National Research University Higher School of Economics

Faculty of Economic Sciences

Master’s Programme
Strategic Corporate Finance

REPORT
on *a professional project* **internship**

Comparative Analysis of ensemble methods for Credit Risk Assessment in the IT Sector

Completed by:
Vivas Teran Katherine Carolina
____
(signature)

Reviewed by:

Senior Research Fellow, Sergei Grishunin

____
(signature)

____
(date)

Content

1. General Description of the Project

- **Project Initiator, Client, and Supervisor**:
  - **Initiator**: National Research University (NRU).
  - **Client**: The project is academic and research-oriented, aimed at developing ensemble methods for credit quality assessment.
  - **Supervisor**: Alena Astakhova and Sergey Grishunin.
- **Project Type**:
  - **Research Project**: The project focuses on developing and comparing ensemble machine learning models for predicting credit ratings of non-financial companies, particularly in the technology sector.
- **Place of Project Implementation**:
  - The project is implemented within the academic framework of the National Research University, specifically in the field of corporate finance and machine learning.

2. Main Body

2.1. Description of the Project's Implementation

- **Objective**: The project aims to develop ensemble machine learning models to predict credit ratings for non-financial companies, particularly in the technology sector. The goal is to create a more accurate and reliable credit rating prediction tool compared to traditional methods.
- **Tasks**:
  - Conduct a literature review on credit rating prediction and ensemble learning methods.
  - Collect and preprocess data, including financial metrics, macroeconomic variables, and market indicators.
  - Develop and test ensemble models (Random Forest, Gradient Boosting, Bagging).
  - Compare the performance of ensemble models with traditional statistical methods.
  - Create a credit rating scale and classification system for non-financial companies.
- **Timeline**:
  - Data collection and cleaning: January 28 - February 8.
  - Model development and testing: February 12 - March 5.
  - Final analysis and report preparation: March 11 - March 24.

2.2. Description of Project Outcomes

Outcome: The project resulted in the development of ensemble machine learning models that can predict credit ratings for non-financial companies with higher accuracy than traditional methods. A credit rating scale and classification system were also created.

- **Deliverables**:
    - A set of ensemble models (Random Forest, Gradient Boosting, etc.).
    - A credit rating scale for non-financial companies.
    - A final report and presentation summarizing the findings and recommendations.

2.3. Description of Methods and Technologies Applied During the Project

- **Methods: Ensemble Learning Techniques**

  Ensemble learning methods combine multiple base models to produce a stronger, more robust predictive model. The following techniques were employed to enhance prediction accuracy:

  Random Forest

  Random Forest is an ensemble learning method based on decision trees trained via bagging (bootstrap aggregating), where each tree is built on a random subset of the training data and features, reducing overfitting through prediction averaging while handling high-dimensional data effectively due to feature randomness. It provides feature importance scores for interpretability and offers advantages such as robustness to noise, versatility in both classification and regression tasks, and minimal hyperparameter tuning requirements.

  Gradient Boosting Machines (GBMs)

  During the project, Gradient Boosting Machines (GBMs) were employed as a pivotal component of the ensemble learning approach to predict credit ratings for non-financial companies. GBMs operate by sequentially building decision trees, where each new tree corrects errors from previous iterations using gradient descent optimization. Three specific variants were utilized: XGBoost, LightGBM, and CatBoost. XGBoost was optimized for speed and performance, incorporating L1 and L2 regularization to prevent overfitting, and efficiently handled missing values while supporting parallel processing. LightGBM leveraged histogram-based algorithms for faster training and reduced memory usage, making it suitable for large-scale datasets and seamlessly integrating categorical features. CatBoost automatically managed categorical variables without extensive preprocessing, reducing overfitting through ordered boosting and symmetric tree structures. These models collectively offered high predictive accuracy and flexibility in handling diverse data types, proving effective in both competitive environments and real-world applications. Each model underwent rigorous hyperparameter tuning and cross-validation to ensure robust performance, and their ensemble nature allowed for a comprehensive evaluation of credit risk factors.

