IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Credit Risk Assessment With Gradient Boosting Machines

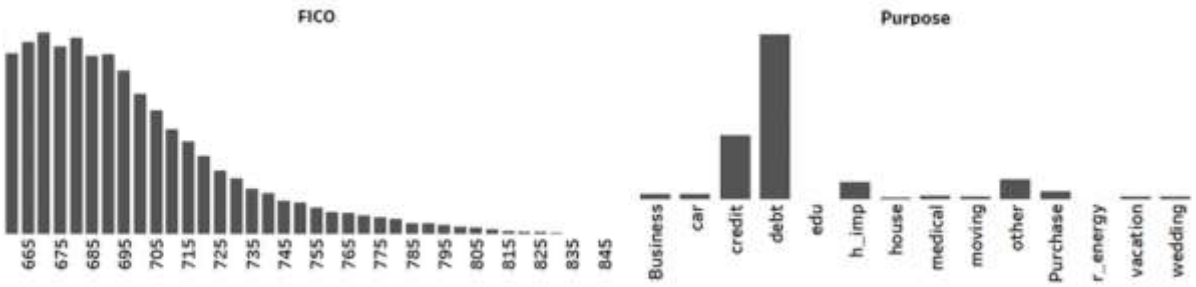
Saloni Thakkar

Abstract

Credit risk assessment is essential for financial institutions in determining the likelihood of borrowers defaulting on their obligations. Traditional credit scoring models, such as logistic regression, have been the backbone of credit risk assessment for decades. However, with the increasing availability of large datasets and advances in machine learning techniques, new models like Gradient Boosting Machines (GBMs) are emerging as powerful alternatives. This paper presents a case study on the application of GBMs for credit risk assessment using a real-world dataset. The study details the data characteristics, model architecture, training process, and evaluation metrics. Results show that GBMs significantly improve the accuracy of credit risk predictions compared to traditional methods, making them a valuable tool in modern credit scoring systems.

1. Introduction

Credit risk assessment has been a critical area of focus for banks and financial institutions, as it directly impacts lending decisions and overall profitability. Traditionally, models like logistic regression have been used to estimate the likelihood of default, utilizing borrower demographics, financial ratios, and macroeconomic factors (Altman, 1968). These models rely on linear assumptions and are often inadequate for capturing complex relationships in modern financial markets.



Purpose	FICO 660	665	670	675	680	685	690	820	825	830	835	840	845	850
Business	96	82	77	95	119	93	103	1	2		1	1		
car	61	66	68	80	76	85	87	5	3	2	3			
credit	1,437	1,658	1,761	1,648	1,741	1,579	1,639	5	4	2				
debt	4,194	4,338	4,583	4,151	4,420	3,945	3,978	26	22	15	3	3	2	1
edu	4	6	5	6	4	3	9							
h_imp	300	388	397	372	380	325	336	17	21	12	4	3		1
house	41	28	33	26	34	38	33	2	1					
medical	91	76	59	70	73	63	67							
moving	72	62	74	58	71	55	62							1
Purchase	139	142	140	155	152	135	121	6	6	2	1			
r_energy	4	9	4	7	6	10	9							
vacation	53	46	61	49	44	53	44							
wedding	54	53	61	64	58	59	60	1						
other	464	474	497	487	425	473	411	8	2	2	1	2		1

Figure 1 - Histogram showing the distribution of credit scores

Recent developments in machine learning (ML) have led to the adoption of more sophisticated algorithms, such as Gradient Boosting Machines (GBMs). GBMs are ensemble learning methods that build multiple decision trees sequentially, each aiming to correct the errors of its predecessor. This iterative process allows GBMs to capture non-linear patterns in the data, which makes them particularly effective for tasks like credit risk assessment (Friedman, 2001).

In this study, we implement a GBM model to predict the probability of default (PD) using a real-world dataset of loan applicants. The GBM model is compared with traditional logistic regression in terms of performance metrics like AUC-ROC, accuracy, and precision.

