International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 786-794 www.ijaem.net ISSN: 2395-5252

Loan Approval Prediction

Pallapothu Nishita, Bipasha Bhowmik , Poli Reddy Gayani, Avantika R , Sardar Puneet Prasad Singh

Department of CSE (Data Science), Jain (Deemed-to-be-University) Bengaluru, Karnataka, India
Department of CSE (DataScience), Jain (Deemed-to-be-University), Bengaluru, Karnataka, India
Department of CSE (Data Science), Jain (Deemed-to-be-University), Bengaluru, Karnataka, India
Department of CSE (Data Science), Jain (Deemed-to-be-University), Bengaluru, Karnataka, India
Assistant Professor Department of CSE, Jain (Deemed-to-be-University), Bengaluru, Karnataka, India

---

Date of Submission: 05-04-2023 Date of Acceptance: 15-04-2023

ABSTRACT—It is customary to discover to whom a loan might be given that will be the bank's safer alternative., because several individuals have applied for bank loans due to the advancements in the banking sector, however the bank has its limited resources and can only grant to a small number of people.Unnecessary openness to threatcan lead to dissatisfaction with banks and distress a great many persons because of the dimensions of certain banks. Legislatures can make better principles to advance mindful administration and dynamic by better comprehension the dangers presented to banks. The selections of financial backers are likewise influenced by a bank's ability to oversee risk. Despite the fact that a bank can deliver critical incomes, unfortunate gamble the board could bring about decreased benefit due to credit misfortune openness. Proficient financial bankers are more disposed to support a bank that can create benefits and doesn't represent a critical gamble of monetary misfortune.There are times where high reputed banks even fail in identifying the right person while giving a loan. Thus, the project focuses on identifying the right person at a very early stage.To save a considerable amount of time and money for the bank, we consequently take steps to diminish the possible risks connected with choosing the safe individual in this project. This is accomplished by obtaining information from the databases of the borrowers who have already received loans. Based on these histories, a machine was trained using a Python and ML model that yields the best accurate result. Foreseeing whether it will be protected to relegate the credit to a particular individual is the significant objective of this exploration.

Keywords—Loan, Loan prediction, credit, credit score,cibilscore Machine Learning, Probit model, logistic regression, mortgages, subprime mortgages, Decision tree, Datasets.

I. INTRODUCTION

A. Overview

Nearly every bank's main line of business is the distribution of loans. A large portion of the bank's assets were directly derived from the money the bank made from the loans it distributed. In a banking setting, placing assets in trustworthy hands is the main goal. Many banks and finance organizations now offer loans following a drawn-out identification process. and validation process, but there is no guarantee that the selectedcandidate is the worthiest candidate among all candidates. This technique allows us to forecast if a certain candidate is secure or not. We can determine whether a certain candidate is prudent or not using this approach. Both claimants and bank workers definitely gain from loan prediction. The goal of this technique is to offer a rapid, easy, and instant manner to select the deserving candidates.

In the credit area, the two most significant inquiries are:

- How unsafe is the borrower?
- Would it be a good idea for us to loan to the borrower given the gamble in question?

The answers to these questions take a lot of time and still the risk factor is high or low can bedeterminedcompletely.Financial institutions lend money to borrowers in return for the assurance that they would pay it back with interest. So, the lender only receives payment (interest) if the borrower pays the obligation in full. Yet if the borrower doesn't pay back the loan, the bank suffers a loss. The goal of this project is to improve the accuracy and efficiency of the loan approval process by using logistic regression, probit model and Artificial Neural Networks and making a risk model to assess the risk of lending money to

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 786-794 www.ijaem.net ISSN: 2395-5252

subprime mortgages.

For the prediction of loan, there are a number of variables to take into account, including borrower and loan characteristics, in order to create a predictive model that can accurately anticipate loan acceptance. One of the most crucial elements is the borrower's credit score, which provides a gauge of their financial standing and default risk. Other aspects of the borrower, such as their income, job situation, and debt-to-income ratio, are crucial indicators of their capacity to repay the loan. Loan features, in addition to borrower characteristics, are crucial factors in loan approval decisions. The borrower's capacity to repay the loan and the overall risk of default can be impacted by the loan's quantity, period, interest rate, and fees. The loan's objective, such as financing a home purchase or schooling, may also be taken into account, as different types of loans may have different risk profiles.

