

A PROJECT REPORT

ON

Approval using Prediction of Loan Machine Learning

Submitted in partial fulfillment for the requirement of the award of

TRAINING

IN

Data Analytics, Machine Learning and AI using Python



Submitted By

Gorle.Jeevanashish(bhilai institute of technology)

Under the guidance of

Mr. Bipul Shahi

ABSTRACT :-

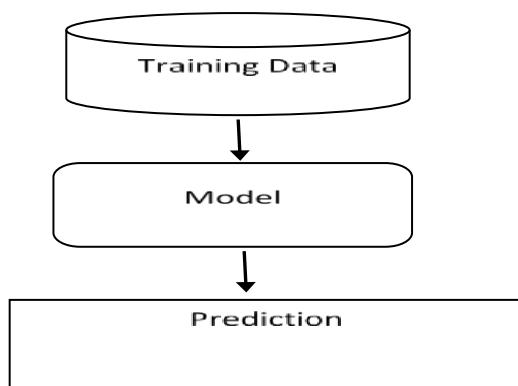
Loan approval is a very important process for banking organizations. The system approved or reject the loan applications. Recovery of loans is a major contributing parameter in the financial statements of a bank. It is very difficult to predict the possibility of payment of loan by the customer. In recent years many researchers worked on loan approval prediction systems. Machine Learning (ML) techniques are very useful in predicting outcomes for large amount of data. In this paper three machine learning algorithms, Logistic Regression(LR), Decision Tree (DT) and Random Forest (RF) are applied to predict the loan approval of customers. The experimental results conclude that the accuracy of Decision Tree machine learning algorithm is better as compared to Logistic

1.Introduction Now a day's people rely on bank loans to fulfill their needs. The rate of loan applications increases with a very fast speed in recent years. Risk is always involved in approval of loans. The banking officials are very conscious about the payment of the loan amount by its customers. Even after taking lot of precautions and analyzing the loan applicant data, the loan approval decisions are not always correct. There is need of automation of this process so that loan approval is less risky and incur less loss for banks

Artificial Intelligence AI is an emerging technology now a day. The application of AI solves many problems of the real world. Machine Learning is an AI technique which is very useful in prediction systems. Figure 1 is showing a basic model of machine learning. It creates a model from a training data. While making the prediction the model which is developed by training algorithm (which is machine learning) is used. The machine learning algorithm trained the system using a fraction of the data available and test the remaining data.

The machine Learning techniques can be applied on a sample test data first and then can be used in making prediction related decisions. This paper applied the machine learning approaches in solving loan approval problem of banking sector. Next section discusses the literature survey.

Then proposed work, results and analysis are discussed. Finally, conclusion and future scope is discussed which is followed by the references used in this paper.



2.Literature Survey

A. Vaidya proposed a method for approval of loan prediction using logistic regression [1]. Logical regression is a machine learning technique which is very useful in prediction system. The approval of loan is a very important process in banking system.

A. Vaidya solves the problem by applying machine learning in a sample data set for loan approval applications. It also opens other areas on which machine learning is applicable. A. Li and Q. Sun [2] find a method to calculate risk involved in loan approvals for SMEs. A concept of loan consuming radius was introduced which was based upon supply chain in consumer market. F. M. Isik et al. develop a loan approval system using Business Process ExecutionLanguage BPEL [3].