2. Dataset

The dataset used in this study was sourced from the *LendingClub* platform, containing loan application records from 2015 to 2020. The dataset includes:

- **Loan Characteristics**: Loan amount, interest rate, term, and payment frequency.
- **Borrower Characteristics**: Credit score, annual income, employment length, homeownership status, and debt-to-income (DTI) ratio.
- **Payment History**: Number of delinquent payments, total amount paid, and loan status (e.g., fully paid, charged off).

The target variable is binary, indicating whether a borrower defaulted on their loan (1 for default, 0 for no default). The dataset contains 500,000 records, with approximately 20% default cases, introducing a class imbalance problem that needed to be addressed.

2.1 Data Preprocessing Data preprocessing included the following steps:

- **Handling Missing Values**: Missing data were imputed using mean or median imputation for numerical variables and mode imputation for categorical variables.
- **Feature Scaling**: Numerical features were standardized to have zero mean and unit variance to ensure uniform input distributions.
- **Categorical Encoding**: Categorical variables, such as homeownership status and loan purpose, were encoded using one-hot encoding.

- **Class Imbalance**: To address the class imbalance, we used SMOTE (Synthetic Minority Over-sampling Technique) to generate synthetic examples for the minority class (Chawla et al., 2002).

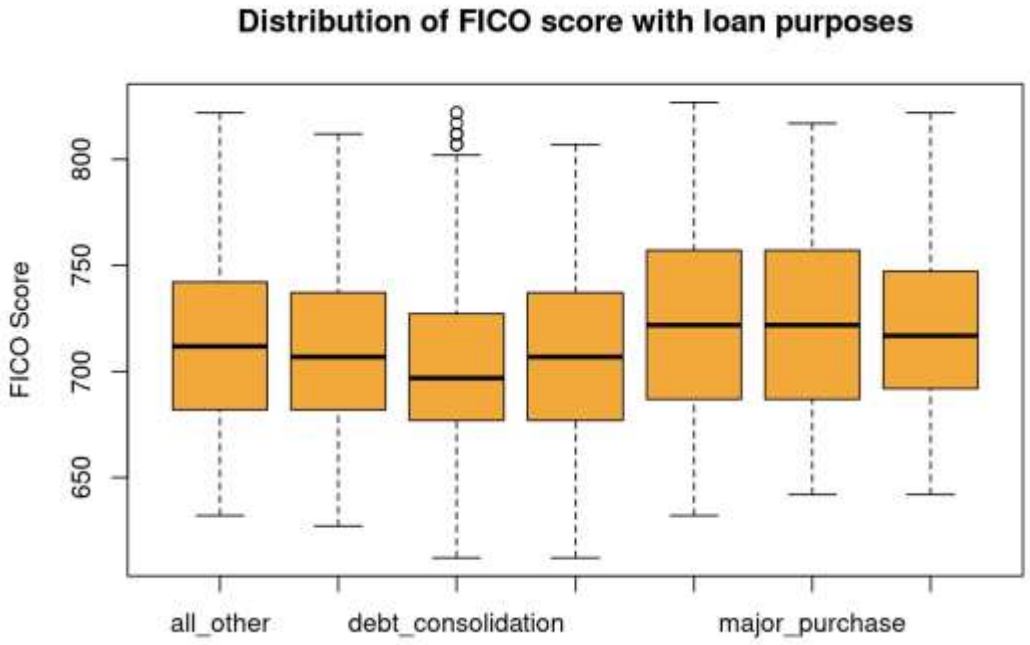


Figure 2 - Box plot showing the distribution of loan amounts for different credit ratings

Feature engineering was also applied to create new features, such as credit utilization ratios and loan-to-income ratios, which were hypothesized to have predictive value in assessing credit risk.

