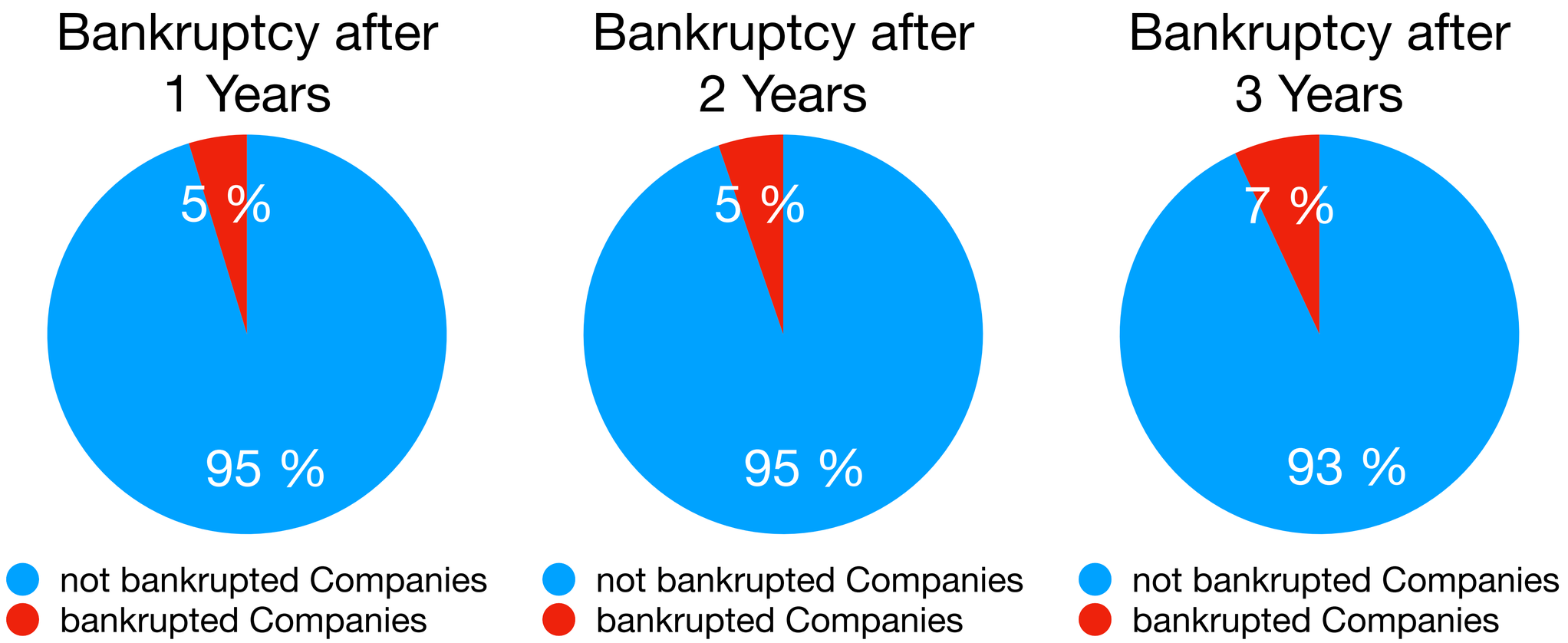
Bankruptcy Prediction

Vanessa Klebe, 09.02.22

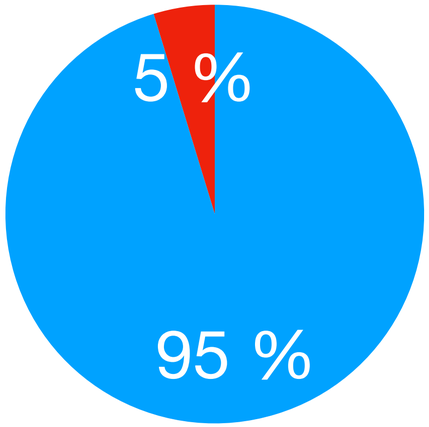
Dataset


Bankruptcy after
1 Years
5 %
95 %
not bankrupted Companies
bankrupted Companies
Bankruptcy after
2 Years
5 %
95 %
not bankrupted Companies
bankrupted Companies
Bankruptcy after
3 Years
7 %
93 %
not bankrupted Companies
bankrupted Companies

