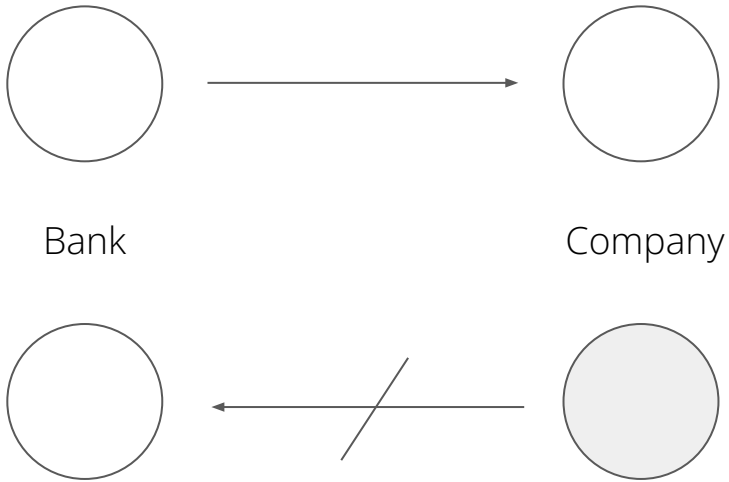
Predicting Bankruptcy in Polish Companies

Mohammed Ali | Daniel Chung | Ahmad Hussain

Problem

Lending money to companies comes with a risk of not being paid back if the company goes bankrupt



Predicting which companies go bankrupt helps inform lending and investment decisions

Data

Rows (10,503): Individual Polish companies

Features (64): Financial ratios

Label: 1 if company went bankrupt, 0 otherwise

Source: Emerging Markets Information Service (EMIS), UCI Machine Learning Repository

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

Mohammed Ali | Daniel Chung | Ahmad Hussain

Data

"total assets"

Rows (10,503): Individual Polish companies

Features (64): Financial ratios

Label: 1 if company went bankrupt, 0 otherwise

Source: Emerging Markets Information Service (EMIS), UCI Machine Learning Repository

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

Mohammed Ali | Daniel Chung | Ahmad Hussain

Data

"gross profit"

Rows (10,503): Individual Polish companies

Features (64): Financial ratios

Label: 1 if company went bankrupt, 0 otherwise

Source: Emerging Markets Information Service (EMIS), UCI Machine Learning Repository

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

Mohammed Ali | Daniel Chung | Ahmad Hussain

Data

Rows (10,503): Individual Polish companies

Features (64): Financial ratios

Label: 1 if company went bankrupt, 0 otherwise

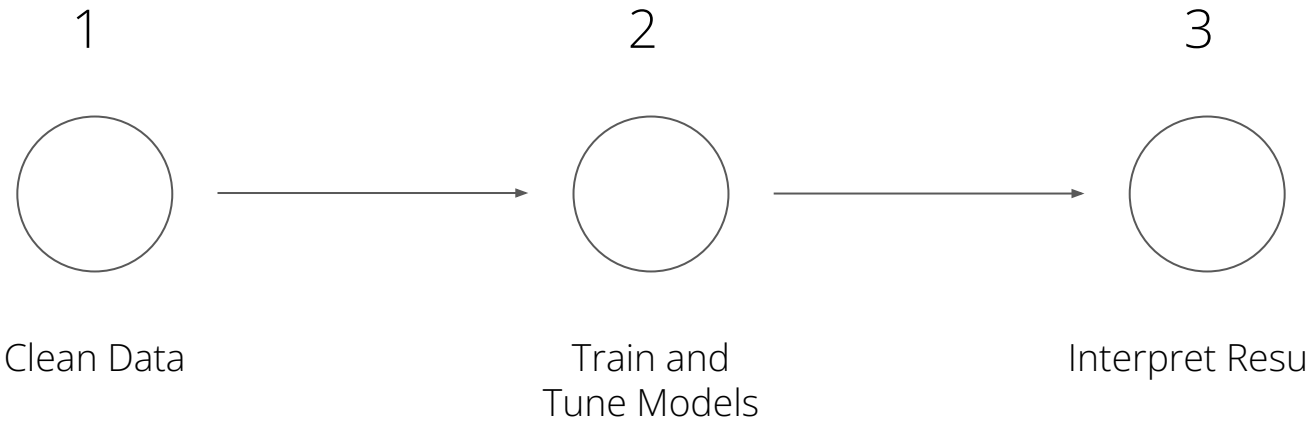
Source: Emerging Markets Information Service (EMIS), UCI Machine Learning Repository

