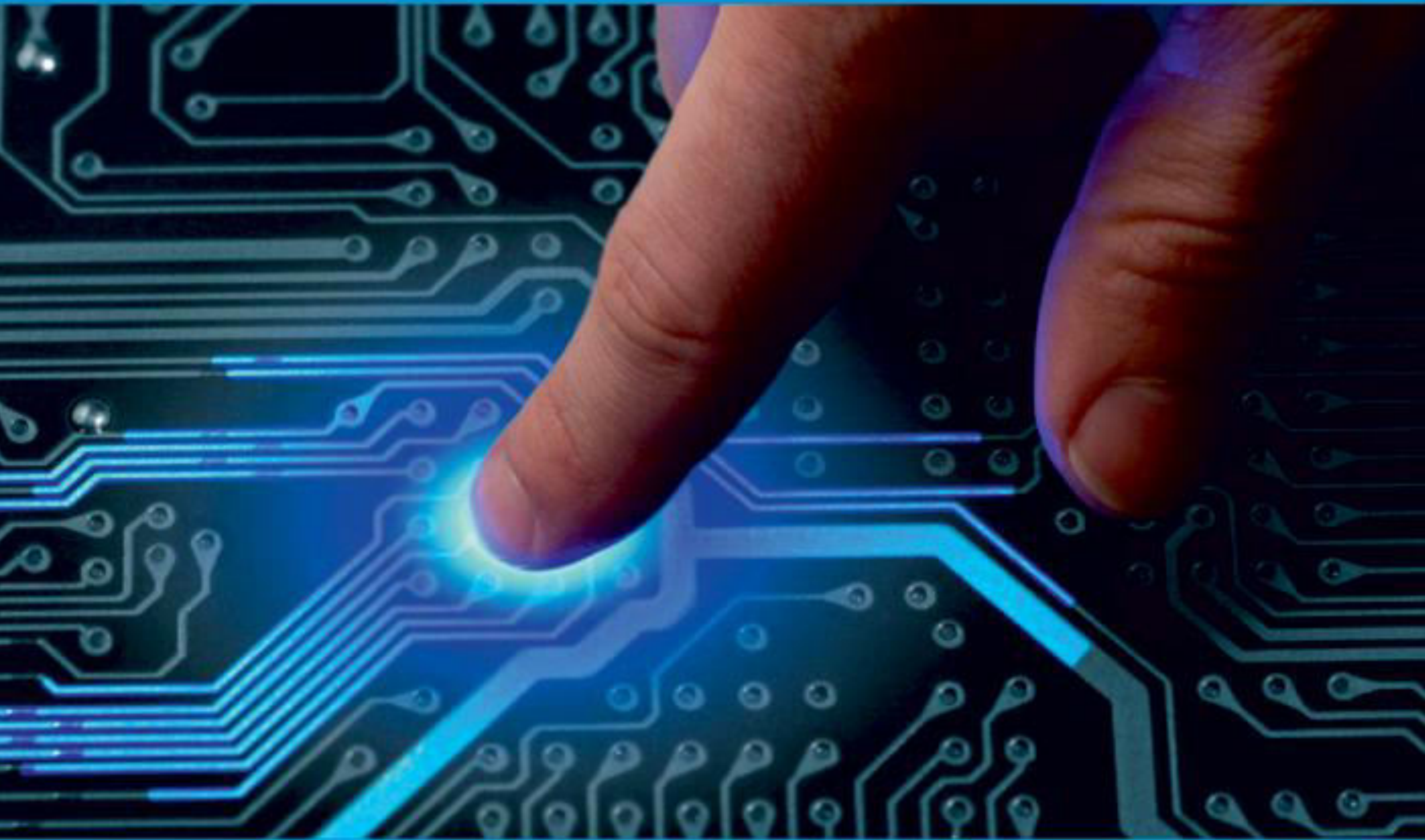
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 11, November 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.625

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Corporate Credit Rating Analysis using Ensemble Technique and Sentiment Analysis