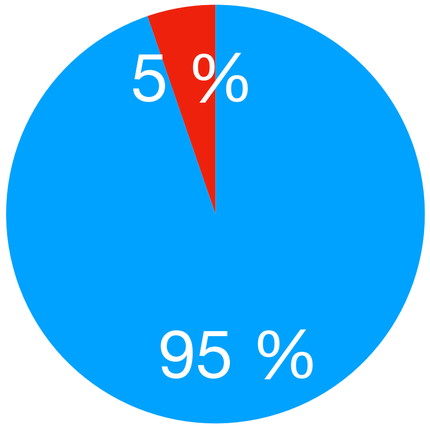
Dataset Preparation - Removing NaN Values

1. Feature Deletion: Features that contain more than 10% of NaN Values
2. Removing rows with NaN values

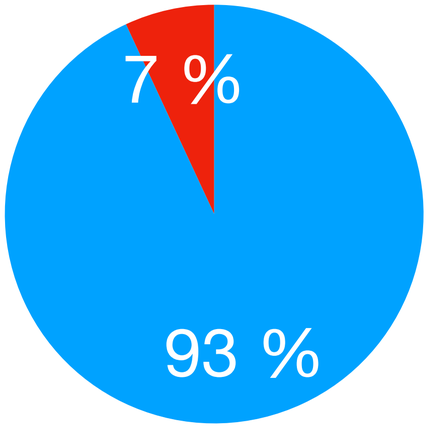
Bankruptcy after 1 Years



Bankruptcy after 2 Years



Bankruptcy after 3 Years

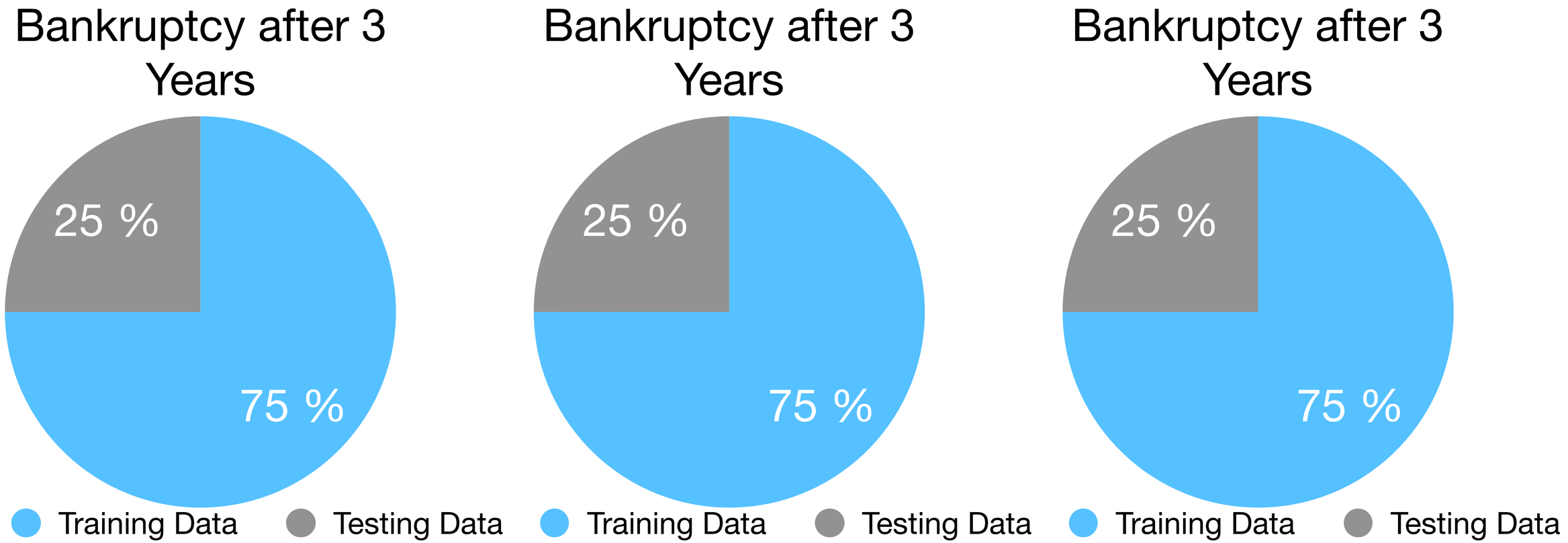


not bankrupted Companies
bankrupted Companies

not bankrupted Companies
bankrupted Companies

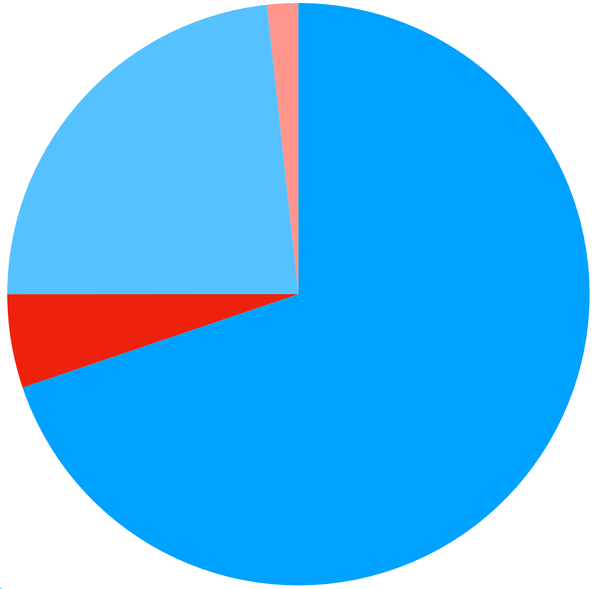
not bankrupted Companies
bankrupted Companies

