Computer Science
E-ISSN No : 2581-6195 | Volume : 2 | Issue : 1 | Jan-Feb 2019

ENHANCED CREDIT WORTHINESS OF BANK CUSTOMER IN NIGERIA USING MACHINE LEARNING AND DIGITAL NERVOUS SYSTEM

Amanze, B.C.[1], Asogwa, D.C[2], Chukwuneke C.I.[3]

ABSTRACT

This paper presents an enhanced model of machine learning and digital nervous system based on credit worthiness of bank customer in Nigeria. The machine learning have the capability to determine the relevant features and customer's credit worthiness in efficient manner. The Digital Nervous System (DNS) approach will enable the Credit Risk Management System (CRMS) to capture credit risk information and provide it where it is needed for decision making and when it is needed. The paper aimed at developing an enhanced model for detecting a credit worthiness in a bank. In recent years, customer's credit worthiness is becoming more crucial for financial organizations. To deal with major problems like noise, data incompleteness and lack inconsistency in loan while building predictive models, this paper proposes a predictive model for detection of credit worthiness. For developing and testing of model a large, real and most recent dataset of credit card, obtained from UCL repository, is proposed. The paper makes it possible for the really availability of more comprehensive credit risk management information and to minimizes decision-making errors with respect to loans. The key focus of the work is on detection of credit worthiness which is defined as the probability of default on the loan or credit from financial organizations like banks. The efficiency of model is demonstrated using confusion matrix on the basis of prediction accuracy and other metrics, against benchmark classifies.

KEYWORDS: Machine Learning, DNS, Predictive Modeling, Credit Risk, Credit Worthiness

[1]Department of Computer Science, Faculty of Physical Sciences, Imo State University, Owerri, Nigeria.
[2,3]Department of Computer Science, Faculty of Physical Sciences, Nnamdi Azikwe University, Awka, Nigeria.

HOW TO CITE THIS ARTICLE:
Amanze, B.C., Asogwa, D.C., & Chukwuneke C.I. (2019). Enhanced Credit Worthiness Of Bank Customer In Nigeria Using Machine Learning And Digital Nervous System., International Educational Journal of Science and Engineering (IEJSE), Vol: 2, Issue: 1, 01-07

Research Paper

1. INTRODUCTION

In the banking sector, lending which may be on short term, medium or long term basis is one of the services that commercial banks usually render to their customers. In other words, banks do grant loans; overdrafts and advances to individuals, business organizations as well as government in order to enable them embark on developmental activities as a means of aiding their growth in particular or contributing towards the economic development of a country in general. The customer may be in need of the fund for the various purposes which may spread through new capital venture loan, farming, contract jobs, and business expansion among others. Credit Worthiness (CW) is a valuation performed by lenders that determines the possibility a borrower may default on his debt obligations [1]. It is represented as a credit score by Financial Organizations. The creditworthiness of a company or individual is determined by using credit rating systems. A high credit score grants high CW. Payment history or credit history, health status and credit score depicts how a person meets debt obligations, which establishes credit worthiness or the financial character of a person. Payment history counts generally counts for 35% of a person's credit score [1]. Lending institutions also consider the amount of available assets and the amount of liabilities to determine the probability of a customer's default.

In addition, it sees other factors such as age, health status, income, employment status, financial obligations, debt owed, accounts, length of payment history and the capability to repay debt. Banks also determines the interest rate, loan and other fees and fines, terms and conditions of a credit or loan on the basis of score. In case of unavailability of history of defaults some banks models for predicting credit risk with Moody's KMV [2].

In fact, banks have to take suitable actions to lessen credit risks to decrease costs as much as possible. The need of cloud platforms for analysis and prediction from credit datasets can be seen as the shortcomings of present methods of credit scoring that runs over a single server and faces various problems including Non availability of customer's loan information with other banks electronically as a result of non-integration of bank database server across banks, inability to electronically authenticate the collaterals presented by the customer leading to time delay in document verification, the presence of sentiments in decision making by the banker which normally leads to bad and doubtful debts and customer's behaviors in terms of previous loan repayment are not monitored electronically and therefore does not form part of the decision process in the existing system. As commercial bank embarks on this loan venture, they are interested in giving out loans and advances to their numerous customers bearing in mind, the three principles guiding their operations which are profitability, liquidity and solvency.

