Enhancing Credit Scoring Models with Bidirectional Dense Transformer and XG Boost

Andre Herndon
Thunderbird School of Global Management
Arizona State University
tdherndo@asu.edu

Ayush Choudhary
Ira A. Fulton Schools of Engineering
Arizona State University
aytxay@asu.edu

Priyanshi Verma
Thunderbird School of Global Management
Arizona State University
pverma27@asu.edu

Thor Abbasi
Thunderbird School of Global Management
Arizona State University
tabbasi@asu.edu

Tathagata Saha
Thunderbird School of Global Management
Arizona State University
tsaha3@asu.edu

ABSTRACT

In credit underwriting, the choice of algorithm used significantly impacts the accuracy and reliability of credit assessments. This paper presents a novel approach to credit scoring using bidirectional dense transformer models with XGBoost as the boosting algorithm and Adam as the optimizer. The proposed bidirectional dense transformer model offers a unique architecture with 16 inputs, followed by dense layers of 128, 256, and 128 neurons, culminating in two outputs: a credit score type and its corresponding probability. We discuss the implementation details and highlight the differences between our approach and existing algorithms in this paper.

1 Introduction

In modern finance, access to credit is often guarded by the algorithms that determine the creditworthiness of an individual. The Federal Reserve has identified access to standard rate credit as a key driver of economic mobility (Davidson 2018). The usage of traditional algorithms such as the FICO Score and VantageScore have allowed millions to access credit and are instrumental in the financial lives of many, yet they are not without flaws. A recent study by Experian and OliverWyman estimated that 91 million Americans are unable to access standard rate credit, in part due to the limitations in traditional scoring algorithms (Hepinstall et al. 2022). These models, while robust, often fail to encapsulate the complexity and nuance of consumer financial behavior, excluding a large swath of prospective borrowers from accessing credit.

However, this same study found that the adoption of enhanced scoring techniques could potentially allow an additional six million Americans to access standard rate credit (Hepinstall et al. 2022), underscoring the urgent need for innovation in credit scoring techniques.

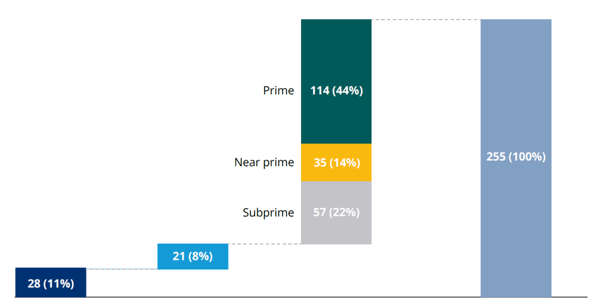

Figure 1: Current state of access to credit (Hepinstall et al. 2022)

Amid this backdrop, advances in machine learning and artificial intelligence present a promising avenue for improving how we do credit scoring. In this study, we aim to employ machine learning techniques, such as bidirectional dense transformer models to gain insights into how these techniques can be used to improve scoring with commonly used underwriting data.

2 Related Work

Prior to deciding which models we were going to employ we examined previous work related to the issue we were aiming to address. Here is what we found:

Ensemble and Boosting Methods:
Boosted decision tree ensembles using techniques like XGBoost have shown superior performance on credit scoring tasks due to their ability to handle large, complex datasets effectively. For instance, Xia et al. (2017) and Mokheleli & Museba (2023) showed the advantages of such models in improving prediction accuracy for credit scoring.

Feature Selection Techniques:
Effective feature selection, as explored by Liang et al. (2015), was found to enhance model performance, which is critical in managing the dimensionality and complexity of data used in credit scoring.

Extensions of Logistic Regression:
Modifications to logistic regression, such as introducing random coefficients to capture non-linear effects, have been proposed to retain interpretability while improving flexibility and accuracy, similar to the aims of our project using a transformer based model.

Systematic Reviews and Comparisons:
Reviews by researchers like Louzada et al. (2016) highlight the importance of novel methods that can process the complexities of modern data sets more effectively than traditional models used in credit scoring, thus enabling the expansion of the scorable credit universe.

Our project extends these findings by implementing advanced machine learning models (bidirectional dense transformers) to enhance the accuracy and economic efficiency of credit scoring, addressing the shortcomings of standard algorithms. This approach promises improvements in predictive performance and broader credit access, reflecting the evolving landscape of financial technology.

3 Data

For our project, we utilized a public credit scoring dataset sourced from Kaggle, titled "Credit Score Classification" that was compiled by Rohan Paris. This dataset was designed to simulate a real-world scenario and contains features commonly used in predictive models when determining credit scores.