  Bagging (Bootstrap Aggregating)

  Bagging (Bootstrap Aggregating) was employed as a fundamental ensemble learning technique to enhance the robustness and accuracy of credit rating predictions for non-financial companies. Bagging operates by generating multiple versions of a base model, such as decision trees, using different bootstrap samples of the training data. The predictions from these models are then aggregated, typically through voting for classification tasks or averaging for regression tasks. A key example of this method is Random Forest, which extends the bagging approach by introducing additional randomness in the feature selection process at each split. This technique effectively reduces variance and improves the stability of the model, making it particularly well-suited for high-variance models like deep decision trees. By leveraging bagging, the project achieved more reliable and consistent predictions, contributing to a comprehensive evaluation of credit risk factors.

- **Data Preprocessing**

  The initial phase of our project involved the meticulous collection and cleaning of financial data and macroeconomic variables. The financial data encompassed key metrics such as revenue, EBITDA (Earnings Before Interest, Taxes, Depreciation, and Amortization), and various debt ratios. These metrics are crucial as they provide insights into a company's financial health and operational efficiency. For instance, EBITDA serves as an alternative measure of a company's overall performance, allowing for a clearer understanding of profitability by excluding non-operational expenses

  In addition to financial data, we also gathered macroeconomic indicators, including GDP (Gross Domestic Product) and inflation rates. These variables are essential for contextualizing the financial data within the broader economic landscape, as they can significantly influence business performance and investment decisions. The cleaning process involved handling missing values, removing duplicates, and ensuring that the data was in a consistent format, which is vital for accurate analysis and modeling.

- **Model Evaluation**

  Once the data was preprocessed, we proceeded to evaluate the performance of our predictive models. This evaluation was conducted using several key metrics, including accuracy, precision, recall, and the F1-score. Each of these metrics provides unique insights into the model's performance:

  - Accuracy measures the overall correctness of the model's predictions, indicating the proportion of true results (both true positives and true negatives) among the total number of cases examined.
  - Precision assesses the model's ability to correctly identify positive instances, which is particularly important in scenarios where false positives can lead to significant costs or risks.
  - Recall, on the other hand, evaluates the model's ability to capture all relevant positive instances, highlighting its effectiveness in identifying true positives.
  - The F1-score serves as a harmonic mean of precision and recall, providing a single metric that balances both concerns, especially in cases where there is an uneven class distribution.
- **Technologies**:
  - **Programming Languages**: Python (using libraries like Scikit-learn, XGBoost, LightGBM).
  - **Data Sources**: Financial reports (Yahoo Finance, Bloomberg), macroeconomic data (World Bank, IMF), and credit rating agencies (Moody's).
  - **Tools**: Jupyter Notebook, Pandas, NumPy, Matplotlib, and Seaborn for data analysis and visualization.

Additionally, a Business Intelligence (BI) tool, **Yandex DataLens (Fig 5)**, was utilized to visualize and explore the data interactively. This tool enabled the creation of dynamic dashboards, allowing stakeholders to gain insights into the financial and macroeconomic variables in real-time. The integration of Yandex DataLens facilitated a more intuitive understanding of the data, enhancing the overall analytical process.

2.4. Description of Student's Role in a Project Team

- **Role**: Main contributor and responsible for:
  - Conducted a comprehensive literature review on ensemble learning and credit rating prediction.
  - Independently collected and preprocessed financial and macroeconomic data.
  - Developed and tested various ensemble models to predict credit ratings.
  - Analyzed model performance to derive meaningful insights and conclusions.
  - Compiled findings into a detailed final report, showcasing methodologies, results, and implications.

2.5. Detailed Project Methodology

- **Data Preparation**

The initial phase of the project focused on Data Preparation, which is a critical step in ensuring the quality and reliability of the dataset before any analysis or modeling can take place. This phase can also be referred to as Data Cleaning and Feature Engineering.

Clean the Data:

During this stage, several key actions were taken to ensure the dataset was ready for analysis:

Handling Missing Values: We conducted a thorough examination of the dataset to identify any missing values. Various strategies were employed to address these gaps, including imputation techniques to fill in missing data and removal of records when necessary. This step is crucial, as the integrity of the analysis heavily relies on the completeness of the data.

