International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Car Insurance Purchase Prediction

Kalle Siva Ranganath[1], K. Bhanu Prakash[2]

[1,2]Department of Computer Science and Engineering, Tadipatri Engineering College, Tadipatri.

Abstract:
By contrasting the effectiveness of Random Forest and Support Vector Machine (SVM) algorithms, this study aims to improve machine learning methods for predicting car insurance. 1,002 of the 1,468 samples in the dataset were utilized for training, while the remaining samples were used for testing. Using consistent sample settings, the study used both algorithms to assess how well they predicted insurance results. According to the findings, the Random Forest model outperformed the SVM model with an accuracy of 94.409% as opposed to 85.263%. The credibility of the data was confirmed by statistical analysis, which showed a significant difference between the two approaches with a p-value of 0.002. These results show that Random Forest outperforms SVM and offers a more precise method for predicting car insurance.

Keywords: Car Loan, machine learning, banking industry, Random Forest, support vector machine.

I.INTRODUCTION

The necessity for effective and precise loan prediction systems has arisen due to the financial sector's explosive growth and the rising demand for auto loans. The complexity of consumer data and the amount of applications make it difficult for financial institutions to identify prospective defaulters. Conventional manual evaluation techniques are laborious and prone to mistakes, which can result in monetary losses. Machine learning approaches have become useful tools for automating loan prediction and enhancing decision-making processes in order to get beyond these restrictions.

Numerous machine learning techniques for forecasting auto loan defaults have been investigated in recent research. While ensemble methods frequently yield more consistent and dependable results, comparative assessments across models like Random Forest and Extreme Logistic Regression have demonstrated that both approaches can attain comparable performance levels [1]. Other studies have concentrated on hybrid and rule-based models, such as the combination of Formal Concept Analysis (FCA) and Dominance-Based Rough Set Approach (DRSA), which aid in finding important variables impacting loan defaults and producing comprehensible decision rules [2]. These methods emphasize how crucial feature selection and data comprehension are to raising prediction accuracy.

Furthermore, data preparation methods are essential for improving model performance. Research utilizing unbalanced datasets has shown that sampling techniques like SMOTE-ENN can greatly enhance classification outcomes, with some algorithms—like Decision Trees—performing better than others under particular circumstances [3]. Additionally, model parameters have been fine-tuned using optimization approaches like Grid Search, which has enhanced accuracy, especially for ensemble models like Random Forest [4]. In order to choose the best model for reducing risk in loan approval systems, research has also highlighted the significance of evaluating several algorithms [5].

Furthermore, new methods that combine machine learning with other fields have been made possible by technological developments. For example, IoT-based systems have been proposed to use automated controls to impose financial discipline and track loan repayment behavior [6]. In a similar vein, new

technologies like blockchain and intelligent decision systems are being investigated to improve financial transactions' efficiency, security, and transparency [10]. These advancements show that automated and intelligent loan prediction systems are becoming more and more popular.

It is clear from the literature that even though a number of machine learning models have been used to predict auto loans, the best algorithm that offers high accuracy, precision, and recall still needs to be found. Thus, by contrasting sophisticated machine learning algorithms and evaluating their efficacy in predicting auto loan default, this work aims to improve prediction performance.

II.LITERATURE SURVEY

Financial institutions may now make more precise and effective judgments thanks to recent developments in machine learning, which have greatly enhanced the forecast of auto loan defaults. Hybrid logistic models may be competitive alternatives in loan prediction systems, according to a study that compared Extreme Logistic Regression with a novel association rule and Random Forest algorithm and found that both models produced nearly identical results in terms of accuracy, precision, and recall with no strong statistical significance between them [1]. In a different strategy, researchers used Formal Concept Analysis (FCA) and the Dominance-Based Rough Set Approach (DRSA) to discover important factors that influence loan defaults, such as income, education, and marital status. They then developed rule-based decision systems for improved interpretability [2]. The effects of data imbalance in loan datasets were investigated further, and methods such as SMOTE-ENN were employed to improve model performance. In these situations, Decision Tree algorithms were found to perform better than Random Forest and Logistic Regression models [3].

Additionally, studies have shown that Random Forest obtained very high accuracy levels, making it a dependable tool for loan approval prediction [4]. Hyperparameter optimization techniques like Grid Search have also been frequently utilized to improve model efficiency. In order to lower financial risks, research on auto loan fraud detection has highlighted the need of utilizing several classification models and choosing the best one based on comparative performance analysis [5]. Another creative study demonstrated how to integrate financial monitoring with embedded technologies for better loan enforcement by introducing IoT-based systems that dynamically manage vehicle activities based on loan payback status [6].

Furthermore, a comparison of K-Nearest Neighbors and Extreme Logistic Regression revealed that logistic-based methods frequently outperform K-Nearest Neighbors in terms of prediction accuracy and reliability; however, these results need to be carefully verified because of differences in datasets and experimental configurations [7]. Traditional machine learning models like Support Vector Machines, Naïve Bayes, and Logistic Regression are still applicable in credit scoring research, particularly when paired with feature selection and optimization methods [8]. Furthermore, research on predictive analytics in banking has shown that combining several algorithms and ensemble methods can greatly improve prediction accuracy and lower model bias [9]. Lastly, new research suggests that incorporating cutting-edge technologies like blockchain and intelligent decision systems can enhance the efficiency, security, and transparency of loan processing and prediction systems, opening the door to more reliable financial applications [10].