However, commercial banks' decision to lend out loans to customers are influenced by a lot of determinants such as the prevailing interest rate, the volume of deposits, the level of their domestic and foreign investments, banks' liquidity ratio, prestige and recognition.

Looking at the economic situations in Nigeria today, the risk of loan default is particularly high because of many factors. The actual financial position of some customers or enterprises could not be considered creditworthy.

However, working in market environment is unthinkable and not effective without raising funds to cover the needs for financial resources. Many enterprises regardless of their ownership form are based on limited capital which does not allow them to sell their work on the basis of their capital, which in its turn creates the need to overtake credit resources and use them. In the process of conducting a credit strategy at the level of a specific borrower it is important for a bank to determine its credibility. The need to justify the economic methods of managing credit operations focused on the economic boundaries of the credit comes to the fore. It will allow avoiding credit exposures, which are not effective from the point of view of money circulation, to ensure the timely and full repayment of loans, which is essential to improve the utilization of material and financial resources.

Virtually all lending decisions are made under creditors on uncertainties. The risk and uncertainty associated with lending decision requires that the concepts of risk and risk analysis need to be employed by lending bankers in order to determine the credit worthiness of the customer. This statement implies that if credit worthiness of the customer are to be objectively assessed, lending decisions by the money-deposit banks should be based less on quantitative data (the volume of cash the customer is requesting for) but more on principles that can provide sound unbiased judgment. Furthermore, the banks depend heavily on historic information as a basis for decision making.

The increasing trend of provision for bad and doubtful debts in most money-deposit banks is a major source of concern to bank management. Bad debts in banks destroy part of the earning assets of the banks such as loans and advances which have been described as the main source of earning and also determines the liquidity and solvency which generates two major problem; to earn sufficient income to meet its operating costs and to have adequate return on its invesstments. Currently, credit approval for customers is evaluated on certain areas. There are five C's involved in credit evaluation; character, credit report, capacity, cash flow, and collateral [3].

Credit scoring can be defined as a technique that helps credit providers decide whether to grant credit to consumers or customers. The last two decades have seen a rapid growth in both the availability and the use of consumer credit. Until recently, the decision to grant credit was based on human judgment to assess the risk of default. The growth in the demand for credit, however, has led to a rise in the use of more formal and objective methods (generally known as credit scoring) to help credit providers decide whether to grant credit to an applicant [4]. Credit scoring also helps to increase the speed and consistency of the loan application process and allows the automation of the lending process. The character of a person applying for a credit is a big factor to the decision for credit approval. A person with a sound financial objective is likely to be granted a credit approval quickly and are possibly than an individual who is in bad shape, not just on the financial facet, but also on other aspect. Credit history is another important factor considered by lenders in their decision to grant and approve credit applications. The credit report is a record of an individual's past borrowing and reimbursing transactions. It also includes information about late payments and bankruptcy.

Also, collateral is a common term in credit. A lender seeks for security whenever the borrower defaults the credit payment. If no collateral is present as security for a credit, it is likely that the lender will give the borrower a high-interest rate credit. The need for sufficient and reliable information is the foundation of a successful credit worthiness decision. A credit manager may call on references, run background checks, pull a credit report, verify bank accounts or ask questions of the applicant to validate the information on the credit application. Credit managers are challenged with the task of obtaining readily available information to support their decision while sending a timely response to the applicant.

Machine Learning is the process of learning a set of rules from instances or more generally speaking creating a classifier that can be used to generalize from new instances [5]. Creation of a classifier is a two-step process. In the first step the classifier model is constructed using a given data set. This step is called training. In this step classification rules are constructed. The second step called testing determines the accuracy of the classification rules. If the accuracy of the classifier is above an acceptable limit then the classifier model constructed in the first step may be used for classification of new data records. Training the classifier requires the use of machine learning algorithms (MLAs). Choice of algorithm is critical because the accuracy of prediction is often dependent on these algorithms. Over the years several researchers have used varied MLAs to perform classification in various knowledge domains.