Data Formatting: We ensured that all numerical values were correctly formatted, eliminating any instances of text or erroneous entries in numerical columns. This step is vital for maintaining the accuracy of calculations and analyses.

Categorical Variable Conversion: Categorical variables, such as Moody's Credit Rating, were converted into numerical values when necessary. For example, ratings like "Aa3" were transformed into a numerical score to facilitate their use in machine learning models.

Normalized or standardized numerical features to ensure they are on the same scale, especially for models like Gradient Boosting or Neural Networks.

- **Split the Data:**

Split the data into training, validation, and test sets (e.g., 70% training, 15% validation, 15% test). Ensured the split was stratified if the target variable (e.g., Moody's Credit Rating) was imbalanced.

- **Define the Target Variable**

  The target variable was the Moody's Credit Rating (e.g., "Aa3"). Converted the credit ratings into a numerical scale (e.g., "AAA" = 1, "AA" = 2, ..., "D" = 17). Decided to treat this as a classification problem (predicting a rating category). Grouped ratings into broader categories (e.g., "Investment Grade" vs. "Non-Investment Grade").

- **Model Selection**

  Started with a simple baseline model (Logistic Regression) to establish a performance benchmark. And then based on the project, used ensemble methods such as: Random Forest, Gradient Boosting (XGBoost, LightGBM), Bagging, Compared these with traditional models like Logistic Regression.

- **Model Training**

  Trained each model on the training dataset. Used cross-validation (e.g., 5-fold or 10-fold) to evaluate model performance and avoid overfitting.

  Hyperparameter Tuning:

  Used techniques like Grid Search or Random Search to optimize hyperparameters for each model. For example, for Random Forest, tuned parameters like n_estimators, max_depth, and min_samples_split.

- **Model Evaluation**

  Used appropriate metrics for evaluation:

  For classification: Accuracy, Precision, Recall, F1-Score, ROC-AUC.

  Compared the performance of ensemble models with the baseline model.

  Feature Importance:

  Analyzed feature importance scores from models like Random Forest or Gradient Boosting to understand which variables have the most impact on credit ratings.

- **Hypothesis Testing**

  Used the results from the models to test the hypotheses outlined in the project:

  > *Hypothesis 1:* Ensemble models are more robust and accurate than traditional models.
  >
  > *Hypothesis 2:* Macroeconomic variables have a more significant impact than financial ratios(They other way).
  >
  > Hipotesis con IT or with latinoamerica. Macroeconmic variables are more important in latin america
  >
  > *Hypothesis 3*: Combining financial and market-based indicators improves accuracy.
  >
  > *Hypothesis 4*: Irrelevant features reduce predictive power.

  Statistical Tests:

  Performed statistical tests (t-tests, ANOVA) to validate the findings.

- **Develop a Rating Scale**

  Based on the model predictions, developed a rating scale that classifies companies into different credit quality categories ("High Risk," "Medium Risk," "Low Risk").

  Validate the Scale:

  Tested the rating scale on the validation and test datasets to ensure it accurately reflects credit quality.

2.6. Description of Any Issues and Challenges During Project Implementation

Challenges:

i. **Data Quality:** We encountered issues with missing or incomplete financial data for some companies. To address this, we used imputation techniques to fill gaps and ensure the dataset's reliability.

ii. **Model Complexity:** The use of ensemble models, such as Gradient Boosting, required significant computational resources and extensive hyperparameter tuning. Managing these demands was essential for optimizing model performance.

iii. **Interpretability:** While ensemble models achieved high accuracy, they posed challenges in interpretability. Unlike traditional models, these methods can be opaque, making it difficult to explain predictions. We explored techniques like SHAPE values to clarify feature importance and enhance understanding.

3. Project Outcomes

The results of the model evaluation indicate varying levels of performance across different algorithms (Fig. 1). The Tuned Random Forest Model emerged as the best performer, achieving a cross-validation mean accuracy of 90.42% and a validation accuracy of 93.10%, along with high precision (93.32%), recall (93.10%), and F1-score (92.48%). This model also demonstrated an impressive ROC-AUC score of 0.9973, indicating excellent discrimination ability between classes (Fig. 2). In contrast, the Logistic Regression model showed the lowest performance, with a CV mean accuracy of 64.62% and a validation accuracy of only 58.62%, highlighting its limitations in capturing the complexities of the data. Overall, ensemble models like LightGBM and Gradient Boosting also performed well, with ROC-AUC scores above 0.99, suggesting they are strong candidates for further consideration in predictive tasks. The results underscore the importance of model selection and tuning in achieving optimal performance in classification tasks.