Dataset Preparation - Splitting



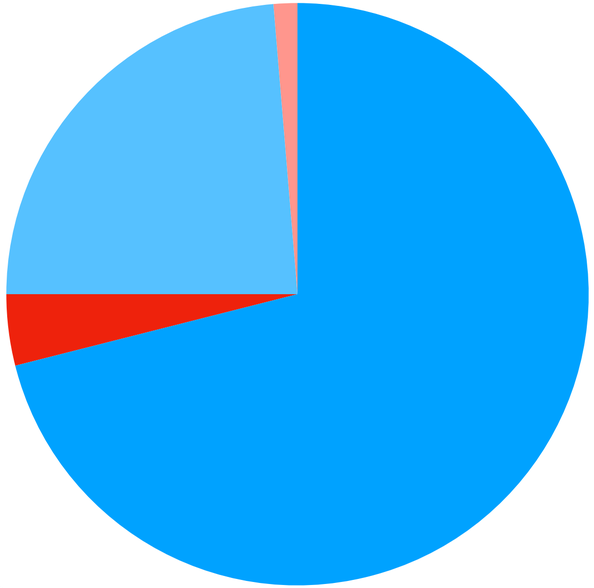
Dataset Preparation - Splitting

Bankruptcy after 1 Years



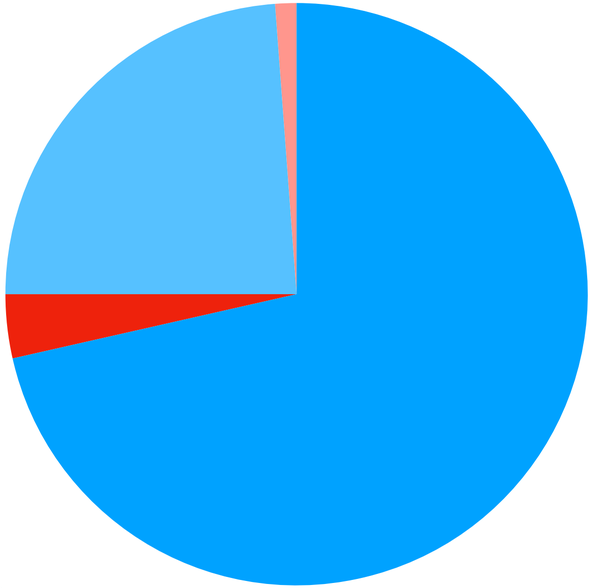
- not bankrupted Companies (Training)
- bankrupted Companies (Training)
- not bankrupted Companies (Testing)
- bankrupted Companies (Testing)

Bankruptcy after 2 Years



- not bankrupted Companies (Training)
- bankrupted Companies (Training)
- not bankrupted Companies (Testing)
- bankrupted Companies (Testing)

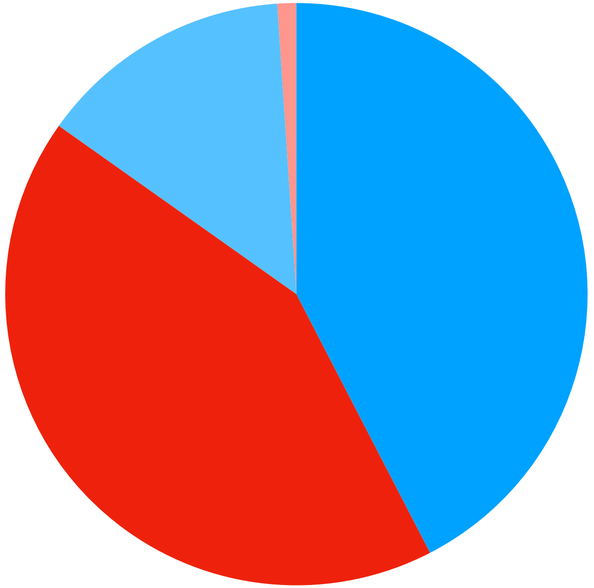
Bankruptcy after 3 Years



- not bankrupted Companies (Training)
- bankrupted Companies (Training)
- not bankrupted Companies (Testing)
- bankrupted Companies (Testing)

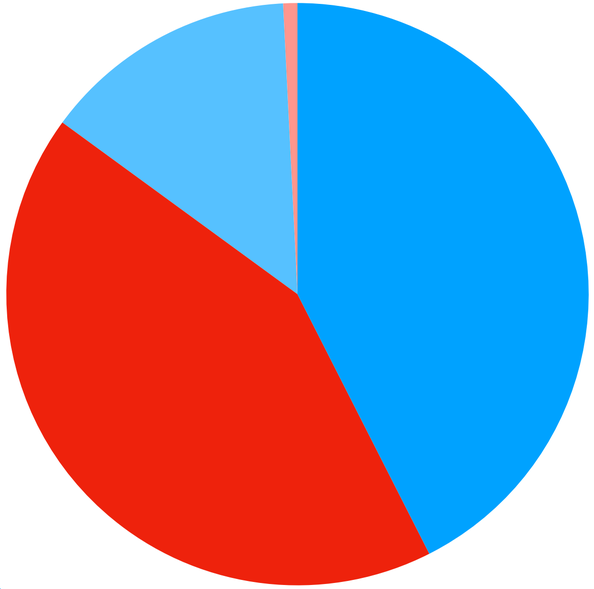
Dataset Preparation - Random oversampling