B. Problem Definition

For financial institutions, predicting loan acceptance is a crucial undertaking since it allows them to figure out applicants' credibility and control the risk of loan defaults. Due to their low credit scores, subprime borrowers frequently encounter considerable difficulties getting the funding from financial institutions. Utilizing different machine learning algorithms, including logistic regression, decision trees, and random forests, previous studies have created models for predicting loan endorsement. These models could, however, be limited in their ability to anticipate outcomes for subprime borrowers due to their inability to fully account for the intricate connections between borrower and loan variablesthe fundamental goal of this project is to guess whether a loan will be accepted or denied for an end-user based on their Cibil score. There are various loan types, including (Personal Loans, Home Loans, Mortgages Loans etc.). By calculating Cibil Score from a variety of historical data and factors, including credit history, the number of credit cards a person has, the loans they have taken out, the number of times they have fallen behind on their payments, credit mix, credit utilization ratio, and credit history age, our model will be able to make educated results about if to accept or reject a loan request by feeding the dataset to the deep learning models like GRU(Gated Recurrent Neural Network) , Bi-LSTM , Logistic Regression and Probit model. We will create a risk model for subprime borrowers (those with low credit scores) in addition to a model for loan acceptance prediction. In simple words a method for assessing the potential risks associated with lending money to borrowers with poor credit score or no credit score at all.

C.Objective

- The goal of loan approval prediction is to use historical information on loan applicants to create a predictive model that can assess a loan applicant's likelihood of loan approval. This might help lenders decide if they should approve a loan application based on information they have.
- A loan approval prediction model can give lenders insights into the risk involved with loan acceptance by examining numerous parameters such credit score, income, employment history, debt-to-income ratio, and other pertinent information. As a result, lenders may be able to approve loans with greater accuracy, lower the risk of default, and ultimately increase the effectiveness and profitability of their lending operations.
- The main objective of loan approval prediction using CIBIL score is to evaluate loan applicants' creditworthiness based on their credit histories and ascertain the chance that they would repay the loan. A company called CIBIL (Credit Information Bureau India Limited) maintains trail of individuals' and companies' credit histories. The three-digit CIBIL score is determined by taking into account a person's credit history, which includes their use of credit cards, loan repayment history, and other financial activities
- From the perspective of the lender, the goal of this loan approval prediction for subprime mortgages is to reduce the risk of loan defaults while maximizing profitability. Borrowers with subprime mortgages usually have average or low credit scores, larger debt-to-income ratios, or other characteristics that make them riskier borrowers.
- In order to increase accuracy and reliability, our loan approval prediction models are likely to include more complex algorithms that include deep learning, neural networks, and other cutting-edge methods.
- Our project can be utilized by Lenders and they will be better equipped to manage lending risks as loan approval prediction models advance, leading to more efficient loan underwriting and a decrease in loan defaults.
- With our project Lenders may provide consumers a more seamless and simple experience by speeding the loan approval

process, which can boost customer happiness and loyalty.

II. LITERATURE REVIEW

[1] The authors provide a method for calculating the likelihood of approving a loan application by using previous information on loan applicants to create a predictive model. The system analyses various factors including credit score, income, employment history, debt-to-income ratio, and other pertinent information to give lenders an understanding of the risk associated with approving a loan. These factors include Logistic Regression, K-Nearest Neighbors (KNN), Decision Trees, and Random Forest. The authors compare the outcomes with conventional credit scoring models to evaluate the efficiency of the suggested loan predictive model using real-world data from a financial institution. The study provides insights into the possible advantages of using such a strategy in the lending business and proposes a thorough approach to loan prediction using machine learning algorithms. The suggested system can assist lenders in making more precise and knowledgeable loan approval decisions, which will decrease loan defaults and increase profitability.

[2] This study predicts customer loan acceptance using two machine learning methods: Random Forest (RF) and Support Vector Machine (SVM). In this Paper, the suggested system for predicting bank loan credibility may assist firms in creation of the finest choice deciding whether to grant a client's request for a loan or deny it. This can help the banking industry set up efficient distribution channels. and mitigating significant financial losses without any doubts. SVM and RF algorithms are being employed for the prediction in this proposed system. The application of alternative methods might perform better.