The concept of BPEL is very useful in business firms. A reasoning engine was developed which removes some services from the BPEL process which are not necessary to complete a process. The system was applied on loan approval which involve many processes. [4]V. C. T. Chan et al. proposed a credit approval system using web services. The system approved credit for the customers. With credit application the customer submits some other useful information's. This information's are processed by Credit Approval System which finally give credit score to the applicant. The paper developed a web services based solution of this problem. J. Lohokare et al. [5] proposed a system which automatically collect data for an applicant and decides the credit score. The system work on the social media to collect information about the user. R. Yang et al. [6] analyzed that whether the credit default behavior of a SME depends upon credit features of its owner or not. The author concluded that features of the owner behaves as valuable parameters to calculate risk of a loan for SMEs. [7]M. Bayraktar et al. [7] proposed a method for credit risk analysis using machine learning. Boltzman machine was used to make the analysis for risk calculation of loan. H. A. P. Pérez et al. [8] introduced fuzzy model for calculation of credit score of the customer. The information collected by the system for calculation of the credit score was converted into gradual values using fuzzy sets. The fuzzy based method performs better for calculation of the credit score of the applicants. S. Yadav and S. Thakur [9] applied Big Data approach for loan analysis. The techniques of big data analysis was applied on customer data to calculate bank loan analysis. Hadoop based method was used in the loan analysis. Y. Lin [10] analysis of the effect of the political approaches effect the loans of state banks. The paper investigated that in state owned banks, the political relationship plays a considerable role. [11]Ruifen Zhao worked on approval of college loans. Education loans are very common among students because of rise in the cost of education. The paper investigated the issues in loan approval of college students. M. Houshmand and M. D. Kakhki [12] proposed an expert system which evaluates the loan approvals. The system used rule base approaches for loan approval decisions. L. Hui-ling [13] analyze the relation between characteristics of the banks, firms and loans approval. The paper investigated that there is a strong relationship between approval of loans and characteristics of business firm who apply the loan and characteristics of the bank. C. Yin [14] apply fuzzy logic to calculate the bank loan risks. A new pattern recognition system using fuzzy logic was developed which evaluate the risks involves in the approval of bank loans for applicants. J. Ma and Y. Cheng [15] proposed Markov Chain based model for risk management of bank loans. A. V. Gutierrez [16] proposed a model for housing loan. The model was worked for green housing loans. J. Chen and W. Guo [17] worked on loan limit of the loan applicants. The model worked on supply chain for financing decision making. G. Arutjothi and C. Senthamarai, [18] used machine learning classifier for prediction of loan approval status in banks. The machine learning based prediction system was applied on commercial banks. The paper conclude that the machine learning approach is very useful in loan status prediction.

:=Machine Learning Algorithms

Machine learning algorithm which are used in this work to make a model are as follows:

1. Logistic Regression

2. Decision Tree
3. Random forest

Logistic Regression

Logistic Regression (LR) is a machine learning technique. The LR is very commonly used to solve binary classification problem. There are following basic postulation:

1. Binary logistic regression has binary dependent variables.
2. In binary regression dependent variables have level 1.
3. The included variables should have meaning. All included independent variables should be selfreliant.
4. The independent variables are related to the log odds linearly.
5. The sample size should be large for LR. **Decision Tree**

Decision TREE is a supervised ML technique which is non parametric in nature. It has predefined target variable which is generally used in problem classification. It is useful for classification and regression both. It works categorical & continuous both for input and output variables.

Random Forest

Random Forest (RF) is a very useful machine learning algorithm. It is mostly used in areas such as classification, regression analysis etc. At the training time RF algorithm creates many decision trees.

RF is a supervised learning approach which need a test data for the model for training. It creates random forests for the problem set and then find the solution using these random forests.

1. Results and Analysis

Three machine learning approaches are applied on the test data to predict the loan approvals of loan requests.

Python programming language is used to implement machine learning algorithms. For training 70 percent data is used

Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property_Area
LP001015	Male	Yes	0	Graduate	No	5720	0	110	360	1	Urban
LP001022	Male	Yes	1	Graduate	No	3076	1500	126	360	1	Urban
LP001031	Male	Yes	2	Graduate	No	5000	1800	208	360	1	Urban
LP001035	Male	Yes	2	Graduate	No	2340	2546	100	360	1	Urban
LP001051	Male	No	0	Not Graduate	No	3276	0	78	360	1	Urban
LP001054	Male	Yes	0	Not Graduate	Yes	2165	3422	152	360	1	Urban
LP001055	Female	No	1	Not Graduate	No	2226	0	59	360	1	Semurban
LP001056	Male	Yes	2	Not Graduate	No	3881	0	147	360	0	Rural
LP001059	Male	Yes	2	Graduate	No	13633	0	280	240	1	Urban
LP001067	Male	No	0	Not Graduate	No	2400	2400	123	360	1	Semurban
LP001078	Male	No	0	Not Graduate	No	3091	0	90	360	1	Urban
LP001082	Male	Yes	1	Graduate	No	2185	1516	162	360	1	Semurban
LP001083	Male	No	1+	Graduate	No	4166	0	40	180	1	Urban
LP001094	Male	Yes	2	Graduate	No	12173	0	166	360	0	Semurban
LP001096	Female	No	0	Graduate	No	4666	0	124	360	1	Semurban
LP001099	Male	No	1	Graduate	No	5667	0	131	360	1	Urban
LP001105	Male	Yes	2	Graduate	No	4583	2916	200	360	1	Urban
LP001107	Male	Yes	1+	Graduate	No	3786	333	126	360	1	Semurban
LP001108	Male	Yes	0	Graduate	No	9226	7916	300	360	1	Urban
LP001115	Male	No	0	Graduate	No	1300	3470	100	180	1	Semurban
LP001121	Male	Yes	1	Not Graduate	No	1888	1620	48	360	1	Urban
LP001124	Female	No	1+	Not Graduate	No	2083	0	28	180	1	Urban
LP001128		No	0	Graduate	No	3909	0	101	360	1	Urban
LP001135	Female	No	0	Not Graduate	No	3765	0	125	360	1	Urban
LP001149	Male	Yes	0	Graduate	No	5400	4380	290	360	1	Urban
LP001153	Male	No	0	Graduate	No	0	24000	148	360	0	Rural
LP001163	Male	Yes	2	Graduate	No	4363	1250	140	360	1	Urban
LP001169	Male	Yes	0	Graduate	No	7500	3750	275	360	1	Urban
LP001174	Male	Yes	0	Graduate	No	3772	833	57	360	1	Semurban
LP001176	Male	No	0	Graduate	No	2942	2382	125	180	1	Urban
LP001177	Female	No	0	Not Graduate	No	2478	0	75	360	1	Semurban
LP001183	Male	Yes	2	Graduate	No	6250	820	192	360	1	Urban
LP001185	Male	No	0	Graduate	No	3268	1683	152	360	1	Semurban
LP001187	Male	Yes	0	Graduate	No	2783	2708	158	360	1	Urban

and 30 percent data is used for testing. The prediction accuracy of the different ML approaches is calculated and compared. The training data set is shown in figure 3.

On the basis of this train data set (shown in figure 3), system analyze rest of 30 percent data and predict the results in term of loan status either accepted or rejected. Results with loan status by applying the logistic regression (shown in figure-4(a)), decision tree (shown in figure-4(b)) and random forest (shown in figure-4(c)).Figure 5(a), 5(b) and 5(c) and 5(d) are demonstrating the histograms generated. Figure 5(a) is showing the histogram for applicant income. Figure 5(b) is showing histogram of co applicant income. Figure 5(c) is showing histogram of loan amount term. Figure 5(d) is showing loan amount