Bankruptcy after 1 Years



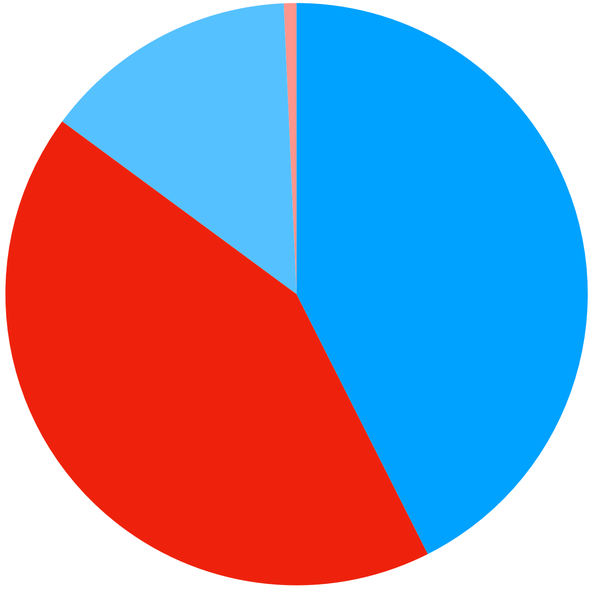
- not bankrupted Companies (Training)
- bankrupted Companies (Training)
- not bankrupted Companies (Testing)
- bankrupted Companies (Testing)

Bankruptcy after 2 Years



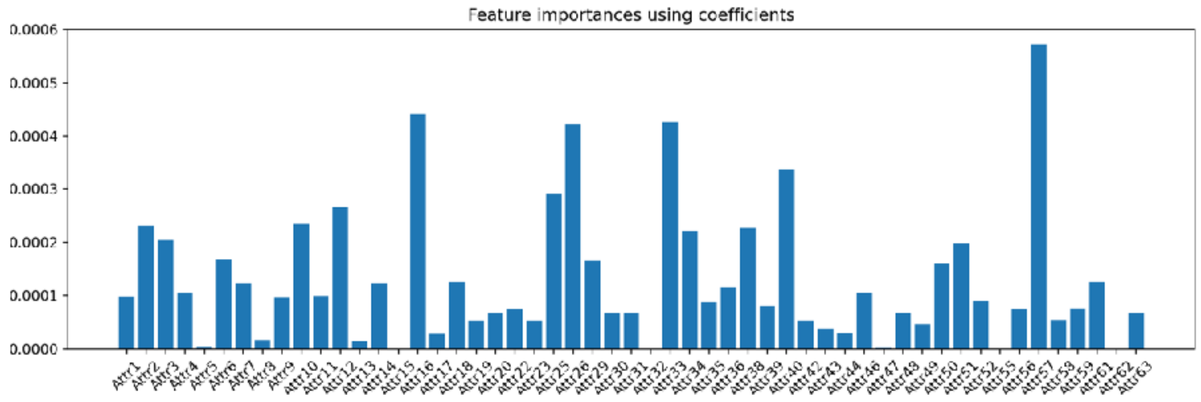
- not bankrupted Companies (Training)
- bankrupted Companies (Training)
- not bankrupted Companies (Testing)
- bankrupted Companies (Testing)