[3] In this study, a dataset of typical loan applicants is analyzed to find trends and forecast financial outcomes using machine learning (ML) algorithms. Those who will be deserving of loans. Previous customer information will be utilized to age, kind of income, loan annuity, most recent credit bureau report, type of employer, and length of employment are all factors that should be studied .The most important characteristics—those that most significantly affect the outcome of the prediction —were found using ML techniques like RF, XGBoost, Adaboost, Lightgbm, DT, & KNN. Using common measures, these algorithms are contrasted and evaluated against one another. The highest accuracy of these was 92%, attained through Logistic Regression. Also, it was proven to be the best model and outperformed alternatives significantly.

[4] The goal of the paper was to forecast loan acceptance using machine learning techniques. The authors pointed out that loan acceptance is a crucial choice made by financial organizations and that machine learning can automate this process to make it more accurate and efficient. The authors used a dataset that included details on the demographics, earnings, credit histories, and other aspects of loan applicants. To forecast loan acceptance, they used a variety of machine learning methods, including logistic regression, decision treesrandom forests. The accuracy, precision, recall, and F1 score were some of the metrics the authors used to evaluate the effectiveness of each algorithm.They demonstrated that, with 85% accuracy, the random forest algorithm outperformed the opposition. The authors came to the conclusion that automating loan approval decisions using machine learning algorithms would boost accuracy and efficiency. They pointed out that future studies might look into the use of additional data sources, such social media data, to enhance predictions of loan acceptance.

[5] In this paper, four algorithms, including RF, DT, Naive Bayes, and Logistic Regression, are employed in this study to forecast whether or not clients would be approved for loans. The same dataset will be utilized for all four methods, and the most accurate algorithm will be chosen to deploy the model. From this point forward, we create a machine learning-based bank loan prediction system that chooses the qualified applicants for loan approval on its own. Exploratory analysis is preceded by data cleaning and missing value handling, then model building and model assessment. We have the highest when we obtain improved accuracy scores and other performance indicators, which will be evaluated.This paper may be beneficial in determining whether or not a candidate will be given a bank loan. The application of alternative methods might perform better.

Here in this research paper the authors have created a technique that allows us to forecast whether the selected applicant will be a deserving application for loan approval or not. The model,

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 786-794 www.ijaem.net ISSN: 2395-5252

which has been built using machine learning methods, serves as the foundation for the system's predictions. Even the accuracy of various machine learning algorithms has been compared. Our accuracy ranged from 75 to 85%, but the greatest result came via logistic regression, coming in at 88.70%. The system has a web application user interface where the user can input the information needed for the model to forecast. This model's disadvantage is that it considers a lot of factors, however in reality, a loan application may occasionally be approved based just on one significant factor.

[7]In this study, patterns are retrieved from a shared dataset of loans sanctioned and extracted using machine learning (ML) techniques to predict future defaulters on loans. The study will be done using historical customer data, including their employment history, loan balance, age, and income. Several machine learning (ML) methods, including Random Forest, Support Vector Machine, K-Nearest Neighbor, and Logistic Regression, were used to identify the most relevant characteristics, or the elements that had the biggest influence on the forecast's result. These algorithms are contrasted and scored according to accepted standards. The random forest method is more accurate.

[8]In this study, Random Forest and Decision Trees are comprehensively compared and contrasted to one another. Both methods used the similar dataset, and analysis of the outcomes demonstrated that the Random Forest algorithm outperformed the Decision Tree methodology and did so with a far higher level of accuracy. The goal of the project was to create a model that uses decision trees and random forest algorithms to forecast loan defaults. This work investigated, analyzed, and created a machine learning algorithm to predict one's likelihood of defaulting on a loan given a collection of characteristics. Using this kind of model, Lender might be able to spot specific financial characteristics in prospective borrowers that might increase their risk of defaulting and making timely payments on their loans.The Random Forest Classifier's accuracy was 80%, whereas the Decision Tree method's accuracy was 73%. As a result, it appears that the Random Forest model is a better choice for these types of data.

[9]As according earlier research from this era, there are many ways to look into the issue of preventing loan default. Nonetheless, since producing accurate projections is essential for optimizing revenues, it is necessary to learn about the severalmethods and compare them. The problem of predicting loan nonpayers is calculated using the logistic regression model, a key predictive analytics method. Data from Kaggle is gathered for predictions and analysis.To estimate the various performance metrics, logistic regression models were used. The models are contrasted based on performance indicators like sensitivity and specificity. The final outcomes have shown that the model produces a range of outcomes. This research offers a machine learning algorithm-based approach to automate this process. To build and implement a system that forecasts whether a user will be approved for a loan from a bank using data mining and machine learningto decrease scams and increase accuracy.

