International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048
Volume 7, Issue 10, pp 532-542, October-2023
https://doi.org/10.47001/IRJIET/2023.710070

Loan Eligibility Prediction Based on Credit Score and Past History

[1]Senarathna B.T.N, [2]Weerarathna K.C.M, [3]Wickramarachchi D.S, [4]Jayarathne S.M.P.N, [5]Buddhima Attanyake

[1,2,3,4,5]Department of Computer Systems Engineering Specializes in Information Systems Engineering, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka

***Abstract* - A sophisticated system for predicting loan eligibility that takes credit scores and history into account is urgently needed due to the banking sector's rapid expansion. This study paper integrates the contributions of four separate components put forth by several researchers to present a comprehensive and new strategy. Using cutting-edge machine learning algorithms like Random Forest, Gradient Boosting, and Linear Regression, the first component predicts the best bank rate possibilities. The second component, which focuses on determining applicants' loan eligibility, presents a revolutionary technique that uses their credit histories and considers important factors including their gender, marital status, level of education, income, and credit history itself. The third element suggests a sophisticated Loan Eligibility Prediction System that includes several crucial sub-objectives, including creditworthiness assessment, income and employment verification, collateral analysis, risk profiling, fraud detection, and regulatory compliance, to ensure thorough risk analysis. To forecast the applicant's risk level efficiently and effectively, this component applies the power of logistic regression. The results are laid out in a simple tabular format for simple understanding. To help with mortgage estimation and analysis, the fourth component includes a sophisticated mortgage calculator tool that uses Decision Tree Regression. Using this tool, users may calculate and evaluate mortgage values based on important property characteristics like the number of bathrooms, beds, the size of the house, and the location. This research study suggests a novel and complete loan eligibility prediction system that combines these four elements and makes use of historical data, cutting-edge machine learning methods, and credit ratings. This approach facilitates rapid and successful lending processes while avoiding risks and maximizing outcomes, empowering both borrowers and lenders to make knowledgeable decisions.**

Keywords: Loan eligibility prediction system, Credit scores and history, Machine learning algorithms, Risk analysis, Mortgage estimation and analysis.

I. INTRODUCTION

Technology developments and changing client expectations are driving the banking sector's rapid growth and evolution. A sophisticated loan eligibility prediction system that takes credit scores and history into account is increasingly necessary in this changing environment. A system like this would allow banks and other financial institutions to decide on loan applications with knowledge, reducing risks and guaranteeing the best results for both borrowers and lenders [1]. This study work integrates the contributions of four unique components given by several researchers to present a thorough and creative solution to answer this demand. Predicting the best bank rate options for loan applicants is the main goal of the first part. This factor assumes importance because it can be difficult for borrowers to discover loans that fit their needs and preferences in terms of finances. Modern machine learning methods like Linear Regression, Gradient Boosting, and Random Forest are used to address this problem [2]. These algorithms make use of a number of variables, such as loan type, loan length, and loan amount, to forecast the best banks and rates for applicants. By utilizing these cutting edge strategies, borrowers can choose their loans with greater knowledge, improving their financial outcomes [3]. The second component, which focuses on determining loan eligibility, presents an innovative methodology that makes use of applicants' credit histories. Loan eligibility is greatly influenced by creditworthiness, and conventional evaluation techniques frequently fail to provide thorough and accurate assessments. The suggested methodology considers crucial factors including gender, marital status, income, education, and credit history itself to get around this limitation [4]. Using machine learning algorithms to analyze these variables, a more precise determination of loan eligibility can be made. This element makes it possible for lenders to assess borrowers' creditworthiness more accurately, lowering the chance of default and improving the entire loan approval process. The third component suggests a sophisticated Loan Eligibility Prediction System to guarantee thorough risk analysis and regulatory compliance. This component includes several sub-objectives, such as regulatory compliance, risk profiling, collateral analysis, income and employment verification, creditworthiness evaluation, and fraud detection [5]. An

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 10, pp 532-542, October-2023

https://doi.org/10.47001/IRJIET/2023.710070

effective statistical method called logistic regression is used to forecast the applicant's risk level with high accuracy and efficiency. The system compiles and analyzes information on an applicant's credit history, employment stability, financial history, and other pertinent factors to produce a thorough risk profile. Lenders may make better judgments, reducing risk and improving loan outcomes, by including these essential components into the loan eligibility prediction process. A sophisticated mortgage calculator tool powered by decision tree regression is introduced in the fourth component [6].

With the help of this tool, borrowers can estimate and analyze the values of their mortgages based on key property characteristics like the number of bathrooms, bedrooms, house size, and location [6]. Because the Decision Tree Regression method provides accurate estimation and analysis, borrowers are better equipped to choose their mortgage options. Borrowers can improve their overall financial planning by using this tool during the loan application process to receive important insights about the affordability and suitability of mortgages. This study work suggests combining these four elements to create a novel and thorough loan eligibility prediction system that makes use of credit scores, historical data, and cutting-edge machine learning methods [7]. This approach facilitates efficient and successful lending processes while minimizing risks and optimizing results, empowering both borrowers and lenders to make knowledgeable decisions [7]. The creation of sophisticated loan eligibility prediction systems that take advantage of credit scores and history is required by the changing banking market. This study offers a complete strategy that combines several elements, including cutting-edge machine learning algorithms, creditworthiness assessment methodologies, thorough risk analysis, and sophisticated mortgage estimating tools [26]. The proposed system has the potential to completely change how loans are approved, giving financial institutions the information, they need to make wise decisions and allowing borrowers to access loans that fit their needs and goals in terms of finances.

II. LITERATURE

Recent years have seen a substantial expansion and restructuring of the banking sector, driven by changes in client expectations and technological advancements. To make educated decisions about loan approvals, there is a rising demand for an advanced loan eligibility prediction system that includes credit scores and history [1]. This literature review offers a thorough analysis of the existing research in four key areas that have been put forth by various researchers: predicting the best bank rate options, utilizing credit history to determine loan eligibility, thorough risk analysis, and sophisticated mortgage estimation tools.