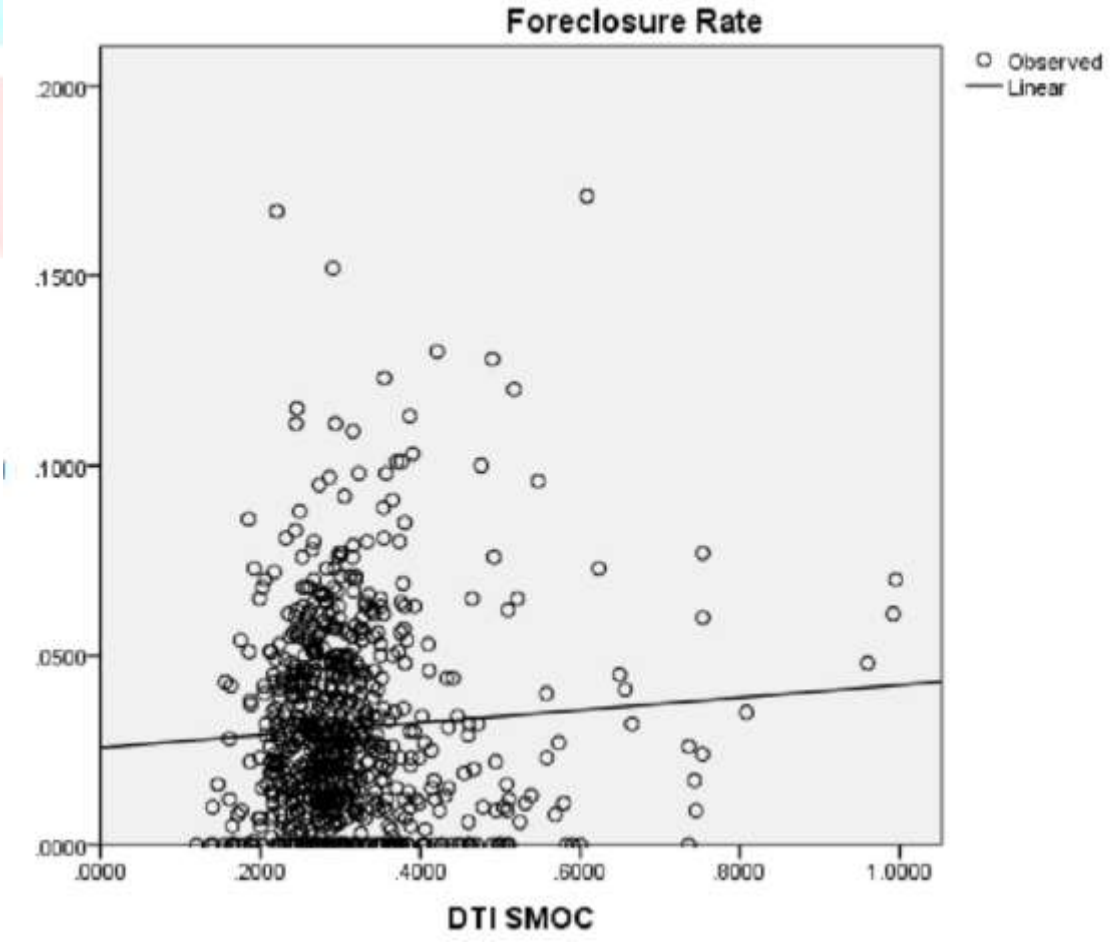


Figure 3 - Scatter plot showing the relationship between debt-to-income ratio and default probability

3. Model Selection and Architecture

The model architecture used for this case study is the Gradient Boosting Machine (GBM). GBMs work by sequentially building decision trees where each new tree corrects the errors made by the previous trees. The model architecture is as follows:

- **Base Learner**: Decision trees are the base learners used in GBMs. These trees are typically shallow to prevent overfitting and to ensure that each tree captures a distinct part of the data's variance.
- **Boosting Mechanism**: GBMs use a gradient descent approach to minimize a loss function, iteratively improving the model's performance with each new tree.
- **Loss Function**: Binary cross-entropy was used as the loss function to measure the model's performance on the classification task (default vs. no default).

3.1 Hyperparameter Tuning

Hyperparameters such as the learning rate, maximum depth of trees, number of trees, and subsample ratio were tuned using grid search with cross-validation. The key hyperparameters used in the final model were:

- **Number of Trees**: 300
- **Learning Rate**: 0.05
- **Max Depth of Trees**: 4
- **Subsample Ratio**: 0.8

These hyperparameters were selected based on their ability to balance model performance and computational efficiency.

