Company **Bankruptcy** Prediction with Naive Bayes Algorithm

A. Aylin Tokuç

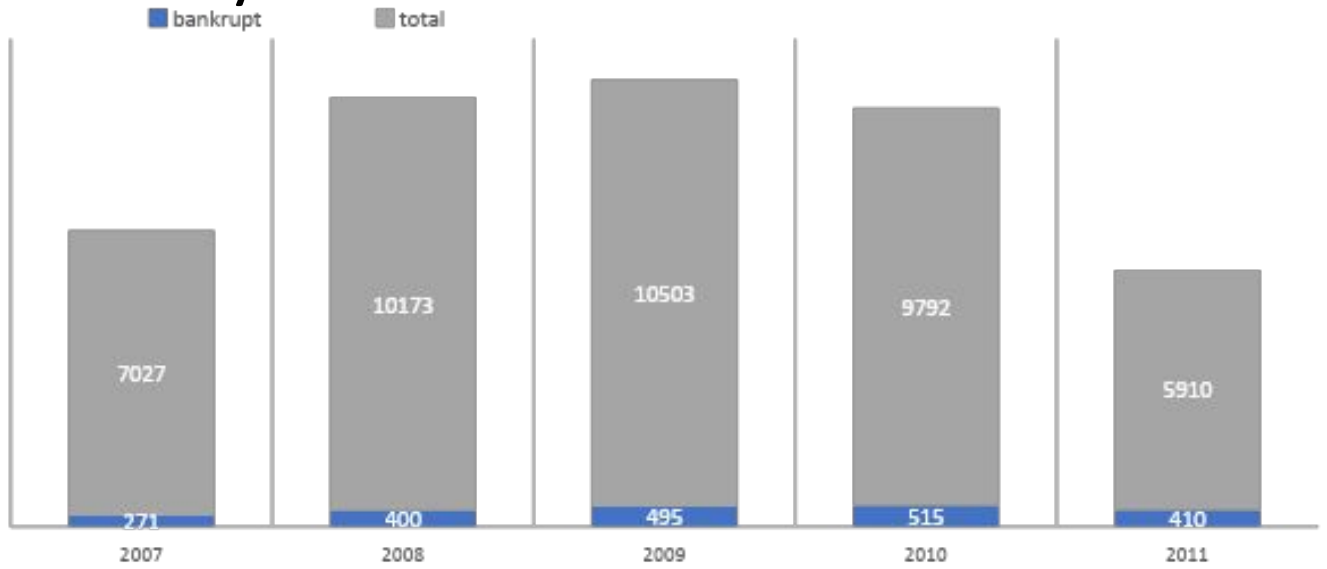
Situation & Project Goals

- Situation
  - Due to economical difficulties, companies are facing bankruptcy
  - Over the 5 years period, **2091** Polish companies have bankrupted among **43405** observations.
- Goals
  - Develop a prediction model to decide whether a company will bankrupt in the following year or not
  - Model's bankruptcy prediction power should be as good as Random Forest classifier
  - Model should be scalable and fast to generate results

Main Findings

- Predicting whether a company will bankrupt in the following year or not will give the investors and banks a better insight, and reduce possible waste of time and money.

- Proposed model has an accuracy score of **92.70%**



- Proposed model can run within 2 seconds.

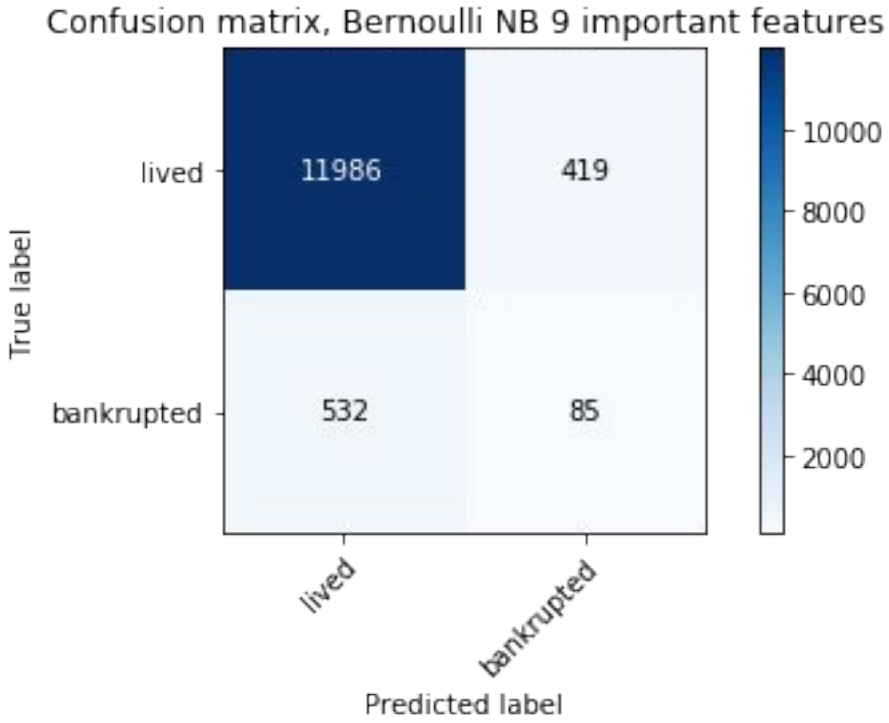
Approach

- Developed a bankrupcy model in Python using Naive Bayes based techniques
  - Easy to build, powerful on even large datasets
  - Classification based on conditional probability
  - Can handle multiple features
- Impact of using different subsets of features is investigated
- Model can be easily deployed and run