[10]The main focus of this work is to accept a customer loan and to provide a quick, easy, and immediate technique for selecting deserving loan applicants using machine learning algorithms. The goal of the project was to create a machine learning algorithm-based model for forecasting loan approval. The authors obtained information from a financial institution about the borrowers' income, employment status, loan amount, and other pertinent criteria. The four classification-based machine learning techniques used in this study are random forests, support vector machines, decision trees, and logistic regression. The most effective method for accurately predicting loan acceptance is support vector machine technology.

[11]The main goal in a lending setting is to place one's money in reliable hands. In order to predict whether or not consumers would be authorized for loans, this study employs three machine learning techniques: Logistic Regression (LR), Decision Tree (DT), and Random Forest (RF). The experimental results show that the Decision Tree machine learning algorithm surpasses Logistic Regression and Random Forest machine learning techniques in terms of accuracy.After a thorough analysis of the product's benefits and drawbacks, it is safe to conclude that the component is a highly effective one. This application meets with all necessary requirements and is operationally sound. It is straightforward to connect this component to a variety of other systems.There have been numerous examples of both technical and content failures in automated prediction systems, which have a predefined weight for the characteristics that matter the most. As a result, it's likely that the infamous technology may soon be enhanced to offer dynamic weight change while being safe. The automated production system can soon be attached to this prediction module.Future software can be developed so that

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 786-794 www.ijaem.net ISSN: 2395-5252

fresh testing data should be used, despite the fact that the system was trained on an outdated training dataset.

[12]This article reduces the risk associated with choosing the perfect borrower who might repay the loan on time and keep the bank's non-performing assets (NPA) on hold. This is achieved by supplying previous information about bank customers.

who have obtained loans into a trained machine learning model, which could produce reliable results. Determining the major purpose of this paper is to investigate if it will be safe to distribute the loan to a particular person. The sections of this study are (I) Data Collecting, (ii) Data Cleansing, and (iii) Performance Assessment.By carefully analyzing positive qualities and limits, it can be concluded from the experimental data presented in this paper that the Nave Bayes model is extremely effective and yields better results when compared to other models.It operates properly, satisfies all bankers' needs, and is interconnected with numerous other systems. There were several computer issues, content mistakes, and weight fixing in automated prediction systems. The banking software may become more dependable, accurate, and dynamic in the near future and may be able to integrate with an automated processing unit

III. METHODOLOGY

A. Proposed Methodology

In the first part of the project, we are using the primary attribute, or cibil score, to decide whether or not a person will be granted a loan. We have calculated the cibil score based on the weights of the attributes like number of credit cards, number of loans taken, number of delayed payments, credit mix, credit utilization ratio, credit history age. The pre-processing of this dataset involves feature extraction, which results in the removal of characteristics like customer name and customer id. The dataset is then standardized and fed into the GRU (Gated Recurrent Unit), BiLSTM(Bidirectional Long Short-Term Memory), and Logistic Regression. The accuracy is contrasted with the current models. Basically, the borrowers can be of four types Subprime (credit scores of 580-619) Near-prime (credit scores of 620-659) Prime (credit scores of 660-719) Super-prime (credit scores of 720 or above). Now the next part of the project focuses on the subprime mortgages (average or below-average credit score). The main factors taken into account for this prediction include the bankruptcy indicator, finance inquiries 24 months, Number trade lines 30 or 60 days 24 months, Time since last inquiry etc. Using this, a risk model is created that forecasts the likelihood of the risk the bank is taking when lending money to subprime borrowers. Machine learning algorithm like Probit Model, which is an extension of logistic regression and deep learning algorithm like ANN is used.

B. Dataset

In our paper, we employed two datasets. Approximately 1000 entries make up the first data set we used, which came from Kaggle. Its columns include customer name, customer id, age, monthly income, number of credit cards, number of loans, type of loan, and several more. The following dataset, which also comes from Kaggle, has about 3000 entries. This dataset mainly includes information on subprime borrowers, namely those with credit scores within 580 and 619. It includes information such as Financial Number of Trade Lines 30 or 60 Days 24 Months, Number of Trade Lines 50% Used, Time Since Last Trade Line, Number of Inquiries 24 Months, 6 Months, etc.

