Prediction Using Data mining" International Journal of Computer Science and Mobile Applications, Vol.2 Issue. 2, 2014, pg. 84-88 ISSN: 2321-8363.