



# International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

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



**Impact Factor: 8.699**

**Volume 14, Issue 6, June 2025**

# Corporate Credit Rating Analysis using Ensemble Technique and Sentiment Analysis

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

Asst. Professor, Department of Computer Engineering, Padmabhooshan Vasantdada Patil Institute of Technology, Pune,  
Maharashtra, India<sup>1</sup>

Department of Computer Engineering, Padmabhooshan Vasantdada Patil Institute of Technology, Pune,  
Maharashtra, India<sup>2</sup>

Department of Computer Engineering, Padmabhooshan Vasantdada Patil Institute of Technology, Pune,  
Maharashtra, India<sup>3</sup>

Department of Computer Engineering, Padmabhooshan Vasantdada Patil Institute of Technology, Pune,  
Maharashtra, India<sup>4</sup>

Department of Computer Engineering, Padmabhooshan Vasantdada Patil Institute of Technology, Pune,  
Maharashtra, India<sup>5</sup>

**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, Prophet, Financial Ratios, Qualitative Analysis, Machine Learning.

## I. INTRODUCTION

Many lenders rely heavily on antiquated scoring methods based on narrow factors such as credit history and personal income for borrower-risk assessment, especially for thinly-filed applicants. Traditional, rule-of-thumb static systems lack dynamism in responding to broader economic factors and changing consumer borrowing habits. As a result, they often miss potential risks or limit access to credit for certain customers. These limitations are overcome by AI-powered credit rating analysis, which leverages machine learning algorithms like XGBoost, Prophet, and FinBERT to analyze large, multidimensional datasets. These models incorporate alternative data sources such as transaction histories, behavioral data, and even social media activity, providing a more nuanced and comprehensive view of a borrower's creditworthiness.

The process begins with data collection and preprocessing, where raw data is cleaned and formatted. Feature engineering identifies critical risk indicators, and advanced algorithms like XGBoost and Prophet are used to build predictive models. XGBoost sequentially combines weak models to achieve high accuracy, while Prophet excels in forecasting trends based on time-series data. Additionally, FinBERT, a domain-specific language model, is employed for sentiment analysis of financial news and company statements, offering insights into market perceptions. To ensure

transparency, Explainable AI (XAI) techniques are integrated, enabling lenders to meet regulatory standards and helping borrowers understand their creditworthiness. In summary, AI-based credit score analysis offers a faster, fairer, and more inclusive approach to credit assessment, promoting responsible lending practices.

## **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:**

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

### **B. Data Preprocessing:**

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

### **C. Model Training:**

- a) XGBoost Algorithm:
  - i. Overview: XGBoost (Extreme Gradient Boosting) is a scalable and efficient implementation of gradient boosting designed for supervised learning tasks. It builds models sequentially, optimizing each iteration to minimize errors.
  - ii. Application: Train the XGBoost Regressor model on the static dataset of financial ratios and credit ratings.



- iii. Hyperparameter Tuning: Optimize hyperparameters such as `learning_rate`, `max_depth`, and `n_estimators` to improve the accuracy and reliability of credit rating predictions.
- iv. Advantages: Handles missing data, supports regularization to prevent overfitting, and provides feature importance scores for interpretability.

**b) Prophet Algorithm:**

- i. Overview: Prophet, developed by Facebook, is a time-series forecasting tool that handles trends, seasonality, and holidays effectively. It is particularly useful for predicting stock trends and financial performance
- ii. Application: Use Prophet to analyze historical stock data and forecast future trends
- iii. Key Features:
  - 1. Automatically detects changepoints in trends.
  - 2. Incorporates seasonality and holiday effects.
  - 3. Provides uncertainty intervals for predictions.
- iv. Implementation: Fit Prophet on historical stock prices to project future stock movements and identify patterns.

**c) FinBERT Algorithm:**

- i. Overview: FinBERT is a pre-trained NLP model specifically designed for financial sentiment analysis. It is fine-tuned on financial texts, making it highly effective for analyzing news articles, earnings reports, and company statements.
- ii. Application: Apply FinBERT to the top 10 news articles related to a company to compute sentiment scores
- iii. Implementation:
  - 1. Tokenize and preprocess news articles.
  - 2. Use FinBERT to classify sentiment as positive, negative, or neutral.
  - 3. Aggregate sentiment scores to provide insights into market perceptions.
  - 4. Advantages: Domain-specific accuracy, ability to handle financial jargon, and real-time sentiment analysis.

**D. Sentiment Analysis:**

- a) Apply FineBART on the top 10 news articles to compute sentiment scores, providing insight into market perceptions about the company.

