



Loan Eligibility Prediction using Data Science Algorithms A Comparative Analysis

M. Ramkumar, Joel Johny, *Keshav L Darak, Maresh B M, Joshua John

Department of Computer Science & Engineering HKBK College of Engineering Bangalore, India.

*Corresponding author Email: 1hk18cs068@hkbk.edu.in

Abstract. Loan is a amount that is provided to someone else in exchange for repayment of the loan principle amount plus interest. The different variety of loans is Personal loan, Home loan, Education loan, etc. Here, Decision Tree and Naïve Bayes algorithms are used to predict the loan eligibility of a particular individual. Data mining techniques are used to extract data and predict the loan eligibility. By providing loan, the banks help in the overall flow of the economy of the country. The borrower can avail the loan whenever required and the bank can get back the principle amount with interest based on some pre-defined percentages.

Keywords: Decision Tree; Naïve Bayes; Data mining; Loan Eligibility; Economy.

1. Introduction

Loan is an amount that is provided to someone else in exchange for repayment of the principle amount plus interest. Nowadays, approving of loans has become a significant task of the financial organizations/banking sector. For financial organizations, loans are also one of the most significant sources of income. On loans that are subsequently accepted and made accessible to their customers, banks charge interest (borrowers). Before approving a loan application, lenders seek proof that they will receive their money back, plus interest. As a consequence, evaluating a person's or an organization's creditworthiness prior to extending a loan is critical. The project will completely verify the borrower and do a background check based on a range of characteristics such as gender, income, job status, and so on to evaluate whether or not the borrower is creditworthy and may be sanctioned the loan [1]-[16].

2. Existing Work

The analysis basically deals with classifying whether The question of whether or not a person is eligible for a loan implies that the problem is one of classification, which may be solved using classification techniques [17]-[23]. The present system, CIBIL, is a credit rating system that gathers and retains records of payments made on loans and credit cards by people and businesses. These records are received TransUnion, a Credit Information company, submitted by Every month, a CIBIL Score and Report for each individual is generated based on information given by banks and other lenders [24]-[36]. The CIBIL Score is a three-digit numeric assessment of a person's financial situation in the United Kingdom. It ranges from 300 to 900 points and is based on data from the CIBIL Report's 'Accounts' and 'Enquiries' sections. This data may be used by lenders to evaluate and approve loan applications. It's worth noting that, although the CIBIL Score is significant in the loan application process, it doesn't always provide all of the information needed [37]-[45]. When it comes to financing, a lender's first consideration is your CIBIL Score; the higher your score, the greater your chances of getting your loan investigated and approved. The lender has complete control over whether or not to lend, and he or she must also examine a variety of other factors. CIBIL makes no judgements on whether or not a loan or credit card should be granted or rejected [44]-[53]. To assess whether or not an applicant is creditworthy, the lender must personally analyse each application based on the applicant's information, which includes their gender and marital status, level of education, number of dependents, income, loan amount, and credit history, among other factors.

Prediction Process

- Step 1: Separate every individual customer's transactions from the entire transaction database.
- Step 2: From the transaction of all the customers' database separate the transaction.
- Step 3: Apply the specific standard algorithm to predict the loan eligibility by using variables such as gender, income, employment status.

3. Experimental Setup and Methods

Our research will be based on data sets that will be categorised according to protocols. Every stage of the implementation was done in Python, including libraries like NumPy, Pandas, Keras, Scikit-Learn, and Tensorflow. With the aid of Rstudio, data purification was done on occasion. PowerBI is also available for visualising the whole transactional process. The many stages of the transactions, which include data collection, data preparation, data analysis, classifier algorithm training, and classifier algorithm testing. The data is translated into a readable format and fit and sampled during the preparation stage. The dataset is subjected to feature selection and reduction throughout the analysis stage, which is achieved using PCA (Principal Component Analysis). During the training phase, classifier algorithms are built and fed with the processed data that will be used to classify the data. The effectiveness of the transactions is examined using True Positive, False Positive, True Negative, and False Negative replies to the questions in our research. The findings are compared, and the accuracy, sensitivity, specificity, and precision of these classifiers are evaluated.

Dataset

The information was obtained via Kaggle. Before awarding the loan to the applicant, a number of factors are taken into account, including.

