



ИПС 61

Разработка ансамблевых методов оценки кредитного качества или прогнозирования кредитных рейтингов **нефинансовых** компаний

Руководитель: Астахова Алёна Андреевна, Гришунин Сергей Вадимович
Автор: Каролина Вивас



Цель – Исследование направлено на сравнение моделей машинного обучения для прогнозирования риска дефолта в зарубежных банках, разработку рейтинговой шкалы для оценки их кредитоспособности и установление критериев для классификации банков по стадиям кредитного качества в соответствии на основе выборки из банков.

Краткое содержание исследования

Будет исследована методология оценки кредитоспособности компаний, применяемая такими рейтинговыми агентствами, как **Moody's, Standard & Poor's и Fitch Ratings**. На основе этой методологии, с дополнением ключевых финансовых показателей, таких как **выручка, EBIT, процентные расходы, нераспределенный денежный поток, общий долг, EBITDA, чистый долг, рентабельность материальных активов, капитализация по балансовой стоимости, денежный поток от операционной деятельности и дивиденды**, будет создана собственная модель.

Техническое задание проекта:

1. Проводить обзор литературы по оценке кредитного качества и методам ансамблевого обучения.
2. Определить проблему и цели исследования.
3. Собрать подготовьте необходимые данные для анализа.
4. Разработать и оцените различные ансамблевые модели для предсказания кредитного качества. (Random Forest, Gradient Boosting, Regression model, ect)
5. Проанализировать результаты и сделайте выводы на основе находок.

Ожидаемые результаты

Разработка точной модели оценки кредитоспособности компаний на основе ансамблевых методов с использованием ключевых финансовых показателей, адаптированной для международных рынков, с созданием рейтинговой шкалы и системы классификации кредитного качества.



Цель – Исследование направлено на разработку и сравнение ансамблевых моделей машинного обучения для прогнозирования кредитного рейтинга нефинансовых компаний в технологическом секторе, создание рейтинговой шкалы для оценки их кредитоспособности и установление критериев для классификации компаний по стадиям кредитного качества на основе данных выборки из технологических компаний.

Цели исследования:

- Разработка ансамблевых методов для более точного прогнозирования кредитных рейтингов технологических компаний.
- Сравнение точности разработанных методов с существующими моделями машинного обучения.
- Создание инструмента для финансовых аналитиков и инвесторов, который поможет минимизировать риски при принятии решений.

Задачи исследования:

- Провести анализ текущих методов машинного обучения, используемых для прогнозирования кредитных рейтингов (включая работу Гришунина и Егоровой).
- Выбрать и адаптировать существующие ансамблевые методы (например, Random Forest, Gradient Boosting, Stacking, Bagging) для анализа данных технологического сектора.
- Провести тестирование моделей на реальных данных, используя выборку технологических компаний, и сравнить результаты с традиционными статистическими методами.
- Разработать методику оценки результатов моделей с точки зрения их применимости в реальных условиях.
- Создать рекомендации по использованию модели для участников финансового рынка.



Актуальность

- Технологический сектор играет ключевую роль в мировой экономике, однако его высокая волатильность усложняет оценку кредитных рисков. Традиционные методы имеют ограничения в условиях большого объема данных. Разработка ансамблевых методов повысит точность и надежность прогнозов, что важно для динамично развивающегося рынка.



Научная новизна

- Применение ансамблевых методов для оценки кредитного качества компаний именно в технологическом секторе, что недостаточно изучено в текущей литературе.
- Разработка гибридных подходов, объединяющих машинное обучение и традиционные финансовые показатели для улучшения прогноза кредитных рейтингов.
- Углубленное сравнение производительности ансамблевых моделей с другими подходами, используемыми для прогнозирования рейтингов, в том числе на основе данных из статьи Гришунина и Егоровой.



