Bankruptcy Prediction for Polish Companies

Flatiron School

Kamile Yagci 5/19/2022

Content

- Business Problem
- Data
- Method
- Analysis and Results
- Interpretation of Results
- Conclusion
- Future Work
- Questions

Business Problem

KPMG, international corporate financial consulting firm, hired me to analyze the financial standing of the Polish companies.

The goal of the analysis is identifying whether the business will go to bankruptcy in 1-5 years or not.

KMPG will use the results of this study to provide an early warning to Polish business clients on their financial standings, so they can take preventive actions.

Data (1)

- contains the financial information and bankruptcy status of Polish companies.
- collected from Emerging Markets Information Service (EMIS).
- collected in the period of
    - 2000-2012 for the bankrupt companies
    - 2007-2013 for the still operating companies

Data (2)

Depending on the forecasting period, the data is classified in five categories/datasets.:

- 1st Year: financial rates from 1st year of the forecasting period AND class label that indicates bankruptcy status after 5 years. (Data 1)
- 2nd Year: financial rates from 2nd year of the forecasting period AND class label that indicates bankruptcy status after 4 years. (Data 2)
- 3rd Year: financial rates from 3rd year of the forecasting period and AND class label that indicates bankruptcy status after 3 years. (Data 3)
- 4th Year: financial rates from 4th year of the forecasting period AND class label that indicates bankruptcy status after 2 years. (Data 4)
- 5th Year: financial rates from 5th year of the forecasting period AND class label that indicates bankruptcy status after 1 year. (Data 5)

Data (3)

- Data has 64 attributes for each company.
- The number of companies in each dataset and class distributions:

Data #	Total	Still Operating (class=0)	Bankrupt (class=1)
Data 1	7027	6756	271
Data 2	10173	9773	400
Data 3	10503	10008	495
Data 4	9792	9277	515
Data 5	5910	5500	410

Method (1)

- This is a binary classification problem, since the project goal is to identify whether the company will bankrupt or not.
- I use Ensemble Method 'XGBoost', eXtreme Gradient Boosting.
- I focus on minimizing the False Negatives on predictions when designing my classifier model.

Method (2)

- Evaluation metrics are used to measure the performance of the models:
  - precision: What percentage of model predictions are true?
  - recall: What percentage of the classes we're interested in were actually captured by the model?
  - f1-score: Harmonic Mean of Precision and Recall
  - accuracy: Out of all the predictions our model made, what percentage were correct?
  - AUC: Area under the ROC Curve, graphical plot that illustrates the true positive rate against the false positive rate.

Analysis and Results (1)

- The model is optimized to
  - increase the ‘recall’ score (higher recall = less false negatives)
  - maintain moderate values on other metrics
  - decrease the performance gap between training and testing (reduce overfitting)
- The model performance is initially studied on Data 3; three best models determined.
- The final model is selected based on the performance of three best models on five datasets.

Analysis and Results (2)

- Table shows the the performance of Final Model on five datasets.
- In each dataset, 80% of data is used for training the model and 20% for testing.
- Common best predictors:
  - X27: profit on operating activities / financial expenses
  - X34: operating expenses / total liabilities
  - X5: [(cash + short-term securities + receivables - short-term liabilities) / (operating expenses - depreciation)] * 365

Data	Sample	precision	recall	f1	accuracy	auc
Data 1	Train	0.8700	1.0000	0.9300	0.9940	1.0000
Data 1	Test	0.6410	0.8040	0.7130	0.9770	0.9610
Data 2	Train	0.8750	1.0000	0.9330	0.9940	1.0000
Data 2	Test	0.5060	0.6200	0.5570	0.9660	0.9280
Data 3	Train	0.7710	1.0000	0.8710	0.9860	1.0000
Data 3	Test	0.5350	0.7200	0.6140	0.9540	0.9230
Data 4	Train	0.8060	1.0000	0.8930	0.9870	1.0000
Data 4	Test	0.5560	0.6800	0.6110	0.9550	0.9490
Data 5	Train	0.8290	1.0000	0.9070	0.9860	1.0000
Data 5	Test	0.6070	0.7890	0.6860	0.9450	0.9600
Average	Train	0.8302	1.0000	0.9068	0.9894	1.0000
Average	Test	0.5690	0.7226	0.6362	0.9594	0.9442

Interpretation of Results

- Model correctly identifies
  - 80.4% of the true bankrupt companies, which will bankrupt 5 years later
  - 62.0% of the true bankrupt companies, which will bankrupt 4 years later
  - 72.0% of the true bankrupt companies, which will bankrupt 3 years later
  - 68.0% of the true bankrupt companies, which will bankrupt 2 years later
  - 78.9% of the true bankrupt companies, which will bankrupt 1 years later
- Among the model predicted bankruptcy companies, on average 57% of them are true bankrupt companies.

Future Work

- Create seperate final models for each dataset; not just one final model that applies on all.

- Search for alternative Classifier methods/tools.

Questions?