Financial Ratios and Bankruptcy Prediction:

An analysis of data from the Taiwan Stock Exchange

Blake McMeekin • 07.01.2022

Our Problem

Can we determine Bankruptcy by analyzing Financial Ratios?

If so, how can this steer our investment decisions and company partnerships?

- Which measures matter?
- How confident can our predictions be?
- Which predictive model performs the best?

Our Problem

Access to large amounts of financial data

- Sourced from the Taiwan Stock Exchange
- Data cleaned and provided by the Taiwan Economic Journal

Can we predict Bankruptcy?

- Insolvent companies are those which are not able to pay outstanding debts or continue providing services

Our Data

About our Data

- Data spanning from 1999 to 2009
- 6819 rows and 96 columns
- Data is provided cleaned and normalized
- All data is numeric

Our Target

- Predict the “Bankrupt” Label
- Only about 3.2% of companies are labeled as “bankrupt” - a considerable class imbalance
- Target is binary

Approaches

Logistic Regression

Our simplest model - Logistic Regression fits a log line representing the odds ratio of our target, given features

XGBoost

XGBoost is an Extreme Gradient-boosted decision tree alg. which iteratively builds shallow decision-trees to minimize past errors

Random Forest

Random Forest models build large numbers of shallow decision trees

Approaches

Hyperparameter Tuning

- RandomizedSearchCV is used to test hyperparameters
- These hyperparameters can be tuned iteratively to improve on our baseline model

Model Evaluation

- Lastly, we check model performance against holdout data and determine its performance. Because of our class imbalance, training data is oversampled while test data is not.

Findings

Findings

1. We found that the best algorithm depends on the stakeholder goal
2. The most accurate model was about 98.5% accurate
3. The most inclusive model correctly labeled 84% of the bankrupt companies