In this paper, a hybrid security model is used to provide an enhanced model for customer's credit worthiness in banks. Machine learning and DNS are used to electronically confirm the credit worthiness of the customer seeking for loan from the bank. This will enhance the security of the loan granted and thereby reduce the risk of bad and doubtful debts in the banking sector. The paper therefore proposed a more efficient CRMS for banks using DNS approach on web-based platform. DNS is not a program nor a hardware product, but a combination of IT infrastructures, different software applications, Internet technology and the web concept, which enables the efficient exchange of information on an organizational network.

Technological development and business development are inseparable, as technology always answers the needs of businesses and businesses adapt to every technological development, such as the Digital Nervous System (DNS). Modeled on the human nervous system, which coordinates each separate system of the human body, the Digital Nervous System coordinates all the internal and external processes of an organization to easily and swiftly obtain information. DNS also revolutionized the traditional information flow and knowledge sharing between the employees and the company. Through the use of intranets (internal network) or internet, employees can access and easily share information on their own, anytime they need it. Personal computers entered the corporate world and lessened the need for manual paper-based procedures.

During the 1970s, large corporations even used Electronic Data Interchange (EDI) as a different approach to send information on networks. Unfortunately, it proved too expensive for small companies and not flexible enough for big companies. Furthermore, it cannot efficiently handle some delicate business activities and transactions. Intranets and extranets have become the modern backbones of effective modern business management and of many e-commerce activities. In the "Digital Era", they constitute the "Digital Nervous System". Microsoft is one of the first companies that adopted DNS and greatly benefited from it. A model of the proposed system is presented in Figure 1 below.

2. RELATED WORK

Available evidence has, however, showed that most of the liquidated banks' officers flouted these provisions with impunity and some still in operation are allegedly not obeying these provisions [6]. Loans were granted without collateral; when taken, not adequate and when adequate, not perfected. [7] Discussed the values Biometrics, Digital Nervous System, Global Positioning System, Artificial Neural Networks intelligent agent and data mining in mitigating the problems of loan fraud. R. Peter discussed the GPS-based mobile GIS equipment used by Ghanaian surveyors to perform a cadastral survey of the property boundaries as described by the occupant, as well as neighbors [8]. They utilized the latest geospatial technologies to create a land titling process and GIS-based land records system. Credit Assessment Software was developed by [9] to help bank assess credit risk. The assessment of credit risk by [9] was viewed from the organization perspective.

Evaluation and prediction of customers credit worthiness is a key for preventing losses for the banking sector. The survey is presented in Table 1

Authors	Work/Methods undertaken	Achievements	Dataset used	Tool used
[10]	Applied 15 different ML methods. Proposed a model based on Linear Regression.	Accuracy between 76 – 80% in all except in 2 methods. Features selected = 5	Taiwan Bank Credit Card dataset [11]	Scikit – Learn [11] and MAT LAB
[12]	Performance comparison of ensemble of classifiers for credit scoring and bankruptcy prediction is done.	Random Subspace (RS) ensemble method. With Neural Net classifier, performed better than other Ems.	Australian credit, German credit and Japanese credit dataset.	Different suitable toolkits
[13]	03 Ems namely Bagging, AdaBoost and Random Forest combined with three ML algorithm, to asses credit risk. Feature selection is applied to derive the important attributes.	Assessment on performance of the ensemble classifiers	German credit dataset [14]	Weka [15,16]
[17]	3 popular Ems, i.e., Boosting, Bagging, and Stacking, based on 4 basic learners: Logistic Regression, Decision Tree (DT), ANN and SVM, are compared.	Performance evaluation shown that ensemble methods improve base learners.	German credit and Australian credit dataset [14]	Weka toolkit [15,16]
[18]	Proposed a Model based Bagging EM with REP Tree, for Credit Risk prediction.	Accuracy of more than 81% with minimum number of features selected, i.e. 3.	Taiwan Bank Credit Card dataset [11]	Weka toolkit [15,16]
[19]	Experimented with batch & incremental classifiers. Pre-processing and feature extraction using ID3 are applied to reduce the dataset and to get highest gain ratio.	Author's concluded that sample and partition size of training and testing has an effect on accuracy of prediction.	Malaysian Bank Dataset.	Weka [15,16].