Dataset Overview:

- Number of Records: 150,000
- Number of Attributes: 27
- Data Split: Pre-split for training and testing

Data Quality and Preprocessing Efforts:
Upon initial examination, we determined the dataset would require significant preprocessing to correct issues such as missing data (approximately 25% of records had incomplete information), data entry errors, and outliers (such as ages exceeding realistic limits). Our preprocessing steps included:

Cleaning Text and Numerical Data: We removed extraneous characters from text and numerical fields.

Handling Missing Data: Missing entries were imputed using statistical methods where feasible, or records were omitted where necessary.

Outlier Management: Outliers were identified through statistical analysis (e.g., z-scores) and were either corrected or removed to prevent skewing the model's performance.

Feature Transformation: Categorical variables were encoded numerically to facilitate their use in machine learning models, and continuous variables were scaled and normalized to uniform scales.

After conducting these steps our preprocessed data was ready to be fed into our models.

4 Methodology

Our team is leveraging two machine learning models: Bidirectional Transformer Models and XGBoost. The mathematical structure behind these models complements our project's objectives exceptionally well.

Bidirectional Transformer Models (BTM), Such as the prominent BERT (Bidirectional Encoder Representations from Transformers), employ self-attention and bidirectional text processing. This empowers them to decipher the complex relationships between words in a sentence (Devlin et al. 2018). Their attention mechanism determines how much weight to give each word based on its relationship to others. This makes them highly effective at handling the nuances and long-range connections within financial data, which is essential for our project (Vaswani et al. 2017). We plan to leverage these models to extract meaningful features from textual data sources like types of current loans, payment history, and more.

For developing the predictive model, we have chosen XGBoost (Extreme Gradient Boosting), an ensemble algorithm based on gradient boosting (Chen and Guestrin 2016). XGBoost combines multiple weak decision tree models into a strong predictive model by iteratively adding new models to the ensemble, minimizing the overall prediction error. Its underlying math involves optimizing an objective function combining the loss function (e.g., mean squared error for regression, logistic loss for classification) and a regularization term to prevent overfitting (Friedman 2001). XGBoost's ability to handle missing data, efficiency in dealing with large datasets and high-dimensional feature spaces, and proven performance in financial applications make it well-suited for our project (Zheng et al. 2018).

Model Building

After the completion of preprocessing and feature extraction we began to build our model. First, we split our dataset into training, validation, and testing sets. After which, we began to train our model. XGBoost trains multiple decision trees iteratively. Each new model progressively reduces errors made by previously trained trees. The final prediction is a weighted sum of the predictions made by all trees, which helps to improve accuracy and control overfitting.

We utilize grid search and cross-validation techniques to find the optimal settings for XGBoost's parameters, such as the learning rate, number of trees, and depth of trees, to improve model performance.

5 Results

In our analysis, the primary variable of interest was the prediction of Credit_Score, which is classified into "Good," "Standard," or "Poor" categories. The distribution of these categories suggested the presence of imbalanced classes—a common challenge in datasets where the frequency of instances in one class significantly exceeds those in others. Such imbalances can skew classification models towards the majority class, compromising the accuracy of minority class predictions. To address this, we implemented and will be implementing strategies aimed at achieving a more accurate and fair analysis.

METHOD	Training Accuracy	Testing Accuracy	Training Loss	Testing Loss
XGBoost	99	73	35	40
ANN	82	79	45	44
Transformer	83	82	34	53
Transformer +XGB	84	81	39	42
Transformer(2056 neuron)	87	85	27	29

Figure 2: Model Results

Initially, our efforts concentrated on enhancing the precision of our model's predictions for each Credit_Score category. By fine-tuning the model's threshold, we sought to ensure the reliability of identifying "Good" credit mixes, a crucial aspect in scenarios where inaccurate predictions could have considerable impacts. Subsequently, our focus shifted towards optimizing recall to accurately capture instances across all Credit_Score categories, especially the "Poor" ones, indicative of potentially risky financial behaviors.

To achieve a balance between precision and recall, we employed the F1 score as a metric, guiding us to a point where the model effectively balanced both aspects without undue bias towards either. This equilibrium was vital for fair and accurate category prediction.

Furthermore, we utilized the ROC-AUC metric to gauge the model's proficiency in distinguishing between the different Credit_Score categories. A higher AUC value indicated superior model performance in differentiating between categories. The plotting of ROC curves facilitated the selection of models that best balanced true positive rates against false positives.

Given the dataset's imbalanced nature, we placed a particular emphasis on Precision-Recall AUC. This metric was instrumental in assessing our model's performance in predicting less represented Credit_Score categories, ensuring that "Poor" or "Standard" mixes were accurately identified despite their lower occurrence. Adjustments based on precision-recall curves helped improve this metric, aiming for predictions that were not only precise but also equitable across categories.

