

The background is a complex digital-themed collage. On the left, there are white circuit-like lines on a dark blue background. In the center, there are several data visualizations: a green bar chart, a red line graph, and a red bar chart. Overlaid on these are large, glowing blue binary digits (0s and 1s). The right side of the image is a dark grey gradient where the text is located.

# FIE453 - BIG DATA WITH APPLICATIONS TO FINANCE

GROUP 8 FINAL PROJECT  
PRESENTATION

# AGENDA

- Motivation
- Data
- Algorithms
- Preliminary Results
- Looking Forward

# MOTIVATION

- we want to create an indicator for selling a stock with a one month outlook
  - predict whether the return of a stock is positive or negative in one month
  - predict if  $\text{return}^2 > 0.01$  or not in one month

# DATA

- 134.083 observations
- 37 variables
- Training and Test Split (70%/30%) splitted by date (31.12.2017)

## Company-specific

- Stock return
- Volume
- Market capitalization
- Book to market ratio
- Earnings per share
- Revenue
- Equity
- Dividends
- Goodwill
- Taxes

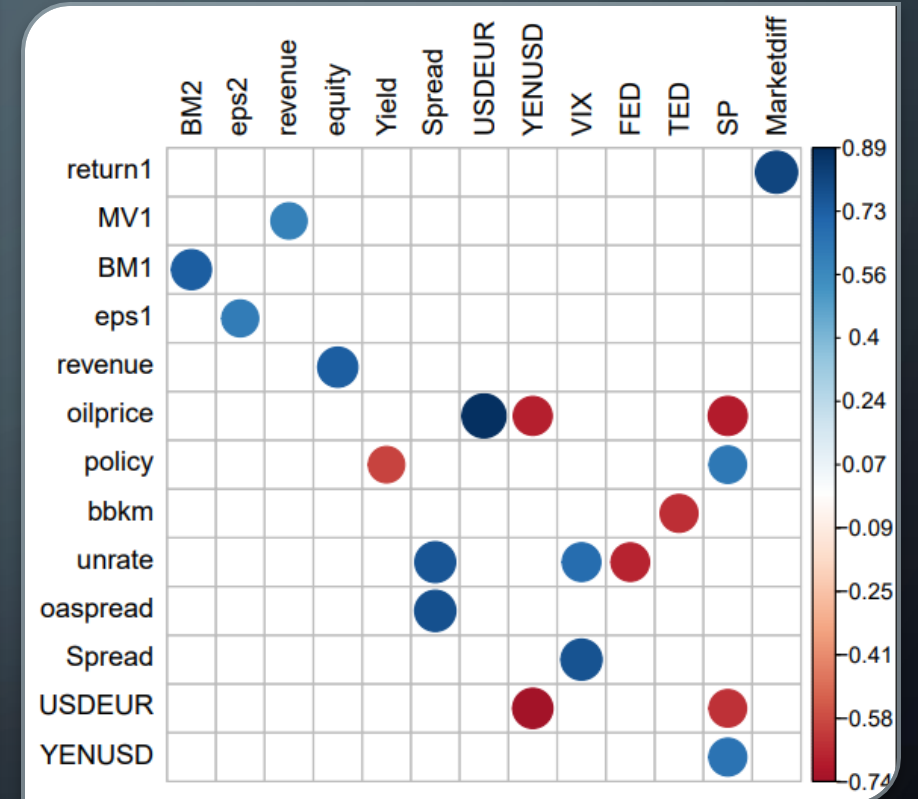
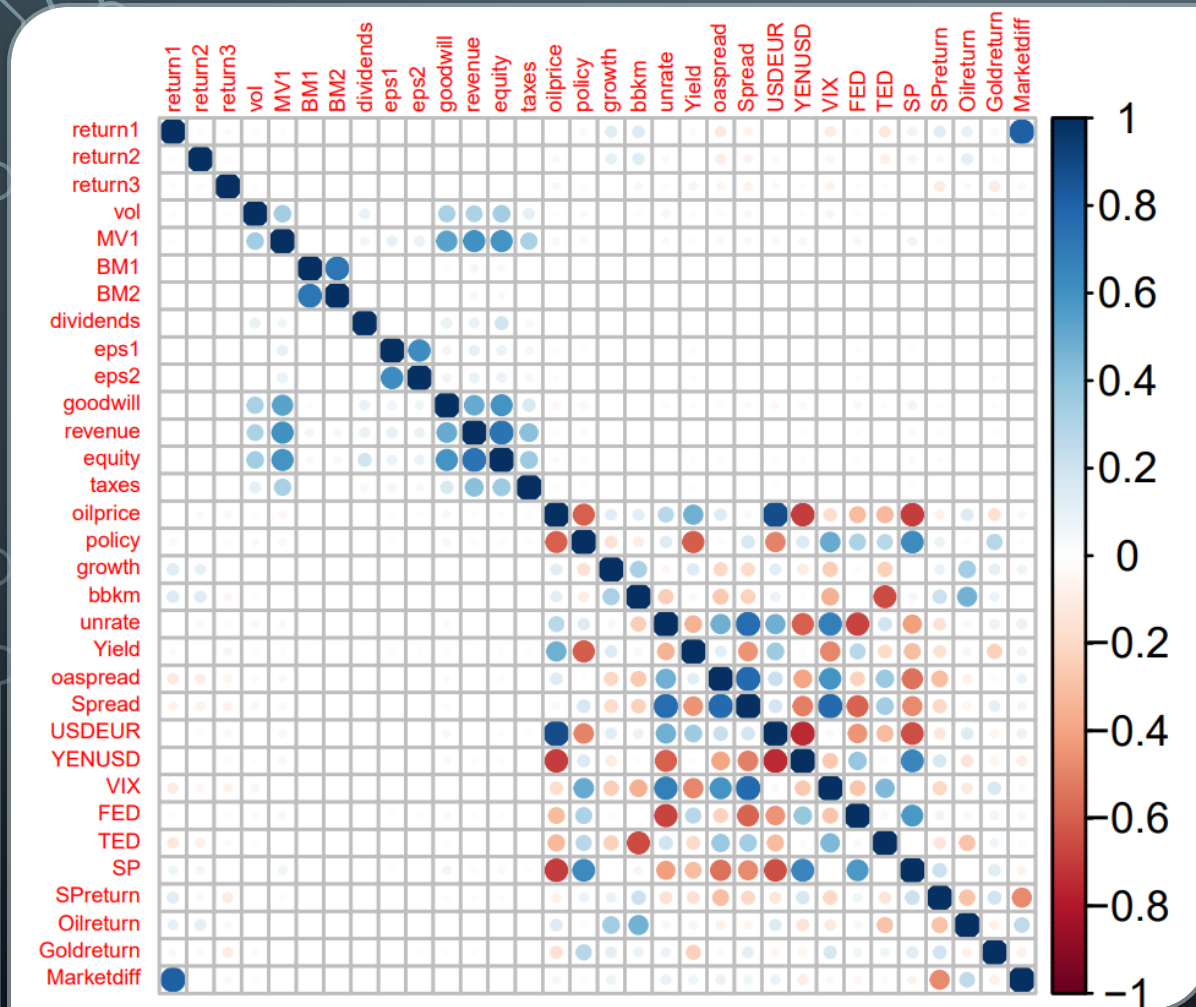
## Macroeconomic