Bankruptcy after 3 Years

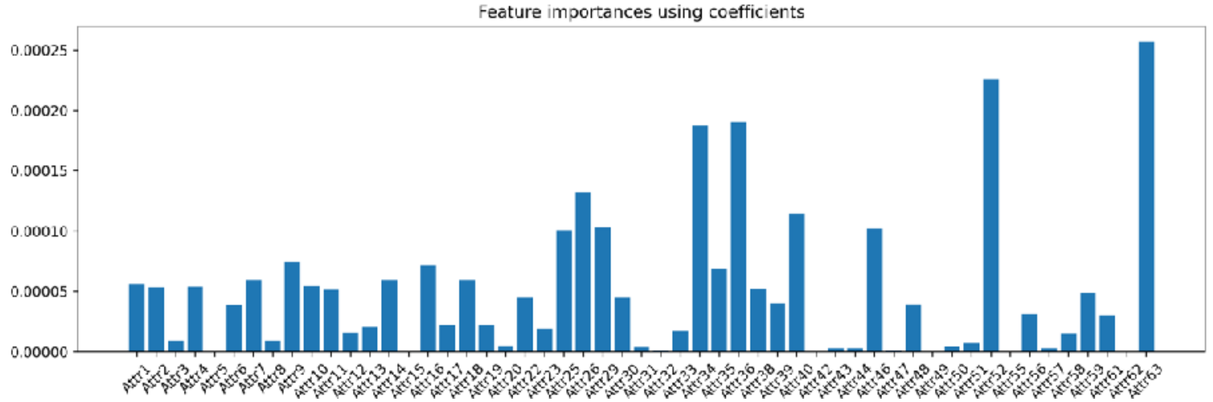


- not bankrupted Companies (Training)
- bankrupted Companies (Training)
- not bankrupted Companies (Testing)
- bankrupted Companies (Testing)

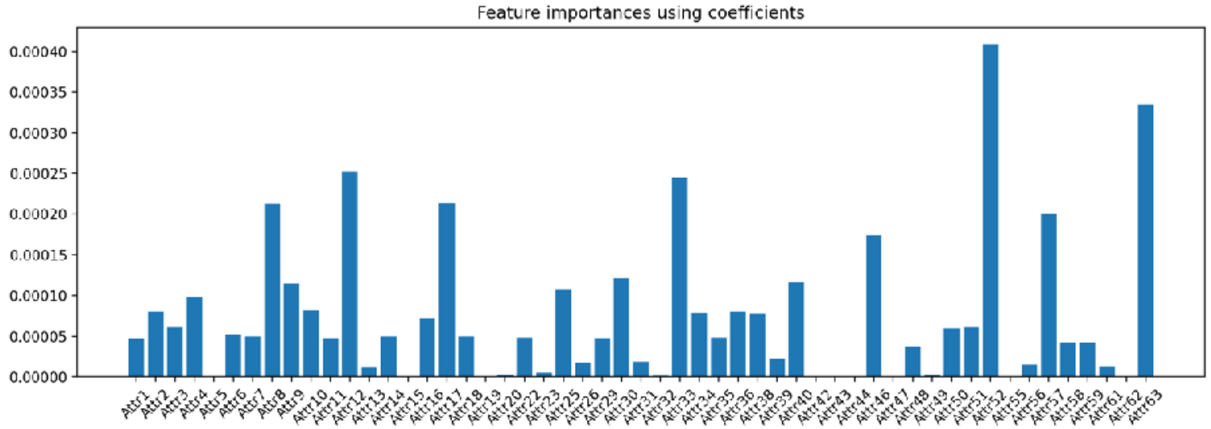
Dataset Preparation - Feature Selection (Correlation)