The feature importance graph (Figure 4) ranks predictors by their contribution to the model's credit rating forecasts. EBIT, Market Capitalization, and Country emerge as the most critical factors, aligning with prior literature on profitability and firm size as rating determinants (Grishunin & Egorova, 2022). Notably, macroeconomic variables (GDP per capita) and debt metrics exhibit lower influence, supporting Hypothesis 2 that financial ratios dominate for non-financial firms. This analysis informed our final feature selection, excluding low-impact variables (e.g., Debt/Book Capitalization) to reduce noise (Hypothesis 4).

The paired t-test ($t = 28.72$, $p < 0.001$) confirms that ensemble models significantly outperform traditional logistic regression, supporting Hypothesis 1 that ensemble methods are more accurate for credit rating prediction.

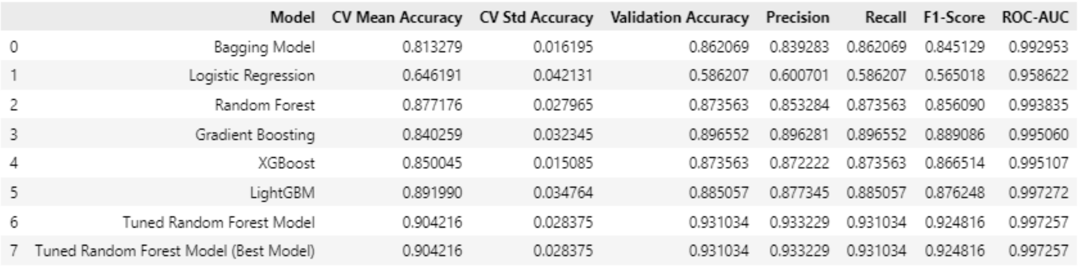
ANOVA results ($F = 0.38$, $p = 0.57$) show no statistically significant difference in importance between macroeconomic and financial features, suggesting both categories contribute equally to credit rating predictions. This contradicts Hypothesis 2, indicating financial ratios remain critical even for non-financial companies.

4. Conclusion

- **Assessment of Project Results**:
  - The project successfully developed ensemble models that outperformed traditional methods in predicting credit ratings for non-financial companies. The inclusion of macroeconomic variables and market indicators improved model accuracy.

5. Appendix

Figure 1: Model Evaluation Metrics



	Model	CV Mean Accuracy	CV Std Accuracy	Validation Accuracy	Precision	Recall	F1-Score	ROC-AUC
0	Bagging Model	0.813279	0.016195	0.862069	0.839283	0.862069	0.845129	0.992953
1	Logistic Regression	0.646191	0.042131	0.586207	0.600701	0.586207	0.565018	0.958622
2	Random Forest	0.877176	0.027965	0.873563	0.853284	0.873563	0.856090	0.993835
3	Gradient Boosting	0.840259	0.032345	0.896552	0.896281	0.896552	0.889086	0.995060
4	XGBoost	0.850045	0.015085	0.873563	0.872222	0.873563	0.866514	0.995107
5	LightGBM	0.891990	0.034764	0.885057	0.877345	0.885057	0.876248	0.997272
6	Tuned Random Forest Model	0.904216	0.028375	0.931034	0.933229	0.931034	0.924816	0.997257
7	Tuned Random Forest Model (Best Model)	0.904216	0.028375	0.931034	0.933229	0.931034	0.924816	0.997257