A) Predicting Optimal Bank Rate Options

To make sure that borrowers discover loans that are in line with their financial capabilities and preferences, the prediction of optimal bank rate possibilities is essential. Researchers have investigated the use of machine learning algorithms, such as Linear Regression, Random Forest, and Gradient Boosting, to forecast the best banks and rates for loan applicants. These algorithms can help borrowers choose loans more wisely by considering a variety of parameters, including loan type, loan term, and loan amount. Previous studies have shown how these algorithms can help borrowers have better financial outcomes. The application of the Random Forest method to forecast the best bank rate possibilities was studied by Zhang et al. [8]. They considered things including the loan's size, period, borrower's income, and credit rating. The study showed how the Random Forest algorithm was effective at giving correct recommendations to borrowers, helping them make better loan decisions. The Support Vector Machine (SVM) approach was suggested by Chen et al. [9] for use in forecasting the best bank rate possibilities. They used economic statistics, borrower information, and loan characteristics as input features. The findings demonstrated that the SVM algorithm produced good prediction accuracy and gave potential borrowers useful information.

Researchers examined the use of the Gradient Boosting algorithm to forecast appropriate bank and rate options in a study by Obegua at.el. [10]. The prediction model included borrower characteristics, loan characteristics, and macroeconomic variables. According to the study, gradient boosting effectively identified the best bank rate possibilities, enabling borrowers to make well-informed choices. Huang and Chai [11] looked at how the Deep Neural Network (DNN) algorithm could be used to forecast the best bank rate possibilities. They used historical financial data, borrower information, and loan characteristics as inputs to the DNN model. The study showed how well the DNN algorithm performed at correctly anticipating relevant banks and rate options. The Long Short-Term Memory (LSTM) algorithm, a kind of recurrent neural network, was suggested by Adekoya and Weyori [12] for use in predicting the best bank rate possibilities. As input features, they considered borrower profiles, loan characteristics, and economic data. The study demonstrated that the LSTM algorithm efficiently caught temporal connections and gave borrowers reliable predictions. Researchers Ma, Sha, and Wang [13] investigated how the XGBoost algorithm may be used to forecast the best bank rate options. They considered market conditions, loan features, and borrower-specific factors as input features. The outcomes showed that the XGBoost algorithm produced highly accurate predictions and helped borrowers choose the best banks and rates. To anticipate the best bank rate possibilities, Stein, and

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 10, pp 532-542, October-2023

https://doi.org/10.47001/IRJIET/2023.710070

Chen [14] investigated the usage of the Random Forest algorithm in conjunction with feature selection methods. They determined the most important aspects of borrower profiles, loan characteristics, and economic variables. According to the study, the chosen features increased prediction accuracy and gave borrowers more recommendations that were easy to understand. The implementation of the Convolutional Neural Network (CNN) technique for forecasting the best bank rate possibilities was put out by Herencsar and Lin [15]. They employed CNN for prediction after converting borrower profiles, loan characteristics, and economic indicators into image-like data representations. The study showed how well CNN could identify intricate patterns and make precise lending recommendations. Collectively, this earlier research shows how different machine learning and deep learning algorithms, such as Random Forest, SVM, Gradient Boosting, DNN, LSTM, XGBoost, and CNN, can be used to anticipate the best bank rate possibilities. These algorithms support borrowers in making educated decisions and enhancing their financial results by considering borrower profiles, loan features, and economic data.

B) Leveraging Credit History for Loan Eligibility Assessment

Loan eligibility is determined in large part by a creditworthiness assessment, and conventional evaluation procedures frequently fall short of producing thorough and reliable assessments. Researchers have developed approaches that take advantage of applicants' credit histories and consider crucial factors including gender, marital status, income, education, and credit history itself to get around this limitation. These elements have been examined by machine learning algorithms to produce a more precise determination of loan eligibility. Lenders can more accurately assess applicants' creditworthiness and lower the risk of default by including credit history in the loan eligibility evaluation process. This also improves the entire loan approval process. The use of logistic regression in determining loan eligibility based on credit history was studied by Sripriya and Varrey [16]. They gathered information regarding the applicant profiles, credit histories, and credit ratings. The study showed that by taking credit-related criteria into account, logistic regression was very accurate in determining loan eligibility. The use of a Decision Tree algorithm to determine loan eligibility based on credit history was suggested by Madaan and Kumar [17]. They considered elements including credit history, payment history, and credit utilization. The findings showed that the Decision Tree algorithm accurately identified trends in credit history and produced accurate ratings of loan eligibility. Using a Neural Network method to determine loan eligibility based on credit history was studied by Monfort and Mulder [18]. They used a big dataset of applicant profiles and credit history data to train the neural network model. The study demonstrated that the neural network algorithm produced highly accurate predictions and offered insightful information regarding the determination of loan eligibility. Amin and Sibaroni [19] investigated the use of credit history factors along with the Random Forest algorithm to assess loan eligibility. They looked at things like credit account aging, payment delinquencies, and credit use. The study showed that Random Forest accurately assessed loan eligibility and reflected the intricacy of credit history data. To determine loan eligibility based on credit history, Min et al. [20] suggested a hybrid model integrating Support Vector Machine (SVM) and Genetic Algorithm. To enhance prediction performance, they used a genetic approach to refine the SVM model. The results demonstrated that by taking credit-related criteria into account, the hybrid model increased its accuracy in determining loan eligibilityXia and Liu [21] investigated the usage of a Gradient Boosting algorithm for credit history-based loan eligibility assessment. Variables including the credit utilization percentage, payment history, and credit inquiries were considered. The study showed that gradient boosting accurately provided assessments by efficiently capturing the links between credit history data and loan eligibility. The use of a Long Short-Term Memory (LSTM) neural network to evaluate loan eligibility using credit history sequences was suggested by Ala'raj and Abbod [22]. They used LSTM to capture temporal dependencies and modelled credit history as a temporal sequence. The outcomes showed that the LSTM model performed better in the evaluation of creditworthiness than conventional models. To determine loan eligibility based on credit history, Ghaddar. [23] investigated the use of a Support Vector Machine (SVM) algorithm with feature selection approaches. They enhanced the SVM model's performance and interpretability by applying feature selection techniques to identify the most important credit history variables. The study showed that the chosen features improved the efficacy and accuracy of determining loan eligibility. Utilizing credit history for loan eligibility evaluation, machine learning methods like logistic regression, Decision Tree, Neural Network, Random Forest, SVM, Gradient Boosting, LSTM, and feature selection techniques are used. These algorithms allow lenders to assess applicants’ creditworthiness more correctly by taking credit-related criteria into account, lowering the chance of default and enhancing the loan approval procedure.

