A RESEARCH PAPER ON PREDICTION ANALYSIS: Acceptance of E-Banking

Himanshi verma [1], Rainu Nandal [2], Aarushi [3]
1 Research Scholar, University Institute of Engineering & Technology, Maharshi Dayanand University, Rohtak, India
2 Assistant Professor, University Institute of Engineering & Technology, Maharshi Dayanand University, Rohtak, India
3 Research Scholar, University Institute of Engineering & Technology, Maharshi Dayanand University, Rohtak, India

Abstract - Now-a-days world has become digital. Everything is electronically available to the people. People pay using e-banking service. In that case how can we decide about the people who will use this service in future as well? Two classification techniques of data mining are used to know who will do e-banking transactions in future. The result of prediction will be based on the previous available data (regarding customers like number of transactions done by them, etc.) to the banks. This is called prediction analysis. Comparison analysis between the results of different techniques used has also been done. WEKA data mining tool has been used for the research.

Keywords - WEKA tool, J48 Decision algorithm, naïve bayes algorithm, E-Banking.

1. Introduction

Data mining
Data mining came in view in 1990's. Data mining is method or technique retrieving or extracting the information from the database. It is also called data analyzer source and problem solving.
Data mining techniques are association, classification and prediction, cluster, k-means algorithm.
Data mining support many application like emerging trends education system, market basket analysis, digital library retrieves, customer relationship management, health care, banking fraud detection etc.

E-banking
E-banking means electronic banking. It means financial transaction performed by peoples with the help of electronic device like credit card, debit card or mobile banking etc.

Prediction analysis
Prediction analysis is described as analyzing dataset/ database and finds the important information or result of any problem of present or future on the basis of past information known as prediction.

Classification algorithms used
This section explains the classification algorithms that are used in this paper.

1) **J48 Decision algorithms**
It is a simple decision tree algorithm for classification. This algorithm creates a binary decision tree. This is very useful in solving problem of classification. A model constructed through decision tree. In that case once a tree is create that tree applied on each tuple of database and then provide result. Fig.(a)[33]

```
1: Create a root node N;
2: IF (T belongs to same category C)
       {leaf node = N;
         Mark N as class C;
         Return N;
       }
3: For i=1 to n
       {Calculate Information_gain  (Ai);}
4: ta= testing attribute;
5: N.ta = attribute having highest information_gain;
6: if (N.ta == continuous )
        { find threshold;}
7: For (Each T in splitting of T)
8:       if (T is empty)
            {child of N is a leaf node;}
            else
            {child of N= dtree T)}
10: calculate classification error rate of node N;
11: return N;
```

J48 algorithm Fig.(a)[33]

2) **Naïve bayes algorithm**:-A set of probabilities are calculated using simple naïve bayes classifier. The classifier models are constructed. Various problem instances are classified. This algorithm represents supervised learning method. This algorithm based on Bayesian's theorem and Bayesian formula can be written as:-

P (A/B) = [p(B/A)* P(B)] / P(B)
A- Hypothesis or event
B- Evidence

The basic idea of bayes rule is that the outcome of a hypothesis or an event (A) can be predicted based on some evidence (B) that can be observed from the bayes rule.

WEKA Tool

It stands for "Waikato Environment for knowledge Learning". WEKA support many task of data mining such as classification, clustering, regression, visualization, feature selection, data-pre-processing or their algorithm. WEKA is a tool based on JAVA language. JAVA is an object oriented language and platform independent programming language developed by sun micro systems. Both JAVA and WEKA has GNU (General Public License).Figure (b)

Figure (b) WEKA tool [15]

2. Literature review

Data mining technique is widely used in every sector [17].Data mining have lots of application like banking, health care, Intelligent agencies. So This paper describe the Banking application of data mining. E-banking [3] is assigns the large number of banking condition related to electronic transaction.. On the other hand some authors used techniques of data mining to predict the heart disease [21] dataset, student performance prediction(describe about the student performance that helpful to improve the weak students), Prediction of Diabetes (describe about the patients details of diabetes.)[20] Etc. Some authors worked on fraud transaction (describe about the fraud transaction which help to know about the customer who performed fraud transaction), payment default or predicting payment and some performed classifiers on the bank data set for maximize true positive or negative rate [23].so with the help of reviews or research paper of other author, we consider a problem, and there is a data set of bank. The dataset have records of electronic transaction. In this dataset, factors for decision making are education, total amount , previous record of no. of transaction from bank accounts, that will be accept the electronic transaction in next year. So the problem arises that from large number of records present in data set how can we make prediction.