"sales"

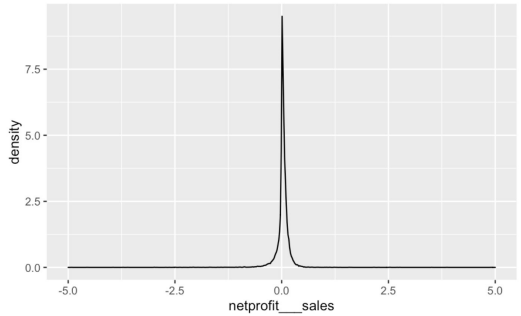
X1 net profit / total assets
X2 total liabilities / total assets
X3 working capital / total assets
X4 current assets / short-term liabilities
X5 [(cash + short-term securities + receivables - short-term liabilities) / (operating expenses - depreciation)] * 365
X6 retained earnings / total assets
X7 EBIT / 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

Mohammed Ali | Daniel Chung | Ahmad Hussain

Approach:



1. Clean Data



pervasive, extreme outliers

X

```
[na   , 0.98, 0.67, 0.86, 0.02]
[0.85, 0.24, 0.64, 0.16, na   ]
[0.78, 0.09, na   , 0.17, 0.86]
[0.93, na   , 0.50, 0.67, 0.20]
[0.47, na   , na   , 0.10, na   ]
[0.53, 0.57, 0.28, 0.69, 0.07]
[na   , na   , 0.84, 0.39, 0.77]
[na   , na   , 0.53, na   , 0.97]
```

missing data

y

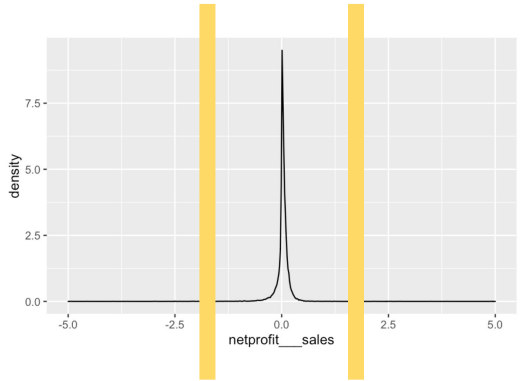
```
[0]
[0]
[1]
[0]
[0]
[0]
[0]
[0]
```

y_train

y_test

imbalanced classes

1. Clean Data



pervasive, extreme outliers

used winsorization

X

```
[na   , 0.98, 0.67, 0.86, 0.02]
[0.85, 0.24, 0.64, 0.16, na   ]
[0.78, 0.09, na   , 0.17, 0.86]
[0.93, na   , 0.50, 0.67, 0.20]
[0.47, na   , na   , 0.10, na   ]
[0.53, 0.57, 0.28, 0.69, 0.07]
[na   , na   , 0.84, 0.39, 0.77]
[na   , na   , 0.53, na   , 0.97]
```