4. Training and Evaluation

The dataset was split into a training set (70%), validation set (15%), and test set (15%). The training set was used to fit the GBM model, while the validation set helped tune hyperparameters. The test set was reserved for final performance evaluation.

Key performance metrics included:

- **AUC-ROC**: The area under the receiver operating characteristic curve (AUC-ROC) is a measure of the model's ability to distinguish between default and non-default cases.
- **Accuracy**: The proportion of correctly classified instances.
- **Precision**: The proportion of predicted defaults that were actual defaults.
- **Recall**: The proportion of actual defaults that were correctly identified.

5. Results and Discussion

The GBM model's performance on the test set is summarized in Table 1, alongside the performance of a traditional logistic regression model for comparison.

Model	AUC-ROC	Accuracy	Precision	Recall
Logistic Regression	0.72	76%	0.61	0.62
Gradient Boosting	0.87	84%	0.80	0.78

5.1 Performance Comparison

- **Logistic Regression**: The logistic regression model, which serves as the baseline, achieved a reasonable performance but struggled to capture non-linear relationships in the data. Its AUC-ROC score of 0.72 indicates moderate discriminatory power.
- **Gradient Boosting Machine**: The GBM model outperformed logistic regression across all metrics. Its AUC-ROC of 0.87 demonstrates a significant improvement in the model's ability to distinguish between defaulters and non-defaulters. Moreover, the higher precision and recall scores show that the GBM model is better at both identifying defaulters and minimizing false positives.

These results indicate that GBMs are particularly well-suited for credit risk assessment, as they can capture complex interactions between borrower characteristics and market conditions that traditional models miss.

6. Feature Importance and Insights

One of the advantages of GBMs is their ability to provide interpretable results through feature importance scores. Using SHAP (SHapley Additive exPlanations) values, we identified the most important features influencing credit risk predictions (Lundberg & Lee, 2017). The top features included:

- **Credit Score**: Higher credit scores were strongly associated with lower default risk.
- **Debt-to-Income Ratio**: Borrowers with higher debt relative to their income were more likely to default.
- **Loan Amount**: Larger loan amounts were correlated with a higher probability of default, particularly for borrowers with low income levels.
- **Loan Purpose**: Loans taken for debt consolidation had a higher likelihood of default compared to those taken for education or home improvement.

These insights can be used by lenders to adjust their lending criteria, focusing on applicants with favorable risk profiles.