3. Results

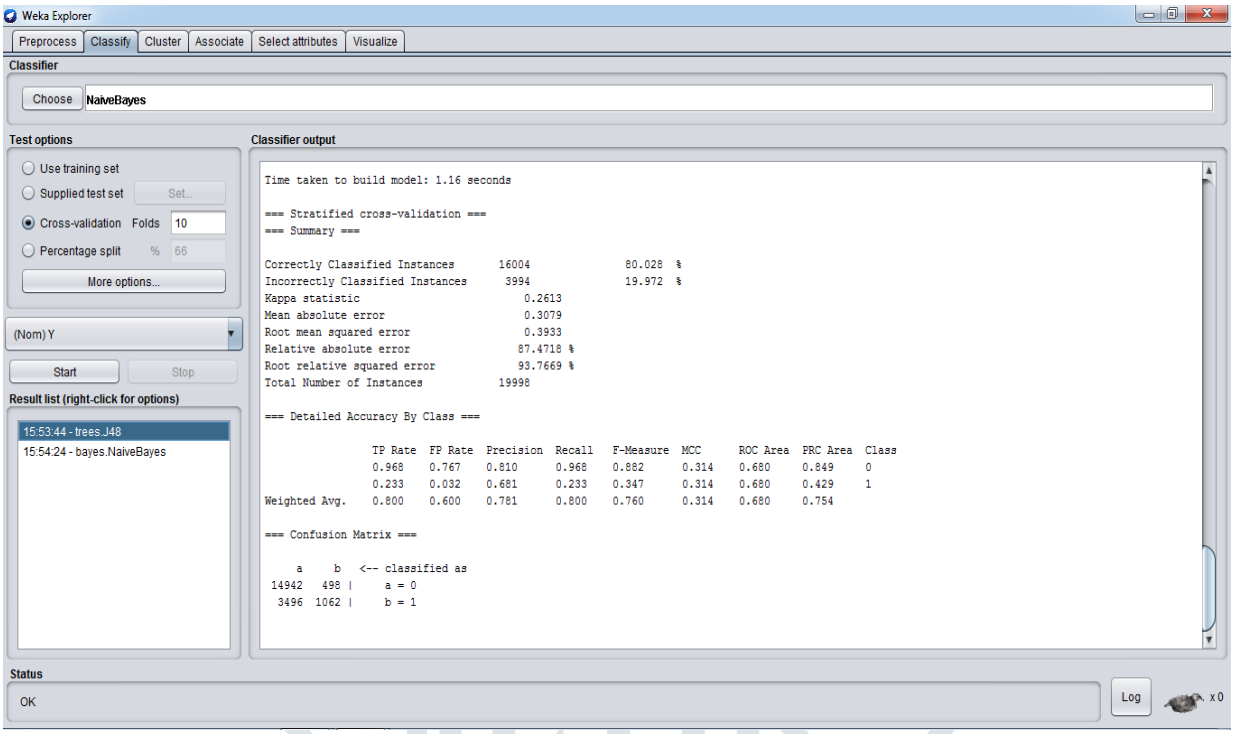
Dataset information

In this study, dataset [32] of banking e-transactions is taken from UCI Machine Learning Repository [31] that has 19998 instances with the 13 attributes that contain records of user who does e-transaction. The decision that next time is made on the basis of, that person will be use digital transaction next year means that person accept the digital transaction.