Bankruptcy in 1 Year

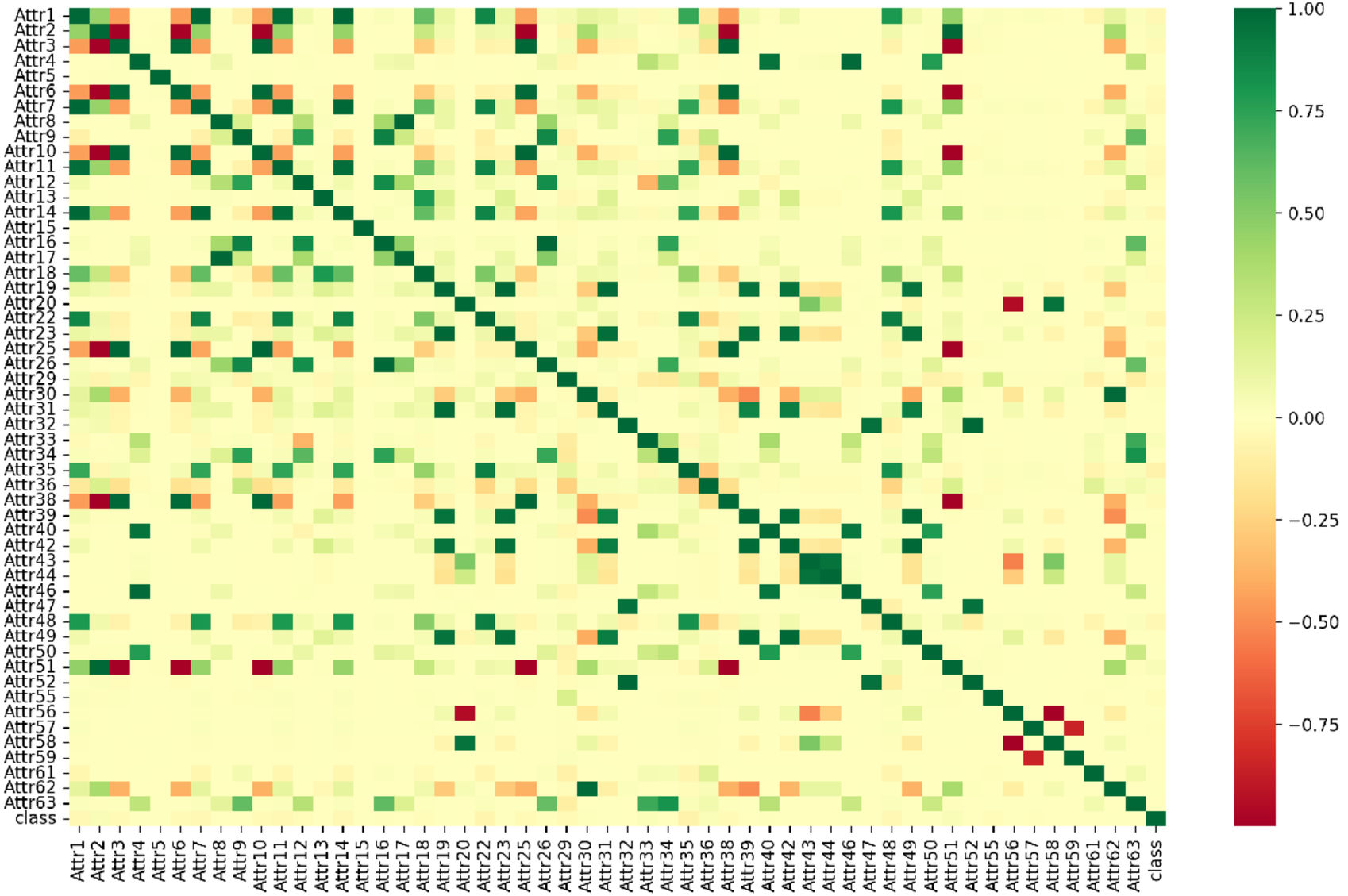


Bankruptcy in 2 Year



Bankruptcy in 3 Year

Dataset Preparation - Feature Selection (Heat Map)



Dataset Preparation - Feature Selection

Using: Mutual Information
15 Features were selected

Bankruptcy in 1 Year	1	7	12	13	14	15	16	18	19	22	23	26	35	39	42