Prof. Hitesh E. Chaudhari[#1], Chitrang R. Potdar[#2], Pranav K. Nimje[#3], Prayag A. Korekar[#4], Harshal S. Bartakke[#5]

Head, Department of Computer Engineering, Padmabhooshan Vasantdada Patil Institute of Technology, Pune, India[1]

Department of Computer Engineering, Padmabhooshan Vasantdada Patil Institute of Technology, Pune, India[2]

Department of Computer Engineering, Padmabhooshan Vasantdada Patil Institute of Technology, Pune, India[3]

Department of Computer Engineering, Padmabhooshan Vasantdada Patil Institute of Technology, Pune, India[4]

Department of Computer Engineering, Padmabhooshan Vasantdada Patil Institute of Technology, Pune, India[5]

ABSTRACT: The credit rating analysis system uses advanced machine learning algorithms such as XGBoost and Ensemble Model to determine creditworthiness based on both traditional and alternative data sources. Boosting algorithms are good at picking up complex patterns because they iteratively build small, sequential models that correct previous errors to result in a highly accurate predictive model. After collecting and preprocessing the data, the system applies feature engineering to extract critical attributes, such as payment history, credit utilization, and transaction behaviors. XGBoost and Ensemble Model then iteratively train multiple rounds, enhancing predictions with each iteration and minimizing errors. The final model outputs a credit score, categorizing borrowers into risk levels (low, medium, high), and continuously updates with new data to adapt to evolving financial behaviors, ensuring regulatory compliance and transparency through explainable AI techniques.

KEYWORDS: Credit Rating, Ensemble Model, Sentiment Analysis, XGBoost, FinBERT, Financial Ratios, Qualitative Analysis, Machine Learning.

I. INTRODUCTION

Many lenders rely heavily on more antiquated scoring methods based simply on the rather narrow confines of credit history and personal income in determining borrower-risk assessment, especially among thinly-filed applicants. Thus, traditional, rule-of-thumb static systems lack much dynamism when responding to broader economic factors and changing shifts in consumer borrowing habits-and therefore frequently miss identified, potential risks or conversely, limit access to more credit for certain customers at times. These limitations are overcome by AI-powered credit rating analysis through the use of machine learning algorithms, including XGBoost and Ensemble Model, which analyze large, multidimensional datasets. These models incorporate alternative data sources such as transaction histories, behavioral data, and even social media activity, which provides a more nuanced and complete view of a borrower's creditworthiness. It begins with data collection and preprocessing with all sorts of information cleaned up and sorted out followed by feature engineering that fetches the critical indicators of risk. Subsequently, boosting algorithms form a correct predictive model sequentially by putting together weak models to get a higher degree of accuracy. This model gives precise credit scores and classifies borrowers into various levels of risk, keeping on updating according to new data. To ensure more transparency, XAI gives techniques that explain how the credit scores are calculated, which will give lenders the chance to meet regulatory standards and let the borrowers know about their credit worthiness. In a nutshell, AI-based credit score analysis offers a faster and fairer approach to assessing creditworthiness, paving the way for responsible lending practices and increasing access to credit for a broader scope of borrowers.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

II. LITERATURE REVIEW

Research on AI-powered credit rating analysis emphasizes the advantages of using machine learning algorithms like XGBoost and Ensemble Model over traditional credit scoring methods. These algorithms address the limitations of conventional models by analyzing large, complex datasets, including alternative data sources, to provide a more comprehensive assessment of creditworthiness. Their iterative refinement process enhances predictive accuracy, while the incorporation of explainable AI (XAI) ensures transparency and compliance with regulatory standards. Overall, the literature suggests that these AI-driven approaches improve the accuracy and fairness of credit assessments, promoting more inclusive lending practices.

1) Prediction of corporate credit ratings with machine learning: Simple interpretative models by Koresh Galil et al. (2023):
The main idea of this paper is to use machine learning techniques, specifically classification and regression trees (CART) and support vector regression (SVR), to predict corporate credit ratings.