Table1: Survey of related works

3. PROPOSED WORK

Machine learning techniques will help to distinguish borrowers who repay loans promptly from those who do not. It also helps to predict when the borrower is at default, and in determining the credit worthiness of borrower by analyzing the behavior and reliability of the customers. With data mining techniques, banks can do a thorough profiling and ranking of their branches with respect to loan fraud risk. Central Bank of Nigeria (CBN) in the same manner can profile and rank commercial banks. To accomplish this, relevant information can be gathered from the credit risk information service databases. These files contain all the essential information pertaining to a loan. That includes characteristics such as identity of loaner and borrower, location of the branch/bank where the loan was issued and changes that were made to the loan. This data is the cornerstone from which the search for any irregularities in the loan process begins. These are specific sets of instructions the bank personnel must comply with. An example of one such rule is whether a loan has been issued without consultation with the CBN credit guidelines. We need to ascertain if clients have loans at other banks before bank A can confidently issue one. Another rule serves to determine whether the pay back account really belongs to the credit owner.

However, the application will go much further than just data mining. Fraud rule results are converted into risk scores and then displayed by the systems reporting application. The reporting application gathers all the information from the rules and transforms these absolute numbers in percentages and relative scores. This data is then combined to create total risk scores for each branch/bank, countrywide. The higher this score, the more likely irregularities occurred at that specific branch/bank.

In addition, the system should have the capability of generating a report or an offer letter after loan approval to a borrower in a clearly tabularized manner on all the interest and charges involved such that from the day one , the customer have a clear knowledge of the repayment plan. It should also generate a report or a Rejection Letter outlining the reasons for rejecting a loan application that did not pass credit check test.

The system will have a dedicated database for all the banks where summarized records of every approved loan will be stored respectively including the expiry date of the facility. It will also keep information on the status of any loan that has expired (whether it has been cleared or turned bad). The system should timely update the CBN bureaus with the information in the form of a credit report; Bank must be mandated to report all liquidated loans to CBN and should be ready to report the status of any unreported loan when such request arises from the CBN. This will deal with the problem of banks not reporting their bad loans to the CBN. Stiff penalty will be awarded to a bank that fails to report a bad loan good to a degree of revoking the license. It presents how the DNS ties each separate financial institution to the Credit Risk Management System. Utilizing a web browser, commercial loan applications from loan officers at various branches of any bank can be analyzed concurrently. An application is checked for completeness before processing. Both the external and CBN credit bureaus are automatically accessed, as well as pre-selected databases such as identity records, business incorporation record for corporate business and collaterals records. All credit reports are processed automatically through a score card, which highlights any deviations from pre-set standards. This is followed by the credit risk computation which include computation of the default risk, risk weighted assets and capital requirement. The fig. 2 described the model phases of predictive framework.