C.Sequence Diagram

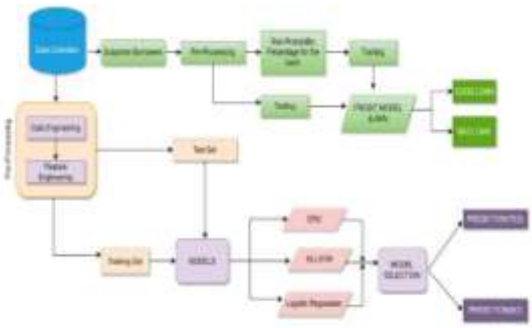

Fig1:System architecture

C. Data Pre-processing

Preparing raw data for analysis through cleaning and transformation is known as data pre-processing. In data science and machine learning, it is a key phase since the quality of the data and the manner in which it is prepared can have a significant impact on the accuracy of the analysis and the effectiveness of the models based on it.

- Initially, during pre-processing, we replaced any missing values in the "Age" column with the average age of all the entries that were present.
- Following that, distinct columns that are unique are removed from the Data Frame. The names of the columns that are dropped using the drop () method are "ID," "Name," "Customer ID," and "SSN."

- To improve the machine learning modelperformance, feature engineering involves developing new features or modifying current features. The categorical variables are Label Encoder was employed in order to transform the values to numerical values.

D. Model Selection and Model Evaluation

We used a total of three approaches in this research paper's first section, which predicts loan eligibility based on Cibil score. Those areGRU (Gated Recurrent Unit), Bi-LSTM (Bidirectional Long Short-Term Memory), and Logistic Regression.

a) GRU (Gated Recurrent Unit):

Recurrent neural networks (RNNs) using GRUs (Gated Recurrent Units) are a typical choice for sequence modelling tasks including language modelling and time series prediction. It is a type of recurrent neural network and also an adaptation of the LSTM (Long Short-TermMemory) architecture, but with less computational complexity and fewer parameters. The loan application data is pre-processed to normalise thecontinuous variables, transform the categorical variables to numerical values, and divide the data into training and testing sets. The GRU model is made up of a number of GRU cells that are sequentially coupled. The data from the current loan application and the output from the prior cell are entered into each GRU cell. Based on the input and the previous hidden state, the cell's hidden state is updated. The GRU cell's output is a forecast of the likelihood that the present application will be approved for a loan.Training and Testing the model:Backpropagation over time is used to train the GRU model on the training set of loan applications (BPTT). The model gains the ability to reduce the loss function between the true outcome and the anticipated loan approval outcome during training.Using the testing set of loan applications, the trained GRU model is assessed. For each application in the testing set, the model forecasts whether a loan will be approved, and the accuracy of the model is measured as the proportion of accurate forecasts. In comparison to other models, the GRU model has a number of advantages, including the capacity to handle variable-length sequences and the ability to identify long-term connections in the data. However, it could be sensitive to the selection of hyperparameters and may need a lot of data to train effectively. Consequently, before utilising the model in production, it is crucial to properly adjust the hyperparameters and assess how the model performs on various datasets.

b) Bi-LSTM (Bidirectional Long Short-Term Memory):

An example of a recurrent neural network (RNN) that can be used to forecast loan acceptance is Bidirectional Long Short-Term Memory (Bi-LSTM). Bi-LSTM can handle variable-length input sequences and capture sequential dependencies, which is its key benefit over other models. We first pre-process and prepare the input data before using Bi-LSTM for loan approval prediction. Normally, this required encoding categorical variables using label encoder and feature engineering. Following that, the input data would be divided into training, validation, and test sets. We also used the minmax scaler to scale the data. The Bi-LSTM model architecture was then defined, typically consisting of an embedding layer to transform categorical variables into continuous representations, one or more Bi-LSTM layers to process the sequential data, and one or more fully connected layers to output the prediction of loan acceptance. The model would then be trained using an appropriate loss function and optimisation method on the training data. On the validation set, the model's performance would be tracked, and hyperparameters would be adjusted as necessary. The model would then be evaluated on the test set to determine how well it generalised. In general, Bi-LSTM can be an effective technique for loanapproval prediction, especially when the input datacontains sequential relationships. To achieve the greatest performance, it is crucial to thoroughly pre-process and design the input data as well as tweak the model hyperparameters.

c) Logistic Regression:

