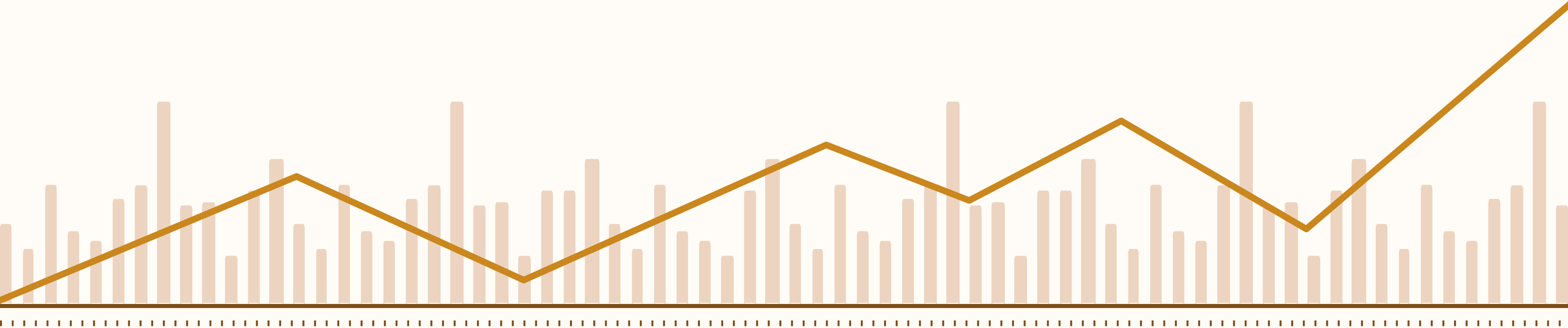


Predicting the direction of the DAX by using indicators

Table of Contents

How does the data of a stock market look?

- The typical representation of market prices is the candlestick chart
- Each candle represents a period (in this example, one day)
- Each candle contains four prices (open, high, low, close)
- The direction can be defined by the difference between the close and open prices
- When the difference is positive, the direction is upward; when the difference is negative, the direction is downward

How is the DAX calculated?

- The DAX is calculated from the price of the 30 largest German companies
- The price development of individual companies is generally influenced by supply and demand
- There are many different factors that influence the market behaviors

Data Collection

- I have collected data from 1987 to the present by using the yfinance library
- The data includes open, high, low, close (OHLC) prices and the volume of the market
- The target variable is the direction, calculated by the difference between close-price and open-price
- As input variables, I have calculated multiple indicators across different timeframes

Indicators

- Use 10 different types of indicators
  - Simple Moving Average
  - Relative Strength Index
  - Average True Range
  - Bollinger Bands
- Each indicator was calculated for three different time periods
- For each indicator that returns a market price, i created a categorical column
  - 1 if the price is above the indicator price
  - 0 if the price is lover then the indicator price

Data Cleaning

- Before 2000 there are missing Values for the volume of the market
  - started from 2000
- By computing an Indicator with a timeperiod of 14 day, the first 13 rows has missing values
  - I dropped the first 13 days

Time Dependency

	open	low	high	close	moving average	direction
01.01.2024	~~5975~~	~~6048~~	~~5974~~	~~6048~~	6050	Up
02.01.2024	~~6043~~	~~6058~~	~~6015~~	~~6032~~	6042	Down
03.01.2024	~~6032~~	~~6047~~	~~5997~~	~~6034~~	6040	Up