By applying these metrics, we conducted a thorough evaluation of our model, tackling the task of predicting Credit_Score categories amidst the challenges posed by imbalanced data. This approach ensured that our predictions were accurate and reliable.

6 Discussion

Limitations of FICO:

FICO scores, while widely used in credit assessment, have inherent limitations. They rely heavily on historical financial data, such as payment history and credit utilization, which may not fully capture an individual's current financial situation or future creditworthiness. Moreover, FICO scores may not consider non-traditional data sources or account for socioeconomic factors that could impact an individual's creditworthiness.

Role of Neural Networks in Credit Scoring: Neural networks offer a promising approach to overcome the limitations of traditional credit scoring models like FICO. By leveraging advanced machine learning techniques, neural networks can analyze vast amounts of data, including both traditional financial metrics and alternative data sources, to develop more accurate and comprehensive credit risk models.

Handling Losses with Dense Architectures: The use of dense architectures, such as the bidirectional transformer model with a complex configuration of 2048 activation nodes, can effectively handle losses in credit scoring models. Dense architectures enable the neural network to learn intricate patterns and relationships within the data, leading to better predictive performance and reduced loss.

Potential of Complex Networks: While the dense architecture has shown promising results, there is a possibility that even more complex neural networks could further enhance credit scoring models. A more extensive and sophisticated network could capture even subtle patterns and nuances in the data, potentially improving the accuracy and reliability of credit assessments.

7 Conclusion

In conclusion, in this paper we have explored the potential of leveraging advanced machine learning techniques, specifically bidirectional dense transformer models and XGBoost to enhance the accuracy of credit scoring. Our research showcases the advantages of unique model architectures in addressing the shortcomings of standard credit scoring algorithms. Our results demonstrate machine learning's crucial role in widening credit availability, advancing the field of financial technology. Our future efforts will center on enhancing these models, testing their efficacy across broader data collections, and confirming their ability to revolutionize credit scoring systems.

REFERENCES

1. Andreas Blöchlinger and Marcel Leippold. 2006. Economic benefit of powerful credit scoring. *Journal of Banking & Finance*, 30, 3 (2006), 851–873. https://doi.org/10.1016/j.jbankfin.2005.07.011
2. Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. ACM, New York, NY, USA, 785–794. https://doi.org/10.1145/2939672.2939785
3. Charles Davidson. 2018. Lack of access to financial services impedes economic mobility. *Federal Reserve Bank of Atlanta*. [Online]. Available: https://www.atlantafed.org/economy-matters/community-and-economic-development/2018/10/16/lack-of-access-to-financial-services-impedes-economic-mobility.
4. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805* (2018).
5. Jerome H. Friedman. 2001. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Statist.*, 29, 5 (2001), 1189–1232. https://doi.org/10.1214/aos/1013203451
6. Mike Hepinstall et al. 2022. Experian and Oliver Wyman find expanded data and advanced analytics can improve access to credit for nearly 50 million credit invisible and unscoreable Americans. *Experian plc*. [Online]. Available: https://www.experianplc.com/newsroom/press-releases/2022/experian-and-oliver-wyman-find-expanded-data-and-advanced-analytics-can-improve-access-to-credit-for-nearly-50-million-credit-invisible-and-unscoreable-americans.
7. Fabríccio Louzada, Ana Ara, and Guilherme B. Fernandes. 2016. Classification methods applied to credit scoring: A systematic review and overall comparison. *Surveys in Operations Research and Management Science*, 21, 2 (2016), 117–134. https://doi.org/10.1016/j.sorms.2016.10.001
8. Ts'eliso Mokheleli and Tranos Museba. 2023. Machine learning approach for credit score predictions. *Journal of Information Systems and Informatics*, 5, 2 (2023), 497–517.
9. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems 30*. Curran Associates, Inc., 5998–6008.
10. Aiyi Zheng, Ayman Zaharna, Nicholas L. Mendenhall, Wai K. Wong, Heather D. Moore, and Vilma Kenausis. 2018. XGBoost Machine Learning for Classification and Regression: Applications in Credit Scoring. *arXiv preprint arXiv:1809.04372* (2018).
11. Yang Xia, Chengqi Liu, Yu Li, and Neng Liu. 2017. A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. *Expert Systems with Applications*, 78 (July 2017), 225-241. https://doi.org/10.1016/j.eswa.2017.02.017
12. Liang, D., Tsai, C.-F., and Wu, H.-T. 2015. The effect of feature selection on financial distress prediction. *Knowledge-Based Systems*, 73 (Jan. 2015), 289-297. https://doi.org/10.1016/j.knosys.2014.10.010