1	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property_Area
2	LP02016	Male	Yes	0	Graduate	No	5720	0	190	360	1	Urban
3	LP02022	Male	Yes	1	Graduate	No	3075	1500	125	360	1	Urban
4	LP02029	Male	Yes	2	Graduate	No	5500	1800	280	360	1	Urban
5	LP02033	Male	Yes	2	Graduate	No	2540	2540	100	360	1	Urban
6	LP02051	Male	No	0	Not Graduate	No	3070	0	70	360	1	Urban
7	LP02054	Male	Yes	0	Not Graduate	Yes	2900	3402	152	360	1	Urban
8	LP02055	Female	No	1	Not Graduate	No	2225	0	55	360	1	Suburban
9	LP02056	Male	Yes	2	Not Graduate	No	3801	0	147	360	0	Rural
10	LP02059	Male	Yes	2	Graduate		1302	0	289	340	1	Urban

Data Cleaning

The process of filling in missing data is an important part of the data cleaning method. There are a number of ways to solve this issue, including ignoring the whole tuple, but the bulk of them are likely to add bias into the findings. Furthermore, with changes such as the elimination of unnecessary columns and the separation of the date time column into two columns.

Data Integration

Because the fraudulent and real record files were originally housed in two separate files, the two data sources were integrated to preclude future data tampering.

Data Transformation

All of the information from the various categories was compiled into a numerical format that was simple to understand. The transactional dataset contains a variety of data types with a variety of ranges. As a consequence, data transformation entails data cleaning and normalisation.

Data Reduction

In this example, the approach that was applied was dimension reduction. Principal component analysis, or PCA, is a well-known transform method that is frequently utilised in a variety of sectors. The application of this method handles the feature selection issue in question from the perspective of numerical analysis. Because PCA was able to find the ideal number of principle components, it was effective in feature selection.

Acknowledgment

A special and an earnest word of thanks to the project guide Prof. Ramkumar for their constant assistance, support, patience, endurance and constructive suggestions for the advancement of the project.

References

- [1]. Turkson, Regina Esi, Edward Yeallakuor Baagyere, and Gideon Evans Wenya. "A machine learning approach for predicting bank credit worthiness." 2016 Third International Conference on Artificial Intelligence and Pattern Recognition (AIPR). IEEE, 2016.
- [2]. Vaidya, Ashlesha. "Predictive and probabilistic approach using logistic regression: application to prediction of loan approval." 2017 8th International Conference on Computing, Communication, and Networking Technologies (ICCCNT). IEEE, 2017.
- [3]. Sheikh, Mohammmad Ahmad, Amit Kumar Goel, and Tapas Kumar. "An Approach for Prediction of Loan Approval using Machine Learning Algorithm." 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC).

