Can Bankruptcy be Predicted using Machine Learning?

An analysis for Investors and Business Owners

Blake McMeekin ● 07.01.2022

Why do we care?

Bankruptcy is the death of an organization.
When an organization goes bankrupt, their mission ends. Stakeholders lose money and time, while key people can have reputations and confidence ruined. When it comes to bankruptcy:

- What can we measure?
- What can we predict?
- What can be done?

What can we measure?

For this analysis, we'll dive into data released by the Taiwan Stock Exchange and Taiwan Economic Journal.

About our Data

- Organizational Financial data

- ~7000 rows, 96 columns

- From 1999 to 2009

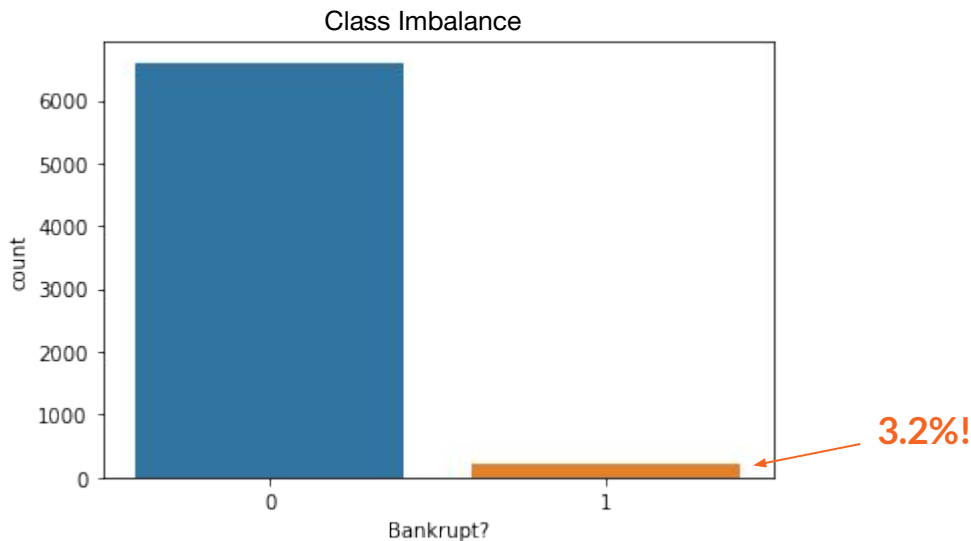- Provided cleaned, normalized, and numeric

Our Goal

- Understand how bankruptcy might be predicted

- Understand causes

- Actionable advice for organizations

Main Problem: Class Imbalance

Only about 3% of the recorded organizations are actually bankrupt. This class imbalance will be the main problem in our analysis.

Class Imbalance

3.2%!

Our Analysis

1.) Exploratory Data Analysis → **2.) Predictive Models** → **3.) Evaluation & Implementation**

- Correlations

- Data Prep

- SMOTENC

- Logistic Regression

- XGBoost

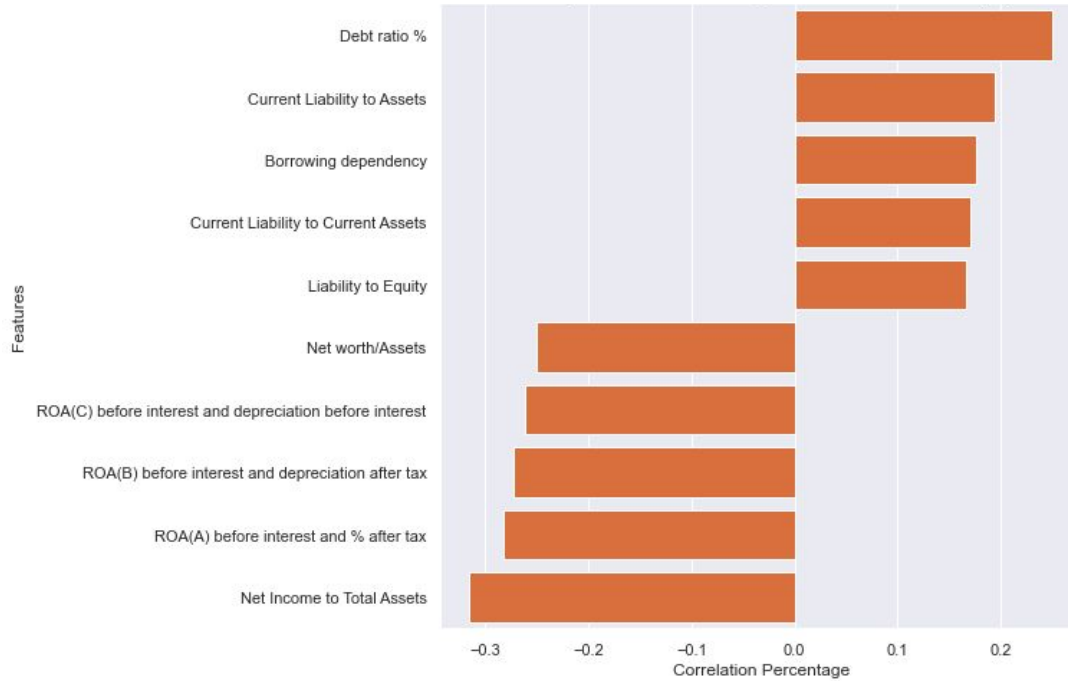- Random Forest

- Precision

- Recall

- Integration

- Uses

Findings

Findings

1. No single financial feature explains every organizational failure (at most 31% correlated)

2. Machine Learning techniques usefully interpret the interactions of many features

3. Our most accurate model (XGBoost) was about 98.5% accurate* (F1 of 0.51)

———

Single Features Related to Bankruptcy

Most of these should be fairly obvious, but now we can quantify their importance.

Findings - What can we predict?

Logistic Regression creates Inclusive predictions

Marks ~80% of bankrupt companies in unseen data; marks represent about a *1 in 5 chance of bankruptcy. Useful for early warnings.

XGBoost creates Precise predictions

Marks ~50% of bankrupt companies in unseen data; marks represent about a *50/50 chance of bankruptcy. Useful for second-stage confirmation.

*Remember that 3% of companies in the dataset are bankrupt. With this sort of imbalance, it's common to have many misclassifications.

Findings - What can be done?

> Find Organizational risks

Financial data can now be fed to our model to score organizations for solvency and identify complex risks that might be missed by humans.

> Score Budgets

Different budgeting plans could be scored by feeding in predicted financial data and assessing model confidence.

> Business Opportunities

Early identification of affordable assets or IP, identifying and consulting at-risk companies, avoiding dangerous partnerships

What's next?

Collect More Data

Larger financial datasets may increase the value and robustness of our predictive model.

Improve Existing Data

Our input data is quite complex. Can it be simplified? Are our models struggling with any distortions introduced to our data before it's in our hands?

Build a dashboard

A dashboard may make the analysis of new data more accessible for more people.

Thank You!
Questions?

Analysis by Blake McMeekin

Contact:

github.com/thegrandblooms

linkedin.com/in/blakemcme/

blakemcme@gmail.com