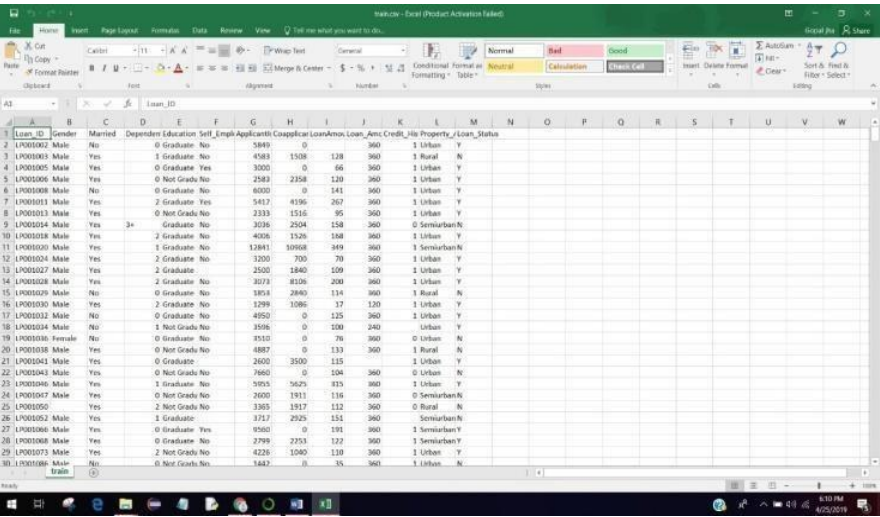
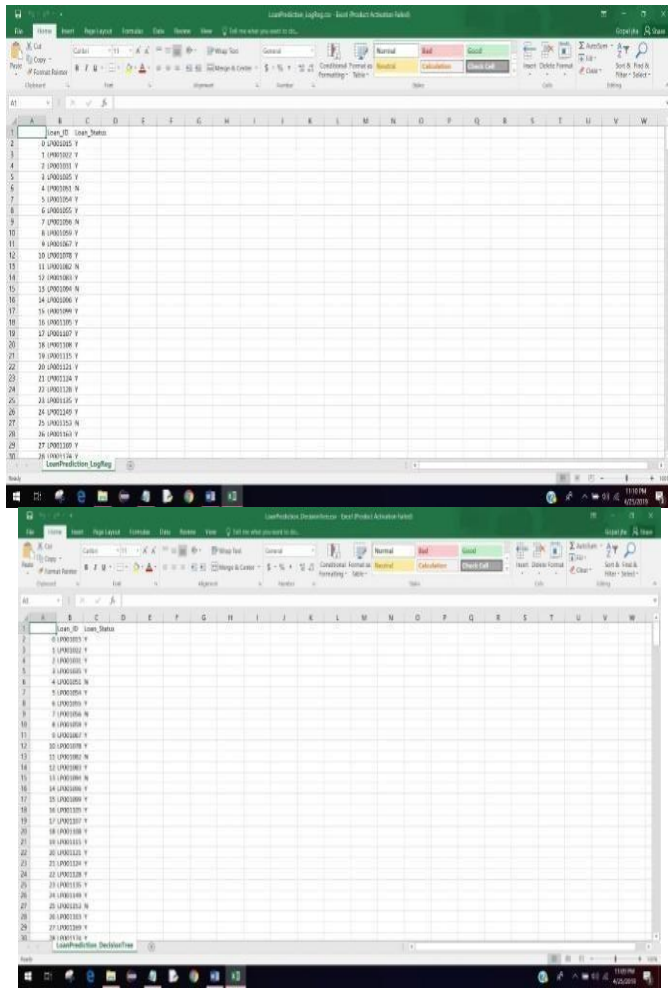


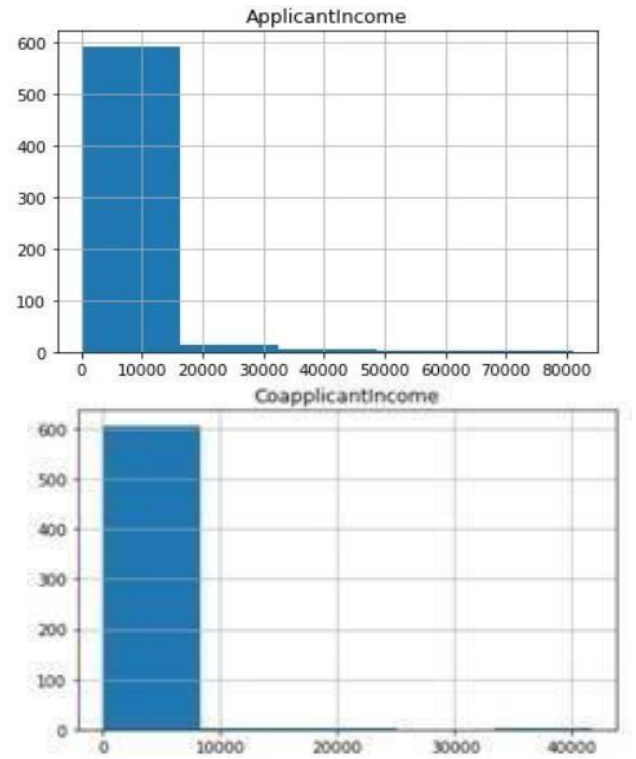
Figure-3: Trained Data Set

S.No.	Machine learning Algorithm	Prediction Accuracy Percentage
1	Logistic Regression	93.04
2	Decision Tree	95.0
3	Random Forest	92.53



Logistics Regression Result with Loan Status.

