Bankruptcy Prediction for Polish Companies

Flatiron School

Kamile Yagci 5/19/2022

Outline

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

Business Problem

KPMG, international corporate financial consulting firm, hired me to analyze the financial standings 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)
  - downloaded from UCI
- collected in the period of
  - 2000-2012 for the bankrupt companies
  - 2007-2013 for the still operating companies
- has 64 attributes for each company
- does not contain company ID

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)

- 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

- About 5% of data belong to bankrupt companies.

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 predicted bankrupt companies are truly bankrupt?
  - recall: What percentage of true bankrupt companies were actually captured by the model?
  - f1-score: Harmonic Mean of Precision and Recall
- Each metric range from 0-1; higher the better

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.
- In each dataset, 80% of data is used for training the model and 20% for testing.

Analysis and Results (2)

Model Prediction Performance for Bankrupt Companies

Data	Sample	precision	recall	f1
Data 1	Train	0.8700	1.0000	0.9300
Data 1	Test	0.6410	0.8040	0.7130
Data 2	Train	0.8750	1.0000	0.9330
Data 2	Test	0.5060	0.6200	0.5570
Data 3	Train	0.7710	1.0000	0.8710
Data 3	Test	0.5350	0.7200	0.6140
Data 4	Train	0.8060	1.0000	0.8930
Data 4	Test	0.5560	0.6800	0.6110
Data 5	Train	0.8290	1.0000	0.9070
Data 5	Test	0.6070	0.7890	0.6860
Average	Train	0.8302	1.0000	0.9068
Average	Test	0.5690	0.7226	0.6362

Model Prediction Performance for Still Operating Companies

Data #	sample	precision	recall	f1
Data 1	Train	1.00	0.99	1.00
Data 1	Test	0.99	0.98	0.99
Data 2	Train	1.00	0.99	1.00
Data 2	Test	0.99	0.98	0.98
Data 3	Train	1.00	0.99	0.99
Data 3	Test	0.98	0.97	0.98
Data 4	Train	1.00	0.99	0.99
Data 4	Test	0.98	0.97	0.98
Data 5	Train	1.00	0.99	0.99
Data 5	Test	0.98	0.96	0.97
Average	Train	1.00	0.99	0.99
Average	Test	0.98	0.97	0.98

Interpretation of Results (1)

- 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 bankrupt companies, on average 57% of them are true bankrupt companies.
- For still operating companies, model is highly successful in identification.
- There is a performance gap between the training and testing results.

Interpretation of Results (2)

- Example: A Polish company contacts KMPG and requests a report on whether it will bankrupt 1 year later.
  - Bankruptcy status after 1 year: Data 5
  - Case 1: It is a true bankrupt company (status after 1 year)
    - Model, with 78.9% of probability, will identify it as ‘bankrupt’
  - Case 2: It is a true still operating company (status after 1 year)
    - Model, with 96% of probability, will identify it as ‘still operating’

Future Work

- Model needs to be improved !
- Create separate final models for each dataset; not just one final model that applies on all.
- Search for alternative classifier methods.
- Gather more data that spans 3-5 years of financial rates for each company. One year period of financial data is not enough to make predictions on company standing.

Questions?

Thank You!