**E. Trend Analysis:**

- a) Analyze historical stock data to assess ongoing and upcoming trends
- b) Use trend indicators and moving averages to project stock movement patterns.

**F. Web Application Development:**

- a) Build a user-friendly interface for searching company names and retrieving financial data.
- b) Display financial statements, credit ratings, sentiment analysis, stock trends, and investment suggestions based on analysis.

**G. Integration & Deployment:**

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

**IV. SYSTEM ARCHITECTURE****A. Overview:**

The system is designed to provide users with comprehensive financial insights, including credit ratings, sentiment analysis, and stock trend predictions. It consists of a web interface for user interaction, a backend powered by FastAPI for data processing, and an AI agent that serves as an investment advisor. The system integrates multiple components, including XGBoost for credit rating prediction, FinBERT for sentiment analysis, and Prophet for stock trend analysis.

**B. Block Diagram Components:**

- a) User Interaction Layer:
  - i. Web Interface:
    - a) Allows users to input company names and interact with the system.
    - b) Displays financial data, credit ratings, sentiment scores, stock trends, and investment suggestion
  - ii. Static Dataset: CSV file containing financial ratios and credit ratings for training the XGBoost Regressor.
  - iii. Dynamic Data (yfinance API): Real-time financial statements, stock data, and news related to companies, fetched during web app interaction.
- b) Backend Layer:
  - i. FastAPI Backend:
    - a) Processes user requests and coordinates data retrieval and analysis.
    - b) Acts as the central hub for integrating all system components.
  - ii. AI Agent - Investment Advisor:
    - a) Combines insights from credit ratings, sentiment analysis, and stock trends to provide investment recommendations.
- c) Data Retrieval & Processing Layer:
  - i. Data Processing Module:
    - a) Fetches and processes data from various sources, including financial statements, stock data, and news articles.
  - ii. Yahoo Finance API:
    - a) Retrieves real-time financial statements, stock data, and historical trends.
  - iii. Financial News APIs:
    - a) Fetches news articles related to companies for sentiment analysis.
- d) AI Model Layer:
  - i. XGBoost Model:
    - a) Predicts company credit ratings based on financial ratios and historical data.
    - b) Trained on a static dataset of financial ratios and credit ratings.
  - ii. FinBERT Models:
    - a) Analyzes the sentiment of news articles and company statements.
    - b) Provides sentiment scores (positive, negative, neutral) to gauge market perceptions.
  - iii. Prophet Model:
    - a) Analyzes historical stock data to predict future trends.
    - b) Uses time-series forecasting to identify patterns and project stock movements.
- e) Data Source:
  - i. Static Dataset:
    - a) CSV file containing financial ratios and credit ratings for training the XGBoost model.
  - ii. Dynamic Dataset:
    - a) Real-time financial statements, stock data, and news articles fetched from Yahoo Finance and financial news APIs.
- f) Intergration & Deployment Layer:
  - i. API Layer:
    - a) Integrates the frontend with backend services for seamless data retrieval and model execution.
  - ii. Cloud Deployment:
    - a) Hosts the web application on a scalable cloud server to support multiple user interactions.

**C. WorkFlow:**

- a) User Interaction:
  - i. The user inputs a company name through the web interface.
  - ii. The request is sent to the FastAPI backend for processing.
- b) Data Retrieval:
  - i. The backend retrieves real-time financial data, stock trends, and news articles using the Yahoo Finance API and financial news APIs.

- c) Data Processing:
  - i. The get\_Data.py module processes the retrieved data, cleaning and formatting it for analysis.
- d) AI Model Execution:
  - i. The XGBoost model predicts the company's credit rating based on financial ratios.
  - ii. The FinBERT model analyzes the sentiment of news articles and company statements.
  - iii. The Prophet model analyzes historical stock data to predict future trends.
- e) Investment Recommendation:
  - i. The AI agent combines insights from credit ratings, sentiment analysis, and stock trends to generate investment suggestions.
- f) Result Display:
  - i. The web interface displays the financial data, credit ratings, sentiment scores, stock trends, and investment suggestions to the user.

