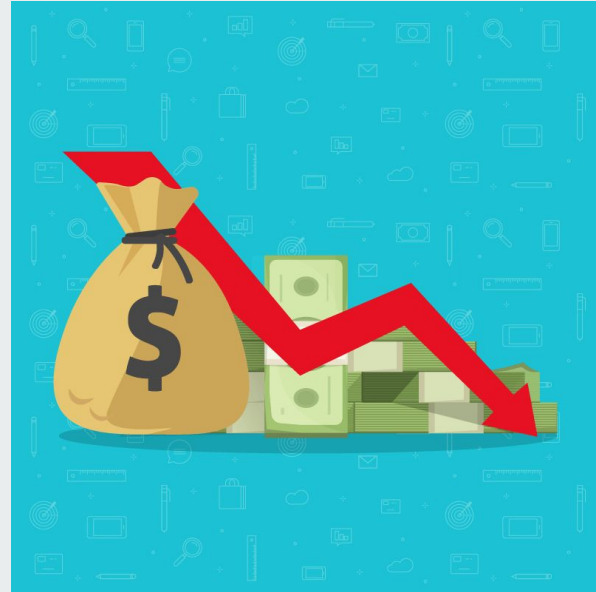# Company Bankruptcy Prediction

Masha Volkova

Company Bankruptcy Dataset

Source: Kaggle

- Data collected from Taiwan Economic Journal for the years 1999- 2009.
- Bankruptcy: defined by business regulations of Taiwan Stock Exchange

Number of instances: 6819, Number of attributes: 96

Response Variable: Bankrupt (yes/no)

Explanatory Variables: 95 attributes

- Cash Turnover Rate
- Borrowing dependency
- Net value per share
- Net income to total assets
- Operating expense rate

Analysis Overview

Analysis:

- Logistic regression
- Decision Tree (CART model)
- Random Forest

Programming Language: R

Training/Testing set split: 0.6

Logistic Regression

Coefficients: (4 not defined because of singularities)

	Estimate
(Intercept)	7.763e+08
ROA.C..before.interest.and.depreciation.before.interest	-1.904e+01
ROA.A..before.interest.and...after.tax	2.818e+00
ROA.B..before.interest.and.depreciation.after.tax	1.768e+01
Operating.Gross.Margin	1.829e+04
Realized.Sales.Gross.Margin	1.092e+02
Operating.Profit.Rate	-1.449e+06
Pre.tax.net.Interest.Rate	1.212e+06
After.tax.net.Interest.Rate	6.284e+01
Non.industry.income.and.expenditure.revenue	-6.930e+05
Continuous.interest.rate..after.tax.	-1.071e+03
Operating.Expense.Rate	7.562e-11
Research.and.development.expense.rate	-1.465e-11
Cash.flow.rate	-3.435e+01
Interest.bearing.debt.interest.rate	8.178e-10
Tax.rate..A.	-2.057e-01
Net.Value.Per.Share..B.	-1.415e+02
Net.Value.Per.Share..A.	-1.543e+02
Net.Value.Per.Share..C.	2.681e+02
Persistent.EPS.in.the.Last.Four.Seasons	-2.452e+01
Cash.Flow.Per.Share	-1.939e+01
Revenue.Per.Share..Yuan...	2.051e-08
Operating.Profit.Per.Share..Yuan...	-2.046e+02
Per.Share.Net.profit.before.tax..Yuan...	2.005e+01
Realized.Sales.Gross.Profit.Growth.Rate	3.053e+00
Operating.Profit.Growth.Rate	-2.367e+01
After.tax.Net.Profit.Growth.Rate	-1.647e+01
Regular.Net.Profit.Growth.Rate	2.311e+01
Continuous.Net.Profit.Growth.Rate	3.726e+00
Total.Asset.Growth.Rate	-4.713e-11
Net.Value.Growth.Rate	-8.198e-10
Total.Asset.Return.Growth.Rate.Ratio	1.338e+01
Cash.Reinvestment..	2.905e+00
Current.Ratio	-1.283e+01
Quick.Ratio	-3.983e-10
Interest.Expense.Ratio	-5.007e+00
Total.debt.Total.net.worth	6.681e-09
Debt.ratio..	1.260e+01
Net.worth.Assets	NA
Long.term.fund.suitability.ratio..A.	2.629e+00
Borrowing.dependency	1.098e+02
Contingent.liabilities.Net.worth	4.486e+01
Operating.profit.Paid.in.capital	2.266e+02
Net.profit.before.tax.Paid.in.capital	-2.260e+01
Inventory.and.accounts.receivable.Net.value	-2.911e+01
Total.Asset.Turnover	-2.021e+00
Accounts.Receivable.Turnover	-4.480e-08
Average.Collection.Days	-1.102e-08
Inventory.Turnover.Rate..times.	-2.027e-11
Fixed.Assets.Turnover.Frequency	5.696e-11
Net.Worth.Turnover.Rate..times.	-1.884e+01
Revenue.per.person	9.403e-09
Operating.profit.per.person	-7.460e+00
Allocation.rate.per.person	-4.279e-09
Working.Capital.to.Total.Assets	-1.031e+09
Quick.Assets.Total.Assets	6.313e-01
Current.Assets.Total.Assets	2.825e+08
Cash.Total.Assets	-3.202e+00
Quick.Assets.Current.Liability	-2.064e-09
Cash.Current.Liability	-7.527e-11
Current.Liability.to.Assets	-9.200e+08
Operating.Funds.to.Liability	-1.349e+01
Inventory.Working.Capital	-1.437e-01
Inventory.Current.Liability	5.463e-11
Current.Liabilities.Liability	-1.589e-01
Working.Capital.Equity	2.173e+01
Current.Liabilities.Equity	-8.063e+00
Long.term.Liability.to.Current.Assets	1.417e-10
Retained.Earnings.to.Total.Assets	1.426e+01
Total.income.Total.expense	-2.175e+02
Total.expense.Assets	-1.013e+00
Current.Asset.Turnover.Rate	6.633e-11
Quick.Asset.Turnover.Rate	3.215e-11
Working.capitcal.Turnover.Rate	-2.141e+02
Cash.Turnover.Rate	-1.323e-10
Cash.Flow.to.Sales	4.462e+02
Fixed.Assets.to.Assets	4.356e-09
Current.Liability.to.Liability	NA
Current.Liability.to.Equity	NA
Equity.to.Long.term.Liability	-1.411e+01
Cash.Flow.to.Total.Assets	4.592e+00
Cash.Flow.to.Liability	-2.369e+01
CFO.to.Assets	7.889e+00
Cash.Flow.to.Equity	-1.607e+01
Current.Liability.to.Current.Assets	-4.381e+00
Liability.Assets.Flag	2.782e+01
Net.Income.to.Total.Assets	-2.404e+01
Total.assets.to.GNP.price	5.894e-10
No.credit.Interval	-4.938e-01
Gross.Profit.to.Sales	-1.839e+04
Net.Income.to.Stockholder.s.Equity	2.041e+01
Liability.to.Equity	-6.532e+01
Degree.of.Financial.Leverage..DFL.	3.913e-01
Interest.Coverage.Ratio..Interest.expense.to.EBIT.	-4.011e+00
Net.Income.Flag	NA
Equity.to.Liability	-4.695e+01

