Corporate Bankruptcy Prediction with Ensemble Methods

Authors: Ben Spilsbury and
Jake Oddi

Overview

Objective: Analyze Polish corporate financial data spanning 2000-2013 to predict bankruptcy

Goals of Analysis:

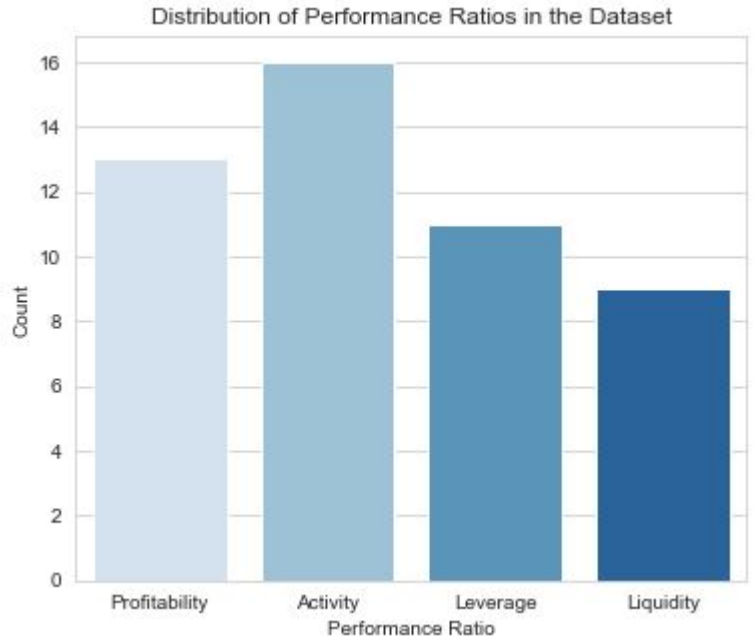
1. Make a good-integrity model to predict corporate bankruptcy
2. Determine the most important features as decided by our model

Data

- Set from Emerging Markets Information Service via UCI Machine Learning Repository
- 10,503 data-points (Polish companies)
- 64 features (various performance ratios)

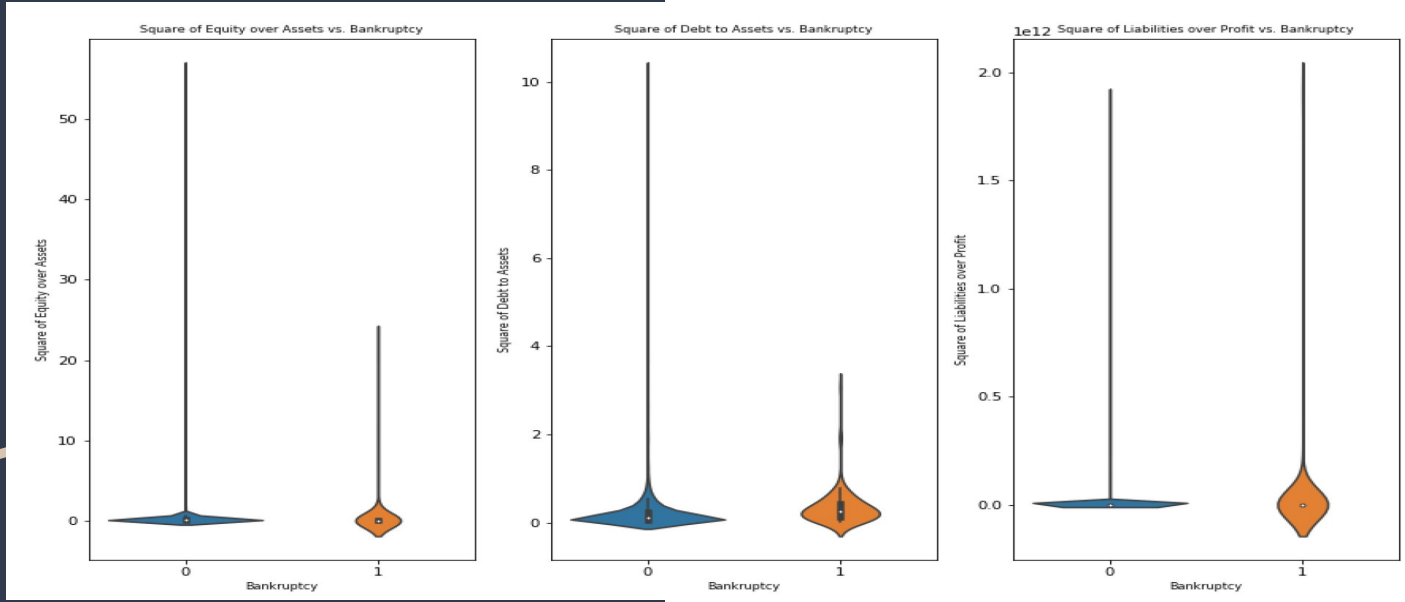
Understanding Performance Ratios

- Profitability
  - Asses revenue relative to costs, assets, and equity
- Activity
  - Measure of how efficiently a company is using its assets to generate sales
- Leverage
  - Asses debts relative to capital and assets
- Liquidity
  - Asses how liquid a company’s assets are relative to its debts



Data Cleaning and Feature Engineering

- ***Cleaning:*** Drop null values, convert all features to numerics, rename columns
- ***Feature Engineering:*** Squares of top three important features (based on correlation to target)