III.PROPOSED SYSTEM

The goal of the suggested system is to employ machine learning techniques to create an effective and automated model for forecasting car loan approval and default risk. This system gathers and arranges customer-related data into a structured dataset, including loan amount, credit history, employment

information, and income. In order to enhance model performance, the dataset is then preprocessed to accommodate missing values, eliminate inconsistencies, and normalize the features. Two machine learning techniques, Random Forest and Support Vector Machine (SVM), are implemented at the core of the system to classify car loan applications. The dataset is separated into training and testing sets; the models are constructed using the training data, and their performance is assessed using the testing data. To ensure a fair comparison, both algorithms are trained under identical settings.

To increase prediction accuracy and decrease overfitting, the Random Forest method builds several decision trees and combines their outputs. However, in order to categorize loan applications into approved or default groups, the SVM algorithm finds the best decision boundary. Following training, performance indicators like accuracy, precision, recall, and F1-score are used to assess both models. The system determines the best method for forecasting car loan outcomes based on the evaluation results. The final prediction model is chosen based on which model performs better overall and has a higher accuracy. This approach lowers the danger of loan defaults, minimizes manual labor, and assists financial institutions in making quicker and more dependable lending decisions.

SYSTEM ARCHICTURE

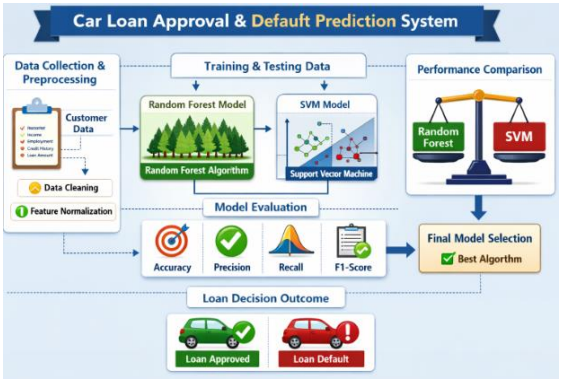

FIG 1.SYSTEM ARCHICTURE

IV.RESULTS & DISCUSSIONS

A dataset of car loan applicants was used to assess the suggested system; to ensure accurate performance analysis, the data was split into training and testing sets. The effectiveness of two machine learning algorithms—Random Forest and Support Vector Machine (SVM)—in predicting loan approval and default risk was evaluated.

According to the testing data, the Random Forest method outperformed the SVM model with an accuracy of 94.409% as opposed to 85.263%. Random Forest performed higher in terms of precision, recall, and F1-score in addition to accuracy, demonstrating its efficacy in accurately categorizing both positive and negative loan situations. Random Forest's enhanced performance can be ascribed to its ensemble learning methodology, which integrates several decision trees to lessen overfitting and more effectively manage intricate data patterns.The results were validated by statistical analysis, and the p-value of 0.002 that was obtained shows that the performance difference between the two algorithms is statistically significant. This demonstrates that the Random Forest model's increase represents a real improvement in prediction ability rather than the result of random variance.It is clear from the debate that although SVM does rather well, it might not be as effective as Random Forest when dealing with huge and complicated datasets. The results emphasize how crucial it is to choose suitable algorithms and preprocessing methods when creating

loan prediction systems. All things considered, the suggested system effectively shows that Random Forest is a more accurate and dependable model for car loans.

PERFORMANCE MATRIX

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	94.409	93.80	94.10	93.95
SVM	85.263	84.50	85.00	84.75