Bankruptcy in 2 Year	1	7	8	13	14	15	18	22	23	25	35	39	42	46	57
Bankruptcy in 3 Year	5	7	11	14	18	22	26	26	33	35	39	42	44	46	52

Customer Model Requirements

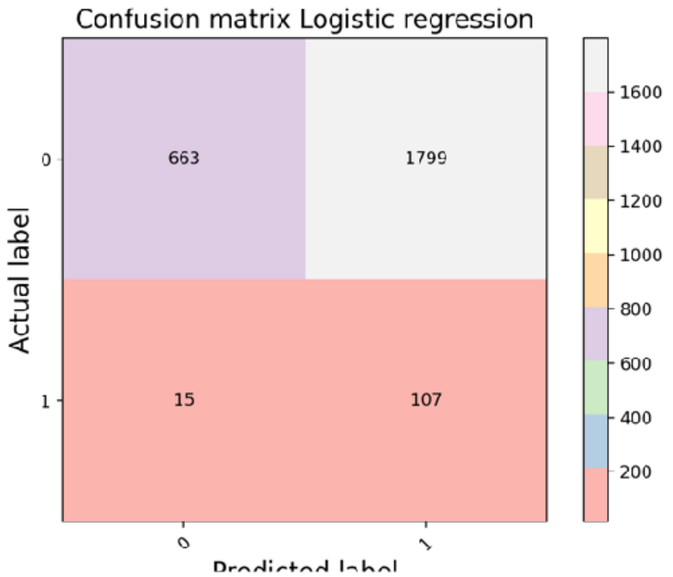
- Dataset Complexity
- Model Explainability
- Real-Time Calculation
- Fast Retrainability

