Using Classification Learning to Predict Bankruptcy

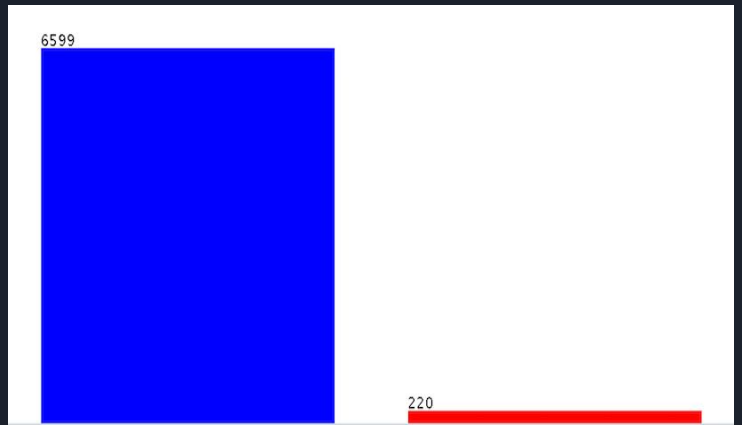
Davis Gilmore and Walker Browne

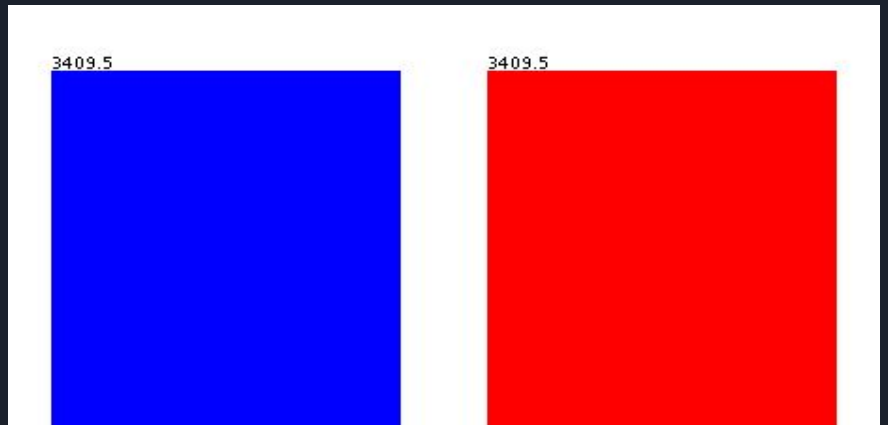
Dataset

- Our dataset was obtained from Kaggle
  - Made up of financial data from the Taiwan Economic Journal from 1999-2009
- Consists of 96 attributes and 6819 instances
  - 93 Numeric Attributes and 3 Nominal Attributes
- Key Attributes
  - inventoryOverWorkingCapital - numeric
  - liabilityAssetsFlag - nominal
  - totalLiabilityEquityRatio - numeric
  - workingCapitalToTotalAssets - numeric
  - debtRatio - numeric
  - borrowingDependency - numeric

Data Preparation

- Unbalanced Data
  - We ran ZeroR on the original data and got an accuracy of 96.7737%
    - Indicator of unbalanced class attribute values
  - We applied the *ClassBalancer* filter to our dataset
    - Evenly weighs the class attribute values in a 1:1 Ratio
    - ZeroR then produced an accuracy of 49.9932%
  - **The charts below show the class attribute weights before and after *ClassBalancer* was applied.**





Glossary of Frequently Used Terms

Bankruptcy: The financial state of being unable to pay outstanding debts

Asset: Any resource owned or controlled by a company

- Cash
- Equipment
- Buildings

Liability: Anything owed by a company, usually a borrowed sum of money

Liquidity: An asset's ease of conversion to cash:

- Cash → Most Liquid
- Building → Not Very Liquid - Takes a long time to convert to cash

Algorithms Used in Analysis

- PART
  - Similar to PRISM
  - Uses a “separate and conquer” technique
    - Identifies a rule
    - Separates all of the rules it covers
    - “Conquers” the remaining instance

Results of PART

- Tested with and without ClassBalancer
- Generated 38 rules
  - Some including:
    - totalIncomeOverTotalExpense > 0.002828: 0 (44.95)
    - currentRatio <= 0.013305 AND operatingProfitGrowthRate <= 0.850845 AND totalExpenseOverAssets > 0.041434 AND currentLiabilityToAssets <= 0.201594: 0 (39.27)
    - interestCoverageRatio <= 0.563998: 1 (32.03/1.03)

Algorithms Used in Analysis

- IBK (Nearest Neighbor Algorithm)
  - Lazy learning
  - Implements the Euclidean Distance formula
    - Assumes attributes have been normalized
    - Assumes attributes have equal importance

Results of IBK

- Tested with Class Balancer filter on each k value
- Tested with various k values
  - When k = 1, it had an accuracy of 61.0655% with 1371.7215 correctly identified instances.
  - When k = 3, it had an accuracy of 66.1246% with 1485.3652 correctly identified instances
  - When k = 19, it had an accuracy of 79.3004% with 1781.3345 correctly identified instances

Algorithms Used in Analysis

- Naive Bayes
  - Classification algorithm that assumes each attribute is independent of each other
  - Predicts the probability of a class attribute, given the value of a specific attribute
  - When applied in Weka, the algorithm produces a confusion matrix of correctly and incorrectly classified instances, as well as the mean, standard deviation, precision, and weight sums for each attribute.