Практическая значимость

- Разработанные методы могут применяться финансовыми аналитиками и инвестиционными компаниями для точной оценки кредитного риска технологических компаний.
- Результаты исследования помогут минимизировать инвестиционные риски и повысить обоснованность финансовых решений.
- Создание модели для обработки больших объемов данных упрощает и автоматизирует анализ кредитоспособности, учитывая



Что сделано

- Проведен обзор литературы
- Определены цели исследования, задачи и методология разработки ансамблевых моделей
- Составлен список данных, необходимых для анализа, включая финансовые метрики и макроэкономические переменные
- Разработаны гипотезы



Что планируется сделать

- Сбор и подготовка данных
- Разработка и тестирование моделей
- Анализ результатов и выводы
- Разработка рейтинговой шкалы
- Подготовка итогового отчета





Inria (2023). Predicting Firms' Credit Ratings Using Ensembles of Artificial Immune Systems and Machine Learning – An Over-Sampling Approach.

- This paper explores artificial immune systems combined with machine learning to predict credit ratings. It emphasizes the benefits of ensemble learning and data over-sampling techniques to improve classification accuracy.
- **Method:** Artificial Immune Systems (AIRS) with K-Nearest Neighbors, Support Vector Machines (SVM)

Zhang, M., & Wang, Z. (2023). Credit rating prediction with supply chain information: A machine learning approach

- This study applies ensemble machine learning models, particularly LightGBM, to improve credit rating predictions by incorporating additional supply chain information from suppliers and customers.
- **Method:** LightGBM, Ensemble Learning



- **Relevance:** The prediction of credit ratings is important because it allows managing financial risks by assessing the probability of default and allows making more informed decisions on loans or interest rates. Ensemble methods are relevant because they combine several models and this helps to improve accuracy and automate the evaluation process which reduces operational costs and speeds up decision making.
- **Innovation:** The literature generally focuses on the use of traditional statistical models for predicting credit ratings (such as Altman's Z-Score model or logistic regression model).
- **Ensemble methods:** Ensemble methods are machine learning techniques that combine multiple individual models to obtain a more accurate and robust prediction while avoiding errors such as overfitting and variance.

Methods and advantages

Random Forest	Gradient Boosting (XGBoost, LightGBM, CatBoost)	Bagging	Regression model
<ul style="list-style-type: none">● It is simple to implement using libraries such as scikit-learn in Python.● It handles feature-rich data well and does not require a lot of parameter tuning.	<ul style="list-style-type: none">● LightGBM es especialmente útil para grandes conjuntos de datos.● CatBoost maneja automáticamente variables categóricas, lo que puede ser útil en modelos financieros.	<ul style="list-style-type: none">● Reduce la varianza del modelo al promediar múltiples predicciones.● Es una técnica básica que se puede implementar fácilmente con scikit-learn	<ul style="list-style-type: none">● Los coeficientes de las variables ofrecen una interpretación clara sobre cómo cada factor afecta el resultado● La regresión es efectiva cuando se trabaja con conjuntos de datos relativamente pequeños



- **Credit ratings** are assessments of the ability of a borrower (such as a company or government) to meet its financial obligations, such as repayment of loans or bonds. These ratings are assigned by credit rating agencies such as **Standard & Poor's, Moody's and Fitch**. They range from **AAA (highest creditworthiness) to D (default)**. These evaluations influence borrowing costs, as entities with lower ratings face higher interest rates due to higher perceived risk, while those with higher ratings can access financing at lower rates.
- **Traditional Credit Evaluation Methods**
Traditional credit evaluation methods are usually statistical techniques and approaches based on expert judgment.
 - **Statistical Methods:**
Based on historical financial data (e.g., debt to equity ratios, profitability, liquidity) and employ statistical methods to predict the creditworthiness of a borrower.
Z-Score Model: Developed by Edward Altman, uses financial ratios to predict the probability of bankruptcy. This model combines five financial ratios (liquidity, profitability, leverage, activity and market value) into a single score.

$$Z = 1.2 \times \frac{\text{Working Capital}}{\text{Total Assets}} + 1.4 \times \frac{\text{Retained Earnings}}{\text{Total Assets}} + 3.3 \times \frac{\text{EBIT}}{\text{Total Assets}} + 0.6 \times \frac{\text{Market Value of Equity}}{\text{Total Liabilities}} + 1.0 \times \frac{\text{Sales}}{\text{Total Assets}}$$

Методы и преимущества

Bagging

Bagging works by generating multiple bootstrap samples (random subsets with replacement) from the training dataset. Each model is trained on one of these subsets, and the final prediction is aggregated by averaging or voting

