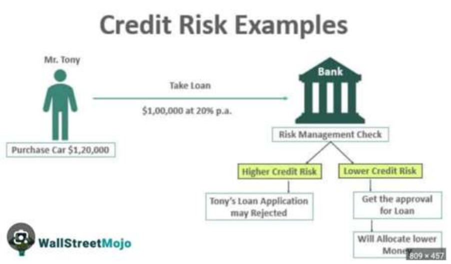
Credit Risk Analysis

Ram Reddy, Narendra, Mehanaz, Lasya

University of Missouri - Kansas City

Introduction

Lending services companies, including banks, credit unions, and online lenders, operate by providing loans to individuals and businesses. However, these companies encounter the challenge of determining the risk associated with each loan application. Assessing the risk level of loans is crucial for making informed decisions regarding approval or denial. If lenders approve loans with high risk levels, it can lead to increased defaults, losses, and financial instability. On the other hand, rejecting potentially profitable loans due to overly conservative risk assessment can hinder business growth and customer satisfaction



Objectives

The main objectives of this Project are:

1. Build and Compare Models:
Create models for Random Forest Classifier and Logistic Regression to forecast loan approval.
Apply supervised machine learning methods to both models.

2. Evaluate the performance of the model:
Analyze how well each model classified the various degrees of loan risk.
To evaluate the efficacy of the model, measure recall, accuracy, precision, and other pertinent parameters.

3. Providing Accuracy:
Examine and contrast the usefulness of Random Forest Classifier and Logistic Regression in evaluating credit risk.
Determine the advantages and disadvantages of each model.
Provide an analysis of the suitability of various machine learning algorithms for controlling credit risk and forecasting loan acceptance.

Methodology

RANDOM FOREST:

Random Forest for credit risk prediction entails creating a collection of decision trees trained on bootstrapped data samples. Each tree forecasts individually and the results are merged by voting or average. Random Forest handles high-dimensional data effectively, provides feature significance insights, and is resistant to overfitting.

Adaboost:
Adaboost for credit risk prediction entails iteratively training weak learners on weighted data while prioritising misclassified samples. Predictions from weak learners are aggregated using weighted voting. Adaboost adapts to complicated data patterns, and prioritises features, but may require preprocessing for noisy data. Overall, it is reliable and effective for credit risk assessment.

Knn:
"KNN for credit risk prediction computes distances between fresh and historical data and classifies them using the majority vote from k-nearest neighbours. It is simple to use, non-parametric, and makes no assumptions about data distribution. However, it necessitates careful scaling and may be sensitive to high-dimensional data."

MLP:
MLP for credit risk prediction entails teaching a neural network with hidden layers to recognise complicated patterns in borrower data. It uses backpropagation to alter weights, learns nonlinear relationships, and necessitates appropriate preprocessing. MLP provides versatility, but it can be vulnerable to hyperparameters and overfitting."

Gradient boosting:

"Gradient Boosting for credit risk prediction builds an ensemble of weak learners progressively, with each rectifying the mistakes of the preceding one. It responds effectively

Results

The main findings of this project are:

1. Analyzing and Comparing Model Scores

Display a side-by-side comparison of the Random Forest Classifier and Logistic Regression models' scores.
Show measures like recall, accuracy, precision, F1-score, and ROC AUC to give a thorough analysis of model performance.

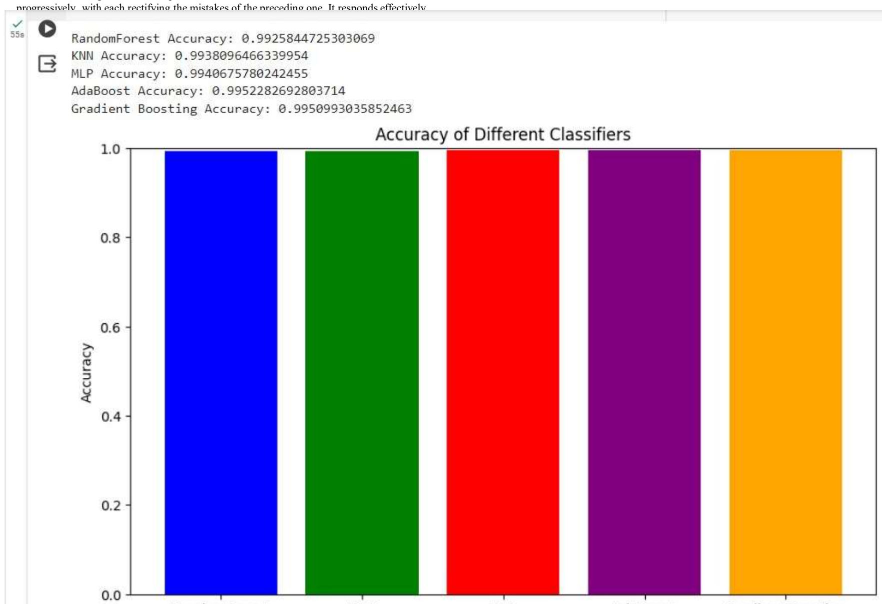
2. Model Performance Metrics Visualization:

Provide graphical depictions of the performance indicators for both models, such as line graphs or bar charts.
To highlight the advantages and disadvantages of each model, visualize the accuracy, precision, and recall ratings.
The trade-off between true positive rate and false positive rate for various categorization criteria may be illustrated using ROC curves.

3. Examination of Projected Loan Approval Results:

Talk about the results that each model predicts for loan approval.
Draw attention to any patterns or trends seen in the categorization findings.

Describe the efficiency of each model in terms of correctly identifying loan risk levels and forecasting loan acceptance.
In conclusion, the difficulty lending services providers have determining loan risk emphasizes how crucial it is to use supervised machine learning methods to create prediction models. These models help to effectively manage financial risks and the general stability of lending institutions by correctly anticipating loan approval and classifying loan risk levels.



Caption

Significance and Innovation of Results

1. Model Accuracy and Performance:
Analyze how well machine learning models like Random Forest Classifier and Logistic Regression predict loan approval. Examine each model's accuracy, precision, recall, and F1 score and contrast them with industry standards.

2. Qualitative Significance and Comprehensibility:
Examine the significance of the many variables that the models employ to forecast loan acceptance. Determine which features most significantly affect the model's predictions, and then talk about how interpretable they are about credit risk assessment.

3. Robustness and Generalization of the Model:
Evaluate the models' resilience by putting them to the test on several datasets or using cross-validation methods. Talk about how effectively the models generalize to new data and whether they remain stable over time or in changing economic environments.

Conclusion

To forecast loan approval, models for Random Forest Classifier and Logistic Regression were used and compared. A thorough examination of the models' performance provided insights into how well they estimate credit risk. The comparison brought to light the advantages and disadvantages of each model, offering lending services firms helpful direction in choosing the right algorithms for predicting loan acceptance.

This research might have a positive effect on loan approval procedures and reduce lending institutions' financial risks. Accurate prediction models help businesses make better decisions, which lowers the risk of losses and defaults. In the end, this promotes a better financial climate for lenders and borrowers alike by increasing the general stability and profitability of loan services organizations.

References

https://www.finalyse.com/blog/credit-risk-analysis-with-machine-learning
https://journalofbigdata.springeropen.com/articles/10.1186/s40537-024-00882-0
https://www.researchgate.net/publication/373108628_Machine_Learning_for_Credit_Risk_Prediction_A_Systematic_Literature_Review

Acknowledgments

Syed Jawad Shah