Abstract

This study investigates the use of gradient boosting and random forest algorithms in the categorization of credit scores. Data on consumer creditworthiness, including payment history, credit use, and other elements, are included in the dataset utilised in this study. To effectively categorise clients into various credit score categories, the study evaluates the effectiveness of gradient boosting and random forest algorithms. Findings indicate that both systems predict credit ratings accurately, with gradient boosting producing somewhat superior outcomes. The study also looks at the significance of various indicators in predicting credit ratings and offers insights into the elements that influence creditworthiness the most. The results of this study have consequences for how credit scores are assessed and can assist financial organisations in making better lending decisions.

Introduction

An individual's creditworthiness is represented numerically by their credit score, which is based on their credit history. Lenders use it to evaluate the danger of giving a borrower a loan. Many elements, including payment history, credit use, length of credit history, types of credit accounts, and new credit applications, are considered when calculating a credit score.

According to Al-Zoubi, 2018, the practice of classifying people or organisations into various creditworthiness categories based on their credit histories and financial behaviour is known as credit score classification. Credit scores are derived using information on recent credit inquiries, payment history, credit use, length of credit history, and various types of credit accounts. Financial institutions, lenders, and other organisations use these ratings to determine the risk of providing money to a person or company. The more creditworthy a person or firm is deemed to be, the more probable it is that they will be offered advantageous loan terms and interest rates. Conversely, those with lower credit scores can be viewed as higher-risk borrowers and subjected to higher interest rates or perhaps denied credit entirely because of this.

Often, credit ratings are divided into several groups or ranges. Depending on the credit bureau and scoring algorithm being utilised, these categories might change. Better loan terms, such as lower interest rates and bigger credit limits, might result from having a higher credit score, which is typically regarded as more favourable. On the other hand, a lower credit score could lead to less advantageous loan terms or perhaps getting turned down for credit entirely.

A chronological order may be created by grouping the transactions in a person's transaction history. Predicting a category for the sequence based on inputs that span space or time is difficult by this. In feature-based sequence classification, each transaction record is represented by a feature vector, making it the most often used method. The performance of the models is limited to the utmost extent feasible by the correctness of the representation, which depends on the user's past knowledge. The characteristics of a single transaction and the order in which transactions occur must inevitably be reduced. (Gu, X., 2018).

Anybody trying to increase their creditworthiness or apply for credit should understand how credit scores are categorised. People may evaluate their credit risk and, if required, take action to raise their score by understanding the range in which their score falls. Individuals may also make wise choices when applying for credit or negotiating loan conditions by learning how lenders interpret various credit score ranges.

The Real-World Problem

Individuals' lack of understanding and knowledge of credit ratings and their significance can result in unfavourable loan terms, lost credit opportunities, and money issues. People must be made aware of credit score classification and how it affects financial performance to remedy this problem.

By leveraging historical credit data to forecast creditworthiness, machine learning algorithms can automate the categorization of credit scores. Machine learning algorithms can identify patterns and characteristics linked to successful borrowers after being trained on a variety of data, including income, assets, and credit history. They can also spot irregularities and possible fraud, making them a more precise and dependable way to assess creditworthiness.

Machine learning's use in credit score categorization has the potential to completely change how lenders assess prospective borrowers. It can aid lenders in making more knowledgeable judgements, lower the likelihood of defaults, and enhance the borrowers' entire loan experience. It is crucial that financial institutions use this technology and capitalise on its promise to deliver more precise and dependable credit evaluations.

Project Aim and Objectives

Develop an accurate, fair, and transparent credit scoring system using the Random Forest Classifier algorithm.

- Provide a reliable credit scoring system that can categorise people into various credit score ranges in accordance with several criteria, including payment history, credit usage, length of credit history, types of credit accounts, and new credit applications.
- Create a machine learning model that can accurately and precisely categorise credit ratings using the Random Forest Classifier technique.
- To guarantee that the resultant credit ratings are not skewed based on criteria like ethnicity, gender, or age, train the model on a broad, diversified dataset that contains a range of demographic and financial information.
- By assessing the model's performance and looking for any potential bias against any group or individual based on demographic or financial attributes, evaluate the model's fairness and equality.