Given the original training data $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$

we generate multiple bootstrap samples D_i ($i=1, \dots, M$)

Each model $f_i(x)$ is trained on D_i , and the final prediction is:

$$\hat{y}_{\text{bagging}} = \frac{1}{M} \sum_{i=1}^M f_i(x) \quad (\text{for regression})$$

$\hat{y}_{\text{bagging}} = \text{mode}(\{f_1(x), f_2(x), \dots, f_M(x)\})$
(for classification)

where M is the number of models

Gradient Boosting (XGBoost, LightGBM, CatBoost)

Gradient Boosting is an iterative method where models are trained sequentially, each correcting the errors of the previous model.

At iteration t , the new model $f_t(x)$ is trained to minimize the following loss function

$$L(\theta) = \sum_{i=1}^N \text{loss}(y_i, f_{t-1}(x_i) + f_t(x_i)) + \Omega(f_t)$$

where $\Omega(f_t)$ is a regularization term to prevent overfitting.

The final prediction is the sum of all model outputs:

$$\hat{y} = f_0(x) + \sum_{t=1}^T f_t(x)$$

where $f_0(x)$ is the initial prediction, and T is the number of iterations.

Random Forest

Random Forest is an extension of Bagging, but with a key difference: it introduces randomness in the feature selection process at each split in the decision trees.

Like Bagging, multiple decision trees $f(x)$ are trained on bootstrap samples of the data, but at each node, only a random subset of m features is considered for splitting.

$$\hat{y}_{\text{rf}} = \frac{1}{M} \sum_{i=1}^M f_i(x) \quad (\text{for regression})$$

or

$\hat{y}_{\text{rf}} = \text{mode}(\{f_1(x), f_2(x), \dots, f_M(x)\})$
(for classification)

Regression model

$$y = \beta_0 + \sum_{i=1}^n \beta_i x_i + \epsilon$$

Where y is the predicted credit rating, x_i are explanatory variables (e.g., financial ratios), β_i are coefficients representing their impact, and ϵ is the error term.



- Taking into account the paper Comparative Analysis of the Predictive Power of Machine Learning Models for Forecasting the Credit Ratings of Machine-Building Companies we propose the use of the following data:
- **Moody's ratings:** converted to a numerical scale to simplify the analysis (e.g., “AAA” to “1”, “C” to “17”).
- Caa - C
- **Explanatory variables:** Financial ratios, macroeconomic factors, corporate governance factors and risks

Основные финансовые показатели

1. Total asset (натуральный логарифм)
2. Выручка (натуральный логарифм)
3. EBITDA Margin
4. Мультипликатор D/E
5. Чистый маржин = Чистая прибыль / Выручка
6. Рыночная капитализация
7. Оборачиваемость активов = Выручка / Общие активы

Рентабельность

8. Return on Equity (ROE) = Net Income / Equity
9. ROIC
10. ROA
11. WACC

Покрывание процентов

12. EBIT/Interest Expense
13. (EBITDA – CapEx) / Interest Expense

Долговая нагрузка

1. Долг к EBITDA = Общий долг / EBITDA
2. Коэффициент долга = Общие обязательства / Общие активы
3. Debt/Book Capitalization
4. Debt/Total Capitalization

Показатели ликвидности

5. Быстрый коэффициент
6. (Деньги + Рыночные ценные бумаги) / Долг

качественные показатели

7. Все, что можно выгрузить с баз данных
8. Международное присутствие (0 – в одной стране, 1 – также за рубежом)
9. Развитая/ развивающаяся экономика
10. Количество лет на рынке
11. Наличие комитета по ESG

Рыночные показатели

1. Волатильность акций
2. Бета рынка
3. Средний объем торгов
4. Коэффициент цена/балансовая стоимость (P/B)
5. Соотношение капитализации рынка к выручке

Макроэкономические переменные страны

6. ВВП на душу населения
7. Годовая инфляция
8. Темп роста реального ВВП

Технические показатели

9. Рост числа активных пользователей в месяц
10. Выручка на пользователя = Выручка / Общее количество пользователей