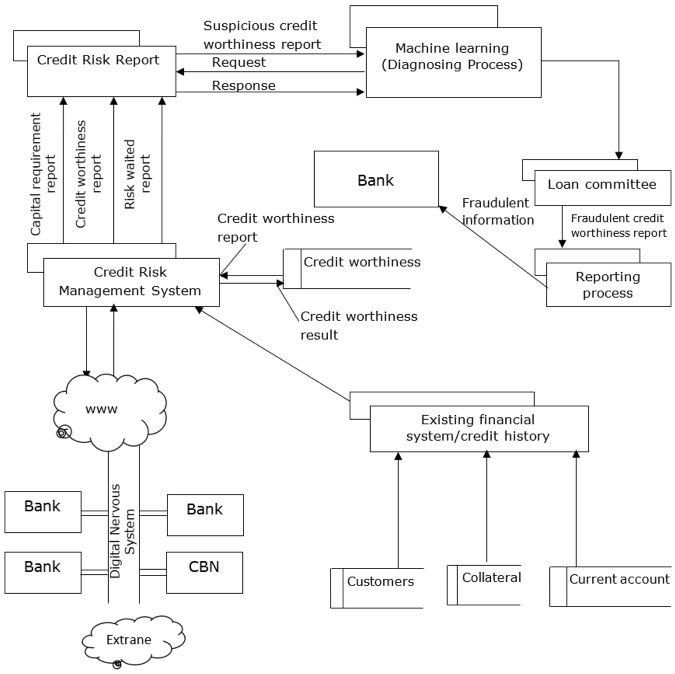

Fig. 1: Data flow diagram of the proposed system

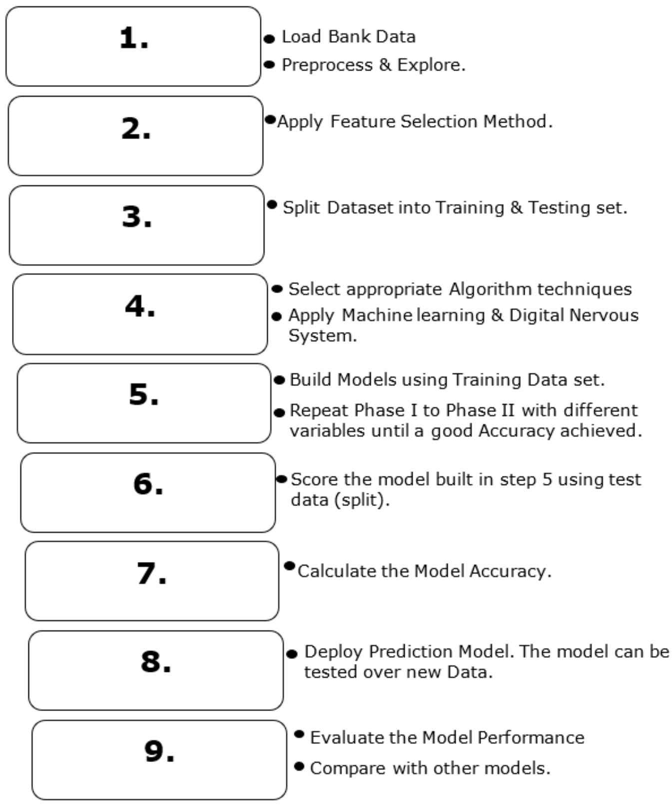

Fig. 2: Proposed Predictive Framework

4. EXPERIMENTAL SETUP AND METHODOLOGY

***a*. Tool Used:**

Weka 3.8.2 tool [15] [16], is used for simulating the proposed work, it is a famous data mining (DM) tool that includes tools for all knowledge Discovery steps. It also allows design and development of new ML and DNS schemes.

***b.* About Credit Dataset**

For this work the dataset of credit card clients, in Taiwan from April, 2005 to September, 2005, is used. The dataset and description is available at UCI – ML repository [20]. The number of reliable (1 = yes) and number of defaulters (0 = no) out of total instances in dataset are shown in Table 2.

Dataset	N0. 0f Features	Total Instances	N0. Of Instances (Yes)	N0. Of Instances (N0)
Credit Card	23	30000	23364 (77.88%)	6636 (22.12%)

Table 2: Details of Dataset

5. SIMULATION RESULTS AND ANALYSIS

The simulation runs are performed several times with different configurations to increase the performance evaluation with base classifiers, and then with new work. To evaluate the performance using confusion matrix, the classifier is evaluated on the basis of correct predictions and several other parameters. The matrix that is interesting to measure when studying machine learning techniques and DNS are Accuracy, Recall, Prediction Rate, and False Alarm. The results of following algorithms / models are compared with that of proposed model (N0. of instances).