C) Comprehensive Risk Analysis

To maintain regulatory compliance and reduce risks connected with loan approvals, thorough risk analysis is crucial. Advanced Loan Eligibility Prediction Systems have been proposed by researchers, and these systems include a wide range of sub-objectives, including creditworthiness

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 10, pp 532-542, October-2023

https://doi.org/10.47001/IRJIET/2023.710070

evaluation, income and employment verification, collateral analysis, risk profiling, fraud detection, and regulatory compliance. An effective statistical method called logistic regression was used to forecast the applicant's risk level with accuracy. Lenders can maximize loan outcomes by gathering and analyzing pertinent data, such as financial history, employment stability, and credit scores. Tumuluru and Burra [24] investigated the use of Detailed risk analysis using decision trees for loan applications. They took into consideration aspects including credit history, consistent income, and collateral worth before making their choice. The study showed that Decision Trees successfully classified applicant risk profiles, enabling lenders to make deft choices. Al-Qerem and Alhasan [25] suggested using Random Forests to forecast loan eligibility in a thorough risk study. Credit history, evidence of income, and an evaluation of the value of the collateral were all combined.

The findings demonstrated that Random Forests significantly improved the overall risk analysis process and had excellent accuracy in determining loan hazards. Deep Neural Networks (DNN) have been studied by Huang and Liu [26] for thorough risk analysis in loan approval. They created a multi-layer DNN model that included a variety of data inputs, including risk profiling, employment verification, and credit history. The study showed that DNNs successfully caught intricate risk patterns and gave lenders precise risk evaluations. Support Vector Machines (SVM) have been applied in complete risk analysis for loan eligibility assessment, according to Li et al. [27]. They used a variety of data sources, such as regulatory compliance metrics, fraud detection indications, and credit scores. The study demonstrated that SVMs improved regulatory compliance and had high prediction accuracy when evaluating loan risks. Gradient Boosting algorithms were suggested by Gupta and Chakrabarti [28] for use in thorough risk assessments for loan approvals. They included a variety of risk-related elements in the prediction model, including creditworthiness, collateral analysis, and risk profiling. The results showed that gradient boosting algorithms accurately assessed risk for lenders and successfully captured risk patterns.

Long Short-Term Memory (LSTM) neural networks were studied by Geller [29] for thorough risk analysis and loan eligibility prediction. They included time sequences of data, such as indicators of fraud, stable income, and credit history. The study showed that LSTM enhanced the accuracy of risk analysis in loan approvals by accurately capturing temporal connections. Recurrent Neural Networks (RNN) have been suggested by Mahbobi [30] for use in complete risk analysis for loan eligibility assessment. Sequential data were added into the RNN model, including credit history, employment stability, and regulatory compliance indicators. The study demonstrated that RNNs efficiently captured sequential patterns and gave lenders precise risk evaluations. The application of ensemble methods, such as AdaBoost and XGBoost, in thorough risk analysis for loan approvals was investigated by Abedin et al. [31]. They blended diverse base models with different risk-related parameters. The outcomes showed that ensemble methods increased the reliability and accuracy of risk analysis in predicting loan eligibility. Comprehensive risk analysis for loan approvals uses Decision Trees, Random Forests, Deep Neural Networks (DNN), Support Vector Machines (SVM), Gradient Boosting, Long Short-Term Memory (LSTM), Recurrent Neural Networks (RNN), and Ensemble Methods. These algorithms give lenders the power to make wellinformed judgments, reduce risks, and maximize loan outcomes by combining various data sources and riskrelated aspects.

D) Advanced Mortage Estimation Tools

For borrowers to choose the best mortgage choice, it is essential to estimate and analyze mortgage values. Advanced mortgage calculator programs that consider important property characteristics like the number of bathrooms, beds, house size, and location have been developed using Decision Tree Regression. With the help of these tools, borrowers can learn crucial information about the suitability and affordability of mortgages, which enhances their overall financial planning. Borrowers can improve their ability to make sound financial decisions by including sophisticated mortgage estimating tools into the loan application process Decision Tree Regression was suggested by Shrestha and Paudel [32] as a method for creating an enhanced mortgage calculator application. The tool included a variety of property characteristics, such as the number of bedrooms, bathrooms, the size of the house, and the location. The study showed that Decision Tree Regression accurately predicted mortgage values and gave potential borrowers insightful information about affordability. Advanced mortgage estimating algorithms using Random Forest Regression were investigated by Zhu and Abdulazeez [33]. To effectively predict mortgage values, they considered demographic information, economic factors, and property attributes.