ESG рейтинги и показатели

11. ESG рейтинги от таких провайдеров, как Sustainalytics и MSCI.
12. Наличие комитета по ESG.
13. Vigeo-EIRIS, MSCI, ISS-Oekom, CDP, RAEX-Europe, все рейтинги Bloomberg



- **Model**

Dependent Variable (Y) = $\beta_0 + \beta_1 \cdot \log(\text{Total Assets}) + \beta_2 \cdot \text{ROE} + \beta_3 \cdot \text{Debt Ratio} + \beta_4 \cdot \text{Net Margin} + \beta_5 \cdot \text{Market Capitalization} + \beta_6 \cdot \text{Volatility of Stock Price} + \beta_7 \cdot \text{Market Beta} + \beta_8 \cdot \text{Average Trading Volume} + \beta_9 \cdot \text{GDP per capita} + \beta_{10} \cdot \text{Inflation Rate} + \dots + \epsilon$

- **Hypothesis 1:** Ensemble models are more robust and accurate than traditional models when predicting credit ratings for non-financial companies with missing or incomplete financial data.
- **Hypothesis 2:** For non-financial companies, macroeconomic variables such as GDP growth, inflation rates, and interest rates have a more significant impact on credit rating predictions than financial ratios like debt-to-equity or profit margins.
- **Hypothesis 3:** Ensemble models that combine financial metrics and market-based indicators provide more accurate credit rating predictions for non-financial companies.
- **Hypothesis 4:** Adding irrelevant or redundant features to ensemble models can reduce their predictive power for credit rating forecasting in non-financial companies.



- **Model**

Dependent Variable (Y) = $\beta_0 + \beta_1 \cdot \log(\text{Total Assets}) + \beta_2 \cdot \text{ROE} + \beta_3 \cdot \text{Debt Ratio} + \beta_4 \cdot \text{Net Margin} + \beta_5 \cdot \text{Market Capitalization} + \beta_6 \cdot \text{Volatility of Stock Price} + \beta_7 \cdot \text{Market Beta} + \beta_8 \cdot \text{Average Trading Volume} + \beta_9 \cdot \text{GDP per capita} + \beta_{10} \cdot \text{Inflation Rate} + \beta_{11} \cdot \text{Interest Rate} + \beta_{12} \cdot \text{Sovereign Rating} + \epsilon$

- For **Total Assets** ($\log(\text{Total Assets})$), **Return on Equity (ROE)**, **Debt Ratio**, **Net Margin**, and **Market Capitalization**, financial reports such as balance sheets and income statements are key sources, and these can be accessed via platforms like **Yahoo Finance**, **Morningstar**, or **SEC's EDGAR Database**. **Volatility of Stock Price**, **Market Beta**, and **Average Trading Volume** can be found on **Yahoo Finance**, **Bloomberg**, or **NASDAQ**, which offer historical stock data and market statistics. **GDP per capita**, **Inflation Rate**, and **Interest Rate** are available from international organizations like the **World Bank**, **IMF**, and **OECD**. For **Sovereign Rating**, credit rating agencies such as **Standard & Poor's**, **Moody's**, and **Fitch Ratings** provide the necessary data.



- Sergei Grishunin, Alexandra Egorova, Comparative Analysis of the Predictive Power of Machine Learning Models for Forecasting the Credit Ratings of Machine-Building Companies , Journal of Corporate Finance Research / Корпоративные Финансы | ISSN: 2073-0438: Том 16 № 1 (2022)
- АстаховаА., ГришунинС. и Поморцев Г. (2023) «Разработка скоринговой рейтинговой модели на основе методологии международных рейтинговых агентств», Journal of Corporate Finance Research / Корпоративные Финансы | ISSN: 2073-0438, 17(1), сс. 5-16. doi: 10.17323/j.jcfr.2073-0438.17.1.2023.5-16.
- Ильиных М. В. Обзор подходов к моделированию кредитных рейтингов // Инновации и инвестиции. 2019. №7. URL: <https://cyberleninka.ru/article/n/obzor-podhodov-k-modelirovaniyu-kreditnyh-reytingov> (дата обращения: 05.11.2024).)

Спасибо за внимание!



NATIONAL RESEARCH
UNIVERSITY