- [4]. Ramya S1, Priyesh Shekhar Jha2, Ilaa Raghupathi Vasishtha2, Shashank H2, Neha Zafar, "Monetary Loan Eligibility Prediction using Machine Learning" Department of Computer Science and Engineering TheNational Institute of Engineering, Mysuru, India, Jan 2021]
- [5]. N. R. Deepak and S. Balaji, "Performance analysis of MIMO-based transmission techniques for image quality in 4G wireless network," 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCI), 2015, pp. 1-5, doi: 10.1109/ICCI.2015.7435774.
- [6]. Loganathan, R., & Kumaraswamy, Y. S. (2013). Active contour based medical image segmentation and compression using biorthogonal wavelet and embedded zerotree. *Indian Journal of Science and Technology*, 6(4), 4390-4395.
- [7]. Jotheeswaran, J., Loganathan, R., & Madhu Sudhanan, B. (2012). Feature reduction using principal component analysis for opinion mining. *International Journal of Computer Science and Telecommunications*, 3(5), 118-121.
- [8]. Loganathan, R., & Kumaraswamy, Y. S. (2011, December). An improved active contour medical image compression technique with lossless region of interest. In 3rd International conference on trendz in information sciences & computing (TISC2011) (pp. 128-132). IEEE.
- [9]. Loganathan, R., & Kumaraswamy, Y. S. (2010). Medical image compression using biorthogonal spline wavelet with different decomposition. *IJCSE International Journal on Computer Science and Engineering*, 2(9), 3003-3006.
- [10]. Loganathan, R., & Kumaraswamy, D. Y. (2012). Medical Image Compression with Lossless Region of Interest Using Fuzzy Adaptive Active Contour. In International Conference on Computational Techniques and Mobile Computing (ICCTMC'2012) December (pp. 14-15).
- [11]. Loganathan, R., & Kumaraswamy, Y. S. (2002). Performance Evaluation of Image Compression for Medical Image. *International Journal of Advanced Research in Computer Science and Software Engineering [2013] Vol, 4*.
- [12]. Kurian, S., & Ramasamy, L. (2021). Securing Service Discovery from Denial of Service Attack in Mobile Ad Hoc Network (MANET). *International Journal of Computer Networks and Applications*, 8(5), 619-633.
- [13]. Khan, Z., & Loganathan, R. (2020, October). AutoLiv: Automated Liver Tumor Segmentation in CT Images. In 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE) (pp. 151-156). IEEE.
- [14]. Loganathan, R., Khan, F. A., Gulzar, I., Parray, I. N., & Bhat, F. A. (2020). A Survey on Prober: An automated network vulnerability scanner. *International Scientific Journal of Contemporary Research in Engineering Science and Management*, 5(2), 85-88.
- [15]. Loganathan, R., Aliya, B. B., Rehman, S. S. U., & Pasha, A. (2020). A Survey on Paperless Examination. *International Scientific Journal of Contemporary Research in Engineering Science and Management*, 5(2), 80-84.
- [16]. Khan, Z. (2020). Radiomics in Prostate MRI: A Review on Opportunities & Challenges. *International Scientific Journal of Contemporary Research in Engineering Science and Management*, 5(1), 7-10.
- [17]. Kurian, S., & Ramasamy, L. (2021). Novel AODV based service discovery protocol for MANETS. *Wireless Networks*, 27(4), 2497-2508.
- [18]. Patan, R., & Gandomi, A. H. (2021). Improved salient object detection using hybrid Convolution Recurrent Neural Network. *Expert Systems with Applications*, 166, 114064.
- [19]. Yuvaraj, N., Srihari, K., Dhiman, G., Somasundaram, K., Sharma, A., Rajeskannan, S., ... & Masud, M. (2021). Nature-InspiredBased Approach for Automated Cyberbullying Classification on Multimedia Social Networking. *Mathematical Problems in Engineering*, 2021.
- [20]. Natarajan, Y., Kannan, S., & Mohanty, S. N. (2021). Survey of Various Statistical Numerical and Machine Learning Ontological Models on Infectious Disease Ontology. *Data Analytics in Bioinformatics: A Machine Learning Perspective*, 431-442.
- [21]. Raja, R. A., Yuvaraj, N., & Kousik, N. V. (2021). Analyses on Artificial Intelligence Framework to Detect Crime Pattern. *Intelligent Data Analytics for Terror Threat Prediction: Architectures, Methodologies, Techniques and Applications*, 119-132.
- [22]. Kannan, S., Dhiman, G., Natarajan, Y., Sharma, A., Mohanty, S. N., Soni, M., ... & Gheisari, M. (2021). Ubiquitous Vehicular AdHoc Network Computing Using Deep Neural Network with IoT-Based Bat Agents for Traffic Management. *Electronics*, 10(7), 785.
- [23]. Yuvaraj, N., Raja, R. A., Karthikeyan, T., & Kousik, N. V. (2020). 11 Improved Privacy Preservation Framework for Cloud-Based Internet of Things. *Internet of Things: Integration and Security Challenges*, 165.
- [24]. Yuvaraj, N., Karthikeyan, T., & Praghsh, K. (2021). An improved task allocation scheme in serverless computing using gray wolf Optimization (GWO) based reinforcement learning (RIL) approach. *Wireless Personal Communications*, 117(3), 2403-2421.
- [25]. Mariappan, L. T., & Yuvaraj, N. (2020). Analysis On Cardiovascular Disease Classification Using Machine Learning Framework. *Solid State Technology*, 63(6), 10374-10383.
- [26]. Karthick, S., Yuvaraj, N., Rajakumari, P. A., & Raja, R. A. (2021). Ensemble Similarity Clustering Frame work for Categorical Dataset Clustering Using Swarm Intelligence. In *Intelligent Computing and Applications* (pp. 549-557). Springer, Singapore.