Findings - F1 vs Accuracy

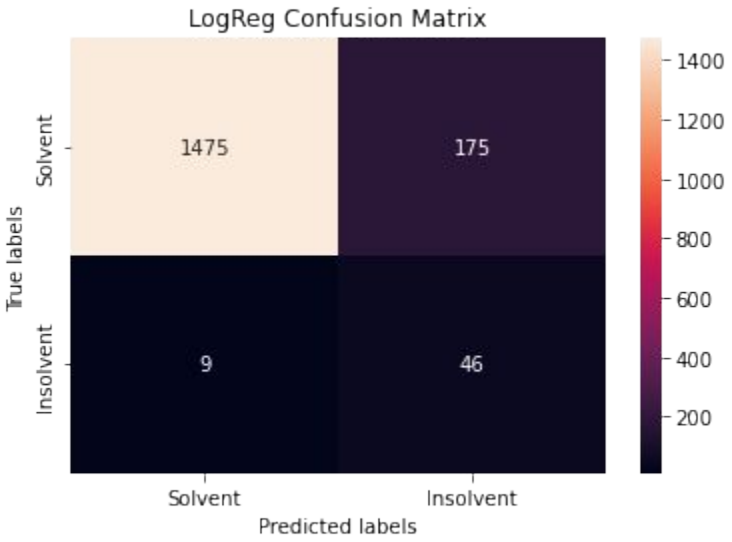
Accuracy

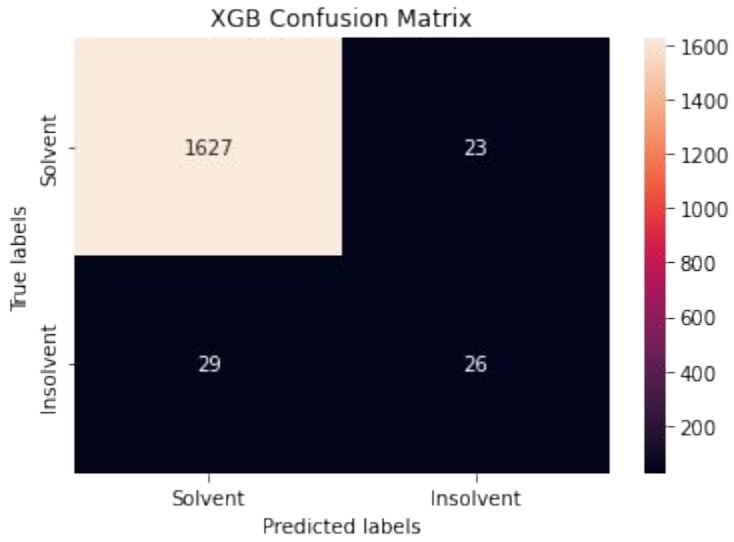
- Because of our significant class imbalance, accuracy is a bad indicator of model performance. A model that predicts no bankruptcies at all would be about 96.8% accurate.

F1

- F1 is a better metric that combines precision and recall. Precision references how few errors are made, while recall indicates how much of our target data is correctly labeled.

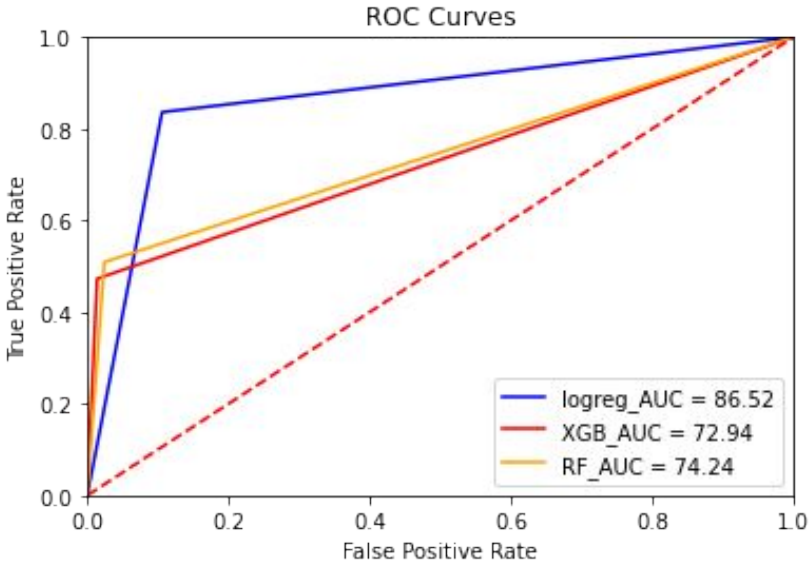
Findings - Confusion Matrices





Logistic Regression has better Recall, while XGB has better Precision

Findings - ROC Curve



Logistic Regression had the best Area Under the Curve

Findings - Important Features

Our three most important features:

- “Fixed Assets to Assets” - Strong positive correlation
- “Accounts Receivable Turnover” - Strong negative correlation
- “Interest-bearing debt interest rate” - Strong negative correlation

More data related to these features may help us with our bankruptcy prediction models, or inform business decisions to avoid bankruptcy.

Findings - Recommendations

Use Logistic Regression to flag companies

Our logistic Regression model found 80% of the bankrupt companies, but had many false positives.

Use XGBoost for a confident confirmation

XGBoost was very precise, but missed about 50% of bankrupt targets.

Bankruptcy is not entirely predicted by numbers

We can make estimations that are a lot better than chance, but bankruptcy was not totally predicted by the available data.

Next steps

More data about the data

Is the financial data from the same time as the bankruptcy label? Can we identify where the noise or discrepancy is coming from? Is it a consequence of the way the data was processed before our analysis?

More Feature Engineering

Are a high number of features reducing the accuracy of our models?

More Access to Raw Data

There may be more information which could be recovered.

Thank You!
Questions?

Analysis by Blake McMeekin

Contact:

github.com/thegrandblooms

linkedin.com/in/blakemcme/

blakemcme@gmail.com