Attribute 1: Balance (it is the total balance in the account of respective customer)
Attribute 2: Education (describes academic qualification of the customer, 1 - graduate school; 2 - university; 3 - high school; 4 - others.)
Attribute 3: Marriage (describes marital status of the customer, 1 - married; 2 - single; 3 - others.)
Attribute 4: age (describes age of the customer in years)
Attribute 5-11: Usage_1 to Usage_6 (indicates previous record of number of e-banking transactions customer has done in an year, Usage_1 denotes time period from Jan-Feb, Usage_2 denotes time period from March-April and so on (0 - no transaction ; 1 - 0-10 transactions; 2 - 10- 20 transactions)).
Attribute 12: Y target prediction variable (0 - no transaction in future; 1- will do transaction in future.)

Result of prediction analysis using J48 classification decision tree

For finding the result at first cross validation [folds =10] of the dataset is done in WEKA. Then J48 decision algorithm on dataset is applied. In cross validation, the dataset is divided into 10 parts in this particular case of 10 folds. One part is taken as testing data and remaining 9 parts are taken as training data for one time. This is done 10 times and in every case one part is used as testing data (one by one each one is used) and remaining ones are used for training the classifier. Figure(c) show the confusion matrix by J48 algorithm.



Figure(c) confusion matrix by J48 algorithm

Results

- The correctly classified instances as “a” are 14942, shown on the top left of classification matrix.
- The incorrectly classified instances as “a” is 3496 which actually is “b”, shown at bottom left of classification matrix.
- The correctly classified instances as “b” are 1062, shown on the bottom right of classification matrix.
- The incorrectly classified instances as “b” are 498 which actually are “a”, shown at top right of classification matrix.

Result of prediction analysis using Naives bayes classification decision tree

For finding the result at first cross validation [folds =10] of the dataset is done in WEKA. Then Naïve Bayes algorithm on dataset is applied. figure(d)

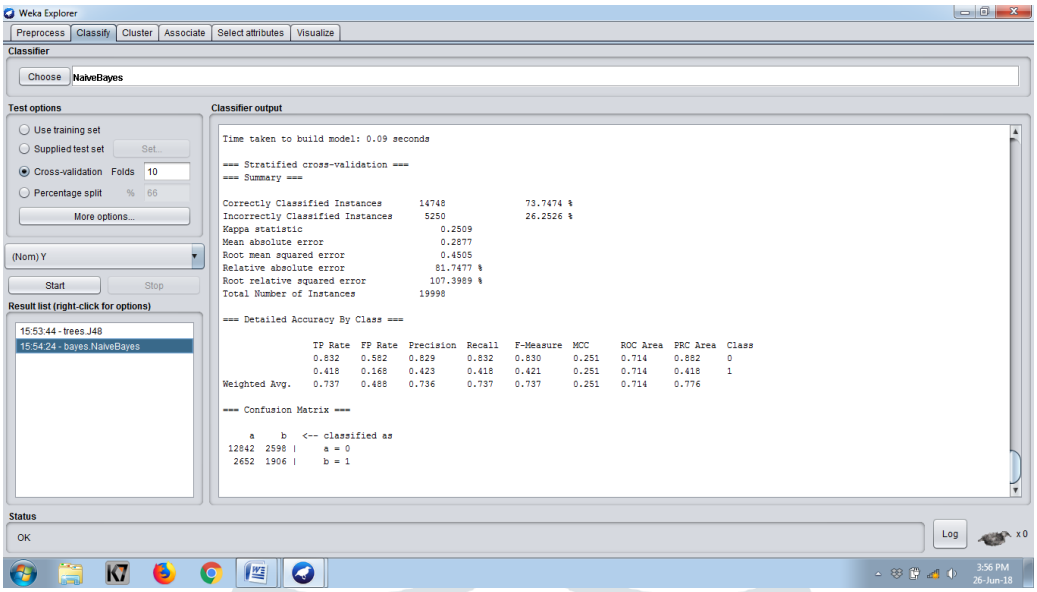


Figure (d) Confusion matrix by Naïve bayes algorithm.

Result

- The correctly classified instances as “a” are 12842, shown on the top left of classification matrix.
- The incorrectly classified instances as “a” are 2498, which actually are “b”, shown at bottom left of classification matrix.
- The correctly classified instances as “b” are 1906, shown on the bottom right of classification matrix.
- The incorrectly classified instances as “b” are 2652, which actually are “a”, shown at top right of classification matrix.