missing data

y

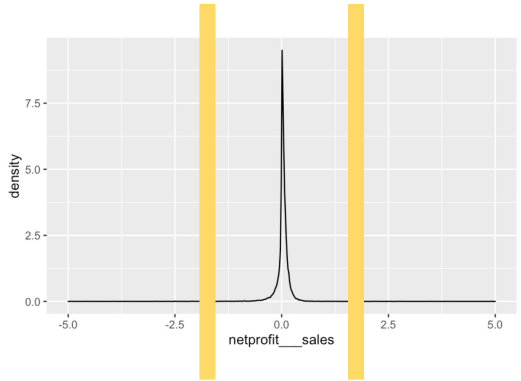
```
[0]
[0]
[1]
[0]
[0]
[0]
[0]
[0]
```

y_train

y_test

imbalanced classes

1. Clean Data



pervasive, extreme outliers

used winsorization

X

y

```
[0.18, 0.98, 0.67, 0.86, 0.02]   [0]
[0.85, 0.24, 0.64, 0.16, 0.40]]  [0]
[0.78, 0.09, 0.66, 0.17, 0.86]]  [0]
[0.93, 0.99, 0.50, 0.67, 0.20]]  [1]
[0.47, 0.30, 0.70, 0.10, 0.22]]  [0]
[0.53, 0.57, 0.28, 0.69, 0.07]   [0]
[0.27, 0.62, 0.84, 0.39, 0.77]]  [0]
[0.76, 0.63, 0.53, 0.50, 0.97]]  [0]
```

```
[0]
[0]
[1]
[0]
[0]
[0]
```

y_train

```
[0]
[0]
```

y_test

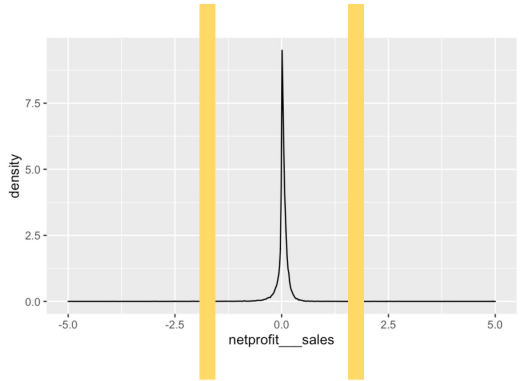
missing data

used KNNimpute

imbalanced classes

Mohammed Ali | Daniel Chung | Ahmad Hussain

1. Clean Data



pervasive, extreme outliers

used winsorization

X

y

```
[0.18, 0.98, 0.67, 0.86, 0.02]   [0]
[0.85, 0.24, 0.64, 0.16, 0.40]]  [0]
[0.78, 0.09, 0.66, 0.17, 0.86]]  [0]
[0.93, 0.99, 0.50, 0.67, 0.20]]  [1]
[0.47, 0.30, 0.70, 0.10, 0.22]]  [0]
[0.53, 0.57, 0.28, 0.69, 0.07]   [0]
[0.27, 0.62, 0.84, 0.39, 0.77]]  [0]
[0.76, 0.63, 0.53, 0.50, 0.97]]  [0]
```

```
[1]
[0]
[1]
[0]
[0]
[0]
```

y_train

```
[0]
[1]
```

y_test

missing data

used KNNimpute

imbalanced classes

used oversampling/ undersampling

2. Train and Tune Models

Model:	Motivation:
Logistic regression	Interpretable coefficients
CART	Interpretable splits
Random Forest	Strong out-of-sample performance
XGBoost	Strong out-of-sample performance
OCT	Interpretable splits, good performance
Neural Nets	Strong out-of-sample performance