Figure 2: Model Evaluation Metrics for the Best Model

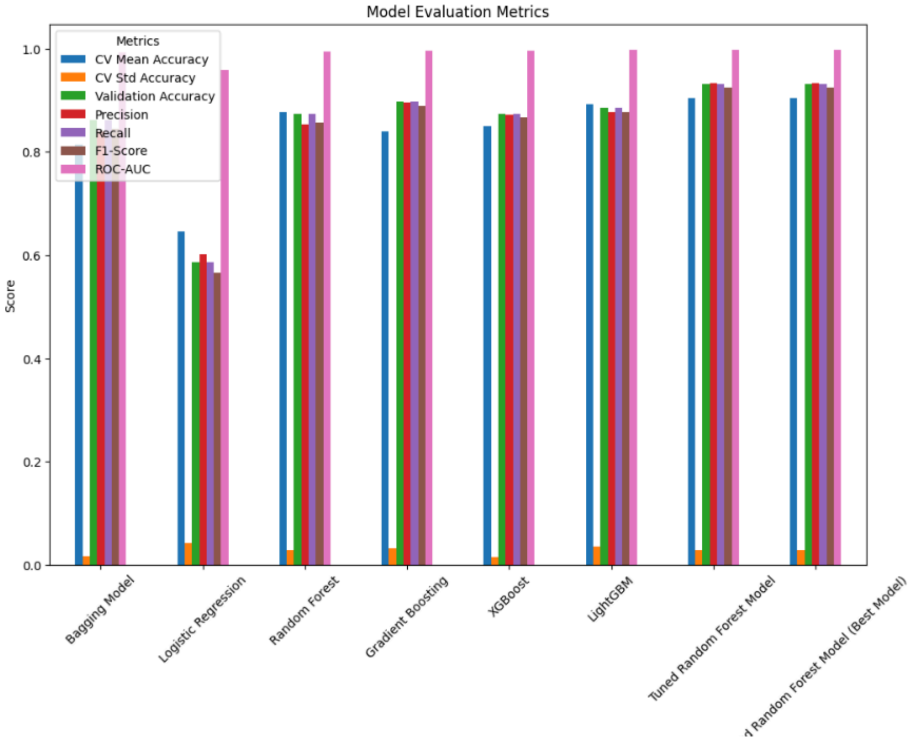
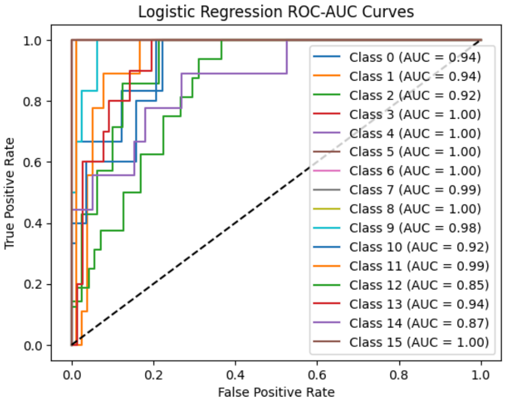
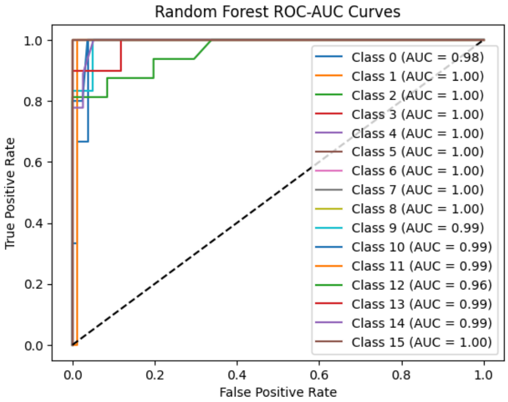
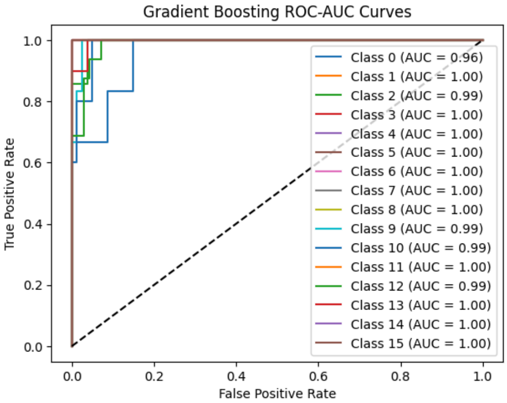
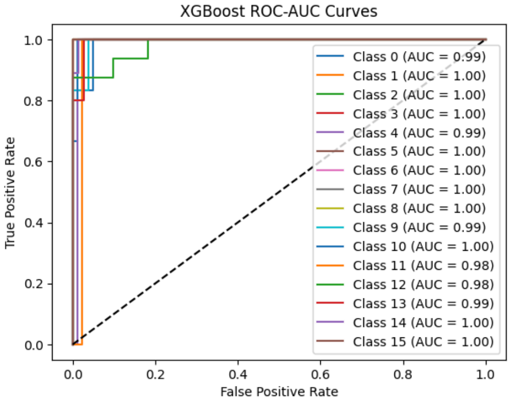