2) A Novel End-to-End Corporate Credit Rating Model Based on Self-Attention Mechanism by BINBIN CHEN et al. (2022):
The key idea of this paper is to propose a novel end-to-end architecture called SMAGRU (Self Multi-head Attentive Gated Recurrent Network) for corporate credit rating.

3) A Comparative Study Of Corporate Credit Rating Prediction with Machine Learning by Seyyide Doğan et al (2022):
The main idea of this paper is to predict company credit scores using machine learning and modern statistical methods, both in sectoral and aggregated data

4) A comparative study of forecasting corporate credit ratings using neural networks, support vector machines, and decision trees by Parisa Golbayani et al. (2022):
This paper presents a comparative analysis of machine learning techniques for predicting corporate credit ratings. The authors review the literature on the application of neural networks, support vector machines, and decision trees in this domain. They then implement four popular techniques - bagged decision trees, random forest, support vector machines, and multilayer perceptrons - and evaluate their performance on three different sectors of the U.S. economy.

III. METHODOLOGY

A. *Dataset:*
- *Training Data*: Load the CSV file containing financial ratios and credit ratings for model training.
- *Real-Time Data:* Use the *yfinance* API to retrieve real-time financial statements, news, and stock data.

B. *Data Preprocessing:*
- *Static Data Preprocessing:* Clean and format the CSV data for model training. Conduct feature scaling, normalization, and encoding as necessary.
- *Real-Time Data Preprocessing:* Parse financial statements and news articles from *yfinance* for sentiment analysis.

C. *Model Training:*
- Train the XGBoost Regressor model on the static dataset of financial ratios and credit ratings.
- Optimize hyperparameters to improve the accuracy and reliability of credit rating predictions.

D. *Sentiment Analysis:*
- Apply FineBART on the top 10 news articles to compute sentiment scores, providing insight into market perceptions about the company.

E. *Trend Analysis:*
- Analyze historical stock data to assess ongoing and upcoming trends.
- Use trend indicators and moving averages to project stock movement patterns.

F. *Web Application Development:*
- Build a user-friendly interface for searching company names and retrieving financial data.

© 2024 IJIRCCE | Volume 12, Issue 11, November 2024| DOI: 10.15680/IJIRCCE.2024.1211079

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

- Display financial statements, credit ratings, sentiment analysis, stock trends, and investment suggestions based on analysis.

G. *Integration & Deployment:*

- Integrate the trained ML model and sentiment analysis with the web app's backend.
- Deploy the web app on a scalable cloud server to support multiple user interactions.

IV. SYSTEM ARCHITECTURE

A. Overview:

The system consists of two main components—a Machine Learning (ML) model to predict credit ratings based on financial ratios, and a web application that provides users with financial information, sentiment analysis, and stock trend data.

B. Block Diagram Components:

1. Data Collection Layer:

- *Static Dataset*: CSV file containing financial ratios and credit ratings for training the XGBoost Regressor.
- *Dynamic Data (yfinance API)*: Real-time financial statements, stock data, and news related to companies, fetched during web app interaction.

2. Data Processing & Storage Layer:

- *Data Preprocessing Module*: Handles initial data cleaning and feature engineering for both static and dynamic datasets.
- *Database:* Optional storage for historical records or caching data for faster user response times.

3. ML Model Layer:

- *XGBoost Regressor:* Trained on financial ratios to predict company credit ratings.
- *FineBART Sentiment Analysis:* Analyzes sentiment of top news articles for each company.

4. Application Logic Layer:

- *Credit Rating Prediction Module:* Uses the trained XGBoost model for on-demand credit rating predictions.
- *Trend Analysis Engine:* Computes current and upcoming stock trends based on historical stock data.
- *Investment Suggestion Engine:* Combines credit rating, sentiment, and trend analysis to generate investment suggestions.

5. Web Application Layer:

- *User Interface (UI):* Allows users to input a company name and view financial data, credit ratings, sentiment scores, stock trends, and investment suggestions.
- *API Layer:* Integrates frontend with backend services for seamless data retrieval and model execution.