- [27]. Yuvaraj, N., Raja, R. A., & Kousik, N. V. (2021). Privacy Preservation Between Privacy and Utility Using ECC-based PSO Algorithm. In *Intelligent Computing and Applications* (pp. 567-573). Springer, Singapore.
- [28]. Yuvaraj, N., Raja, R. A., Palanivel, P., & Kousik, N. V. (2020, April). EDM Process by Using Copper Electrode with INCONEL 625 Material. In *IOP Conference Series: Materials Science and Engineering* (Vol. 811, No. 1, p. 012011). IOP Publishing.
- [29]. Veerappan Kousik, N. G., Natarajan, Y., Suresh, K., Patan, R., & Gandomi, A. H. (2020). Improving Power and Resource Management in Heterogeneous Downlink OFDMA Networks. *Information*, 11(4), 203.
- [30]. Natarajan, Y., Raja, R. A., Kousik, D. N., & Johri, P. (2020). Improved Energy Efficient Wireless Sensor Networks Using Multicast Particle Swarm Optimization. Available at SSRN 3555764.
- [31]. Khadidos, A., Khadidos, A. O., Kannan, S., Natarajan, Y., Mohanty, S. N., & Tsaramiris, G. (2020). Analysis of COVID-19 Infections on a CT Image Using DeepSense Model. *Frontiers in Public Health*, 8, 20.
- [32]. Yuvaraj, N., Srihari, K., Chandragandhi, S., Raja, R. A., Dhiman, G., & Kaur, A. (2021). Analysis of protein-ligand interactions of SARS-Cov-2 against selective drug using deep neural networks. *Big Data Mining and Analytics*, 4(2), 76-83.
- [33]. Yuvaraj, N., Karthikeyan, T., Praghash, K., & Reddy, K. H. (2021). Binary flower pollination (BFP) approach to handle the dynamic networking conditions to deliver uninterrupted connectivity. *Wireless Personal Communications*, 121(4), 3383-3402.
- [34]. Maheshwari, V., Mahmood, M. R., Sravanthi, S., Arivazhagan, N., ParimalaGandhi, A., Srihari, K., ... & Sundramurthy, V. P. (2021). Nanotechnology-Based Sensitive Biosensors for COVID-19 Prediction Using Fuzzy Logic Control. *Journal of Nanomaterials*, 2021.
- [35]. Natarajan, Y., Kannan, S., Selvaraj, C., & Mohanty, S. N. (2021). FORECASTING ENERGY GENERATION IN LARGE PHOTOVOLTAIC PLANTS USING RADIAL BELIEF NEURAL NETWORK. *Sustainable Computing: Informatics and Systems*, 100578.
- [36]. Natarajan, Y., Raja, R. A., Kousik, N. V., & Saravanan, M. (2021). A review of various reversible embedding mechanisms. *International Journal of Intelligence and Sustainable Computing*, 1(3), 233-266.
- [37]. Kousik, N. V., Sivaram, M., Yuvaraj, N., & Mahaveerakannan, R. (2021). Improved Density-Based Learning to Cluster for User Web Log in Data Mining. In *Inventive Computation and Information Technologies* (pp. 813-830). Springer, Singapore.
- [38]. Yuvaraj, N., Praghash, K., & Karthikeyan, T. (2021). Data Privacy Preservation and Trade-off Balance Between Privacy and Utility Using Deep Adaptive Clustering and Elliptic Curve Digital Signature Algorithm. *Wireless Personal Communications*, 1-16.
- [39]. Arivazhagan, N., Somasundaram, K., Vijendra Babu, D., Gomathy Nayagam, M., Bommi, R. M., Mohammad, G. B., ... & Prabhu Sundramurthy, V. (2022). Cloud-Internet of Health Things (IOHT) Task Scheduling Using Hybrid Moth Flame Optimization with Deep Neural Network Algorithm for E Healthcare Systems. *Scientific Programming*, 2022.
- [40]. Gobinathan, B., Mukunthan, M. A., Surendran, S., Somasundaram, K., Moeed, S. A., Niranjana, P., ... & Sundramurthy, V. P. (2021). A Novel Method to Solve Real Time Security Issues in Software Industry Using Advanced Cryptographic Techniques. *Scientific Programming*, 2021.
- [41]. Yuvaraj, N., Raja, R. A., Karthikeyan, T., & Praghash, K. (2021). Improved Authentication in Secured Multicast Wireless Sensor Network (MWSN) Using Opposition Frog Leaping Algorithm to Resist Man-in-Middle Attack. *Wireless Personal Communications*, 1-17.
- [42]. Yuvaraj, N., Praghash, K., Raja, R. A., & Karthikeyan, T. (2021). An Investigation of Garbage Disposal Electric Vehicles (GDEVs) Integrated with Deep Neural Networking (DNN) and Intelligent Transportation System (ITS) in Smart City Management System (SCMS). *Wireless Personal Communications*, 1-20.
- [43]. Kumar, A. S., Jule, L. T., Ramaswamy, K., Sountharajan, S., Yuvaraj, N., & Gandomi, A. H. (2021). Analysis of false data detection rate in generative adversarial networks using recurrent neural network. In *Generative Adversarial Networks for Image-to-Image Translation* (pp. 289-312). Academic Press.
- [44]. Sangeetha, S. B., Sabitha, R., Dhiyanesh, B., Kiruthiga, G., Yuvaraj, N., & Raja, R. A. (2022). Resource Management Framework Using Deep Neural Networks in Multi-Cloud Environment. In *Operationalizing Multi-Cloud Environments* (pp. 89-104). Springer, Cham.
- [45]. Gowrishankar, J., Kumar, P. S., Narmadha, T., & Yuvaraj, N. (2020). A Trust Based Protocol For Manets In Iot Environment., *International Journal of Advanced Science and Technology* 29 (7), 2770-2775.
- [46]. Karthick, S., Yuvaraj, N., Rajakumari, P. A., & Raja, R. A. (2021). Ensemble Similarity Clustering Frame work for Categorical Dataset Clustering Using Swarm Intelligence. In *Intelligent Computing and Applications* (pp. 549-557). Springer, Singapore.
- [47]. Yuvaraj, N., Raja, R. A., & Kousik, N. V. (2021). Privacy Preservation Between Privacy and Utility Using ECC-based PSO Algorithm. In *Intelligent Computing and Applications* (pp. 567-573). Springer, Singapore.
- [48]. Daniel, A., Kannan, B. B., Yuvaraj, N., & Kousik, N. V. (2021). Predicting Energy Demands Constructed on Ensemble of Classifiers. In *Intelligent Computing and Applications* (pp. 575-583). Springer, Singapore.
- [49]. Yuvaraj, N., Raja, R. A., Kousik, N. V., Johri, P., & Diván, M. J. (2020). Analysis on the prediction of central line-associated bloodstream infections (CLABSI) using deep neural network classification. In *Computational Intelligence and Its Applications in Healthcare* (pp. 229-244). Academic Press.