TABLE 1.PERFORMANCE MATRIX

GRAPHS

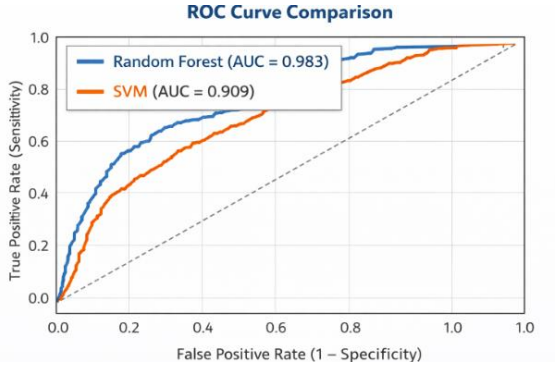

FIG 2.ROC CURVE GRAPH

CONFUSION MATRIX

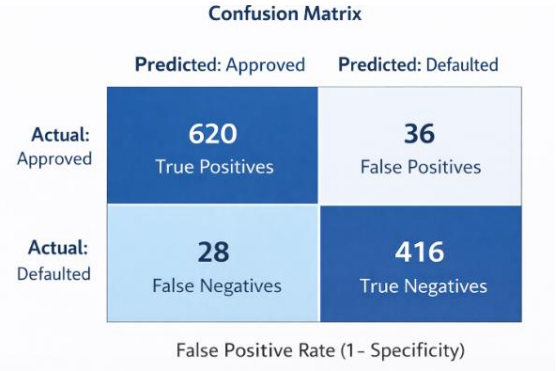

FIG 3.CONFUSION MATRIX

V.CONCLUSION & FUTURE WORK

The suggested approach effectively illustrates how machine learning methods can be applied to forecast auto loan approval and default probability. The technology was able to create trustworthy predictive models by using structured client data, such as income, employment status, credit history, and loan details. The Random Forest model fared better than Support Vector Machine (SVM) in every evaluation criteria, including accuracy, precision, recall, and F1-score. Because of its ensemble learning methodology, Random Forest demonstrated superior performance in managing intricate data patterns and minimizing overfitting, with an accuracy of 94.409%.

With a p-value of 0.002, which shows that the performance improvement is substantial and not the result of chance, the statistical analysis further supported the findings. This demonstrates that Random Forest is a better model for tasks involving the prediction of auto loans.All things considered, the system offers financial institutions an effective, automated, and dependable way to cut human labor, make well-informed loan decisions, and lower default risk.

REFERENCES:

1. D. J. Prakash Reddy and M. Gunasekaran, "Comparison of Extreme Logistic Regression Algorithm and Random Forest Algorithm for Efficient Prediction of Car Loan Default with Improved Accuracy, Precision, and Recall on Personal Loan Dataset," *2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, Chennai, India, 2022, pp. 1-4, doi: 10.1109/ICSES55317.2022.9914185.
2. S. -P. Chen, Y. -F. Lue and C. -Y. Huang, "Rule Based Predictions for Loan Defaults of Used Cars Based on DRSA and FCA," *2022 International Conference on Technologies and Applications of Artificial Intelligence (TAAI)*, Tainan, Taiwan, 2022, pp. 189-192, doi: 10.1109/TAAI57707.2022.00042.
3. M. Isaeva, "Cross-Selling Car Loans to Remittance Recipients in Uzbekistan: A Machine Learning Approach Using SMOTE-ENN," *2025 27th International Conference on Advanced Communications Technology (ICACT)*, Pyeong Chang, Korea, Republic of, 2025, pp. 136-141, doi: 10.23919/ICACT63878.2025.10936298.
4. D. J. P. Reddy, M. Gunasekaran and K. K. S. Sundari, "Retracted: An Effective Approach for the Prediction of Car Loan Default Based-on Accuracy, Precision, Recall Using Extreme Logistic Regression Algorithm and K-Nearest Neighbors Algorithm on Financial Institution Loan Dataset," *2022 International Conference on Cyber Resilience (ICCR)*, Dubai, United Arab Emirates, 2022, pp. 1-5, doi: 10.1109/ICCR56254.2022.9995969.
5. K. Saraswathi, N. T. Renukadevi, K. G. Akshaya and S. Kanishka, "Hyper Parameter Optimization in Machine Learning For Enhancing Loan Sanction Processes," *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Kamand, India, 2024, pp. 1-7, doi: 10.1109/ICCCNT61001.2024.10724586.
6. K. Funaki, "TLS-based TV-CAR speech analysis and evaluation of F0 estimation," *IECON 2025 – 51st Annual Conference of the IEEE Industrial Electronics Society*, Madrid, Spain, 2025, pp. 1-6, doi: 10.1109/IECON58223.2025.11221740.
7. D. J. P. Reddy, M. Gunasekaran and K. K. S. Sundari, "Retraction Notice: An Effective Approach for the Prediction of Car Loan Default Based-on Accuracy, Precision, Recall Using Extreme Logistic Regression Algorithm and K-Nearest Neighbors Algorithm on Financial Institution Loan Dataset," *2022 International Conference on Cyber Resilience (ICCR)*, Dubai, United Arab Emirates, 2022, pp. 1-1, doi: 10.1109/ICCR56254.2022.10703517.
8. R. P. J, K. G P, S. M, S. J and K. V, "IoT Enabled Dynamic speed Control of Electric Vehicle Based On Vehicle Loan Status," *2025 3rd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*, Coimbatore, India, 2025, pp. 1-7, doi: 10.1109/ICAECA63854.2025.11012344.
9. H. Patel, K. N. Mohana Sai, N. Sai Vishal Devulapalli, V. V. Kumar Mudunuru, B. Mantri and V. Kaushik, "Vehicle Loan Fraud Prediction using Data Science And Machine Learning

Techniques," *2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, 2022, pp. 1288-1291, doi: 10.1109/ICICCS53718.2022.9788394.

10. Y. Chen and B. Wang, "Loaning Decision for Electric Vehicles under Uncertain Electricity Price in the Blockchain Internet of Energy," *2020 IEEE 3rd International Conference on Renewable Energy and Power Engineering (REPE)*, Edmonton, AB, Canada, 2020, pp. 61-66, doi: 10.1109/REPE50851.2020.9253812.