2. Train and Tune Models - Results:

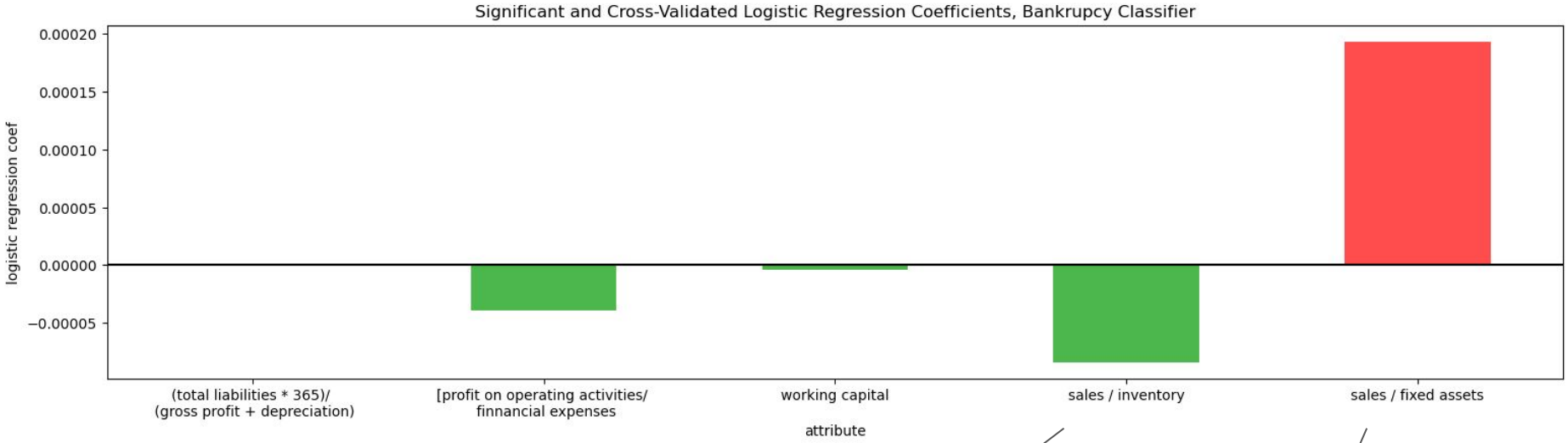
AUC	Logistic Regression	OCT	CART	Neural Network	Random Forest	XGBoost
Original Data	0.53	0.64	0.63	0.75	0.80	0.94
Undersampled	0.60	0.64	0.70	0.73	0.79	0.90
Oversampled	0.58	0.63	0.73	0.75	0.80	0.95

2. Train and Tune Models - Results:

Recall	Logistic Regression	OCT	CART	Neural Network	Random Forest	XGBoost
Original Data	0.61 (threshold=0.084)	0.67	0.30	0.68	0.58	0.65
Undersampled	0.65 (threshold=0.515)	0.68	0.66	0.73	0.71	0.67
Oversampled	0.58 (threshold=0.4982)	0.70	0.67	0.64	0.72	0.72

3. Interpret Results: Logistic Regression

(oversampled, cross-validated, sparsified)