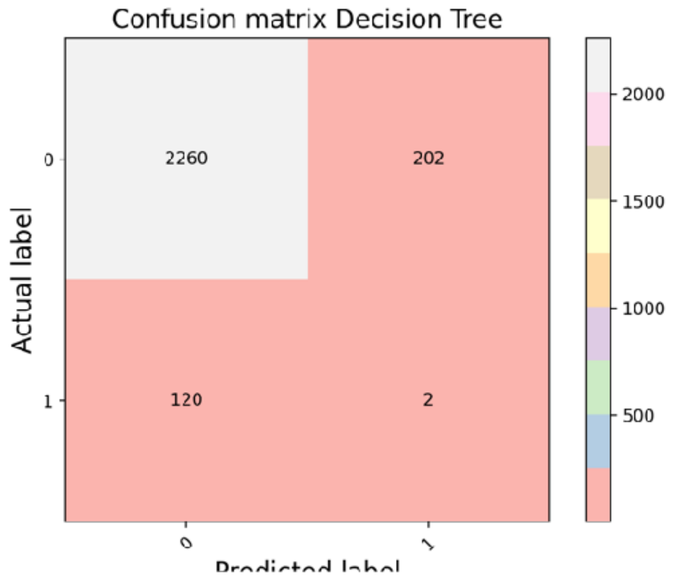
- Logistic Regression
- Decision Trees
- Random Forest
- Neural Networks

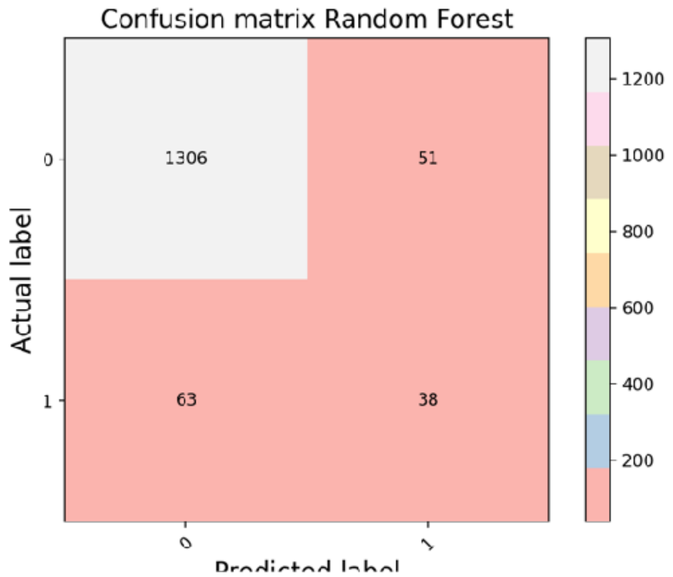
Model Selection - Accuracy

	Logistic Regression	Decision Tree	Random Forest	Neural Network
Bankruptcy in 1 Year	50.2%	83.5%	91.5%	60.2%
Bankruptcy in 2 Year	50.7%	86.4%	94.7%	78.8%
Bankruptcy in 3 Year	60.9%	92.3%	95.2%	86.3%

Model Selection



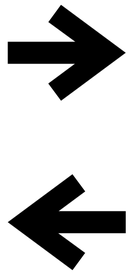




Model Serving

Backend (Kubernetes):

- ML Model



Frontend:

- Discuss with Client
- WebApp

Bib

- https://archive.ics.uci.edu/ml/datasets/Polish+companies+bankruptcy+data#