Actual Values

Predicted Values	Legal	Fraud
Legal	1263	2049
Fraud	641	11047

Table 3: Confusion matrix of decision tree classifier

Actual Values

Predicted Values	Legal	Fraud
Legal	1375	1972
Fraud	753	10900

Table 4: Confusion matrix of Neural Network Classifier

Actual Values

Predicted Values	Legal	Fraud
Legal	568	2466
Fraud	474	11492

Table 5: Confusion matrix of Bayes Point Classifier

Actual Values

Predicted Values	Legal	Fraud
Legal	1362	1834
Fraud	656	11148

Table 6: Confusion matrix of a proposed machine learning and DNS Classifier

The parameters are presented below:

a. Accuracy: The accuracy of model is measured generally on basis of correctly classified instances. The comparison is depicted in figure 3.

$$Accuracy = \frac{TP + TN}{N0.of\ Instances}\ x\ 100$$

b. True Positive: 9/20/2024 It represents number of correctly identified instances from among the total number of correct instances. The comparison is depicted in figure 4

c. Prediction Rate: Prediction rate refers to the percentage of correct prediction among all test data, and is defined as follows:

$$Prediction\ Rate = \frac{TP}{TP + TN}\ x\ 100$$

The comparison is depicted in figure 6

d. Recall: It is also called Sensitivity. It is defined as number of positive cases that are correctly identified. The comparison is depicted in figure 5

$$Recall = \frac{TP}{TP + FN}$$

e. Comparison of Results: The results evaluated are mentioned in Table 7. The graphs based on obtained results along with evaluation and analysis is presented below.

Algorithm	Accuracy	True Positive	False Positive	Recall	Prediction Rate
Decision Tree (DT)	82.00	1263	641	0.38	10.3
Neural Network (NN)	81.83	1375	753	0.41	4.7
Bayes point (BT)	80.40	568	474	0.18	10.5
Proposed Model (PN)	83.40	1362	656	0.43	10.9