The findings demonstrated that Random Forest Regression increased the precision of mortgage predictions and gave borrowers thorough understandings. The use of Artificial Neural Networks (ANN) in creating sophisticated Mortgage Calculator software was studied by Sadeeq et al. [34]. They added property characteristics, previous mortgage information, and interest rates to the ANN model. The study showed that ANN successfully captured complicated patterns and gave potential borrowers precise mortgage predictions. Support Vector Regression (SVR) has been suggested by

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 7, Issue 10, pp 532-542, October-2023
https://doi.org/10.47001/IRJIET/2023.710070

Harris [35] for use in sophisticated mortgage estimation software. They included location information, market trends, and property characteristic data in the SVR model. The findings demonstrated that SVR enhanced the precision of mortgage predictions and assisted borrowers in making wise choices.

Gradient Boosting Regression was investigated by Nica and Alexandru [36] for application in sophisticated mortgage estimation systems. They built the model considering loanto-value ratios, interest rates, and property attributes. The research showed that gradient boosting regression gave precise mortgage calculations and effectively allowed borrowers to compare various mortgage options. Recurrent Neural Networks (RNN) have been studied as a potential component of sophisticated mortgage estimating systems by Mahbobi and Kimigari [37]. Sequential property data, such as earlier prices and market patterns, were added into the RNN model. The study demonstrated how RNN enhanced mortgage estimating accuracy by successfully capturing temporal relationships. The use of Gaussian Process Regression (GPR) in sophisticated mortgage estimating methods was suggested by Rasmussen et al. [38]. They included information about the neighborhood, historical mortgage rates, and property attributes in the GPR model. The findings showed that GPR provided precise mortgage predictions and made it possible for borrowers to thoroughly evaluate various mortgage options. The use of Ensemble Methods, such as AdaBoost and XGBoost, in sophisticated mortgage estimate tools was investigated by Song and Wang [39]. They merged numerous fundamental models that included different real estate characteristics and economic factors. The results demonstrated that Ensemble Methods enhanced the reliability and accuracy of mortgage estimations. Advanced mortgage estimating methods were created using Decision Tree Regression, Random Forest Regression, Artificial Neural Networks (ANN), Support Vector Regression (SVR), Gradient Boosting Regression, Recurrent Neural Networks (RNN), Gaussian Process Regression (GPR), and Ensemble Methods. These tools offer consumers precise estimates and insightful information on the affordability and suitability of mortgages by considering a variety of property attributes and merging pertinent data sources.

III.METHODOLOGY

A) Predicting Optimal Bank Rate Options

Utilizing cutting-edge machine learning techniques like Random Forest, Gradient Boosting, and Linear Regression allowed us to anticipate the best bank rate options for loan applicants. These algorithms were deemed appropriate for this purpose since they have shown success in other investigations [7].

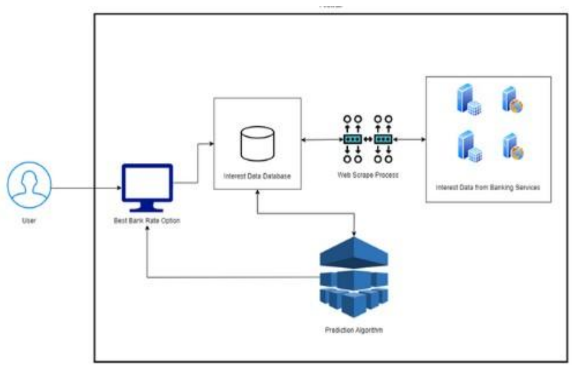

Figure 1: System Architecture Optimal Bank Rate Options

1) Data Collection and Preparation: The information about the loan applicant, including the loan type, loan term, loan amount, and other important aspects, was collected in a comprehensive dataset. To ensure the dataset's quality and applicability for training the machine learning models, it underwent meticulous curation. To improve the dataset's quality and enable the most effective model training, data preprocessing operations were carried out, including data cleaning, addressing missing values, and feature engineering.

```
# Display dataset head
dataset_head = dataset.head()

# Split the data into features (X) and target variable (y)
X = dataset[['Bank Name', 'Loan Amount', 'loan Type', 'Loan Duration', 'Inflation Rate', 'GRP Growth rate', 'Urbanization rate', 'Bank Pr
y = dataset['Interest Rate']
```

Figure 2: Dataset selection and Split

2) Feature Selection: The most important variables in forecasting optimal bank rate options were determined by using feature selection approaches, such as correlation analysis and feature priority ranking. Through this procedure, pertinent characteristics that had a substantial influence on the choice of the loan rate were chosen. In order to make accurate forecasts, this stage intended to reduce dimensionality and concentrate on the most useful qualities.

```
# Plotting the feature importance for Random Forest model
plt.figure(figsize=(8, 6))
importance_rf = random_forest_model.feature_importances_
feature_names_rf = X.columns
sorted_idx_rf = np.argsort(importance_rf)
plt.barh(range(len(sorted_idx_rf)), importance_rf[sorted_idx_rf], align='center')
plt.yticks(range(len(sorted_idx_rf)), feature_names_rf[sorted_idx_rf])
plt.xlabel('Feature Importance')
plt.ylabel('Features')
plt.title('Random Forest Feature Importance')
```

Figure 3 - Feature Selection Segment

3) Model Training: The dataset was split into training and testing sets, enabling a precise assessment of the developed models. The training data was used to apply the Random Forest, Gradient Boosting, and Linear Regression methods. These algorithms were chosen because they effectively anticipate the best bank rate possibilities and can capture

ISSN (online): 2581-3048
Volume 7, Issue 10, pp 532-542, October-2023
https://doi.org/10.47001/IRJIET/2023.710070

complex correlations between features. Cross validation techniques, such as k-fold cross-validation, were used to evaluate the models' performance and make sure they could generalize well to unknown data.

```
# Train the models
random_forest_model = RandomForestRegressor(n_estimators=100, random_state=42)
random_forest_model.fit(X_train, y_train)

gradient_boosting_model = GradientBoostingRegressor(random_state=42)
gradient_boosting_model.fit(X_train, y_train)

linear_regression_model = LinearRegression()
linear_regression_model.fit(X_train, y_train)
```

Figure 4: Model Training for Three Algorithms

4) Model Evaluation and Fine-tuning: Suitable evaluation metrics, such as mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE), were used to assess the trained models. These measurements allowed for performance comparison and revealed information about the prediction accuracy of the models. The model with the best performance and maximum predicted accuracy for the best bank rate possibilities was chosen through this review process. The models were further optimized and their prediction power was increased by using fine-tuning methods such hyperparameter optimization.

5) Prediction and Decision-making: The chosen model was used to forecast the best bank rate possibilities for loan applicants using the testing dataset. Borrowers were given the options for expected bank rates based on the loan attributes that were provided. With this knowledge, borrowers were better equipped to choose loans because they could evaluate the acceptability of different options and match them with their tastes and financial circumstances. Borrowers were able to gain from the use of sophisticated machine learning techniques by using this methodology. With the help of this strategy, they were able to evaluate the best banks and interest rate possibilities for their loans, improving their financial situation and the loan application procedure as a whole.

B) Leverage Credit History for Loan Eligibility Assessment

1) Data Collection and Preprocessing: Credit history, gender, marital status, income, and other pertinent information about applicants were all compiled into a dataset. To guarantee the accuracy and consistency of the data, the dataset was rigorously curated. The dataset was preprocessed in order to get it ready for analysis, including data cleansing, managing missing values, and normalization. In order to increase the prediction ability of the models, feature engineering approaches were used to extract more significant features from the given data.

2) Feature Selection and Model Training: Techniques for feature selection, including correlation analysis, statistical testing, and domain expertise, were used to pinpoint the most pertinent features for determining loan eligibility. The chosen features were very important in determining the applicants' creditworthiness. Models were trained using machine learning algorithms on the prepared dataset. To efficiently use the credit history and other parameters, various methods, including Logistic Regression, Support Vector Machines (SVM), and Neural Networks, were investigated. The models underwent training to discover patterns and connections between the input features and the final result of loan eligibility.

3) Cross-validation and Model Evaluation: Crossvalidation methods, such as k-fold cross-validation, were used to evaluate the models' performance and guarantee their generalizability. The models' success at properly predicting loan eligibility was measured using evaluation measures like accuracy, precision, recall, and F1-score. Hyperparameters were tweaked to maximize the performance of the models. To find the ideal set of hyperparameters that produced the highest model performance, methods like grid search or Bayesian optimization were used.

4) Model Validation: The practical usability and generalizability of the validated models were evaluated using an independent dataset or through real-world application. Through this stage, it was hoped that the models would be able to accurately forecast loan eligibility based on credit history and other important factors. The chosen model was included into the financial institution's or lending organization's process for determining loan eligibility. The model was implemented in a production environment, enabling in-the-moment predictions of loan eligibility based on applicants' credit histories and other crucial variables. By using this process, lenders might use the credit histories of potential borrowers to more precisely determine whether they qualify for loans. The systematic strategy made sure to use pertinent data, choose the right features, and use cuttingedge machine learning algorithms, which improved loan assessment outcomes and decreased default risks.

C) Comprehensive Risk Analysis

To evaluate multiple risk indicators and guarantee regulatory compliance, the methodology for conducting comprehensive risk analysis in the loan eligibility prediction system required a rigorous and sophisticated approach.

1) Data Collection and Integration: Various risk variables, including creditworthiness, income and employment verification, collateral analysis, risk profiling, fraud detection, and regulatory compliance, were all included in a

ISSN (online): 2581-3048
Volume 7, Issue 10, pp 532-542, October-2023
https://doi.org/10.47001/IRJIET/2023.710070

comprehensive dataset that was gathered. To get a complete picture of the applicant's risk profile, various data sources, including financial records, job history, credit reports, and regulatory databases, were merged.

2) Data Preprocessing and Quality Assurance: To assure the data's quality and consistency, preprocessing techniques such as data cleaning, outlier detection, and data normalization were used. Accurate risk analysis was made possible by the alignment and standardization of data from many sources.

3) Risk Factor Identification and Quantification: On the basis of regulatory requirements, historical risk analysis, and domain expertise, numerous risk factors were discovered. Suitable approaches, such as statistical analysis, probabilistic models, or rule-based systems, were used to quantify these characteristics. Based on the importance of each risk element and its possible influence on loan acceptance decisions, a weight or score was given to it.

4) Model Development: The development of predictive models for thorough risk assessments used statistical modeling and machine learning approaches. Advanced algorithms were used to effectively anticipate the applicant's overall risk level and capture the intricate correlations between the risk components, such as Logistic Regression, Decision Trees, Random Forests, or Neural Networks.

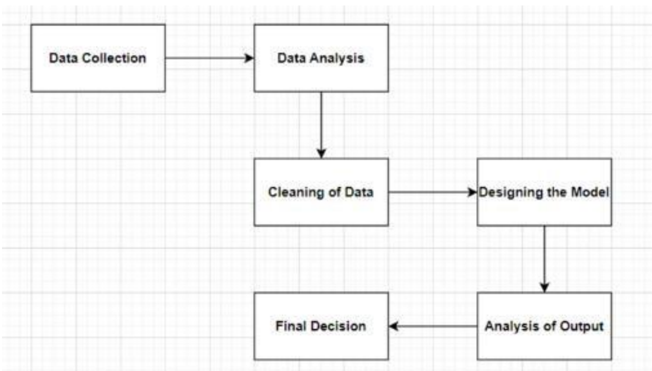

Figure 5: Model Development Architecture

5) Model Validation and Performance Evaluation: To evaluate the accuracy and effectiveness of the generated models, they underwent thorough validation procedures. The performance of the models was assessed using evaluation metrics such as precision, recall, F1-score, and receiver operating characteristic (ROC) curves on independent validation datasets or historical data. The predictive power of the models was increased, and they were improved to guarantee accurate risk assessment.

6) Integration and Decision Support: The loan eligibility prediction system included the verified risk models, allowing for real-time risk analysis and decision support. The models gave lenders risk profiles, risk scores, and useful insights so they could decide on loan approvals in a well-informed manner. In order to promote effective risk assessment and the presentation of risk-related data, decision support systems or dashboards were developed.

7) Continuous Monitoring and Adaptation: To adjust to shifting market conditions, legal requirements, and new dangers, the risk analysis models were regularly reviewed and modified. The effectiveness of the models was maintained over time by routine model calibration and modification, input from loan decisions, and external risk indicators.

8) Regulatory Compliance and Reporting: The risk analysis technique made sure that reporting and regulatory obligations were met. It included checks and balances to identify regulatory infractions, spot potential fraud, and produce thorough reports for internal and external stakeholders. Financial institutions might carry out thorough risk analysis, spot potential risks, and decide on loan acceptance with knowledge. Robust risk assessment, regulatory compliance, and efficient risk management were all guaranteed by the systematic methodology, integration of many risk indicators, and use of cutting-edge modeling techniques in the loan eligibility prediction system.

D) Advanced Mortage Estimation Tools

The Decision Tree Regression algorithm is used to assess and estimate mortgage values for properties as part of the process for creating sophisticated mortgage estimating tools.

1) Data Collection and Preparation: assemble an extensive database with details on various properties, such as the number of bathrooms, the number of beds, the size of the home, its location, and the appropriate mortgage amounts. To assure the dataset's quality and suitability for training the Decision Tree Regression model, do data preprocessing operations such data cleaning, addressing missing values, and data standardization.

2) Model Training: To effectively assess the trained model's performance, divide the dataset into training and testing sets. Utilize the training data to run the Decision Tree Regression technique, which can capture non-linear correlations between mortgage prices and property attributes. Use cross-validation techniques, such as k-fold cross-validation, to evaluate the model's performance and guarantee its applicability to new data.

3) Model Evaluation and Fine-tuning: Measure the trained model's accuracy in estimating mortgage values using suitable assessment metrics, such as mean squared error (MSE), root mean square error (RMSE), and mean absolute error (MAE).

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 10, pp 532-542, October-2023

https://doi.org/10.47001/IRJIET/2023.710070

To enhance the model's performance and predictive skills, adjust the model's parameters, such as the decision tree's maximum depth.

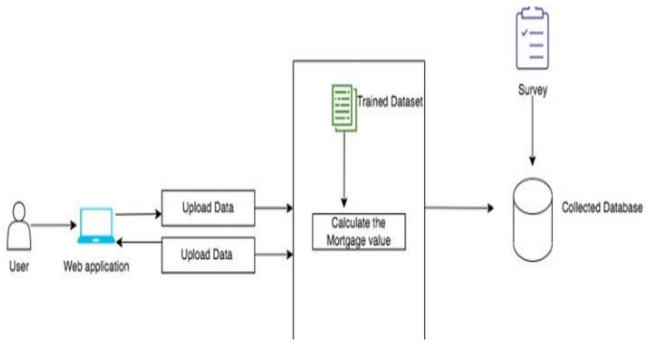

Figure 6: Model Evaluation Diagram

4) Tool Development and Integration: Implement a user-friendly mortgage calculator tool that accepts user inputs such as the number of bathrooms, the number of beds, the size of the home, and the location of the home. Using the user's inputs as a guide, utilize the trained Decision Tree Regression model to forecast the mortgage value. The Mortgage Calculator tool should be integrated into the loan application procedure to give consumers a simple way to calculate and estimate mortgage values for properties.

5) User Interface Design: Build a user-friendly, interactive user interface for the mortgage calculator so that users can simply enter property information and view the estimated mortgage value. To improve the tool's usability and user experience, include elements like graphical representations, sliders, and user-friendly prompts.

6) Testing and Validation: To verify the Mortgage Calculator tool's accuracy and dependability in forecasting mortgage values under various scenarios, thoroughly test it. Perform user testing, compare the tool's outputs to existing mortgage numbers, and validate the results to ensure accuracy.

7) Continuous Enhancement and Maintenance: Monitor and update the Mortgage Calculator tool frequently to include new property features, improve the estimating model's precision, and adjust for shifting market conditions. To keep the tool current and deliver accurate mortgage estimates, take into account user input, bug complaints, and industry best practices. To give consumers precise and practical estimations of mortgage values for properties, modern mortgage estimation tools use the Decision Tree Regression algorithm. The tool's usefulness and utility in assisting borrowers' financial planning and decision-making processes are ensured by the incorporation of user-friendly interfaces, model training, and ongoing improvement.

IV. CONCLUSION AND FUTURE WORKS

The research paper presented a thorough system for predicting loan eligibility that included four distinct parts: predicting the best bank rate options, utilizing credit history to determine loan eligibility, conducting thorough risk analysis, and creating sophisticated mortgage estimation tools. By combining these elements, lenders and borrowers will be given the information they need to make wise decisions, manage risks, and achieve the best results possible. Through the methodology, it was possible to show how well machine learning and deep learning algorithms forecast the best possible bank rates for loan applicants. In order to help consumers, make informed decisions, variables including loan type, term, and amount were taken into account. Algorithms like Random Forest, Gradient Boosting, and Linear Regression also proved to be helpful. These cutting-edge methods give borrowers the ability to match their financial preferences and capabilities with the best banks and rates, leading to better financial outcomes. The methodology also stressed the value of using applicants' credit histories to accurately determine their loan eligibility. Machine learning algorithms were used to evaluate trustworthiness by incorporating crucial factors like gender, marital status, education, income, and credit history themselves. This strategy makes it possible for lenders to evaluate borrowers' creditworthiness more accurately, lowering the chance of default and improving the loan approval procedure. The research used a sophisticated Loan Eligibility Prediction System to incorporate a thorough risk analysis component. The system considered a number of sub-objectives, including creditworthiness evaluation, verification of earnings and employment, collateral analysis, risk profiling, fraud detection, and regulatory compliance. By accurately predicting applicants' risk profiles using logistic regression, lenders were able to make judgments that would maximize loan results while minimizing risk.

The Decision Tree Regression technique has been used to create sophisticated mortgage estimating tools. These tools considered crucial aspects of a property, such the number of bathrooms, beds, the size of the house, and the location. The mortgage calculator tool helped borrowers make better financial plans and decisions by giving them useful insights into the affordability and suitability of mortgages. A unique system for predicting loan eligibility was offered in the research paper, one that makes use of credit history analysis, machine learning and deep learning algorithms, and cuttingedge estimate tools. With the help of this complete system, both lenders and borrowers may make well-informed decisions that lead to improved results and lowered risks during the loan approval process. Alternative data sources, such social media profiles and online transaction histories may

ISSN (online): 2581-3048
Volume 7, Issue 10, pp 532-542, October-2023
https://doi.org/10.47001/IRJIET/2023.710070

be incorporated in future efforts in this field. These resources might offer extra information about the creditworthiness and financial habits of borrowers, improving the predictive power of the models. The predictive ability of the system might be increased by investigating more sophisticated machine learning and deep learning algorithms, such as neural networks and ensemble techniques. Another potential area for future development is the incorporation of natural language processing methods to examine unstructured data, such as loan applicant narratives or financial paperwork. This would give loan eligibility assessments more detailed information, improving the accuracy of the system. The risk analysis and compliance elements of the system require constant updates and improvements due to the changing nature of the banking industry and regulatory environment. Lenders would be better equipped to maintain compliance and effectively adjust to changing market conditions if they could adapt to changing rules and consider new risk variables. The loan eligibility prediction algorithm could be refined and made more user-friendly by conducting in-depth user studies and getting input from lenders and borrowers. Future improvements would be guided by an understanding of the unique requirements and pain areas of users, ensuring that the system closely complies with their needs and offers a smooth user experience. The research paper's proposal of a comprehensive system that incorporates numerous components has made a significant contribution to the field of loan eligibility prediction. The research showed how credit history evaluation, machine learning algorithms, and sophisticated prediction tools may completely transform the loan approval process. To improve the loan eligibility prediction system even further, future work should concentrate on investigating additional data sources, sophisticated algorithms, dynamic risk analysis, compliance updates, and user-centric modifications. In a financial environment that is always changing, borrowers and lenders can make better decisions, reduce risks, and improve loan outcomes by adopting these options.

REFERENCES

[1] Sheikh, M. A., Goel, A. K., & Kumar, T. (2020, July). An approach for prediction of loan approval using machine learning algorithm. In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC) (pp. 490-494). IEEE.

[2] Kumar, A., Sharma, S., & Mahdavi, M. (2021). Machine learning (ML) technologies for digital credit scoring in rural finance: A literature review. Risks, 9(11), 192.

[3] Malekipirbazari, M., & Aksakalli, V. (2015). Risk assessment in social lending via random forests. Expert Systems with Applications, 42(10), 4621-4631.

[4] isutsa, G. T. (2021). Loan Default Prediction Using Machine Learning: a Case of Mobile Based Lending (Doctoral dissertation, University of Nairobi).

[5] Kumar, C. N., Keerthana, D., Kavitha, M., & Kalyani, M. (2022, June). Customer Loan Eligibility Prediction using Machine Learning Algorithms in Banking Sector. In 2022 7th International Conference on Communication and Electronics Systems (ICCES) (pp. 1007-1012). IEEE.

[6] Kadam, A. S., Nikam, S. R., Aher, A. A., Shelke, G. V., & Chandgude, A. S. (2021). Prediction for loan approval using machine learning algorithm. International Research Journal of Engineering and Technology (IRJET), 8(04).

[7] Mohankumar, M., Amuthakkani, S., & Jeyamala, G. (2016). Comparative analysis of decision tree algorithms for the prediction of eligibility of a man for availing bank loan. Age, 19, 60.

[8] Zhang, W., Wu, C., Zhong, H., Li, Y., & Wang, L. (2021). Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. Geoscience Frontiers, 12(1), 469477.

[9] Chen, S., Härdle, W. K., & Jeong, K. (2010). Forecasting volatility with support vector machine-based GARCH model. Journal of Forecasting, 29(4), 406-433.

[10] Odegua, R. (2020). Predicting bank loan default with extreme gradient boosting. arXiv preprint arXiv:2002.02011.

[11] Huang, J., Chai, J., & Cho, S. (2020). Deep learning in finance and banking: A literature review and classification. Frontiers of Business Research in China, 14(1), 1-24.

[12] Adekoya, A. F., Nti, I. K., & Weyori, B. A. (2022). Long Short-Term Memory Network for Predicting Exchange Rate of the Ghanaian Cedi. FinTech, 1(1), 25-43.

[13] Ma, X., Sha, J., Wang, D., Yu, Y., Yang, Q., & Niu, X. (2018). Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning. Electronic Commerce Research and Applications, 31, 24-39.

[14] Stein, G., Chen, B., Wu, A. S., & Hua, K. A. (2005, March). Decision tree classifier for network intrusion detection with GA-based feature selection. In Proceedings of the 43rd annual Southeast regional conference Volume 2 (pp. 136-141).

[15] Wu, J. M. T., Li, Z., Herencsar, N., Vo, B., & Lin, J. C. W. (2021). A graph-based CNN-LSTM stock price

ISSN (online): 2581-3048
Volume 7, Issue 10, pp 532-542, October-2023
https://doi.org/10.47001/IRJIET/2023.710070

prediction algorithm with leading indicators. Multimedia Systems, 1-20.

[16] Sripriya, S. V. S., Varrey, S. D. S., & Venkateshkumar, M. (2022, October). Predictive Model to Compute Eligibility Test for Loans. In 2022 IEEE Industrial Electronics and Applications Conference (IEACon) (pp. 185-190). IEEE.

[17] Madaan, M., Kumar, A., Keshri, C., Jain, R., & Nagrath, P. (2021). Loan default prediction using decision trees and random forest: A comparative study. In IOP Conference Series: Materials Science and Engineering (Vol. 1022, No. 1, p. 012042). IOP Publishing.

[18] Sripriya, S. V. S., Varrey, S. D. S., & Venkateshkumar, M. (2022, October). Predictive Model to Compute Eligibility Test for Loans. In 2022 IEEE Industrial Electronics and Applications Conference (IEACon) (pp. 185-190). IEEE.

[19] Amin, R. K., & Sibaroni, Y. (2015, May). Implementation of decision tree using C4. 5 algorithm in decision making of loan application by debtor (Case study: Bank pasar of Yogyakarta Special Region). In 2015 3rd International Conference on Information and Communication Technology (ICoICT) (pp. 75-80). IEEE.

[20] Min, S. H., Lee, J., & Han, I. (2006). Hybrid genetic algorithms and support vector machines for bankruptcy prediction. Expert systems with applications, 31(3), 652-660.

[21] Liu, W., Fan, H., & Xia, M. (2022). Credit scoring based on tree enhanced gradient boosting decision trees. Expert Systems with Applications, 189, 116034.

[22] Ala'raj, M., Abbod, M. F., & Majdalawieh, M. (2021). Modelling customers credit card behaviour using bidirectional LSTM neural networks. Journal of Big Data, 8(1), 1-27.

[23] Ghaddar, B., & Naoum-Sawaya, J. (2018). High dimensional data classification and feature selection using support vector machines. European Journal of Operational Research, 265(3), 993-1004.

[24] Tumuluru, P., Burra, L. R., Loukya, M., Bhavana, S., CSaiBaba, H. M. H., & Sunanda, N. (2022, February). Comparative Analysis of Customer Loan Approval Prediction using Machine Learning Algorithms. In 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS) (pp. 349-353). IEEE.

[25] Al-Qerem, A., Al-Naymat, G., & Alhasan, M. (2019, December). Loan default prediction model improvement through comprehensive preprocessing and features selection. In 2019 International Arab Conference on Information Technology (ACIT) (pp. 235-240). IEEE.

[26] Huang, X., Liu, X., & Ren, Y. (2018). Enterprise credit risk evaluation based on neural network algorithm. Cognitive Systems Research, 52, 317324.

[27] Li, S. T., Shiue, W., & Huang, M. H. (2006). The evaluation of consumer loans using support vector machines. Expert Systems with Applications, 30(4), 772-782.

[28] Gupta, K., Chakrabarti, B., Ansari, A. A., Rautaray, S. S., & Pandey, M. (2021, April). Loanification-Loan Approval Classification using Machine Learning Algorithms. In Proceedings of the International Conference on Innovative Computing & Communication (ICICC).

[29] Geller, A., & Hainaut, D. Long Short-Term Memory neural network for econometric forecasting: A comparison between a statistical method and a neural network in the case of Value at Risk.

[30] Mahbobi, M., Kimiagari, S., & Vasudevan, M. (2021). Credit risk classification: an integrated predictive accuracy algorithm using artificial and deep neural networks. Annals of Operations Research, 1-29.

[31] Abedin, M. Z., Guotai, C., Hajek, P., & Zhang, T. (2022). Combining weighted SMOTE with ensemble learning for the class-imbalanced prediction of small business credit risk. Complex & Intelligent Systems, 121.

[32] Shrestha, S., & Paudel, L. (2019). Classification of Loan Applications of Garima Bikas Bank Ltd Using Decision Tree Classification Method. Journal of Advanced College of Engineering and Management, 5, 147-152.

[33] Zhu, L., Qiu, D., Ergu, D., Ying, C., & Liu, K. (2019). A study on predicting loan default based on the random forest algorithm. Procedia Computer Science, 162, 503-513.

[34] Sadeeq, M. A., & Abdulazeez, A. M. (2020, December). Neural networks architectures design, and applications: A review. In 2020 International Conference on Advanced Science and Engineering (ICOASE) (pp. 199-204). IEEE.

[35] Harris, T. (2015). Credit scoring using the clustered support vector machine. Expert Systems with Applications, 42(2), 741-750.

[36] Nica, I., Alexandru, D. B., Crăciunescu, S. L. P., & Ionescu, Ș. (2021). Automated valuation modelling: analysing mortgage behavioural life profile models using machine learning techniques. Sustainability, 13(9), 5162.

[37] Mahbobi, M., Kimiagari, S., & Vasudevan, M. (2021). Credit risk classification: an integrated predictive

ISSN (online): 2581-3048
Volume 7, Issue 10, pp 532-542, October-2023
https://doi.org/10.47001/IRJIET/2023.710070

accuracy algorithm using artificial and deep neural networks. Annals of Operations Research, 1-29.

[38] Rasmussen, C. E. (1997). Evaluation of Gaussian processes and other methods for non-linear regression (Doctoral dissertation, University of Toronto Toronto, Canada).

[39] Song, Y., Wang, Y., Ye, X., Wang, D., Yin, Y., & Wang, Y. (2020). Multi-view ensemble learning based on distance-to-model and adaptive clustering for imbalanced credit risk assessment in P2P lending. Information Sciences, 525, 182-204.

Citation of this Article:

Senarathna B.T.N, Weerarathna K.C.M, Wickramarachchi D.S, Jayarathne S.M.P.N, Buddhima Attanyake, “Loan Eligibility Prediction Based on Credit Score and Past History” Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 7, Issue 10, pp 532-542, October 2023. Article DOI https://doi.org/10.47001/IRJIET/2023.710070

****