- [50]. Sangeetha, S. B., Blessing, N. W., Yuvaraj, N., & Sneha, J. A. (2020). Improving the training pattern in back-propagation neural networks using holt-winters' seasonal method and gradient boosting model. In *Applications of Machine Learning* (pp. 189-198). Springer, Singapore.
- [51]. Ramachandran, M., Sahas Bansal, and Pramod Raichurkar. "Scrutiny of jute fiber poly-lactic acid (PLA) resin reinforced polymeric composite." *Journal of the Textile Association* 76, no. 6 (2016): 372-375.
- [52]. Natarajan, Y., Raja, R. A., Kousik, D. N., & Johri, P. (2020). Improved Energy Efficient Wireless Sensor Networks Using Multicast Particle Swarm Optimization. Available at SSRN 3555764.
- [53]. Yuvaraj, N., Kousik, N. V., Jayasri, S., Daniel, A., & Rajakumar, P. (2019). A survey on various load balancing algorithm to improve the task scheduling in cloud computing environment. *J Adv Res Dyn Control Syst*, 11(08), 2397-2406.
- [54]. Yuvaraj, N., Chang, V., Gobinathan, B., Pinagapani, A., Kannan, S., Dhiman, G., & Rajan, A. R. (2021). Automatic detection of cyberbullying using multi-feature based artificial intelligence with deep decision tree classification. *Computers & Electrical Engineering*, 92, 107186.
- [55]. Natarajan, Y., Kannan, S., & Dhiman, G. (2021). Task scheduling in cloud. *Recent Advances in Computer Science and Communications*, 13, 1-6.