4. Comparison between the J48 Classification Tree and the Naïve bay

Evaluation	J48 Algorithm	Naive Bayes
Time to Build Model(in seconds)	1.16	0.09
Correctly Classified Instances	16004	14748
Incorrectly Classified	3994	5250
Accuracy of correctly classified	80.028%	73.7474%
Accuracy of incorrectly classified	19.972%	26.2526%
Kappa statics	0.2613	0.2509
Mean absolute error	0.3079	0.2887
Relative absolute error	87.4718%	81.7477%
Root relative Squared error	93.7669%	107.3989%

Table1.Comparison between the J48 Classification Tree and the Naïve bayes

On the basis of accuracy of J48 algorithm and the Naive bayes algorithm the output is that the J48 algorithm is better than the Naïve bayes.

5. Conclusion

with the help of various classification data mining techniques ,prediction analysis about who do e-transaction in future can be done by taking into account the previous records(having information about number of transaction a customer has done academic qualification, etc) that bank save regarding their customer. In this paper two algorithms names J48 decision tree algorithm and naïve bayes algorithm are used for this prediction analysis which 76.61515% yielded accuracy on an average. This will help bank in predicting the usage of e-banking services by customer

References:-

1. Tipawan Silwattananusarn, KulthidaTuamsu"Data Mining and Its Applications for Knowledge Management: A Literature Review", International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.2, ISSN: 2230 – 9608, [online]; 2231 -007X [print.], 2012.
2. Kazi Imran Moin, Qazi Baseer Ahmed, "Use of Data Mining in Banking", Journal of Engineering Research and Applications (IJERA), Vol. 2. ISSN: 2248-9622, pp.738-742, 2012.
3. Dileep B. Desai, R.V.Kulkarni, "Application of Data Mining Tools in CRM for Selected Banks", International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 4.ISSN: 0975-9646, pp.199 – 201, , 2013.
4. C. Sunil Kumar, P.N. Santosh, "Data Mining Techniques for Banking Applications". International Journal of Research Studies in Computer Science and Engineering (IJRSCSE), Volume 2. ISSN 2349-4840 (Print) & ISSN 2349-4859 (Online), 2013.
5. K. Chitra, B. Subashini,"Data Mining Techniques and its Applications in Banking Sector", International Journal of Emerging Technology and Advanced Engineering(IJETAE), Vol.3, ISSN 2250-2459, ISO 9001, 2013.
6. Rashid Farooqi, Naiyar Iqbal , "Effectiveness of Data mining in Banking Industry: An empirical study", International Journal of Advanced Research in Computer Science(IJARCS) REVIEW ARTICLE, Vol.8, ISSN No. 0976-5697, 2017.
7. Sangeeta Goele, Nisha Chanana,"Data Mining Trend In Past, Current and Future" International Journal of Computing & Business Research, vol.6. ISSN (Online): 2229-6166, 2012.
8. I-Hsien Ting,"Web Mining application in E-commerce and E-services", Springer-Verlag, 2009.
9. Sreekumar Pulakkazhy, R.V.S. Balan,"Data Mining in Banking and its applications-a review", Journal of Computer Science vol. 9, ISSN: 1549-3636, 2013.
10. Cristobal Romero, and Sebasti´an Ventura,"Educational Data Mining: A Review of the State of the Art", Journal IEEE vol. 40, ISSN: 2231-2803, P-ISSN: 2349-0829, 2010.
11. AndrewKusiak, Bradley Dixonb, Shital Shaha,"Predicting survival time for kidney dialysis patients", International conference on biology medicine ELSVIER , vol. 35, 2004.
12. Andrew Kusiak, Jeffrey A. Kern, Kemp H. Kernstine, and Bill T. L. Tseng, "Autonomous Decision-Making Data mining approach",Journal IEEE transaction of information technology in biomedicine, vol.4, 2000.
13. E.W.T. Ngai, Li Xiu, D.C.K. Chau, "Application of data mining techniques in customer relationship management: A literature review and classification", Journal Expert Systems with Applications ELSREVIER vol. 36.ISSN 2592-2602, 2009.
14. Behrouz Minaei-Bidgoli, Deborah A. Kashy, Gerd Kortemeyer, William F. Punch "Predicting Student Performance: AN Application Of Data Mining Methods with an Educational Web-Based System", IEEE Frontiers in education conference, 2003,.
15. https://www.ibm.com/developerworks/library/os-weka1/index.html.
16. Nathaniel Charlton, John Kingston, Miltos Petridis, Ben Fletcher ,"Using Data Mining to Refine Digital Behavior Change Interventions", International Conference on Digital Health., 2017.
17. Sunita Joshi, Bhuwaneshwari Pandey, Nitin Joshi," comparative analysis of naive bayes and J48 classification algorithm", journal IJARCSSE, vol. 5**,** ISSN: 2277 128X, 2015.
18. Sagar Bhise, Seta Kale," Efficient Algorithms to find Frequent Item set Using Data Mining", International Research Journal Of Engineering and Technology(IRJET) vol. 04**,** e-ISSN: 2395 -0056 ,pp:- 2645- 2648, 2017.