A statistical model that can be used to forecast loan approval is logistic regression. The key benefit of adopting logistic regression over other models is that it is straightforward and easy to understand, as it sheds light on the significance and trend of each input variable. Similar to other models, we must first pre-process and prepare the input data before using logistic regression to predict loan acceptance. Moreover, feature engineering and categorical variable encoding are also involved. Then, training, validation, and test sets would be created from the input data.A suitable loss function and optimisation algorithm, such as maximum likelihood estimation or gradient descent, would then be used to train the model using the training data. On the validation set, the model's performance would be tracked, and hyperparameters would be adjusted as necessary. The model would then be evaluated on the test set to determine how well it generalised. Overall, logistic regression can be a valuable method for predicting loan acceptance, especially when the input data and the target variable have a straight, linear relationship. So, for the project's last phase, where we are creating a risk model to help banks and other lenders determine the likelihood of the risk they will assume if they lend money to subprime borrowers and thereby choose whether or not they should do so, we are utilising the Probit model and an ANN.

d) Probit Model:

A particular kind of statistical model called a probit model is employed in regression analysis to examine categorical dependent variables. A probit model can be used to forecast the possibility of default or foreclosure for a specific mortgage in the context of risk modelling for subprime mortgages. The borrower's financial inquiry history (24 months), inquiry history (6 months), number of trade lines (30 or 60 days) (30 months), number of trade lines 50% used (30 months), time since last inquiry (30 months), time since last trade line (30 months), and other pertinent factors would typically be used to build the model. These characteristics would be employed to forecast the probability of default or foreclosure, which might be used to evaluate the risk of the mortgage and guide decisions on lending policies or investing tactics. The ability to estimate the chance of default as a continuous variable rather than a binary result is one potential benefit of employing a probit model in this situation (i.e., default or no default). This can offer more detailed details regarding the degree of risk connected to a specific mortgage, and it can be especially helpful in portfolio-level research.

e)ANN (Artificial Neural Network):

In order to model risk for subprime mortgages, neural networks, also called artificial neural networks, are machine learning techniques that we applied (ANNs). ANNs are built with interconnected nodes or neurons that analyse information and make predictions based on patterns in the data, mimicking the structure and operation of the human brain.An ANN could be used to forecast the likelihood of bankruptcy or loss in the context of subprime mortgage risk modelling using a variety of input factors, such as borrower characteristics, loan attributes, and economic indicators. In order to discover patterns and links between these factors and the desired outcome, the network would be trained using historical data.An ANN may be advantageous in this situation because it can quickly capture complex nonlinear correlations between variables that are difficult for traditional statistical models to do so. This may result in better risk management plans and more accurate predictions.

IV. HARDWARE AND SOFTWARE REQUIREMENTS

Hardware used:

- RAM: 2GB
- Hard Disk: 500GB
- Processor: Intel i5
- Operating System: Windows / Linux / Mac

Software used:

- Programming Language: Python 3.8
- IDE: Jupyter Notebook 6.4.8
- Python Libraries:
- Models – Sequential
- Keras layers – Bidirectional, LSTM, GRU, Dense, Dropout

V. RESULTS

In order to determine which consumers are eligible for loans, we have compared various deep learning algorithms and classification algorithms in this project: Bi-LSTM (Bidirectional Long Short-Term Memory), GRU (Gated Recurrent Unit), Logistic Regression, and Probit Model.Our findings demonstrated that, when compared to other models, GRU had the highest accuracy. GRU is more accurate than Bi-LSTM, which is just about 94% more accurate on average. Although the accuracy of logistic regression is similarly 97%, GRU is more suited for sequence prediction tasks like predicting loan approval because it can capture temporal dependencies, handle nonlinear relationships, manage noisy data, and scale to huge datasets.