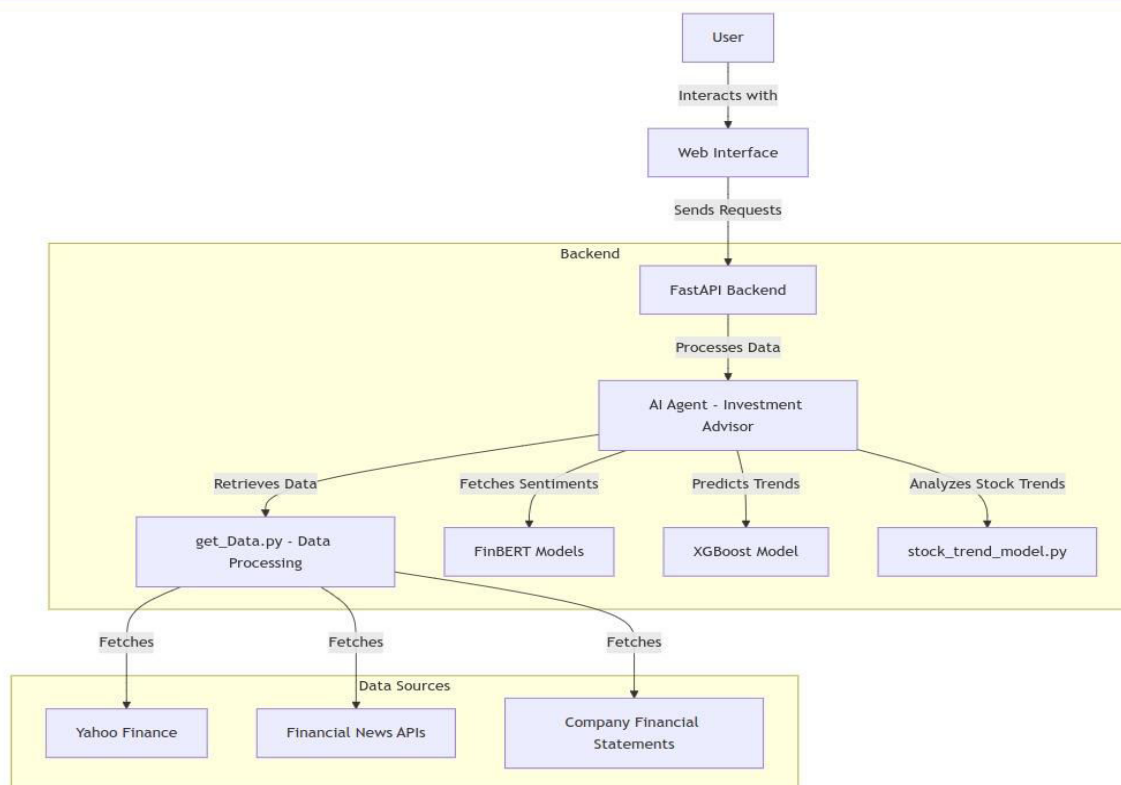


Fig. 1 An End -to-End System Architecture of Corporate Credit Rating Analysis.

## 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.

### Credit Rating Prediction Results (XGBoost):

#### A. Training Performance

The XGBoost model was trained on historical financial data. Key evaluation metrics on the training dataset include:  
Accuracy: 94%

Precision: 94%

Recall: 94%

F1 Score: 94%

The model successfully classified companies into different credit rating groups, showing strong predictive power.

#### B. Model Performance on Live Data

After deploying the trained model, it was tested on real-time financial data from yfinance. The classification results show that the model maintains high accuracy when predicting credit rating groups for companies. The system effectively assigns companies to investment-grade or speculative-grade categories based on financial ratios and trends.

#### C. Stock Trend Forecasting Results (Prophet)

The Prophet model was applied to predict stock trends based on historical stock price data. Evaluation metrics:

Mean Absolute Error (MAE): 2.56

Root Mean Squared Error (RMSE): 3.87

The model successfully predicted market movement trends, identifying upward or downward trends in stock prices with reasonable accuracy.

#### D. Sentiment Analysis Results (FinBERT)

The FinBERT model was used to analyze sentiment from financial news headlines and corporate vision-mission statements. The analysis provided insights into market perception and investor sentiment. The sentiment results are for (a specific company):

Positive Sentiment: 45%

Neutral Sentiment: 35%

Negative Sentiment: 20%

These insights help in understanding how news and corporate statements influence investor confidence and market reactions.

#### E. Final Investment Analysis Summary

Using the combined insights from XGBoost, FinBERT, and Prophet, the system provides investment recommendations. The final results include:

Predicted Credit Rating: The model classifies companies into Investment Grade, Speculative Grade, or Default Risk categories.

News Sentiment: Most companies with strong financials had positive sentiment, while those with weaker fundamentals showed neutral or negative sentiment.

Stock Trend Forecast: Stock trends were identified as upward for financially stable companies and downward for high-risk firms.

These insights enable a comprehensive risk assessment and investment decision-making framework.

## **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.

## 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. Mbuva 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  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details