- GDP growth
- Unemployment rate
- FED rate
- Policy index
- Consumer price index
- \$ to € exchange rate
- ¥ to \$ exchange rate
- VIX
- TED

## Assets

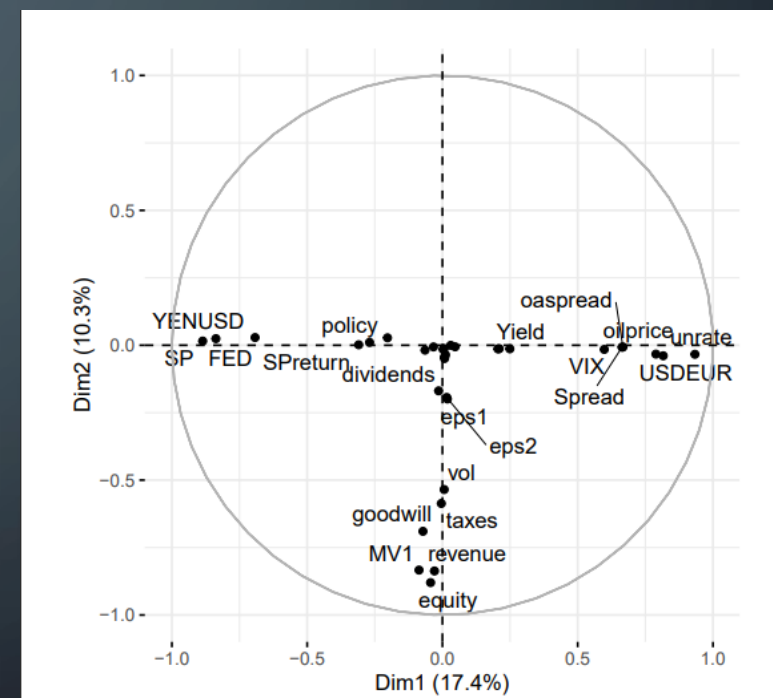
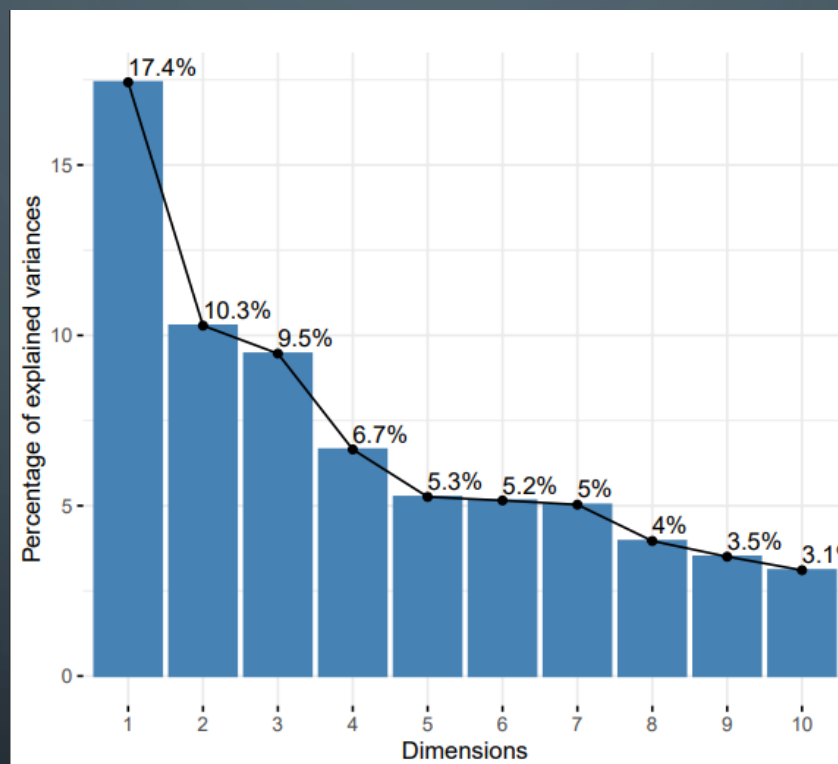
- Gold price
- Gold return
- Oil price
- Oil return
- Corporate bond yields
- Gov. Bond yield (10y)
- SP500 index
- SP500 return
- US High Yield Index  
Option-Adjusted  
Spread