Bi-LSTM

```
Accuracy  :: 0.95
Precision :: 0.81
Recall    :: 0.81
Specificity :: 0.97
F1score   :: 0.81
```

	precision	recall	f1-score	support
0	0.97	0.97	0.97	174
1	0.81	0.81	0.81	26
accuracy			0.95	200
macro avg	0.89	0.89	0.89	200
weighted avg	0.95	0.95	0.95	200

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 786-794 www.ijaem.net ISSN: 2395-5252

GRU

```
Accuracy  :: 0.97
Precision :: 0.89
Recall    :: 0.92
Specificity :: 0.98
F1score   :: 0.91
              precision    recall  f1-score   support

           0       0.99      0.98      0.99       174
           1       0.89      0.92      0.91        26

    accuracy                           0.97       200
   macro avg       0.94      0.95      0.95       200
weighted avg       0.98      0.97      0.98       200
```

Logistic Regression

```
Accuracy  :: 0.98
Precision :: 1.00
Recall    :: 0.88
Specificity :: 1.00
F1score   :: 0.94

[[174   0]
 [  3  23]]
```

Probit

```
              precision    recall  f1-score   support

           0       0.86      0.97      0.91       500
           1       0.62      0.23      0.34       100

    accuracy                           0.85       600
   macro avg       0.74      0.60      0.63       600
weighted avg       0.82      0.85      0.82       600
```

ANN

```
Accuracy  :: 0.85
Precision :: 0.68
Recall    :: 0.19
Specificity :: 0.98
F1score   :: 0.30
              precision    recall  f1-score   support

           0       0.86      0.98      0.92       500
           1       0.68      0.19      0.30       100

    accuracy                           0.85       600
   macro avg       0.77      0.59      0.61       600
weighted avg       0.83      0.85      0.81       600
```

In this paper, we present a thorough investigation and evaluation of the loan approval prediction model using a variety of machine learning algorithms and methods. We processed and analyzed the data using the Kaggle Loan Train dataset using a variety of techniques, including text cleaning, feature engineering, standardization, label encoding, and models including GRU, Bi-LSTM, and Logistic Regression.The study emphasizes the significance of including a wide range of characteristics, including elements like payment history, income, and work status, into the loan acceptance prediction model. The model can offer a more thorough evaluation of loan applicants by include these variables, which can help financial organizations make wise lending decisions.Our findings show that the accuracy levels of our various models are different. GRU has a 97% accuracy rating, whereas Bi-LSTM provides a 95% accuracy rating and logistic regression provides a 97% accuracy rating.

V. CONCLUSION AND FUTURE SCOPE

In conclusion, in this research paper we have done intensive study on and analysis of the risk model of subprime mortgages.Financial institutions may improve by implementing the subprime mortgage risk model to reduce risk and decrease the chances of default. To sustain the risk model's effectiveness in a market that is frequently –, it is crucial to understand the inherent risk associated to subprime mortgages. The risk model must additionally be upgraded periodically. Following are the different accuracy rates: Model for probit:85% ANN:85%.

The lending sector may undergo a revolution as a result of the immense future possibilities of loan approval prediction.Lenders can employ blockchain technology to verify the accuracy of data used to forecast loan acceptance.Chatbots and virtual assistants can be used by lenders to communicate with borrowers, gather information, and respond to inquiries. On the basis of the borrower's financial background, these bots can also offer tailored loan recommendations using machine learning algorithms.Lenders can improve their loan approval prediction models by using alternate data sources, such as social media data. This can assist lenders in locating potential customers who may lack a typical credit history but are nevertheless deserving of financing.

REFERENCES

[1] Loan Prediction System Using Machine Learning Anant Shinde1, Yash Patil2, Ishan Kotian3, Abhinav Shinde4 and Reshma Gulwani5 1,2,3,4Department of Information Technology, RAIT, Nerul, India 5 D.Y. Patil Deemed to be University, Ramrao Adik Institute of Technology, Nerul, Navi Mumbai, India

[2] Loan Eligibility Prediction using Machine Learning Algorithms SachinMagar1, N.S. Nikam2, Nilesh Taksale3, Suprem Hajare4 E&TC Department, SKNCOE, SPPU Pune, India

[3] Predicting Bank Loan Eligibility Using Machine Learning Models and Comparison Analysis Miraz Al Mamun Department of Economics & Decision Sciences, Afia Farjana and Muntasir Mamun Department of Computer Science, The University of South Dakota Vermillion, South Dakota, USA