19. Devinder Singh, Iqbal Singh,"Performance Analysis of Data Mining Techniques over Sleeping Problem Dataset", International Journal of Engineering Science and Computing (IJESC) vol. 7. 2017.
20. Gaganjot Kaur, Amit Chhabra,"Improved J48 Classification Algorithm for the Prediction of Diabetes", International Journal of Computer Applications(IJCA) vol. 98. ISSN (0975 – 8887), pp: 813- 817, 2014.
21. Nidhi Bhatla, Kiran Jyoti," An Analysis of Heart Disease Prediction using Different Data Mining Techniques", International Journal of Engineering Research & Technology (IJERT) vol.1, ISSN: 2278-0181, 2012.
22. Mahendra Tiwari, Manu Bhai Jha, OmPrakash Yadav, "performance analysis of Data Mining algorithms in Weka", IOSR Journal of Computer Engineering (IOSRJCE), ISSN: 2278-0661, ISBN: 2278-8727 Volume 6, Issue 3, PP 32-4, 2012.
23. Tina R. Patil, S. S. Sherekar,"Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification", International Journal Of Computer Science And Applications, ISSN: 0974-1011, vol. 6, No.2, 2013.
24. Sushilkumar Kalmegh,"Analysis of WEKA Data Mining Algorithm Rep Tree, Simple Cart and Random Tree for Classification of Indian News", IJISET - International Journal of Innovative Science, Engineering & Technology, vol. 2 Issue 2, , ISSN 2348 – 7968, 2015.
25. Kalpana Rangra, K. L. Bansal,"Comparative Study of Data Mining Tools, Research Paper", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 6, , ISSN: 2277 128X, 2014.
26. Beant Kaur, Williamjeet Singh,"Review on Heart Disease Prediction System using Data Mining Techniques", International Journal on Recent and Innovation Trends in Computing and Communication, ISSN: 2321-8169 Volume: 2, Issue: 10 3003 – 3008.
27. Trilok Chand Sharma, Manoj Jain,"WEKA Approach for Comparative Study of Classification Algorithm", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 4, ISSN (Print) : 2319-5940ISSN (Online) : 2278-1021, 2013.
28. Y. Ramamohan, K. Vasantharao, C. Kalyana Chakravarti, A.S.K.Ratnam," A Study of Data Mining Tools in Knowledge Discovery Process", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-3, 2012.
29. Neeraj Bhargava, Girja Sharma, Ritu Bhargava, Manish Mathuria,"Decision Tree Analysis on J48 Algorithm for Data Mining, Research Paper", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 6, 2013 ISSN: 2277 128X.
30. https://stackoverflow.com/questions/15214179/how-to-read-the-classifier-confusion-matrix-in-weka.
31. https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients#
32. Yeh, I. C., & Lien, C. H. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. Expert Systems with Applications, 36(2), 2473-2480.
33. https://www.google.com/search?q=j48+algorithm&client=firefox-b&source=lnms&tbm=isch&sa=X&ved=0ahUKEwjrzbawzdLbAhVEqo8KHeumC-YQ_AUICygC&biw=1024&bih=654#imgrc=dSSsAViWRKrEGM: