ISSN: 2278-4632
Vol-10 Issue-6 No. 13 June 2020

Linear Regression on Internet Banking Adoption Dataset Using WEKA

Mrs. Nidhi Nigam Verma1, Dr. Deepika Pathak2
1. Research scholar, Department of computer science and application, Dr A.P J Abdul Kalam University, Indore (MP)
2. Associate professor, Department of computer science and application, Dr A.P.J Abdul Kalam University, Indore (MP)

Corresponding Author: nidhinigam2007@gmail.com, +91 9407122448

Abstract: Data Mining or knowledge discovery in the database (KDD) is an excellent process to find out valuable information from a large collection of data. Data mining has successfully been used in different fields such as Medical, marketing, banking, business, weather forecasting, etc. For the banking industry, data mining, its importance, and its techniques are vital because it helps to extract useful information from a large amount of historical data which enable to make useful decisions. Data Mining is very useful for banking sector for better acquiring and targeting new customers and helps to analyze customers and their transaction behaviors. In the recent era, a new technology that has achieved considerable attention, especially among banks, is Internet banking. Its large scope of applications, its advantages brings an immoderate change in a common human's life. Linear Regression is one of the most commonly used and applied data mining techniques. Linear regression is really a very fast and simple regression algorithm and can give the best performance if the output variable of your data is a linear grouping of your inputs. In this paper the linear regression is applied on internet banking adoption dataset in order to compute the weights or coefficients of linear expression, and provides the predicted class value. The analysis here is done with the help of WEKA tool for data mining.

Keywords: Linear Regression, Data mining, WEKA, Internet Banking Adoption

I. INTRODUCTION

Data Mining is one of the most commonly used techniques nowadays to retrieve useful information from large datasets. The techniques which are used for data mining includes Association rule mining, Classification algorithm, Clustering, Sequential Patterns, Linear Regression and Decision Trees. Classification requires a nominal class value but regression requires a numeric class value. Linear regression is basically a typical statistical method, which computes the weights or coefficients of linear expression, and provides the predicted class value, which is basically the sum of each and every attribute values multiplied by their weight. Linear regression works by estimating line coefficients or hyperplane which best fits the loaded training data. Linear regression is really a very fast and simple regression algorithm and can give the best performance if the output variable of your data is a linear grouping of your inputs. Internet banking has now days become a vital part of the banking system in India. Major benefit of internet banking, a customer can complete multiple banking transactions in the comfort of his home. The advantage of adopting internet banking provides competitive advantages to banks over other competitors. One of the most popular techniques of data mining used by banking industry is Linear Regression. The main aim of this paper is to apply Regression algorithm on internet banking adoption dataset with the help of WEKA tool for data mining.

II. METHODOLOGY

Banking customers having accounts in different bank branches located in various cities of Madhya Pradesh was the target population for this research study. In this study specifically survey method is used where the data is collected by means of a questionnaire to determine the opinion of a target population of Madhya Pradesh [22].

ISSN: 2278-4632
Vol-10 Issue-6 No. 13 June 2020

A. Primary Data for the Research

Primary data for this has been collected using a self structured questionnaire designed purposely for this study. Appropriate secondary sources have as well been relied upon for designing a suitable comprehensive questionnaire to gain deeper insights in this field. Questionnaire also includes questions regarding the satisfaction level of the customers using internet banking on identified factors. For this study data is collected using Google forms and through emailing of the questionnaires to users. Data for the research has been collected from 502 customers which includes both users and non-users of Internet banking. The collected data should to be analyzed by using the appropriate analytical tool or technique in order to understand the various factors and reasons behind Internet banking adoption. For this research study, WEKA tool, a data mining tool, is being used.

B. WEKA tool for Data Mining

"WEKA" stands for Waikato Environment for Knowledge Analysis. Basically WEKA is named subsequent to a flightless bird of New Zealand. It is a set of various machine learning algorithms that can easily be applied on any data set directly or can be called from your Java code. WEKA basically contains various tools for data mining and data pre-processing which are clustering, classification, association rules, regression and visualization [23].It is freely available and an open source software for data mining and its applications under GNU general public license, which is developed by the university of Waikato in New Zealand.

C. Sampling Procedure

A sampling, defines the population from which our research sample is drawn. As there is hardly enough money or time to collect information from everyone in overall population, the goal becomes choosing a representative sample, sometimes called as a subset of that population. Area for this research is Madhya Pradesh, India. Madhya Pradesh is situated at the center part of India, so known as the heart of India. For this research sample which is considered are from major cities of Madhya Pradesh. Research Questionnaire was sent to them online, their opinions and concerns are collected in order to measure the adoption rate of internet banking in Madhya Pradesh.

III. EXPERIMENTAL WORK

Primary Data Collection

For this study data is collected using Google forms and through emailing of the questionnaires to users. Only 502 responses were totally complete and found eligible for our analysis. So for this study, a sample of 502 responses was considered.

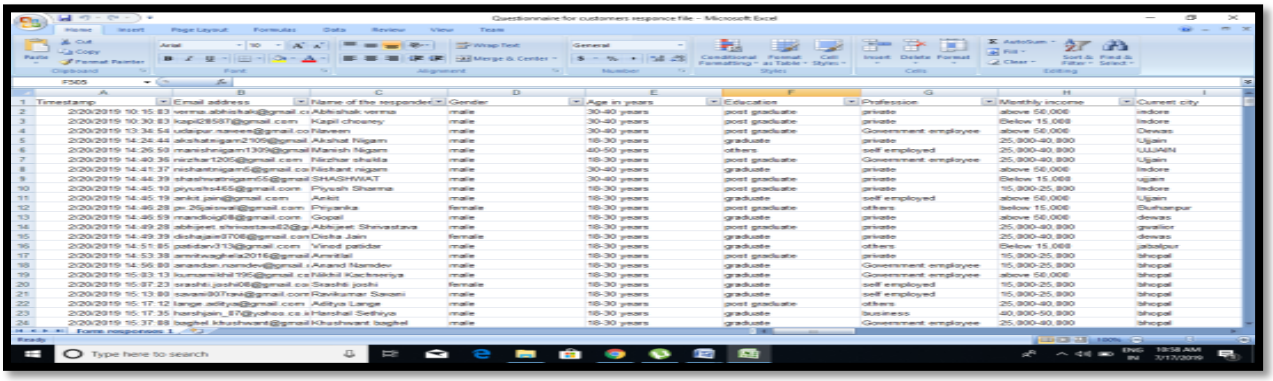

Fig 1: questionnaire for customer's response file containing 502 responses

ISSN: 2278-4632
Vol-10 Issue-6 No. 13 June 2020

Analysis of dataset using WEKA

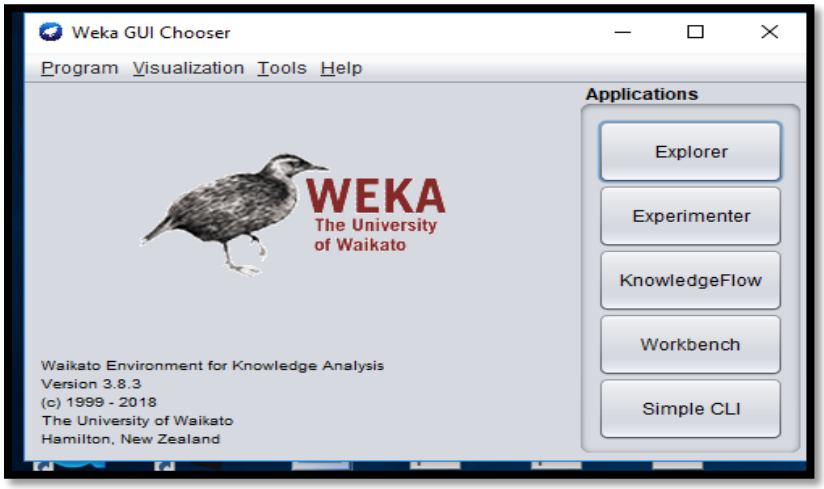

Fig2: Weka first screen while startup

c. Loading Dataset into WEKA Explorer

For this study we are having a dataset on internet banking adoption, which we have collected online through a well structured questionnaire using Google forms. We have received 502 responses which we are going to consider here for the purpose of analysis using WEKA [24]. Here we import our data file which is CSV format and loading it is converted in ARFF file format with the help of save button in WEKA explorer.

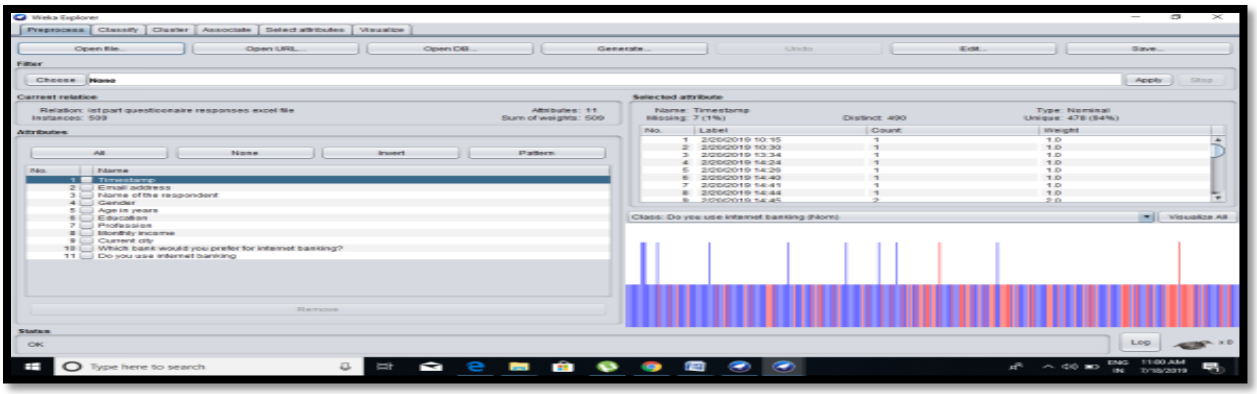

Fig3: Internet Banking Dataset in WEKA Explorer

d. Linear Regression on internet banking dataset

For applying linear regression on our data set it is necessary to first convert our data from nominal to binary form since regression will not work on nominal data set. So we can easily do that by choosing filtering in the preprocess tab and then click on nominal to binary in explorer window.

ISSN: 2278-4632
Vol-10 Issue-6 No. 13 June 2020

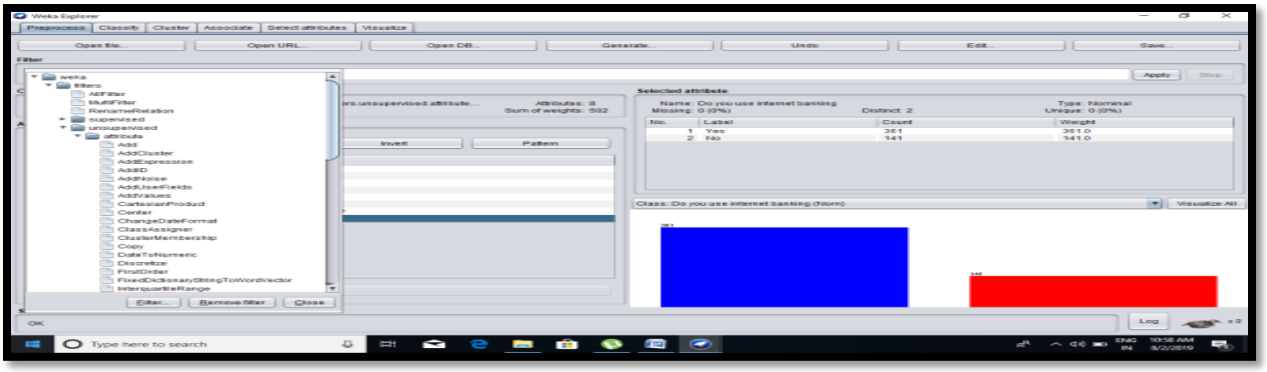

Fig 4: Preprocess to convert data set from nominal to binary using filter

e. Result of Linear regression for internet banking adoption data set

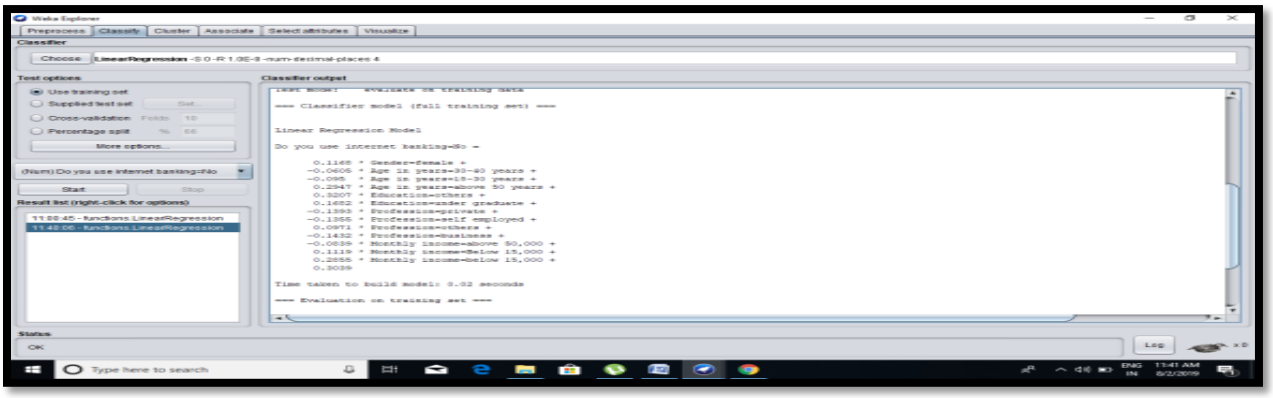

Fig 5: Result of Linear regression for internet banking adoption data set

For a given data set or set of attributes, the linear regression prepares a model, in the form of an equation that is used to calculate the predicted class value. Data mining is not just about providing a single number as an output; it's basically to identify rules and patterns. It's not only used to output an absolute number, but fairly to create a model that makes you discover rules and patterns, output prediction, and brings you to conclusions backed by the data.

Analysis by Linear regression is basically used to predict the value of one variable which is based on the value of some another variable in a dataset. The value of a variable which you want to predict is known as the dependent variable and the variable which is used to predict the value of other variables of the dataset is referred to as the independent variable.

=== Run information ===
Scheme: weka.classifiers.functions.LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4
Relation: ist part questioonaire responses excel file-weka.filters.unsupervised.attribute.Remove-R1-3,9-10-weka.filters.unsupervised.attribute.NominalToBinary-Rfirst-last
Instances: 502 Attributes: 21

ISSN: 2278-4632
Vol-10 Issue-6 No. 13 June 2020

Gender=female
Age in years=30-40 years
Age in years=18-30 years
Age in years=40-50 years
Age in years=above 50 years
Education=post graduate
Education=graduate
Education=others
Education=under graduate
Profession=private
Profession=Government employee
Profession=self employed
Profession=others
Profession=business
Monthly income=above 50,000
Monthly income=Below 15,000
Monthly income=25,000-40,000
Monthly income=15,000-25,000
Monthly income=below 15,000
Monthly income=40,000-50,000
Do you use internet banking=No

Test mode: evaluate on training data

=== Classifier model (full training set) ===

Linear Regression Model

Do you use internet banking=No =

0.1168 * Gender=female +
-0.0605 * Age in years=30-40 years +
-0.095 * Age in years=18-30 years +
0.2947 * Age in years=above 50 years +
0.3207 * Education=others +
0.1682 * Education=under graduate +
-0.1393 * Profession=private +
-0.1355 * Profession=self employed +
0.0971 * Profession=others +
-0.1432 * Profession=business +
-0.0839 * Monthly income=above 50,000 +
0.1119 * Monthly income=Below 15,000 +
0.2855 * Monthly income=below 15,000 +
0.3039

Time taken to build model: 0.02 seconds

Table 1: Linear Regression Model

The output of Linear regression model is regression line which is in the form of y= m*x + b, where y is the dependent variable which is dependent on the value of variable x, m is a slope or gradient which shows how steep the line is and b is a constant. One equation for each output variable. All the equations are composed of the addition of a series, which is composed of an attribute, multiplies a factor that can be either positive or negative. The regression process basically determines the factors, which are based on the data set. The linear regression model for our dataset represents 'Do you use internet banking=no' is the dependent variable, whose values depend on the

ISSN: 2278-4632
Vol-10 Issue-6 No. 13 June 2020

values of various attributes in the dataset like gender, age in years, education, profession, and monthly income. Here 0.3039 is a constant value and time taken to build the model is 0.02 seconds.

III. CONCLUSION

Here in this paper we discussed the results of applying Linear Regression algorithm for data mining on internet banking adoption dataset. Through this paper we have shown how Linear Regression model is useful for predication and can helps to sort out problems regarding internet banking adoption in Madhya Pradesh. Regression is a most popular technique of data mining which is used to predict a range of numeric values called continuous values as well, for given a specific dataset. The output of Linear regression model is regression line which is in the form of y= m*x + b, which clearly shows the relationship between the dependent and independent variables in a dataset. These results will be useful for bank managers and other banking authorities, in order to enhance the internet banking adoption rate in the future.

REFERENCES

[1] Dr. Md. R. Farooqi and N. Iqbal, "Effectiveness of Data mining in Banking Industry: An empirical study", international Journal of Advanced Research in Computer Science, Volume 8, No. 5, MayJune 2017,pp. 827-830.

[2]S.Pulakkazh and R.V.S. Balan," Data Mining In Banking and its applications- A review", Journal of Computer Science, Vol. 9 Issue 10, 2013, pp. 252-1259.

[3]M.Dash and S.K. Nayak," Data Mining Need In Profiling Internet Banking Customer" International Journal of Computer Science and Management Research, 2012

[4]K. Rangra and Dr. K. L. Bansal," Comparative Study of Data Mining Tools ",International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 6, June 2014,pp. 216-223.

[5]S. Raj,"A Review Of Data Mining Applications In Banking ",International Journal of Computer Science and Management Research, June 2015.

[6]V. Jayasree and R.V. Siva Balan, "A Review on Data Mining in Banking Sector", American Journal of Applied Sciences, Vol. 10 Issue 10, 2013, pp. 1160-1165,

[7]Dr.C. Rajabhushanam and S. cumaar, "An Study Of Data Mining Applications In Banking", International Journal of Pure and Applied Mathematics, Volume 116 No. 15, 2017, pp. 265-271.

[8]Mr. N.K. Dokania and Ms. N.Kaur," Comparative Study Of Various Techniques In Data Mining", International Journal Of Engineering Sciences & Research Technology, Vol. 7 Issue 5, May 2018, pp. 202-209.

[9]M.C.S.Geetha, Dr. I. Elizabeth Shanth, and S. Seetha Raman, "A Survey and Analysis on Regression Data Mining Techniques in Agriculture", International Journal of Pure and Applied Mathematics, Vol. 118, Issue.8, Pp. 341-347, 2018

[10] K.I. Moin, Dr. Q.B. Ahmed, "Use of Data Mining in Banking", International Journal of Engineering Research and Applications (IJERA) ,Vol. 2, Issue 2,Mar-Apr 2012, pp.738-742.