[4] PREDICTING LOAN APPROVAL USING ML Nikhil Bansode*1, Adarsh Verma*2, Abhishek Sharma*3, Varsha Bhole*4 *1,2,3,4University Of Mumbai Department of Information Technology, A.C Patil College of Engineering, Navi Mumbai, India
[5] Bank Loan Approval Prediction Using Data ScienceTechnique (ML) (Subhiksha Vaishnavi, Shalini, Mr. N. Manikandan)- May 2022
[6] Loan Approval Prediction Shubham Nalawade1, Suraj Andhe1, Siddhesh Parab1, Prof. Amruta Sankhe2, 1BEStudent, Information Technology, Atharva College of Engineering, Mumbai 2 Assistant Professor, Information Technology, Atharva College of Engineering, Mumbai
[7] Comparative Analysis of Customer Loan Approval Prediction using Machine Learning Algorithms,Praveen Tumuluru,Lakshmi Ramani Burra,M. Loukya,S. Bhavana,H.M.H. CSaiBaba,N Sunanda
[8] Loan default prediction using decision trees and random forest (Mehul Madaan, Aniket Kumar, Chirag Keshri, Rachna Jain and Preeti Nagrath)-OCT, 2021
[9] An Approach for Prediction of Loan Approval Using Machine Learning Algorithm (Ms. Kathe Rutika Pramod, Ms. Dapse Punam Laxman, Ms. Panhale Sakshi Dattatray, Mr. Ghorpade Dinesh B., Ms. Avhad Pooja Prakash)- JUNE,2021
[10] Loan Approval Prediction using Machine Learning Algorithms Approach (Nitesh Pandey, Ramanand Gupta, Sagar Uniyal, Vishal Kumar)-JUNE,2021
[11] Accurate loan approval prediction based on machine learning approach (J. Tejaswini, T. Mohana Kavya, R. Devi Naga Ramya, P. Sai Triveni Venkata Rao Maddumala)- April, 2020
[12] Exploring the Machine Learning Algorithm for Prediction the Loan Sanctioning Process (E. Chandra Blessie, R. Rekha)- NOV,2019
[13] Loan approval prediction using KNN, Decision Tree and Naïve Bayes models (VeeraballiNagajyothi) NOV-DEC,2019
[14] Loan Sanctioning Prediction System (Aditi Kacheria, Nidhi Shivakumar, Shreya Sawkar, ArchanaGupta)- September2016http://www.ijsce.org/wp-content/uploads/papers/v6i4/D2904096416.pdf
[15] Loan Approval Prediction based on Machine Learning Approach (Kumar Arun, Garg Ishan and Kaur Sanmeet) May-June 2016https://www.iosrjournals.org/iosrjce/papers/conf.15013/Volume%203/4.%2018-21.pdf?id=7557
[16] PredictingtheProbabilityofLoan-Default:AnApplicationofBinaryLogisticRegression (Kwofie, Charles & Owusu-Ansah, Caleb & Boadi, Caleb)-2015An Application of Binary Logistic Regression. Research Journal of Mathematics and Statistics. 7. 46-52. 10.19026/rjms.7.2206. https://bit.ly/345awgq
[17] L. Pullagura, N. Kittad, G. Diwakar, V. Sathiya, A. Kumar and M. S. Yalawar, "ML based Parkinson’s Disease Identification using Gait Parameters," 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, 2022, pp. 561-566, doi: 10.1109/ICACRS55517.2022.10029281.
[18] V. Suresh Kumar, LokaiahPullagura, NalliVinaya Kumari, S. Pooja Nayak, B. Padmini Devi, Adnan Alharbi, Simon AtuahAsakipaam, "Internet of Things-Based Patient Cradle System with an Android App for Baby Monitoring with Machine Learning", Wireless Communications and Mobile Computing, vol. 2022, Article ID 1140789, 11 pages, 2022. https://www.hindawi.com/journals/wcmc/2022/1140789/ https://doi.org/10.1155 /2022/1140789.
[19] Rakesh Kancharla, Venkata Rao Maddumala, T. V. N. Prasanna, LokaiahPullagura, Ratna Raju Mukiri, M. Viju Prakash, "Flexural Behavior Performance of Reinforced Concrete Slabs Mixed with Nano- and Microsilica", Journal of Nanomaterials, vol. 2021, Article ID 1754325, 11 pages, 2021. https://www.hindawi.com/journals/jnm/2021/1754325/ https://doi.org/10.1155/2021/1754325.
[20] LokaiahPullagura, Dr. Anil Kumar, “Analysis of Train Delay Prediction System based on Hybrid Model”, Journal Of Advanced Research in Dynamical & Control Systems, Vol. 12, 07-Special Issue,2020. https://www.jardcs.org/abstract.php?id=5860 https://doi.org/10.5373/JARDCS/V12SP7/20202433