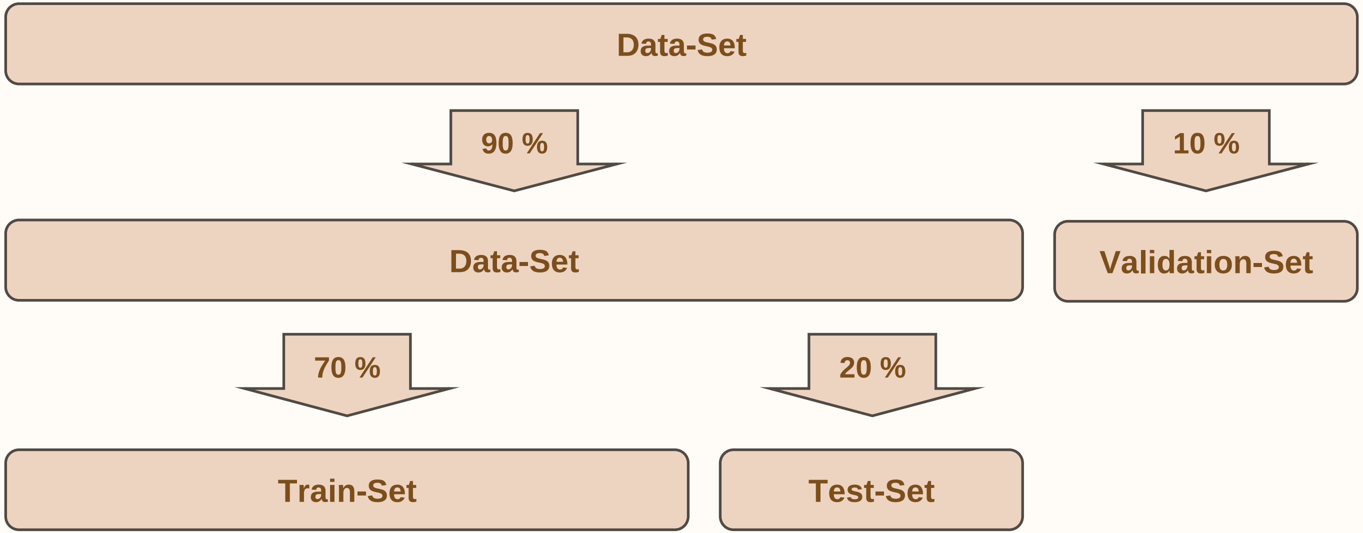
Feature Selection

- Varianace Threshold to drop features with a variance less then 0.02
- Correlation Matrix to identify multicorrelarity with a threshold = 0.99
- End up with 43 features

Transforming, Scaling and Sampling

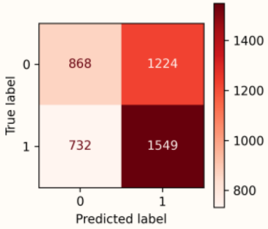
- Power Transformer
- Min Max Scaler for X
- My target is allready in a range between 0 and 1
- Class Imbalance: Train-Set = 0.045 Test-Set = 0.044 -> So I did not use any sampling method

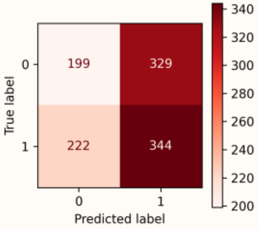
X-y Split

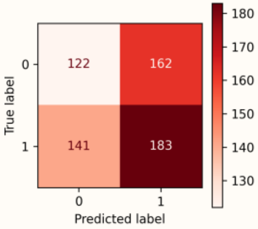

Data-Set
90 %
10 %
Data-Set
Validation-Set
70 %
20 %
Train-Set
Test-Set

Logistic - Regression

Train-Set	Test-Set	Validation-Set

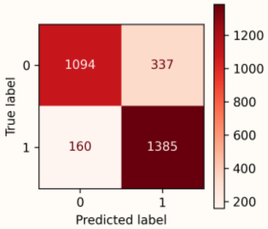
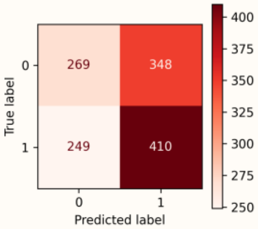
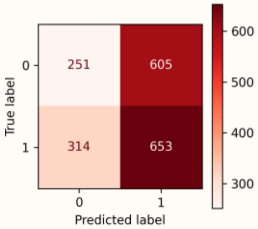