# CORRELATIONPLOTS



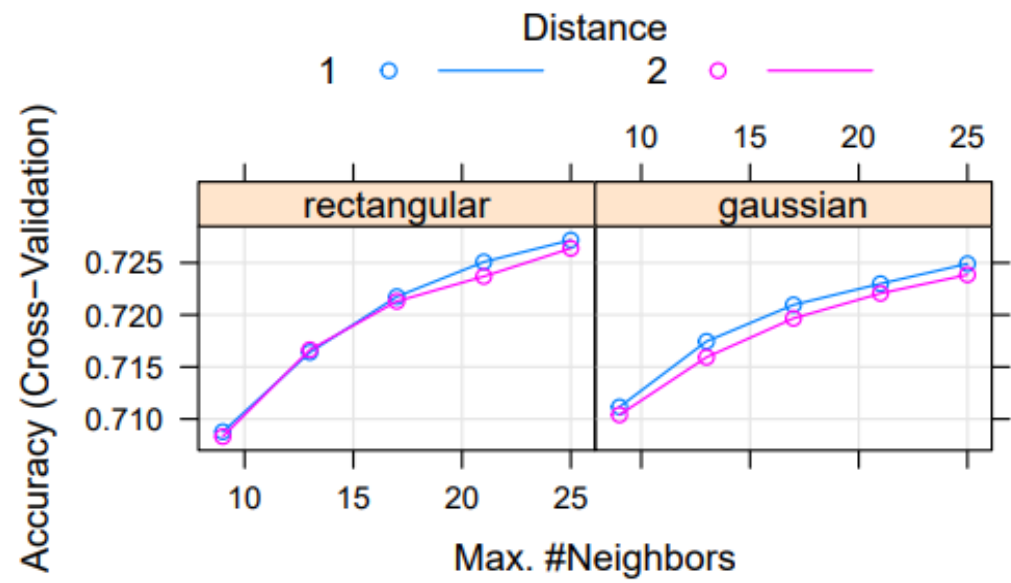
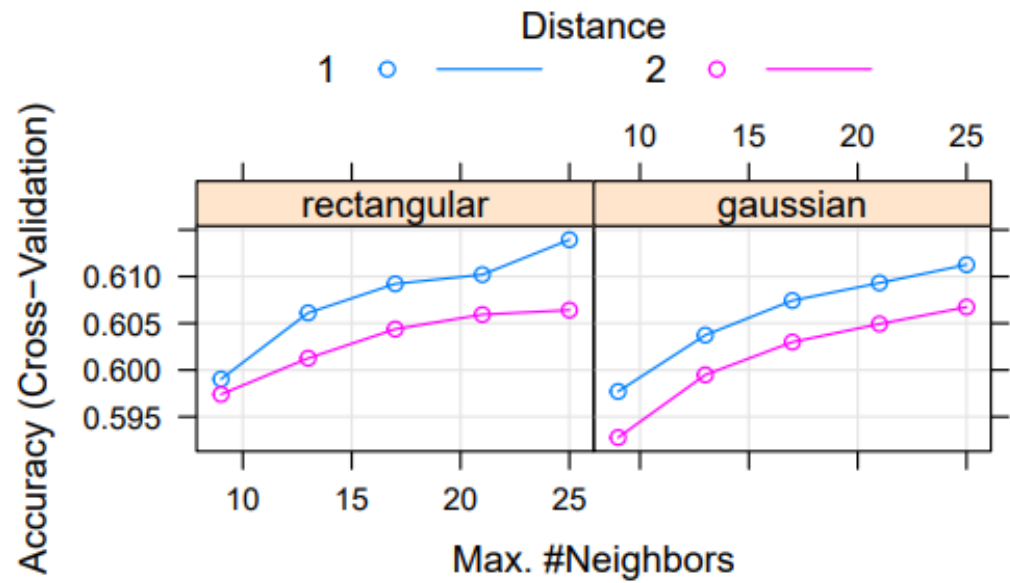
# PCA

- Remove correlated features
- Reduce overfitting
- Improve performance with low cost of accuracy