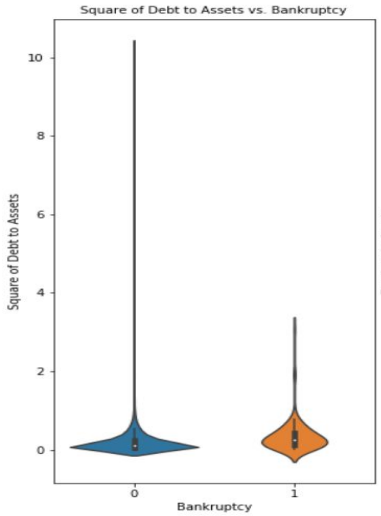


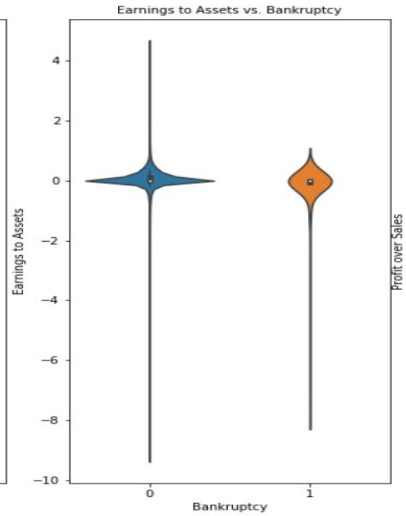
Modeling Methodology

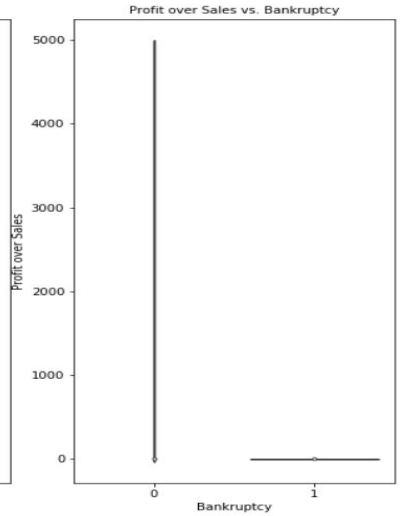
1. Test / train split
2. Create KNN, Logistic Regression, Decision Tree, and Random Forest models
3. Tune hyperparameters
4. Create ensembles
5. Judge models against:
    - F1-Scores (take into account class imbalance)
    - Precision (discourage false positives)

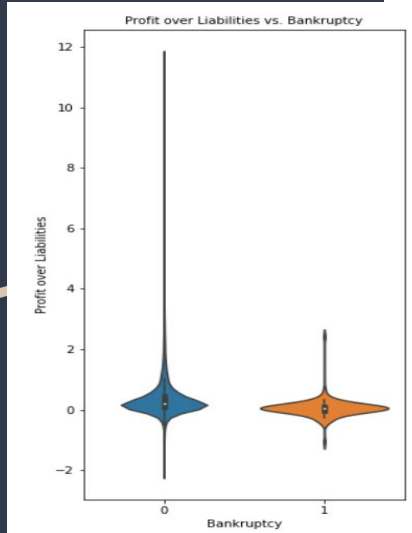
Results

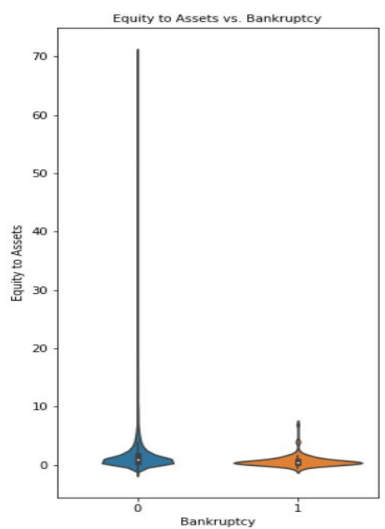
- Logistic Regression, Decision Tree, Random Forest ensemble performs best
  - F1-Score: **0.941**
  - Accuracy: **0.997**
  - Precision: **1.0**
- 7 features found to be most important to overall model

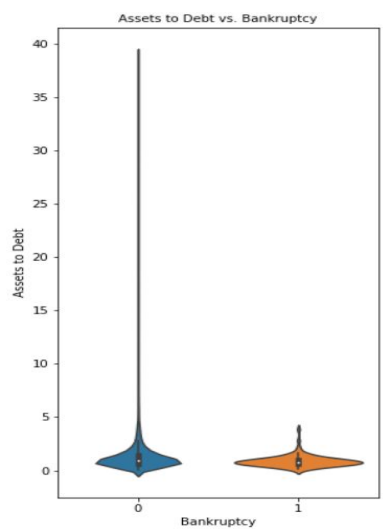


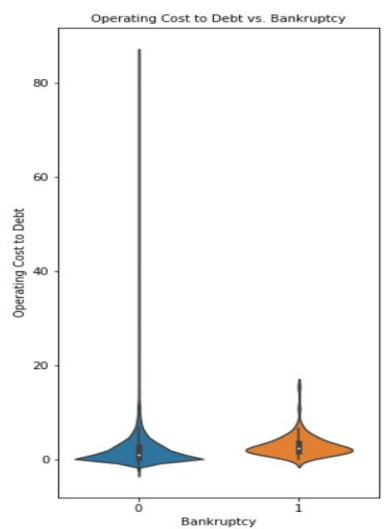












Conclusion and Next Steps

Leverage and liquidity as top bankruptcy predictors.

- 3/7 important features are leverage ratios
- 2/7 important features are liquidity ratios

→ High Leverage + Low Liquidity = High Chance of Going Bankrupt

Looking forward…

- Expand data set globally beyond Poland
- Use more recent data
- Webscrape hard-to-find U.S. bankruptcy data