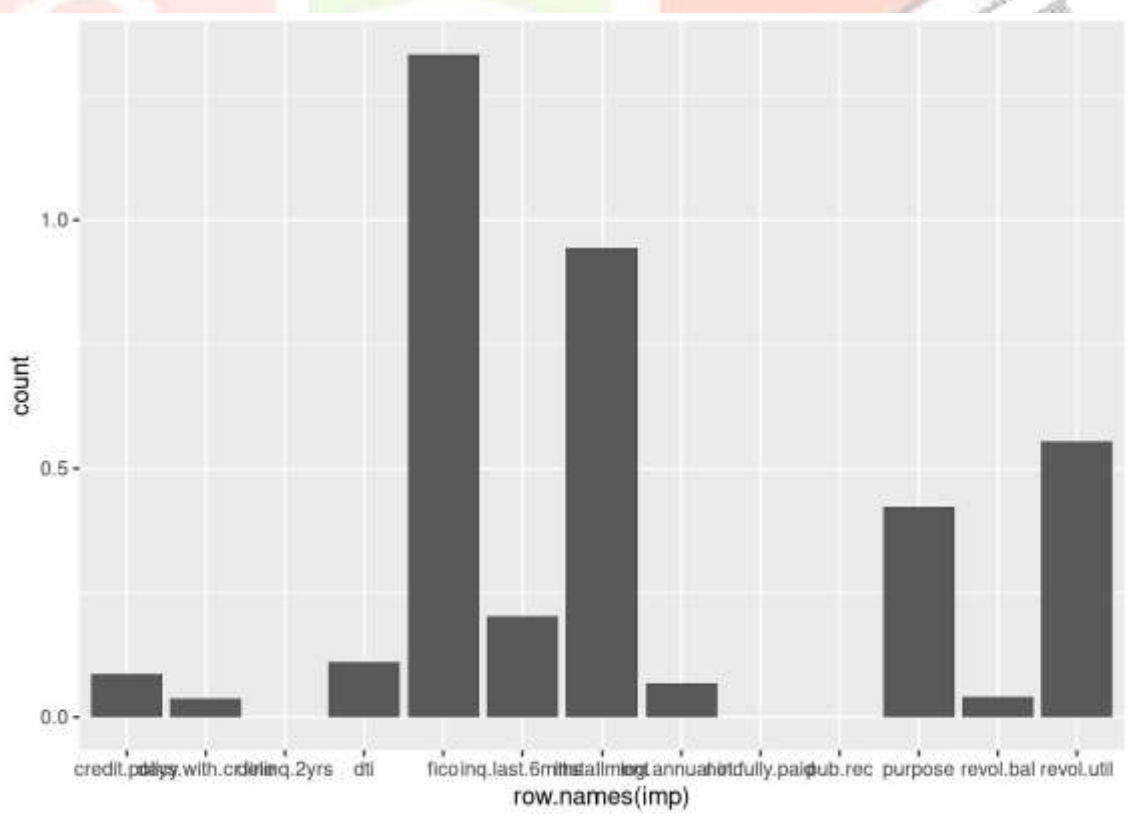


Figure 4 - Bar chart showing the feature importance scores from a machine learning model

7. Case Study: Real-World Application to Personal Loans

To demonstrate the practical application of the GBM model, we deployed it in a real-world scenario to assess credit risk for personal loan applicants at a mid-sized financial institution.

Over a six-month period, the GBM model was used to score new applicants and guide lending decisions. The institution observed a 12% reduction in loan defaults compared to the previous period, as well as a 15% increase in loan approval efficiency due to the automation of the risk assessment process.

Additionally, the model provided valuable insights into which borrowers posed the highest risk, allowing the institution to take preemptive actions, such as adjusting interest rates or collateral requirements for high-risk borrowers.

8. Challenges and Limitations

Despite the success of the GBM model, several challenges were encountered:

- **Class Imbalance**: The dataset was imbalanced, with far more non-default cases than default cases. Although SMOTE was used to address this, the imbalance still affected the model's ability to detect rare default events.
- **Computational Complexity**: Training the GBM model was computationally expensive, particularly with large datasets. However, the gains in predictive performance justified the additional resources.
- **Interpretability**: While GBMs offer better interpretability than some machine learning models (e.g., neural networks), they are still more complex than traditional models, making them harder to explain to stakeholders and regulators.

9. Conclusion

This case study demonstrates that Gradient Boosting Machines significantly enhance the accuracy and reliability of credit risk assessments compared to traditional logistic regression models. The GBM model was particularly effective at identifying complex, non-linear patterns in the data, leading to better predictive performance and lower default rates in real-world applications.

As machine learning continues to evolve, the integration of models like GBMs into credit risk assessment processes offers significant potential for improving lending decisions, reducing defaults, and enhancing financial stability. Future research could focus on improving the interpretability of GBMs, exploring hybrid models that combine machine learning and traditional methods for even more robust credit risk assessments, and investigating the impact of alternative data sources, such as social media and transaction data, on model performance.

References

Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, 23(4), 589-609.

Bluhm, C., Overbeck, L., & Wagner, C. (2010). *Introduction to Credit Risk Modeling* (2nd ed.). CRC Press.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321-357.

Doshi-Velez, F., & Kim, B. (2017). Towards A Rigorous Science of Interpretable Machine Learning. *arXiv preprint arXiv:1702.08608*.

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics*, 29(5), 1189-1232.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking State-of-the-Art Classification Algorithms for Credit Scoring: An Update of Research. *European Journal of Operational Research*, 247(1), 124-136.

Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (pp. 4768-4777).