Results of Naive Bayes

- We split our dataset into 80/20 - Training/Testing and ran Naive Bayes
- Naive Bayes without *ClassBalancer:* 71.6566% Accuracy - 1661 correctly classified instances
- Naive Bayes with *ClassBalancer:* 76.2869% Accuracy - 1713.64 correctly classified instances

- Naive Bayes produces a mean and standard deviation for each attribute in the dataset which gave some interesting insights to the distribution of values

Results of Naive Bayes cont.

Quick Ratio: A company's ability to meet its short term debts/liabilities by utilizing their most liquid assets

- High Quick Ratio: Usually indicates better liquidity and better overall financial health
- Low Quick Ratio: Usually indicates lower liquidity and poor financial health

Instances classified as bankrupt by Naive Bayes: Mean Quick Ratio of ~4.1 Million

- Lower Quick Ratio - Therefore lower liquidity and higher risk of bankruptcy

Instances classified as not bankrupt by Naive Bayes: Mean Quick Ratio of ~ 7.2 Million

- Higher Quick Ratio - Therefore higher liquidity and lower risk of bankruptcy

Results of Naive Bayes cont.

Liability Assets Flag: Nominal indicator for if a company's total liabilities exceed their total assets (0 = no, 1 = yes)

- Out of 3317 instances classified as bankrupt, 94 instances raised the liability assets flag
  - 94/3317 = .028 or 2.8%
- Out of 3409 instances classified as not bankrupt, only 2 instances raised the liability assets flag
  - 2/3409 = .0006 or .06%
- 94 bankrupt flags/ 96 total flags = .979 or 97.9%
  - Therefore if a company's liabilities exceed their total assets, it's highly likely they will be bankrupt

Results of Naive Bayes cont.

Cash Over Total Assets: The ratio of a company's total cash divided by the total value of their assets.

Lower Ratio usually indicates lower liquidity.

Higher Ratio usually indicates higher liquidity.

- Instances Classified as Bankrupt: Mean ratio of .0477
- Instances Classified as Not-Bankrupt: Mean Ratio of .1266
  - .1266/.0477 = 2.654
  - Non-Bankrupt companies had a mean ratio of nearly 3 times greater than bankrupt companies

Algorithms Used in Analysis

J48: Classification algorithm that builds a tree made up of rules produced for the attributes in the dataset

- J48 produces a confusion matrix, a list of the rules produced and amount of instances correctly classified by that rule, as well as a visualizable tree that represents each rule produced.
- Because of our extensive list of attributes, we figured there had to be some noisy attributes we could eliminate in order to produce a smaller and more concise tree.
    - Using the mean and standard deviations produced, we identified several attributes with nearly identical mean and standard deviations (operatingProfitRate, regularNetProfitGrowthRate, and totalAssetReturnGrowthRate)

Results of J48

- After removal of noisy attributes, we ran J48 to compare accuracy and size of the tree produced.
  - J48 with noisy attributes: 70.6552% and 111 leaves
  - J48 without noisy attributes 69.6664% with 93 leaves
- Some Rules Produced by J48:
  - If inventoryOverWorkingCapital > .276357 THEN Bankrupt? = YES (266.21)
  - If borrowingDependency > .3779905 THEN Bankrupt? = YES (98.37)
  - If totalLiabilityEquityRatio > .01093 AND quickRatio <= .00494 THEN Bankrupt? = YES (48)

Results of J48 cont.

- Largest Portion: 5654 Instances with an average Inventory/Working Capital Ratio of .26
- 2nd Largest Portion: 328 Instances with an average Inventory/Working Capital Ratio of .278

Results of J48 cont.

Largest Portion: 5382 instances with an average borrowing dependency of .36

2nd Largest Portion: 514 instances with an average borrowing dependency of .384

Conclusion

Highest Overall Accuracy Achieved:

- Naive Bayes with *ClassBalancer* applied and an 80% split: 76.289% and 1713.64 instances classified correctly

Most Predictive Attributes

- Debt Related Attributes: borrowingDependency and liabilityEquityRatio
  - Higher values generally pointed towards bankruptcy
- Net-Worth Related Attributess: totalIncomeOverTotalExpense and cashOverTotalAssets
  - Lower values generally pointed towards bankruptcy

Important Takeaways

- Manage your debts with the utmost caution
- Don’t take out lines of credit you aren’t sure you can cover
- Financial risk can be both beneficial and detrimental for a company or individual; Make sure your risk is calculated

Questions?