Table-1: Comparison of prediction accuracy of



machine learning algorithms

Figure-5(a): Histogram of Applicant income

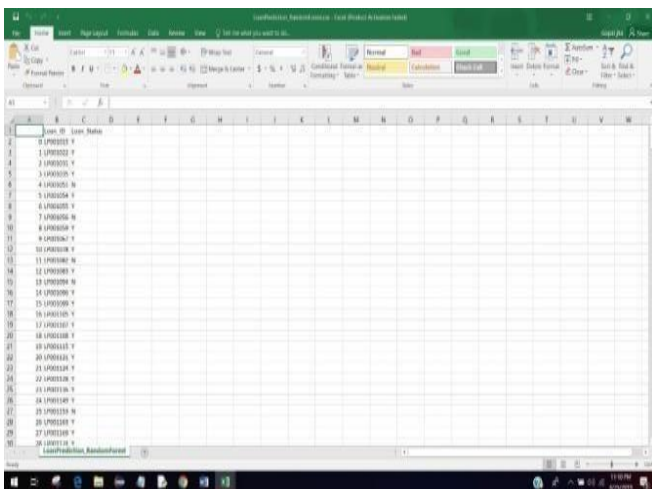
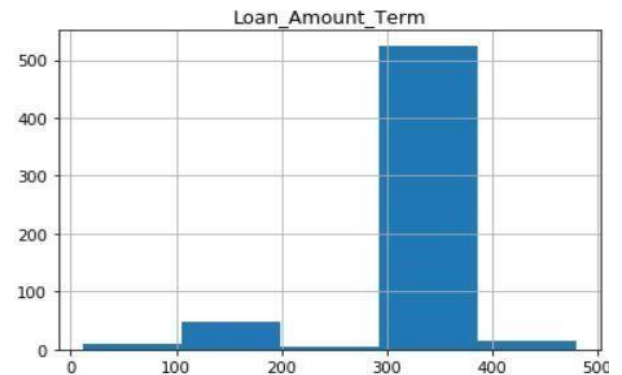
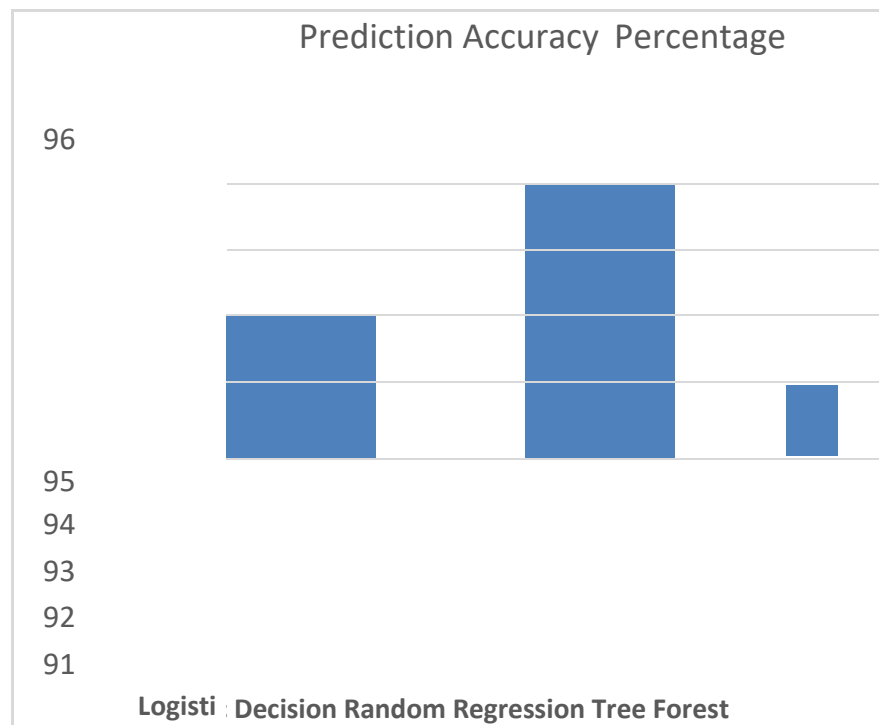
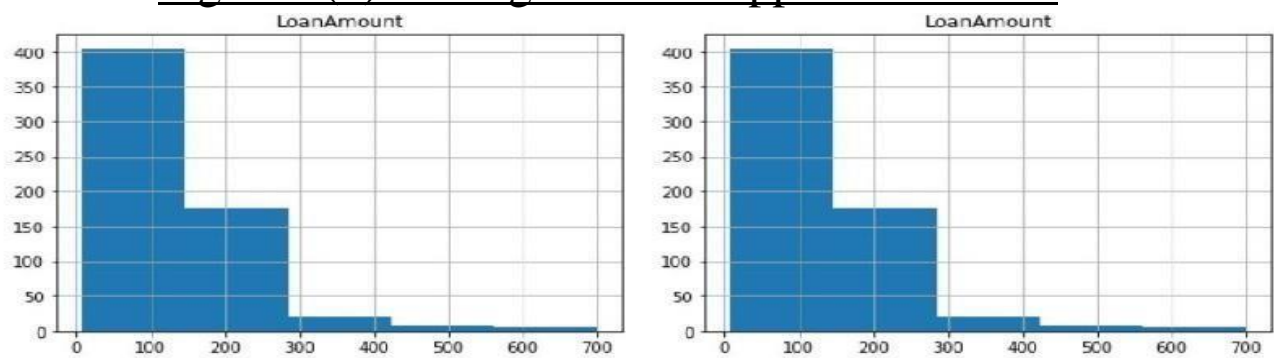