Table 7: Comparison of Results

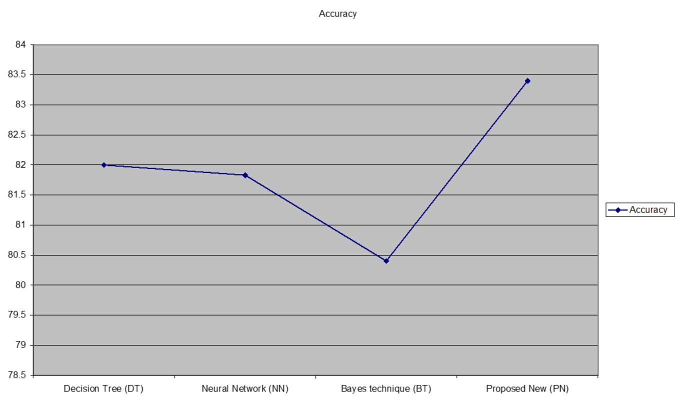

Figure 3: Comparison of Accuracy

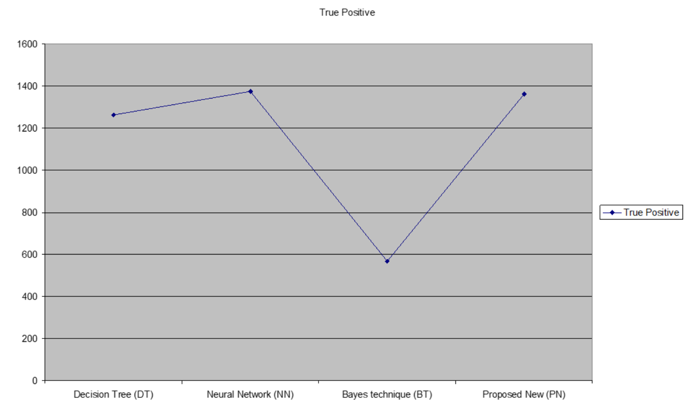

Figure 4: Comparison of True Positive

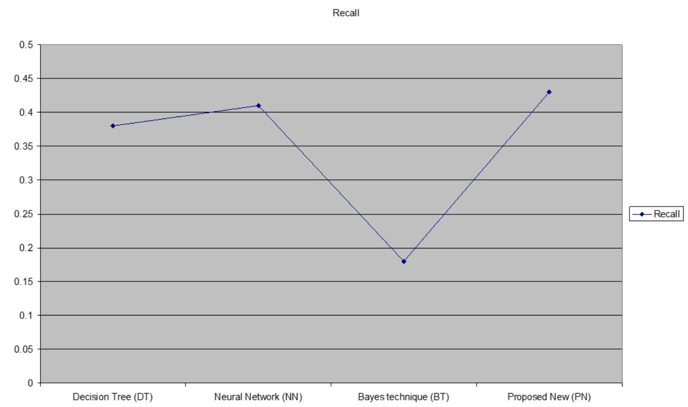

Figure 5: Comparison of Recall

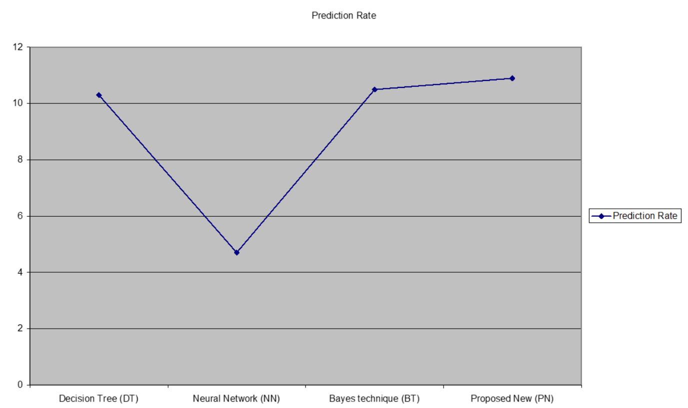

Figure 6: Comparison of Prediction Rate

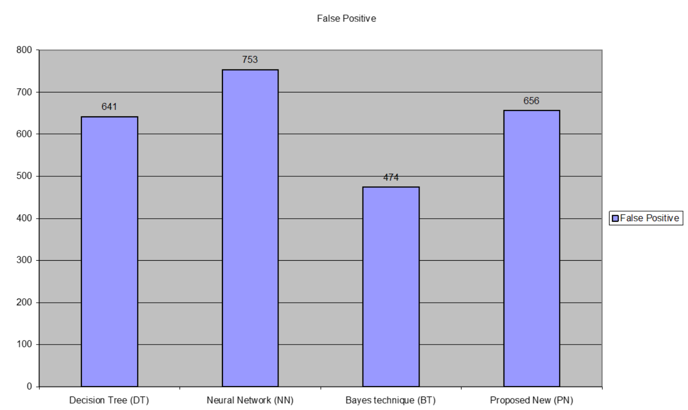

Figure 7: Comparison of False Positive

CONCLUSION

Predictive modeling for detection of credit worthiness of Bank customer is done. The purposed model predicts customer's credit worthiness using enhanced machine learning and digital Nervous system. In this paper, the proposed model for predicting credit worthiness is compared with 4 base classifiers on the basis of parameters of prime importance in ML. The best model based on ML and DNS algorithm is chosen by data scientist to decide many aspects to generate more useful results. Banks and other financial instructions are really facing the challenge of identifying risk factors, which should be considered while advancing the loans or credit to customers. Banks have realize that ML potential can help them to predict their financial decisions and resource utilization. The performance of proposed model is shown by comparing its efficiency against other popular ML algorithms. A customer based information, that is, credit worthiness prediction, and process data sets were produced; presentation of the obtained data is in form of SMS notification on the fraudulent attempt were sent to customer and database. A predictive model was developed, tested for the accuracy using confusion matrix which shown a significantly positive impact of 83.4% in credit worthiness fraud detection system. The work developed a new model approach of solving bank frauds problems especially in the area of credit worthiness. The model is therefore recommended for use by banks, financial agencies and government agencies.

