**Citation Request:**

Zieba, M., Tomczak, S. K., & Tomczak, J. M. (2016). Ensemble Boosted Trees with Synthetic Features Generation in Application to Bankruptcy Prediction. Expert Systems with Applications. [Web Link]

BibTeX:

@article{zikeba2016ensemble,

title={Ensemble Boosted Trees with Synthetic Features Generation in Application to Bankruptcy Prediction},

author={Zi{k{e}}ba, Maciej and Tomczak, Sebastian K and Tomczak, Jakub M},

journal={Expert Systems with Applications},

year={2016},

publisher={Elsevier}

}

**Data Set Information:**

The dataset is about bankruptcy prediction of Polish companies. The data was collected from Emerging Markets Information Service (EMIS, [Web Link]), which is a database containing information on emerging markets around the world. The bankrupt companies were analysed in the period 2000-2012, while the still operating companies were evaluated from 2007 to 2013.

Basing on the collected data five classification cases were distinguished, that depends on the forecasting period:

- **1stYear** the data contains financial rates from 1st year of the forecasting period and corresponding class label that indicates bankruptcy status after 5 years. The data contains 7027 instances (financial statements), 271 represents bankrupted companies, 6756 firms that did not bankrupt in the forecasting period.

- **2ndYear** the data contains financial rates from 2nd year of the forecasting period and corresponding class label that indicates bankruptcy status after 4 years. The data contains 10173 instances (financial statements), 400 represents bankrupted companies, 9773 firms that did not bankrupt in the forecasting period.

- **3rdYear** the data contains financial rates from 3rd year of the forecasting period and corresponding class label that indicates bankruptcy status after 3 years. The data contains 10503 instances (financial statements), 495 represents bankrupted companies, 10008 firms that did not bankrupt in the forecasting period.

- **4thYear** the data contains financial rates from 4th year of the forecasting period and corresponding class label that indicates bankruptcy status after 2 years. The data contains 9792 instances (financial statements), 515 represents bankrupted companies, 9277 firms that did not bankrupt in the forecasting period.

- **5thYear** the data contains financial rates from 5th year of the forecasting period and corresponding class label that indicates bankruptcy status after 1 year. The data contains 5910 instances (financial statements), 410 represents bankrupted companies, 5500 firms that did not bankrupt in the forecasting period.

**Attribute Information:**

**X1** net profit / total assets

**X2** total liabilities / total assets

**X3** working capital / total assets

**X4** current assets / short-term liabilities

**X5** [(cash + short-term securities + receivables - short-term liabilities) / (operating expenses - depreciation)] \* 365

**X6** retained earnings / total assets

**X7** EBIT / total assets

**X8** book value of equity / total liabilities

**X9** sales / total assets

**X10** equity / total assets

**X11** (gross profit + extraordinary items + financial expenses) / total assets

**X12** gross profit / short-term liabilities

**X13** (gross profit + depreciation) / sales

**X14** (gross profit + interest) / total assets

**X15** (total liabilities \* 365) / (gross profit + depreciation)

**X16** (gross profit + depreciation) / total liabilities

**X17** total assets / total liabilities

**X18** gross profit / total assets

**X19** gross profit / sales

**X20** (inventory \* 365) / sales

**X21** sales (n) / sales (n-1)

**X22** profit on operating activities / total assets

**X23** net profit / sales

**X24** gross profit (in 3 years) / total assets

**X25** (equity - share capital) / total assets

**X26** (net profit + depreciation) / total liabilities

**X27** profit on operating activities / financial expenses

**X28** working capital / fixed assets

**X29** logarithm of total assets

**X30** (total liabilities - cash) / sales

**X31** (gross profit + interest) / sales

**X32** (current liabilities \* 365) / cost of products sold

**X33** operating expenses / short-term liabilities

**X34** operating expenses / total liabilities

**X35** profit on sales / total assets

**X36** total sales / total assets

**X37** (current assets - inventories) / long-term liabilities

**X38** constant capital / total assets

**X39** profit on sales / sales

**X40** (current assets - inventory - receivables) / short-term liabilities

**X41** total liabilities / ((profit on operating activities + depreciation) \* (12/365))

**X42** profit on operating activities / sales

**X43** rotation receivables + inventory turnover in days

**X44** (receivables \* 365) / sales

**X45** net profit / inventory

**X46** (current assets - inventory) / short-term liabilities

**X47** (inventory \* 365) / cost of products sold

**X48** EBITDA (profit on operating activities - depreciation) / total assets

**X49** EBITDA (profit on operating activities - depreciation) / sales

**X50** current assets / total liabilities

**X51** short-term liabilities / total assets

**X52** (short-term liabilities \* 365) / cost of products sold)

**X53** equity / fixed assets

**X54** constant capital / fixed assets

**X55** working capital

**X56** (sales - cost of products sold) / sales

**X57** (current assets - inventory - short-term liabilities) / (sales - gross profit - depreciation)

**X58** total costs /total sales

**X59** long-term liabilities / equity

**X60** sales / inventory

**X61** sales / receivables

**X62** (short-term liabilities \*365) / sales

**X63** sales / short-term liabilities

**X64** sales / fixed assets

**Class -** 0 did not get bankrupt/ 1 - got bankrupt

**Problem Statement:**

You have to build a model on the basis given features and target column to predict whether a bank is going to get bankrupt for a given set of features.