Figure-4(b): Decision Tree Result with Loan Status.

Figure-5(b): Histogram of Coapplicant income



2.Conclusion and Future Scope

This paper applied machine learning in prediction of loan approval. Three ML algorithms are used to predict the loan approval status of customers for bank loans. The results shown that the prediction accuracy is 93.04%, 95% and 92.53% for LR, DT algorithmRF algorithms respectively. Among three the accuracy of DT algorithm is best for prediction of loans. In future the Decision Tree algorithm can be applied on other data sets available for loan approvals to further investigate its accuracy. A rigorous analysis of other machine learning algorithms other than these three can also be done in future to investigate the power of machine learning algorithms for loan approval prediction.

. References

- [1] Vaidya, "Predictive and probabilistic approach using logistic regression: Application to prediction of loan approval," 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Delhi, 2017, pp. 1-6.doi: 10.1109/ICCCNT.2017.8203946
- [2] A.Li and Q. Sun, "The Risk of Loan Consuming by SMEs Based on the Supply Chain," 2012 International Conference on Management of eCommerce and e-Government, Beijing, 2012, pp. 356-359.doi: 10.1109/ICMeCG.2012.24
- [3] F. M. Isik, B. Tastan and P. Yolum, "Automatic Adaptation of BPEL Processes Using Semantic Rules: Design and Development of a Loan Approval System," 2007 IEEE 23rd International Conference on Data Engineering Workshop, Istanbul, 2007, pp. 944-951.doi: 10.1109/ICDEW.2007.4401089
- [4] V. C. T. Chan et al., "Designing a Credit Approval System Using Web Services, BPEL, and AJAX," 2009 *IEEE International Conference on eBusiness Engineering*, Macau, 2009, pp. 287-294.doi: 10.1109/ICEBE.2009.46
- [5] J. Lohokare, R. Dani and S. Sontakke, "Automated data collection for credit score calculation based on financial transactions and social media," 2017

International Conference on Emerging Trends & Innovation in ICT (ICEI), Pune, 2017, pp. 134-138.doi: 10.1109/ETIICT.2017.7977024