Adopted Artificial Intelligence Approach

In this respect, we analysed the many most recent advancements in the classification of credit scoring, which has undergone several classification methodologies over the last ten years. A benchmarking analysis using 41 classifiers on eight credit-scoring datasets and several ensemble selection approaches was published in 2015 by Lessmann et al. Six different indications have been used to gauge the scorecard's correctness. They haven't looked at the Taiwan credit data, but they have examined the financial impact of the various scorecards.

Another body of research creates a framework for evaluating and measuring ML models' transparency, explicability, and interpretability. For instance, Murdoch, W. J., 2019, proposes three variables for model evaluation: prediction accuracy, descriptive accuracy, and relevancy, with relevance assessed relative to a human audience. However, Murdoch does not provide quantitative models. The articles include several examples from the actual world to show how practitioners may utilise the framework to assess and comprehend interpretations.

The dataset from an Iranian bank was used by Jafarpour et al. to concentrate on customer relationship management. By examining the interaction between consumers' requirements and banks via a variety of channels, they created a customer relationship management model that resulted in an equation to anticipate loan customers. Banks and loan firms can use this calculation.

Recurrent neural networks (RNN) and convolutional neural networks are becoming increasingly well-liked for model-based sequence classification applications due to the development of deep learning. Based on the individual time steps of the given sequence data, RNN generates predictions. Two improved RNN implementations, Long Short-Term Memory (LSTM) and GRU, have been effectively applied to a variety of natural language processing applications, including speech recognition and sequence tagging.

In the 2018 study by Jurgovsky, J., a static learner (random forest) and a sequence learner (LSTM) were evaluated to identify credit fraud. The study's findings suggested that combining feature-based learning with model-based sequence learning might enhance the detection performance since the frauds detected by one method were consistently distinct from those detected by the other. In our research, we make use of a unique CNN-based structure that can combine a neural network with unprocessed sequences and the feature matrix that results from that to learn features.

Artificial Intelligence Approach Implementation

The offered dataset includes 5 rows and 28 columns of data on creditworthiness. The Credit Score column, which is used to calculate a person's credit score, is the dataset's target variable. Financial firms use credit scores to assess a person's risk of borrowing money from them and to determine whether they will be able to repay the credit or loan they are applying for. To comprehend the elements that affect creditworthiness and create prediction models for credit score categorization, the dataset may be utilised for analysis and modelling. The supplied link will take you to a page where you may download the dataset in zip format.

For the implementation of the proposed model, I have planned to use the Gradient Boost Classifier and Random Forest Classifier as this problem specifically relates to the Classification of credit scores. I have divided the dataset into a ratio of 70:30 for train and test respectively. We have kept the credit score column in the test value and the rest columns are taken into train variables.

Gradient Boosting

According to Saini A, 2018, A well-liked machine learning approach called gradient boosting may be used to classify data, including credit score data. Gradient boosting's main principle is to train a series of ineffective learners (usually decision trees) on the data repeatedly, with each successive learner being taught to fix the mistakes made by the preceding ones.

When classifying credit scores, the input data may include characteristics like an individual's income, job history, credit history, and other pertinent details. The borrower's credit score, a discrete variable that represents the degree of risk involved in lending money to that person, is the output variable.

Typically, the credit score classification gradient boosting algorithm might go like this (Kurama, V., 2020):

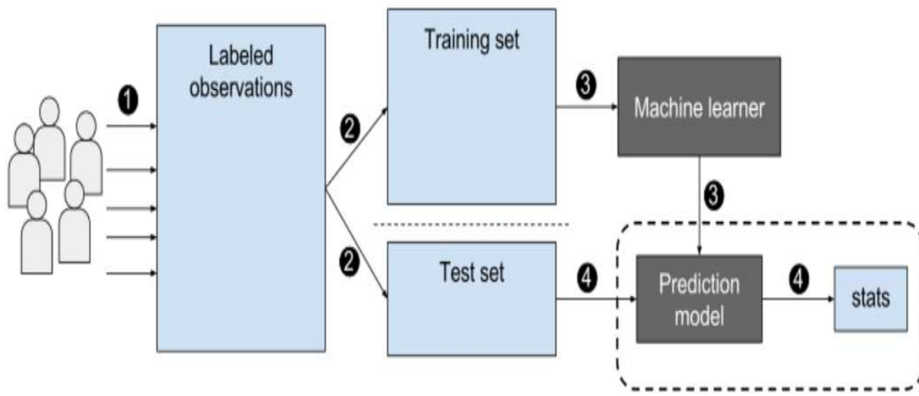

Figure 1: Gradient Boosting Working

Make a training set and a validation set out of the data. This enables you to train the model on a subset of the data and assess its effectiveness on an additional, hidden subset.