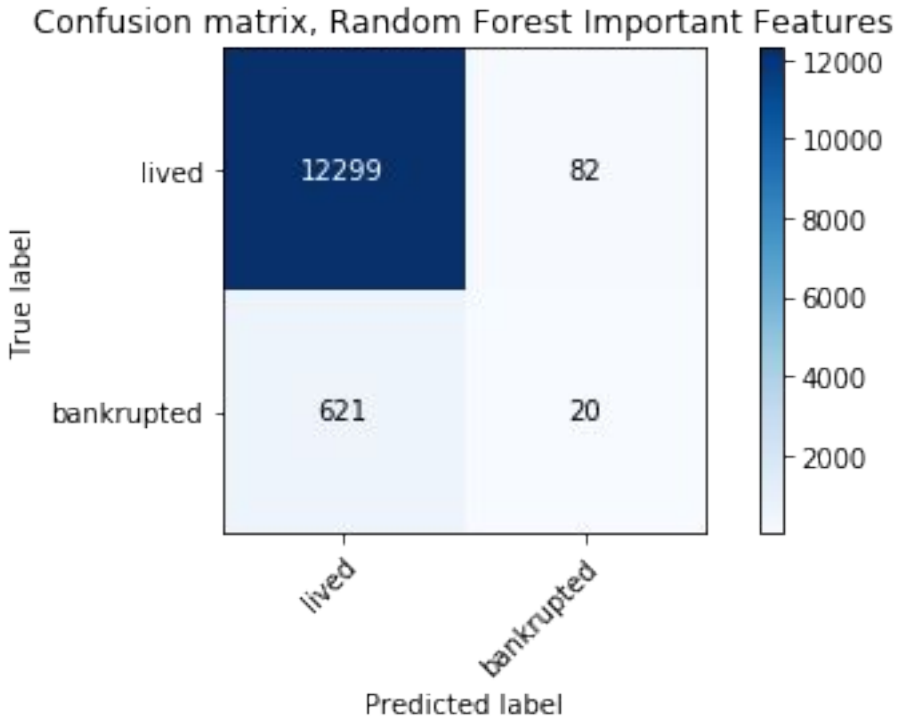
Model Description

- **Overview of the Basic Methodology:** Predict whether the given company will bankrupt in the following year or not
- **Model:** Naive Bayesian Classifier
- **Dependent Variable:** Binary variable, of bankrupt/no bankrupt
- **Scope:**
  - 43405 instances of 5 year's company financial data
  - 64 features and the outcome (bankrupt/no bankrupt)
  - Different feature subsets are selected
- **Sampling:**
  - 70 / 30 stratified train test split (StratifiedShuffleSplit) (30383 – 13022)
- Model's bankruptcy prediction power should be as good as Random Forest
  - Created a baseline model with Random Forest Classifier

Key Points



Accuracy: 92.70%
Precision: 95.75%
Recall: 96.62%



Accuracy: 94.60%
Precision: 95.19%
Recall: 99.34%

Model Details

- Candidate Variables: 64 financial indicators
- Used variables: 9 features according to feature ranking
- Naive Bayes Model built in Python
  - Stratified sampling as the dataset is imbalanced
  - Accuracy, Precision and Recall calculated to measure success

$$P(x_i \mid y) = P(i \mid y)x_i + (1 - P(i \mid y))(1 - x_i)$$

```
from sklearn.naive_bayes import BernoulliNB

bnb = BernoulliNB()
y_pred = bnb.fit(X_train, Y_train).predict(X_test)
```

Recommendations

- Train the model again every year, as more recent data becomes available
  - Otherwise the model can be outdates
- Run the model to predict a company when an indicator is updated
  - A trigger can be set on the DB
- Re-evaluate the feature importances again when a new model is built
  - Using only 9 most important features gave the best result
  - Try different subsets of features

References

Tomczak, S., 2016. **Polish Companies Bankruptcy Dataset**. [Online]

Available at: https://archive.ics.uci.edu/

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, Volume 58, pp. 93-101.

Sci-kit Learn Developers, 2019. Naive Bayes. [Online]

Available at: https://scikit-learn.org/

Model Details: Variables

1	X16 (gross profit + depreciation) / total liabilities
2	X52 (short-term liabilities * 365) / cost of products sold)
3	X32 (current liabilities * 365) / cost of products sold
4	X28 working capital / fixed assets
5	X5 [(cash + short-term securities + receivables - short-term liabilities) / (operating expenses - depreciation)] * 365
6	X40 (current assets - inventory - receivables) / short-term liabilities
7	X9 sales / total assets
8	X11 (gross profit + extraordinary items + financial expenses) / total assets
9	X59 long-term liabilities / equity
10	X23 net profit / sales
11	X25 (equity - share capital) / total assets
12	X55 working capital
13	X17 total assets / total liabilities
14	X14 (gross profit + interest) / total assets
15	X29 logarithm of total assets
16	X13 (gross profit + depreciation) / sales
17	X58 total costs /total sales
18	X30 (total liabilities - cash) / sales
19	X57 (current assets - inventory - short-term liabilities) / (sales - gross profit - depreciation)
20	X56 (sales - cost of products sold) / sales

All Features

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