REFERENCE

1. https://www.investopedia.com/terms/c/credit-worthniness.asp
2. https://www.moodys.com/sites/products/productAttachments/CreditMonitor_brochure.pdf.
3. Seema, P.U., Venkatesh M, & Anjali, N.K. (2012). Credit Evaluation Model of Loan Proposals for Indian Banks. International Journal of Modeling and Optimization, 2(4), 529-534.
4. Huang, S. H. & Chen, M.C. (2015). Credit Scoring and Rejected Instances Reassigning through Evolutionary Computations Techniques. Expert Systems with Applications, 24(4), 433-441.
5. Motwani, A., Chaurasiya, P. & Bajaj (2018). Predicting Credit Worthiness of Bank Customer with Machine Learning and over cloud. International Journal of Computer Sciences and Engineering, 6(7), 1471-1477.
6. P. N. Umoh, "An overview of Risk management Practices in the Nigeria Banking Industry", NDIC Q., 12(4): 36-48, 2002.
7. I. A .Ajah, H.C. Inyiama, "Loan Fraud Detection and IT Combat Strategies", Journal of Internet Banking and Commerce, August 2011, Volume 16, No 2. 2011. http:/www.arraydev.com/commerce/jibcl.
8. R. Peter, "Ghana Project Leverages GIS-Based Title Registration and Microfinance to Alleviate Poverty International Land Systems", The Fall issue of ArcNews magazine. pp. 29- 35, 2008.
9. S. Sudalaimuthu, P. Ramasamy, "Credit Risk Assessment System in Banking Sector", Zenith International Journal of Multidisciplinary Research. Vol.1 Issue 8, December 2011, IS S N 2 231 5780. Retrieved on 10 March, 2011(Online:www.zenithresearch.org.in/images/.../12_VOL%201_ISSUE8_ZEN.pdf).
10. Regina Esi Turkson, Edward Yeallakuor Baagyere, Gideon Evans Wenya, "A Machine Learning Approach for Predicting Bank Credit Worthiness", ISBN: 978-1-4673-9187-0, IEEE 2016
11. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J.

Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
12. Loris Nanni, Alessandra Lumini, "An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring", Elsevier, Expert Systems with Applications 36 (2009). 3028–3033.
13. C.R. Durga devi, Dr. R. Manicka chezian, "A Relative Evaluation of the Performance of Ensemble Learning in Credit Scoring", IEEE International Conference on Advances in Computer Applications (ICACA), 978-1-5090-3770-4/16, 2016
14. Zhang, "Z., Research of credit risk of commercial bank's personal loan based on CHAID decision tree", Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), 2011
15. http://www.cs.waikato.ac.nz/~ml/weka/index.html.
16. Weka, University of Waikato,Hamilton, New Zealand.
17. Zhou, Z. H. (2009). Ensemble. In L. Liu & T. Özsu (Eds.), Encyclopedia of database systems. Berlin: Springer.
18. Anand Motwani, Goldi Bajaj, Sushila Mohane, "Predictive Modelling for Credit Risk Detection using Ensemble Method", International Journal of Computer Sciences and Engineering, Vol.6, Issue.6, pp.863-867, 2018.
19. Ling Kock Sheng, and Teh Ying Wah, "A comparative study of data mining techniques in predicting consumers' credit card risk in banks", African Journal of Business Management Vol. 5 (20), pp 8307-8312, 16 September, 2011
20. http://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients.
21. Singh, M & Dixit, G.K. (2018). Modeling Customer's Credit Worthiness using enhanced Ensemble Model. International Worthiness using enhanced Ensemble Model. International Journal of Computer Sciences and Engineering, 6(7), 1466-1470.
22. Ajah, I.A., & Inyiama, H.C. (2013). A Model of DNS –Based Bank Credit Risk Management System in Nigeria. ARPN Journal of Systems and Software, 3(6), 106-114.