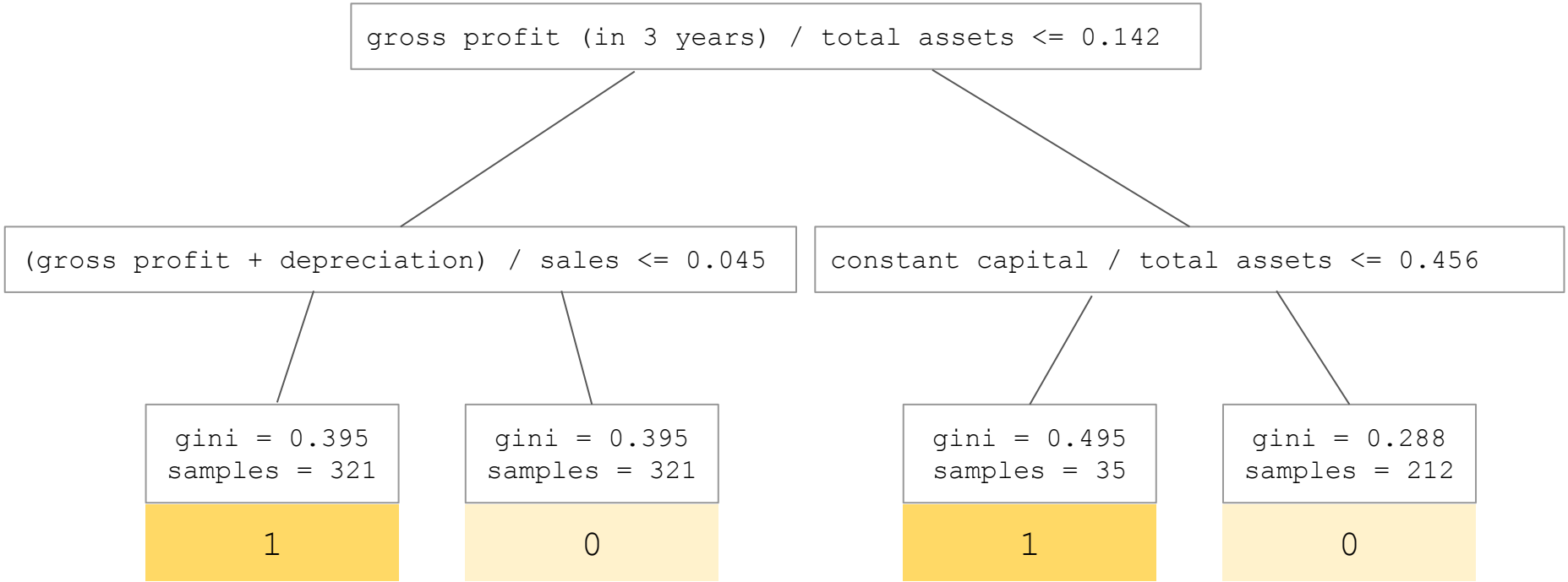


sales / inventory

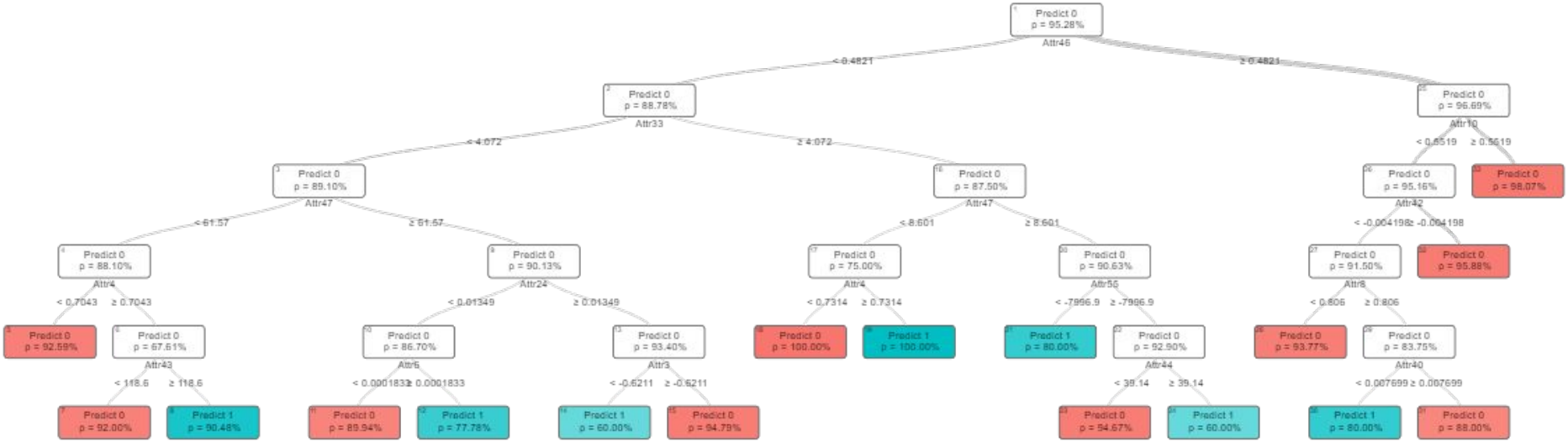
sales / fixed assets

3. Interpret Results: CART

(undersampled, cross-validated)



3. Interpret Results: OCT



Tracing the bottomline for banks

Our models show a **great deal of promise** in terms of predictive power

But

Real world applications mandate further practical considerations across the following

- Fairness & ethical considerations: are we restricting access to credit through algorithmic bias?
- High recall value for our models comes at the cost of precision: even accounting for the conservative nature of traditional banks, we need a cost-benefit analysis to see whether these models add to bank revenue
- Role of exogenous variables: Bankruptcies significantly correlate with external variables such as economic downturns & our model does not include economic variables