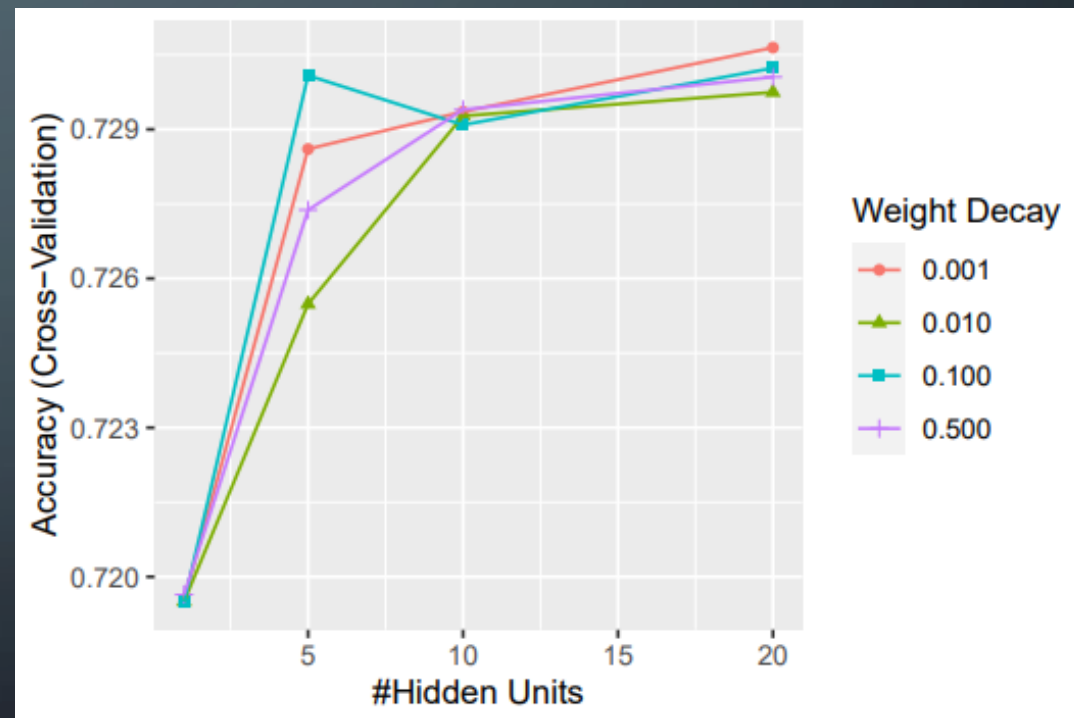
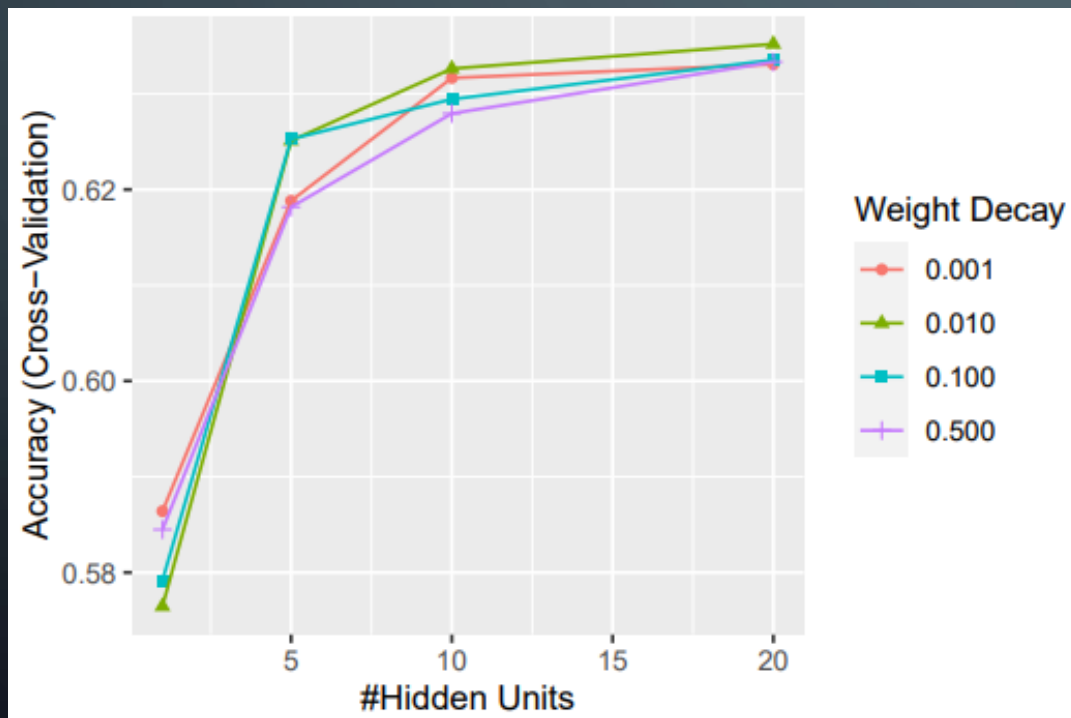