1) With the training data, create a first decision tree. Based on the supplied attributes, this tree will forecast values and provide the projected credit score.
2) Determine the first decision tree's validation set error rate. This provides information on the tree's performance and where errors are being made.
3) Using the training set, create a second decision tree with the goal of fixing the mistakes the first tree made.
4) Use a weighted average to combine the predictions of the initial and new trees. This provides you with a fresh set of predictions that ought to outperform those made using just the original tree.
5) Compute the combined predictions' errors on the validation set. Replicate steps 4-5 with a different tree if mistakes are still too high. If not, halt and produce the last round of forecasts.
6) To categorise borrowers into different credit score groups, use the final set of forecasts.

Since it can manage complicated, non-linear interactions between the input characteristics and the output variable, gradient boosting is an effective method for classifying credit scores. Furthermore, it frequently achieves excellent accuracy with modest quantities of training data, which makes it a helpful tool for banks and other lending organisations.

Random Forest Classifier

According to Sruthi, E. R., 2021, Another well-liked machine learning technique that may be used to categorise credit scores is Random Forest. A huge number of decision trees are constructed using this ensemble learning technique, and their predictions are then combined to get the final result.

When classifying credit scores, the input data may include characteristics like an individual's income, job history, credit history, and other pertinent details. The borrower's credit score, a discrete variable that represents the degree of risk involved in lending money to that person, is the output variable.

Typically, the Random Forest algorithm for credit score categorization might go like this Simplilearn, 2021:

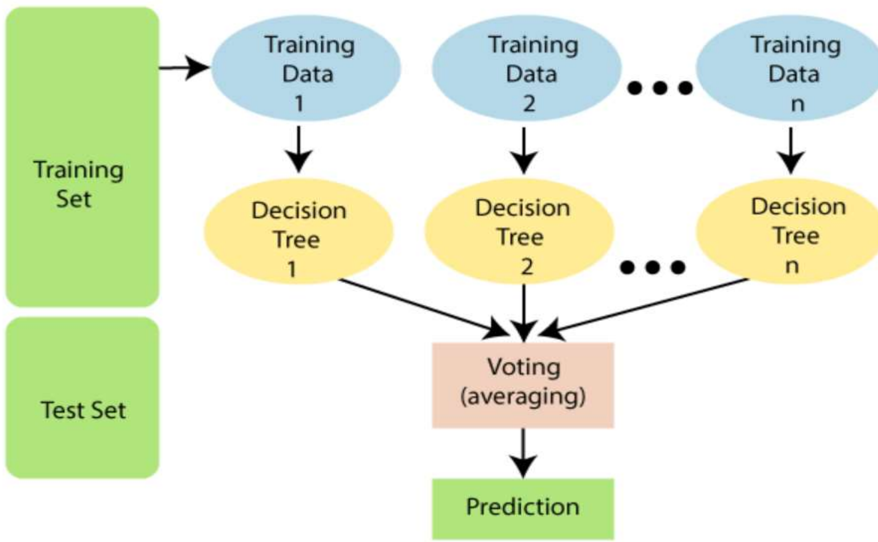

Figure 2: Random Forest Working

1. Make a training set and a validation set out of the data. This enables you to train the model on a subset of the data and assess its effectiveness on an additional, hidden subset.
2. Choose at random a subset of the features and training data to be used in the decision tree. Bagging or bootstrap aggregating is the term for this.
3. Create a decision tree using the chosen subset of characteristics and data.
4. Build several decision trees, each utilising a distinct subset of data and characteristics, by repeating steps 2-3.
5. Run the input data for each new borrower through each decision tree to get a forecast, then average the predictions using a weighted average.
6. Determine the number of prediction mistakes on the validation set. Describes the model's performance and its areas of weakness.
7. To categorise borrowers into different credit score groups, use the final set of forecasts.