[11]V. Bhambri, "Application of Data Mining in Banking Sector", International journal of computer science and technology, Vol. 2, Issue 2, June 2011, pp. 199-202.

[12] N.singh , R. shringar raw, and R.k. chauhan," Data mining with regression technique ",Journal of Information Systems and Communication, Vol.3 Issue.1, 2012 ,pp. 199-202

[13] F. Halili and A Rustemi," Predictive Modeling: Data Mining Regression Technique Applied in a Prototype", International Journal of Computer Science and Mobile Computing, Vol. 5, Issue. 8, August 2016, pp. 207– 215.

[14] S. Jindal," Study of E-Banking Scenario in India", International Journal of Science and Research (IJSR), Volume.5 Issue.12, December 2016, pp. 680-683.

[15] S.N. Hojjati,A. Monavvarian and F.S. Roozbahani," Evaluation of Factors Influencing Internet Banking Acceptance by Internet Users in Iran", International Journal of Advanced Networking and Applications, Volume.6 Issue.6 , 2015, pp. 25032519.

[16] G.A. Weshah," The role of internet banking in continuous improvement areas: Quantitative evidence from Jordanian banks" International Journal of Business Performance Management, January 2013, Vol.14, No.2, pp.181- 196.

[17]R. Farooqi," Impact of Internet Banking Service Quality On Customer Satisfaction", Journal of Internet Banking and Com, April 2017, vol. 22 Issue No. 1, pp. 1-17.

[18] P. Singh and D. Kumar," Impact of Internet Banking on Customer", International Journal of Research (IJR), Vol. 1 issue.4, May 2014, pp. 394-412.

[19] S. Haq and B.M. Khan," E-Banking Challenges And Opportunities In The Indian Banking Sector", Innovative Journal Of Business And Management, vol. 2 issue.4,August 2013, pp. 56 – 59.

[20] K. Nidhi," E-Banking In India: Challenges And Opportunities "International journal of Science, technology and management, Vol. No. 05, Issue no. 08, August 2016, pp. 809-815.

[21] S. Devulapalli and S. k. Orugant, "Challenges and Opportunities of E- Banking in India", OSR Journal of Business and Management (IOSR,JBM) ,pp 56-61.

[22]Dr. G. Sharma and dr. S. malviya," Exploring Usage Pattern of Internet Banking Services Provided By Banks in Indore District of Madhya Pradesh", international journal of advance research in computer science and management studies, Volume 2, Issue 11, November 2014.

[23]S. Palaniappan, A. Mustapha, Md. C.F. Foozy, and R. Atan," Customer Profiling using ClassificationApproachfor Bank Telemarketing", Vol. 4 Issue 2, July 2016

[24]K. Kaur1 and S. Dhiman, "Review of Data Mining with Weka Tool", Vol. 4 Issue. 8, August 2016

[25]Dr. Md. R. Farooqi and N. Iqbal, "Effectiveness of Data mining in Banking Industry: An empirical study", international Journal of Advanced Research in Computer Science, Volume 8, No. 5, MayJune 2017,pp. 827-830.

[26]Dr.C. Rajabhushanam and S. cumaar, "An Study Of Data Mining Applications In Banking", International Journal of Pure and Applied Mathematics, Volume 116 No. 15, 2017, pp. 265-271.

[27]Mr. N.K. Dokania and Ms. N.Kaur," Comparative Study Of Various Techniques In Data Mining", International Journal Of Engineering Sciences & Research Technology, Vol. 7 Issue 5, May 2018, pp. 202-209.

Nidhi Nigam Verma
She is a Research scholar, a PhD student in the field of Computer science and application, Dr. APJ. Abdul kalam University Indore (M.P), having 8 years of experience in teaching, worked as a assistant professor, computer department in Indore till 2019.

Dr. Deepika Pathak
Department of computer science and application, Dr. APJ. Abdul Kalam University, Indore (M.P), Pro vice chancellor in
Dr. APJ Abdul Kalam University, Indore(M.P).