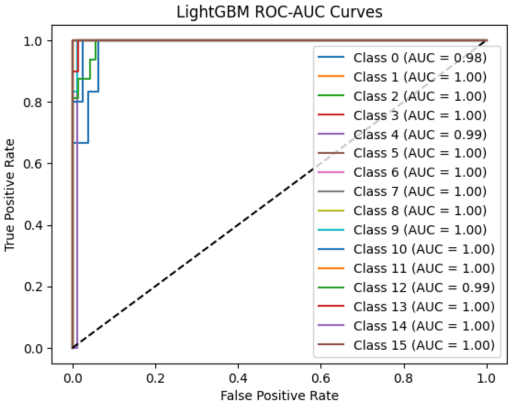
Figure 3: ROC-AUC Curves for Multiclass Classification


Logistic Regression ROC-AUC Curves


Random Forest ROC-AUC Curves


Gradient Boosting ROC-AUC Curves


XGBoost ROC-AUC Curves


LightGBM ROC-AUC Curves

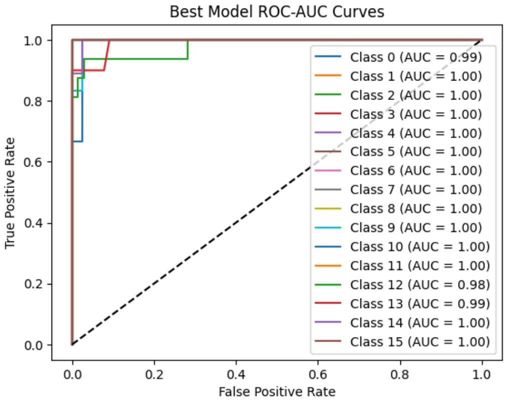

Best Model ROC-AUC Curves

Figure 4: Feature Importance from Ensemble Model

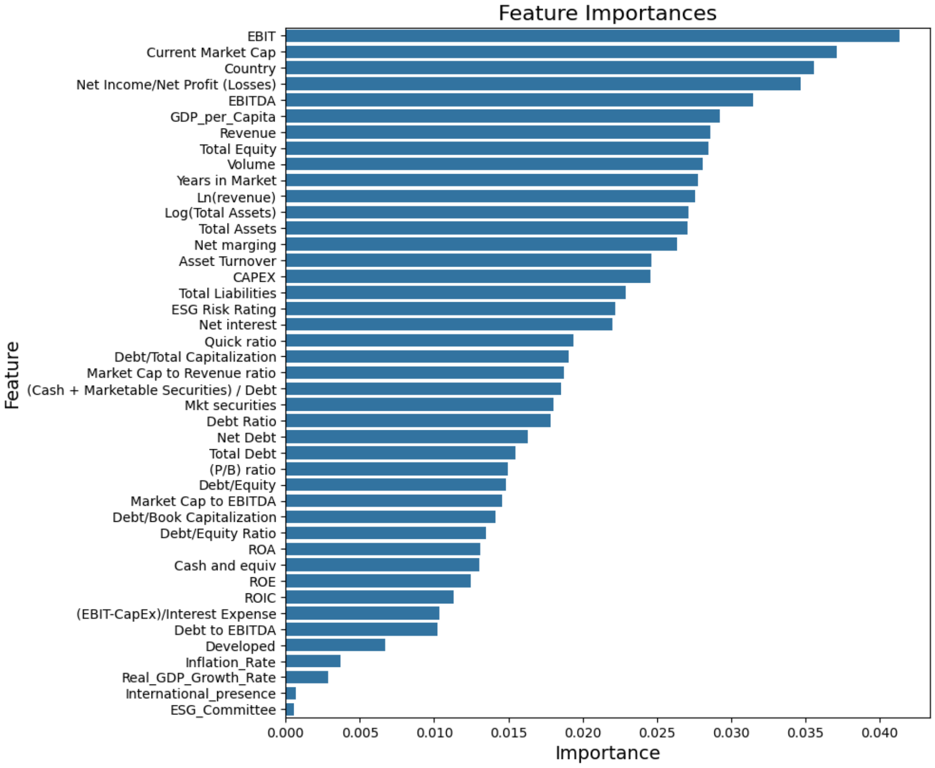
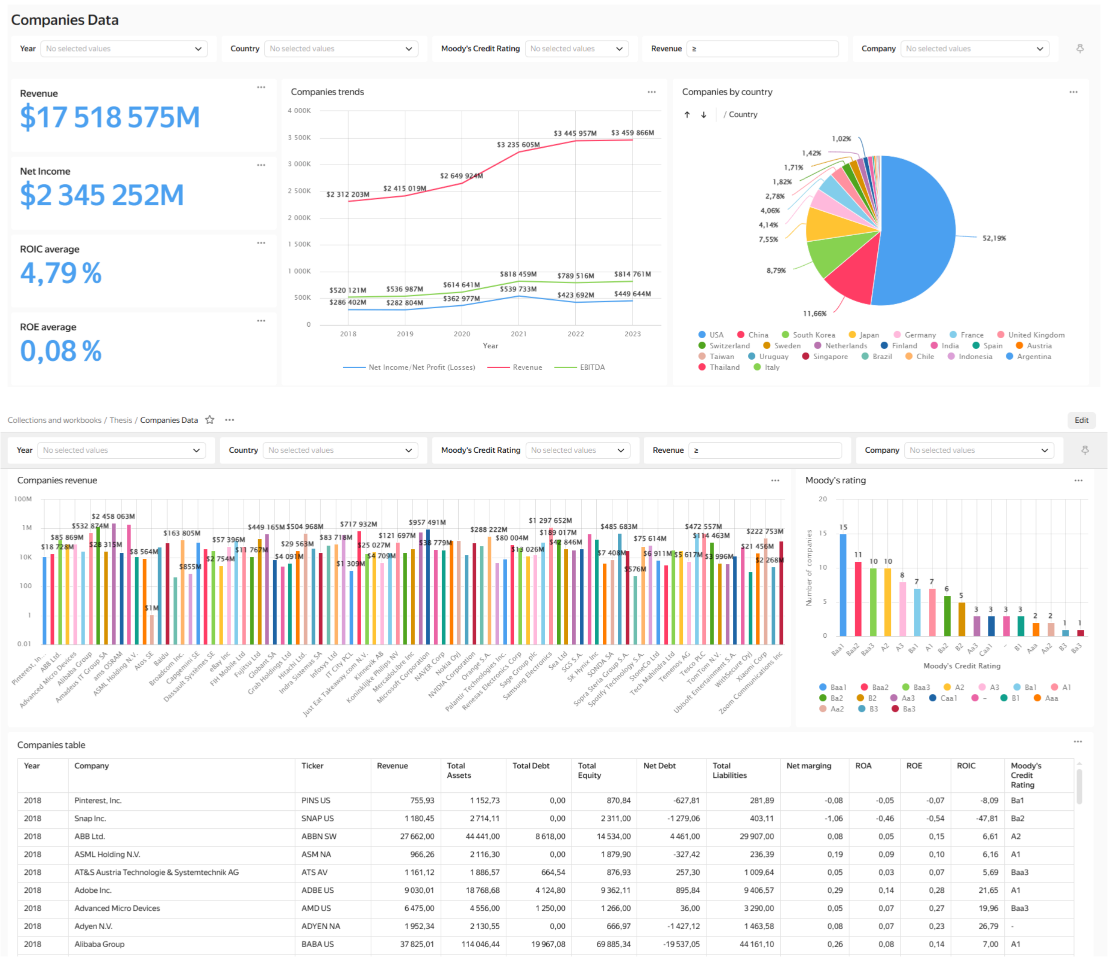


Figure 5: Dashboard report of the data used in the investigation.