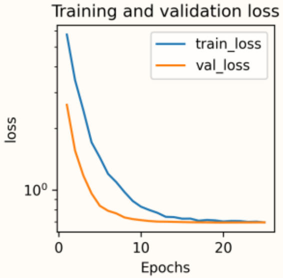


	Train-Set	Test-Set	Validation-Set
Accuracy	0.552	0.496	0.501
Kappa	0.094	-0.015	0.005

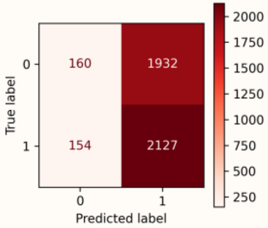
Random Forest with GridSearchCV

	Train-Set	Test-Set	Validation-Set
			
Accuracy	0.833	0.532	0.495
Kappa	0.664	0.058	-0.032

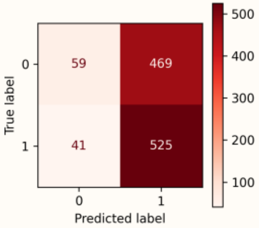
Neural Network



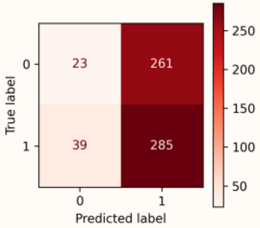
Train-Set



Test-Set



Validation-Set

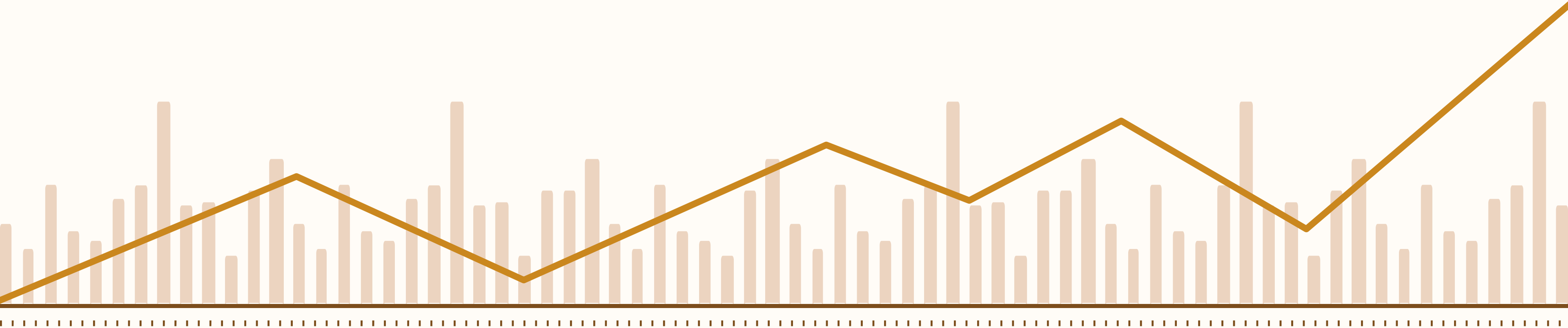


	Train-Set	Test-Set	Validation-Set
Accuracy	0.553	0.533	0.506
Kappa	0.009	0.040	-0.041

Final Results

Validation Set	Accuracy	F1	Recall	Precission	Kappa
Logistic Regression	0.502	0.547	0.565	0.530	-0.006
Random Forest	0.498	0.593	0.685	0.522	-0.030
Neural Network	0.506	0.655	0.879	0.522	-0.041

- The performance of all models is realy bad -> they are not practical for actual use
- Financial markets are known for their unpredictability and constant changes
- The performance of a model heavely relies on the input
- Next Steps: Is there an improvement, if I add features from different sources like more pattern related indicators and fundamental data from the companies?



Thank you for your attention!

- Better trust a goldfish then my model! -