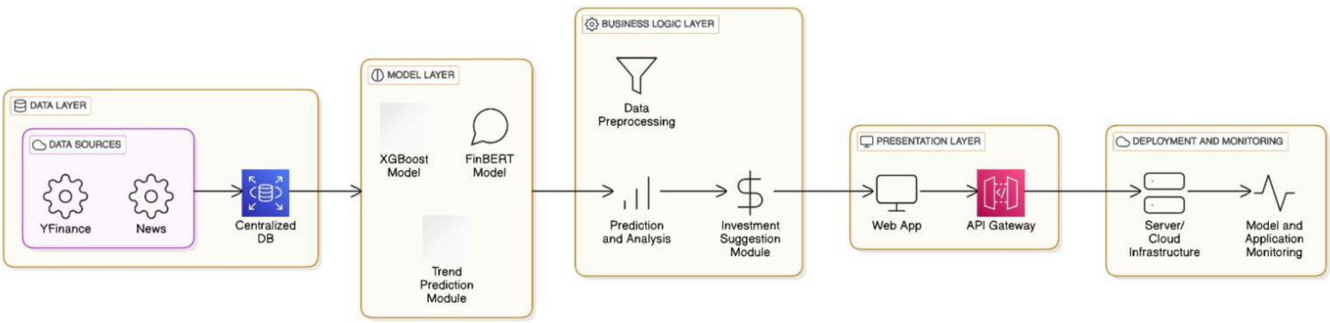


Fig. 1 An End -to-End System Architecture of Corporate Credit Rating Analysis.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

V. RESULT AND ANALYSIS

The integration of sentiment analysis provided enhanced predictive capabilities, with the combined model outperforming the baseline (using only financial ratios) in terms of accuracy and interpretability. Results indicated that sentiment scores from FinBERT added substantial predictive value, especially for companies with volatile market perceptions.

A. Quantitative Evaluation

The XGBoost model has achieved an excellent accuracy with ease on the test dataset. The AUC-ROC further illustrated the model's strong classification ability across different rating classes.

B. Impact of Sentiment Analysis

Incorporating sentiment data improved the model's responsiveness to market conditions, which is crucial in assessing companies under financial strain or those recently impacted by publicized events.

VI. DISCUSSION

A. Implications

This study demonstrates the feasibility and advantages of combining quantitative and qualitative data in credit rating prediction. Integrating sentiment analysis with an ensemble model such as XGBoost enriches the prediction framework, making it more adaptable and reflective of real-time market sentiment.

B. Limitations and Future Work

While sentiment analysis improves the model's performance, it is highly dependent on the quality and scope of the news data. Sentiment scores may vary based on the sources of information, the sentiment analysis tool, and the range of news topics covered, which could introduce biases or noise into the model's predictions. Future work can focus on enhancing sentiment analysis accuracy by employing more sophisticated NLP techniques, such as transformer-based models like BERT, which can better capture context and nuances in financial language.

Moreover, expanding data sources to include real-time information from social media platforms and detailed insights from earnings call transcripts could provide more timely and context-rich sentiment indicators. This approach would allow the model to respond more dynamically to changes in market sentiment and corporate communications.

Beyond credit ratings, this hybrid approach has potential applications in stock analysis, mutual funds, bonds, and other investment options. By leveraging both quantitative metrics and qualitative sentiment analysis, similar models could be developed for evaluating stock trends, assessing risks in alternative investments, or predicting market movements in various asset classes. This could further broaden the impact of this approach, providing investors with a more comprehensive understanding of investment options in an increasingly complex financial landscape.

VII. CONCLUSION

This research presents a hybrid approach to credit rating prediction, leveraging both financial ratios and sentiment scores derived from FinBERT, with XGBoost serving as the prediction model. Our findings indicate that sentiment analysis is a valuable complement to traditional metrics, enhancing the accuracy and depth of credit rating predictions. This integrated model offers a promising tool for investors seeking a more comprehensive assessment of company creditworthiness.