- [6] R. Yang, X. Zhou and W. Wang, "Is the Small and Medium-Sized Enterprises' Credit Default Behavior Affected by Their Owners' Credit Features?," *2011 International Conference on Management and Service Science*, Wuhan, 2011, pp. 1-4.doi: 10.1109/ICMSS.2011.5998460
- [7] M. Bayraktar, M. S. Aktaş, O. Kalıpsız, O. Susuz and S. Bayracı, "Credit risk analysis with classification Restricted Boltzmann Machine," *2018 26th Signal Processing and Communications Applications Conference (SIU)*, Izmir, 2018, pp. 1-4.doi: 10.1109/SIU.2018.8404397
- [8] H. A. P. Pérez, J. A. P. Palacio and C. Lochmuller, "Fuzzy model Takagi Sugeno with structured evolution for determining consumer credit score," *2015 10th Iberian Conference on Information Systems and Technologies (CISTI)*, Aveiro, 2015, pp. 1-6.doi: 10.1109/CISTI.2015.7170485
- [9] S. Yadav and S. Thakur, "Bank loan analysis using customer usage data: A big data approach using Hadoop," *2017 2nd International Conference on Telecommunication and Networks (TEL-NET)*, Noida, 2017, pp. 1-8.doi: 10.1109/TEL-NET.2017.8343582
- [10] Y. Lin, "How Do Political Relations Affect State-Owned Bank Loans?," *2015 9th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*, Blumenau, 2015, pp. 485-488.doi: 10.1109/IMIS.2015.77
- [11] Ruifen Zhao, "Research on college loan problems," *2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic*
- [1] C. Yin, "Fuzzy pattern recognition model of bank loan risk and its application," *2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery*, Yantai, 2010, pp. 1272-1276.doi:

10.1109/FSKD.2010.5569118

- [2] J. Ma and Y. Cheng, "Research on Risk Management of Bank Loan and Loss Provision Based on Markov Chain," *2009 International Conference on EBusiness and Information System Security*, Wuhan, 2009, pp. 1-5.doi: 10.1109/EBISS.2009.5138055
- [3] A. V. Gutierrez, "Green Banking: A Proposed Model for Green Housing Loan," *2016 International Conference on Industrial Engineering, Management Science and Application (ICIMSA)*, Jeju, 2016, pp. 1-4.doi: 10.1109/ICIMSA.2016.7504011
- [4] J. Chen and W. Guo, "An Analysis of Supply Chain Financing DecisionMaking in View of the Loan Limit," *2016 International Conference on Smart City and Systems Engineering (ICSCSE)*, Hunan, 2016, pp. 265-268.doi: 10.1109/ICSCSE.2016.0078
- [5] G. Arutjothi and C. Senthamarai, "Prediction of loan status in commercial bank using machine learning classifier," *2017 International Conference on Intelligent Sustainable Systems (ICISS)*, Palladam, 2017, pp. 416-419.doi: 10.1109/ISS1.2017.8389442
- [6] Y. Shi and P. Song, "Improvement Research on the Project Loan Evaluation of Commercial Bank Based on the Risk Analysis," *2017 10th International Symposium on Computational Intelligence and Design (ISCID)*, Hangzhou, 2017, pp. 3-6.doi: 10.1109/ISCID.2017.60
- [7] R. Zhang, D. Li "Applying Machine Learning and High Performance Computing to Water Quality Assessment and Prediction", *Advances in Science, Technology and Engineering Systems Journal*, vol. 2, no. 6, pp. 285289 (2017).DOI: 10.25046/aj020635
- [8] C. Frank, A. Habach, R. Seetan, A. Wahbeh "Predicting Smoking Status Using Machine Learning Algorithms and Statistical Analysis", *Advances in Science, Technology and Engineering Systems Journal*, vol. 3, no. 2, pp. 184189 (2018).DOI: 10.25046/aj030221

- [9] R. Lopes, M. Ladeira, R. Carvalho, "Use of machine learning techniques in the prediction of credit recovery", *Advances in Science, Technology and Engineering Systems Journal*, vol. 2, no. 3, pp. 1432-1442 (2017).DOI: 10.25046/aj0203179