Since it can manage complicated, non-linear interactions between input characteristics and output variables, Random Forest is an effective method for classifying credit scores. Furthermore, it frequently achieves excellent accuracy with modest quantities of training data, which makes it a helpful tool for banks and other lending organisations. Using bagging or bootstrap aggregation also aids in lowering overfitting and enhancing model generalisation abilities.

Evaluation, Results and Discussions

After the implementation of Gradient Boosting, it is observed that the accuracy of the model is 70%. It means the model has performed average on the chosen dataset.

Gradient Boosting

```
In [13]: gb = GradientBoostingClassifier(n_estimators=100, learning_rate=0.1, max_depth=3, random_state=42)
         gb.fit(ctrain, dtrain.ravel())

Out[13]: GradientBoostingClassifier(random_state=42)

In [14]: dpred = gb.predict(ctest)

In [15]: print(classification_report(dtest, dpred))
         print(confusion_matrix(dtest, dpred))

              precision    recall  f1-score   support

        Good       0.58      0.71      0.64      5866
        Poor       0.70      0.67      0.69      9633
    Standard       0.76      0.72      0.74     17501

    accuracy                           0.70     33000
   macro avg       0.68      0.70      0.69     33000
weighted avg       0.71      0.70      0.71     33000

[[ 4153   122  1591]
 [  732  6437  2464]
 [ 2253  2596 12652]]
```

Figure 3: Gradient Boosting

After the implementation of Random Forest on the model it is observed that the accuracy of the algorithm is 80.69%. From the above results, Random Forest performed better on the dataset compared to Gradient Boosting.

```
In [18]: # Make predictions on the testing data
         y_pred = c_rf_model.predict(ctest)

         # Calculate the accuracy of the model
         accuracy = accuracy_score(dtest, y_pred)
         print("Random Forest accuracy =", accuracy)

         Random Forest accuracy = 0.806969696969697
```

Figure 4: Random Forest

After the implementation of the system to identify the system credit score classification the results obtained are as follows: -

```
In [17]: print("Credit Score Prediction : ")
         z = float(input("Annual Income: "))
         y = float(input("Monthly Inhand Salary: "))
         x = float(input("Number of Bank Accounts: "))
         w = float(input("Number of Credit cards: "))
         v = float(input("Interest rate: "))
         u = float(input("Number of Loans: "))
         t = float(input("Average number of days delayed by the person: "))
         s = float(input("Number of delayed payments: "))
         r = input("Credit Mix (Bad: 0, Standard: 1, Good: 3) : ")
         q = float(input("Outstanding Debt: "))
         p = float(input("Credit History Age: "))
         o = float(input("Monthly Balance: "))

         class1 = np.array([[z, y, x, w, v, u, t, s, r, q, p, o]])
         print("Predicted Credit Score = ", c_rf_model.predict(class1))
```

```
Credit Score Prediction :
Annual Income: 30000
Monthly Inhand Salary: 2500
Number of Bank Accounts: 4
Number of Credit cards: 5
Interest rate: 15
Number of Loans: 2
Average number of days delayed by the person: 10
Number of delayed payments: 2
Credit Mix (Bad: 0, Standard: 1, Good: 3) : 3
Outstanding Debt: 10000
Credit History Age: 12
Monthly Balance: 1000
Predicted Credit Score =  ['Standard']
```

Figure 5: Credit Score Prediction System

Conclusion and Future Work

We may draw the conclusion that, at least for the dataset and characteristics utilised in the investigation, Random Forest is a superior method to Gradient Boosting for classifying credit scores. This is because Random Forest outperformed Gradient Boosting in terms of accuracy, achieving a greater accuracy of 80.69%, albeit it is unable to pinpoint the precise causes of this discrepancy without more investigation.

Yet it's vital to remember that there are other metrics to consider when assessing a classification algorithm's effectiveness in addition to accuracy. Depending on the precise objectives and specifications of the credit scoring application, other metrics such as accuracy, recall, and F1 score may also be pertinent.