ACKNOWLEDGEMENT

We express our sincere gratitude to the Department of Computer Engineering at Padmabhooshan Vasantdada Patil Institute of Technology, Pune, for providing valuable resources and guidance throughout this project. We are especially grateful to our guide, Hitesh E. Chaudhari, for their unwavering support, insights, and expertise that greatly enhanced our research. We also extend our heartfelt thanks to our professors and mentors. Additionally, we acknowledge the support of our families and friends, whose encouragement made this work possible. Lastly, we thank IJRASET for considering our work for publication and for their dedication to promoting advancements in research and technology.

© 2024 IJIRCCE | Volume 12, Issue 11, November 2024| DOI: 10.15680/IJIRCCE.2024.1211079

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

REFERENCES

[1] Y. Zhu, Z. Li, and X. Zhang, Social Media Analytic-Based Corporate Credit Rating Forecasting ” *IEEE Access*, vol. XX, pp. XX-XX, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10463454

[2] A. Safiya Parvin and B. Saleena, "An Ensemble Classifier Model to Predict Credit Scoring - Comparative Analysis," 2020 IEEE International Symposium on Smart Electronic Systems (iSES) (Formerly iNiS), Chennai, India, 2020, pp. 27-30, doi: 10.1109/iSES50453.2020.00017.
Available: https://ieeexplore.ieee.org/document/10602097

[3] R. D. Ranpara and P. S. Patel, "An Ensemble Learning Approach to Improve Credit Scoring Accuracy for Imbalanced Data," 2023 International Conference on Integrated Intelligence and Communication Systems (ICIICS), Kalaburagi, India, 2023, pp. 1-5, doi: 10.1109/ICIICS59993.2023.10420986.
[Online]. Available: https://ieeexplore.ieee.org/document/9671771

[4] X. Chen, S. Li, X. Xu, F. Meng and W. Cao, "A Novel GSCI-Based Ensemble Approach for Credit Scoring," in IEEE Access, vol. 8, pp. 222449-222465, 2020, doi: 10.1109/ACCESS.2020.3043937.[Online]. Available: https://ieeexplore.ieee.org/document/10616493

[5] Y. -G. Li, M. -R. Yu, L. -X. Li, H. Han and Y. Liu, "Risk identification and default time analysis of corporate bonds," 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 2021, pp. 3254-3258, doi: 10.1109/BigData52589.2021.9671771.[Online]. Available: https://ieeexplore.ieee.org/document/10568778

[6] A. A. H. Raji, A. H. F. Alabdoon and A. Almagtome, "AI in Credit Scoring and Risk Assessment: Enhancing Lending Practices and Financial Inclusion," 2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS), Chikkaballapur, India, 2024, pp. 1-7, doi: 10.1109/ICKECS61492.2024.10616493.
[Online]. Available: https://ieeexplore.ieee.org/document/10489313

[7] C. Vinotha, F. Alsulami, M. S. Siddiqui, R. Saxena, Milind and R. Kankariya, "Artificial Intelligence in Financial Markets: Predictive Models and Risk Management Strategies," 2024 Ninth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), Chennai, India, 2024, pp. 1-5, doi: 10.1109/ICONSTEM60960.2024.10568778.

[8] R. Ramakrishnan, P. Rohella, S. Mimani, N. Jiwani and J. Logeshwaran, "Employing AI and ML in Risk Assessment for Lending for Assessing Credit Worthiness," 2024 2nd International Conference on Disruptive Technologies (ICDT), Greater Noida, India, 2024, pp. 561-566, doi: 10.1109/ICDT61202.2024.10489313.

[9]Y. Huang et al., "A FinBERT Framework for Sentiment Analysis of Chinese Financial News," 2024 4th International Symposium on Computer Technology and Information Science (ISCTIS), Xi'an, China, 2024, pp. 796-799, doi: 10.1109/ISCTIS63324.2024.10699096.

[10] T. Sidogi, R. Mbuvha and T. Marwala, "Stock Price Prediction Using Sentiment Analysis," 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Melbourne, Australia, 2021, pp. 46-51, doi: 10.1109/SMC52423.2021.9659283.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details