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
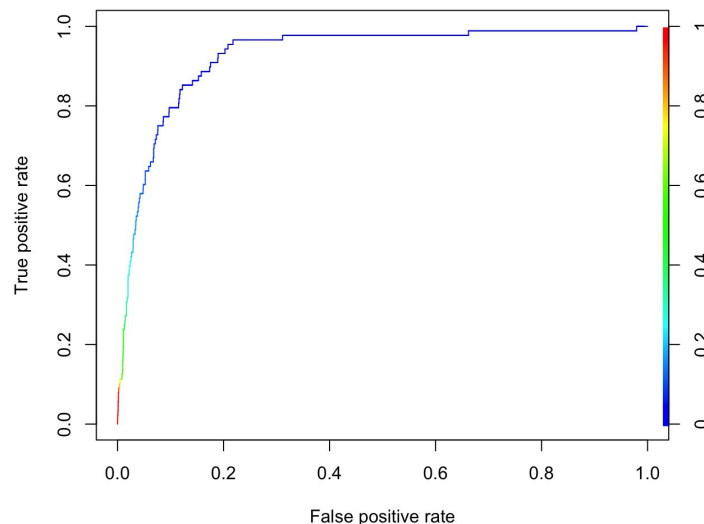
Logistic Regression

ROC Performance (AUC): 0.925297

Accuracy = 0.9637

Significant Variables:

- Continuous interest rate after tax
- Operating Expense Rate
- Net Value Per Share (C)
- Borrowing dependency
- Retained Earnings to Total Assets
- Cash Turnover Rate
- Net Income to Total Assets
- Total assets to GNP price

Logistic Regression- After Feature Selection

ROC Performance (AUC) = 0.8585

Accuracy = 0.9670

Significant Variables:

- Operating expense rate
- Net value per share (C)
- Borrowing dependency
- Retained earnings to total assets
- Net income to total assets

```
Coefficients:
                                    Estimate Std. Error z value Pr(>|z|)
(Intercept)                        -2.620e+00  6.909e+00  -0.379 0.704536
Continuous.interest.rate..after.tax.  1.766e+00  7.196e+00   0.245 0.806180
Operating.Expense.Rate              9.004e-11  2.933e-11   3.070 0.002137 **
Net.Value.Per.Share..C.            -5.765e+01  9.109e+00  -6.328 2.48e-10 ***
Borrowing.dependency                1.515e+01  7.122e+00   2.127 0.033440 *
Retained.Earnings.to.Total.Assets   2.072e+01  5.403e+00   3.834 0.000126 ***
Cash.Turnover.Rate                 -5.257e-11  3.832e-11  -1.372 0.170085
Net.Income.to.Total.Assets         -2.171e+01  3.390e+00  -6.405 1.50e-10 ***
Total.assets.to.GNP.price           6.437e-11  1.574e-10   0.409 0.682654
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
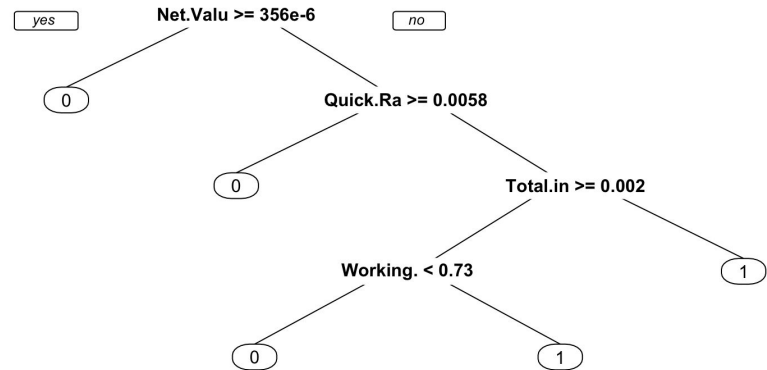
Decision Tree (CART model)
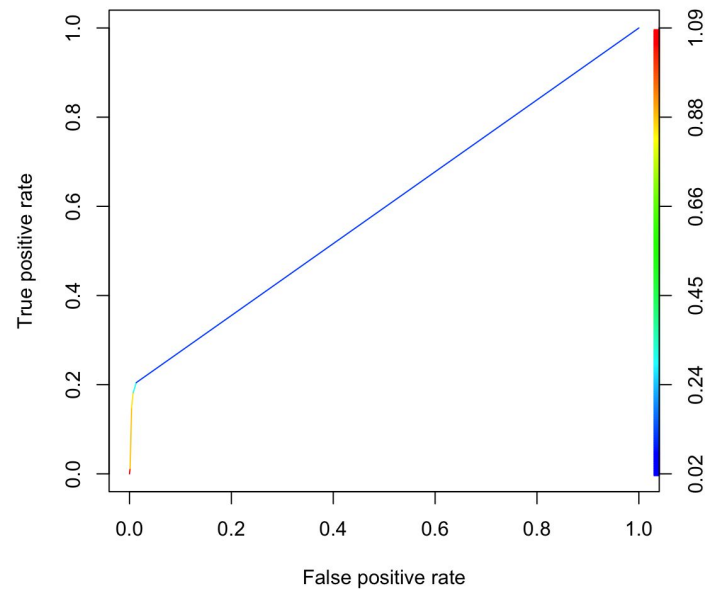
Significant variables:

- Net Value Growth Rate
- Quick Ratio: Acid Test
- Total income / Total expense
- Working Capital / Equity

Decision Tree

ROC performance (AUC)= 0.5962057

Accuracy = 0.9685

Random Forest

Accuracy = 0.9699

Significant Variables:

- Continuous interest rate after tax
- Cash Current Liability
- Net Value Per Share (B)
- Net Income to Total Assets
- Borrowing dependency
- Debt ratio
- Operating Profit Rate

Accuracy Comparison

Logistic Regression:

(2612+17) / (2612 + 17 + 71 + 28) = 0.9637

Logistic Regression (After Feature Selection):

(2628 + 10) / (2628 + 10 + 78 + 12) = 0.9670

Decision Tree:

(2629 + 13) / (2629 + 13 + 75 + 11) = 0.9685

Random Forest:

(2630 + 16) / (2630 + 16 + 72 + 10) = 0.9699

Common Significant Variables

Logistic Regression and Random Forest
- Continuous interest rate after tax (-)
- Net Value Per Share (C) (+)
- Borrowing dependency (+)
- Net Income to Total Assets (-)
- Total assets to GNP price (+)

Decision Tree and Random Forest
- Quick Ratio: Acid Test (-)
- Total income/ Total expense (-)

Conclusion

Company Bankruptcy Prediction:

- Useful for creditors and investors to conduct market analysis
- Forecasting financial loss/ future financial problems
- Could enable companies to take early action to prevent significant financial losses

Thank you!