Overall, the decision between Random Forest and Gradient Boosting (or other classification algorithms) will depend on several variables, including the size and complexity of the dataset, the performance metrics of interest, and the computational resources available for model training and deployment. To discover the optimal model for a given use case, it is always advised to experiment with multiple methods and parameter choices.

Lenders and financial organisations must use the Credit Score Classification System to evaluate potential borrowers' creditworthiness. Future scope and trends in the development of credit score categorization systems include the following as the usage of credit scores becomes more prevalent and significant for financial decisions:

Using machine learning and deep learning, among other Big Data and AI approaches, may help create credit score models that are more reliable and accurate as enormous volumes of data become more readily available.

Real-time Decision-Making: Because of technological improvements, credit score categorization systems may give lenders real-time decision-making skills, enabling them to analyse borrowers' creditworthiness in a timely manner and make wise lending decisions.

References

Al-Zoubi, A. M., Rodan, A. and Alazzam, A., 2018. Classification Model for Credit Data. *In*: *2018 Fifth HCT Information Technology Trends (ITT)*. IEEE, 132–137.

Benyacoub, B., El Bernoussi, S. and Zoglat, A., 2014. Building classification models for customer credit scoring. *In*: *2014 International Conference on Logistics Operations Management*. IEEE, 107–111.

Gahlaut, A., Tushar and Singh, P. K., 2017. Prediction analysis of risky credit using Data mining classification models. *In*: *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. IEEE, 1–7.

Gu, X., Zhou, H. and Fan, L., 2018. Credit scoring based on transaction sequence classification. *In*: *2018 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*. IEEE, 310–3106.

Hindistan, Y. S., Kiyakoglu, B. Y., Rezaeinazhad, A. M., Korkmaz, H. E. and Dag, H., 2019. Alternative credit scoring and classification employing machine learning techniques on a big data platform. *In*: *2019 4th International Conference on Computer Science and Engineering (UBMK)*. IEEE, 1–4.

Jafarpour, H. and Sheikholeslami, H., 2012. New Model of Customer Relationship Management in Iranian Banks. *icbme. yasar. edu. tr*, 1–12.

Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P. E. and He-Guelton, L., 2018. Sequence classification for creditcard fraud detection. *Expert Systems with Applications*, 100.

Kurama, V., 2020. *Gradient boosting for classification* [online]. Paperspace Blog. Available from: https://blog.paperspace.com/gradient-boosting-for-classification/ [Accessed 12 Apr 2023].

Marikkannu, P. and Shanmugapriya, K., 2011. Classification of customer credit data for intelligent credit scoring system using fuzzy set and MC2 — Domain driven approach. *In*: *2011 3rd International Conference on Electronics Computer Technology*. IEEE, 410–414.

Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R. and Yu, B., 2019. Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences of the United States of America* [online], 116 (44), 22071–22080. Available from: http://dx.doi.org/10.1073/pnas.1900654116.

Safiya Parvin, A. and Saleena, B., 2020. An ensemble classifier model to predict credit scoring - comparative analysis. *In*: *2020 IEEE International Symposium on Smart Electronic Systems (iSES) (Formerly iNiS)*. IEEE, 27–30.

Saini, A., 2021. *Gradient boosting algorithm: A complete guide for beginners* [online]. Analytics Vidhya. Available from: https://www.analyticsvidhya.com/blog/2021/09/gradient-boosting-algorithm-a-complete-guide-for-beginners/ [Accessed 12 Apr 2023].

Simplilearn, 2020. *Random Forest Algorithm* [online]. Simplilearn.com. Available from: https://www.simplilearn.com/tutorials/machine-learning-tutorial/random-forest-algorithm [Accessed 12 Apr 2023].

Sruthi, E. R., 2021. *Understand Random Forest algorithms with examples (updated 2023)* [online]. Analytics Vidhya. Available from: https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/ [Accessed 12 Apr 2023].

Sundermeyer, M., Ney, H. and Schluter, R., 2015. From feedforward to recurrent LSTM neural networks for language modeling. *ACM transactions on audio, speech, and language processing* [online], 23 (3), 517–529. Available from: http://dx.doi.org/10.1109/taslp.2015.2400218.