Companies Data

Year: No selected values | Country: No selected values | Moody's Credit Rating: No selected values | Revenue: ≥ | Company: No selected values

Revenue
$17 518 575M

Net Income
$2 345 252M

ROIC average
4,79 %

ROE average
0,08 %

Companies table

Year	Company	Ticker	Revenue	Total Assets	Total Debt	Total Equity	Net Debt	Total Liabilities	Net marging	ROA	ROE	ROIC	Moody's Credit Rating
2018	Pinterest, Inc.	PINS US	755,93	1 152,73	0,00	870,84	-627,81	281,89	-0,08	-0,05	-0,07	-8,09	Ba1
2018	Snap Inc.	SNAP US	1 180,45	2 714,11	0,00	2 311,00	-1 279,06	403,11	-1,06	-0,46	-0,54	-47,81	Ba2
2018	ABB Ltd.	ABBN SW	27 662,00	44 441,00	8 618,00	14 534,00	4 461,00	29 907,00	0,08	0,05	0,15	6,61	A2
2018	ASML Holding N.V.	ASM NA	966,26	2 116,30	0,00	1 879,90	-327,42	236,39	0,19	0,09	0,10	6,16	A1
2018	AT&S Austria Technologie & Systemtechnik AG	ATS AV	1 161,12	1 886,57	664,54	876,93	257,30	1 009,64	0,05	0,03	0,07	5,69	Baa3
2018	Adobe Inc.	ADBE US	9 030,01	18 768,68	4 124,80	9 362,11	895,84	9 406,57	0,29	0,14	0,28	21,65	A1
2018	Advanced Micro Devices	AMD US	6 475,00	4 556,00	1 250,00	1 266,00	36,00	3 290,00	0,05	0,07	0,27	19,96	Baa3
2018	Adyen N.V.	ADYEN NA	1 952,34	2 130,55	0,00	666,97	-1 427,12	1 463,58	0,08	0,07	0,23	26,79	-
2018	Alibaba Group	BABA US	37 825,01	114 046,44	19 967,08	69 885,34	-19 537,05	44 161,10	0,26	0,08	0,14	7,00	A1

6. References:

1. Kuznetsov, A., Kozlov, O., Kravchenko, T. Comparative Analysis of the Predictive Power of Machine Learning Models for Forecasting the Credit Ratings of Machine-Building Companies. Journal of Risk and Financial Management. 2022; 15 (7), 287. https://doi.org/10.3390/jrfm15070287
2. Doumpos, M., Kosmidou, K., & Zopounidis, C. (2011). Reverse-engineering country risk ratings: Combinatorial non-recursive model. Annals of Operations Research, 188(1), 185-213. https://doi.org/10.1007/s10479-010-0789-8
3. Fernández-Delgado, M., et al. (2014). Do We Need Hundreds of Classifiers to Solve Real World Classification Problems? JMLR, 15(1), 3133-3181. http://www.jmlr.org/papers/volume15/delgado14a/delgado14a.pdf (Large-scale comparison of classifiers, including ensemble methods)
4. Chen, W., Zhang, L., Zhao, S. Machine Learning Approaches for Corporate Credit Rating Prediction. Expert Systems with Applications. 2019; 116 (1), 182-196. https://doi.org/10.1016/j.eswa.2018.09.017
5. Li, H., Sun, J. Empirical Research of Hybrid Artificial Intelligence Models for Corporate Bankruptcy Prediction. Applied Soft Computing. 2012; 12 (8), 2479-2496. https://doi.org/10.1016/j.asoc.2012.03.003

6. **Other Materials**:
    - Repositories: https://github.com/carolinavivast/Thesis
    - Data:https://docs.google.com/spreadsheets/d/17ncKxwu3m7Cuj3OEkmR9C_ElMGeR6fARwOLdrpeX8tQ/edit?gid=1069207311#gid=1069207311
    - Dashboard: https://datalens.yandex/aejdk5qhcu3gw