# K-NEAREST NEIGHBORS

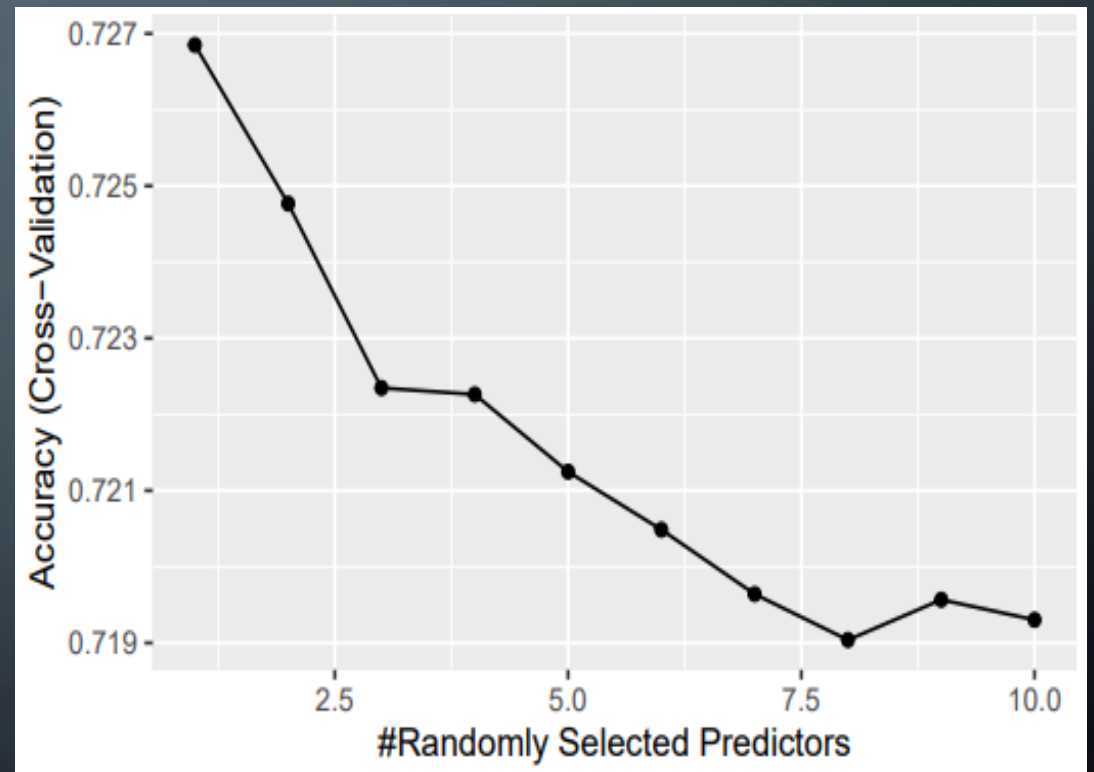
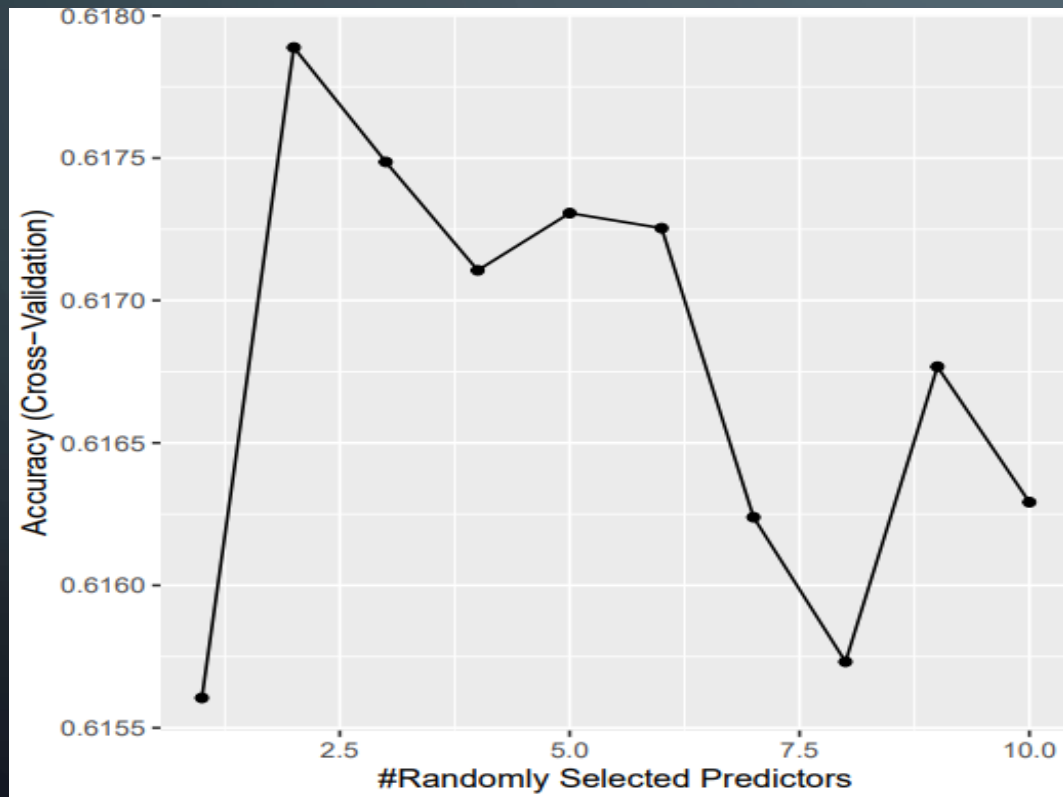


# NEURAL NETWORK

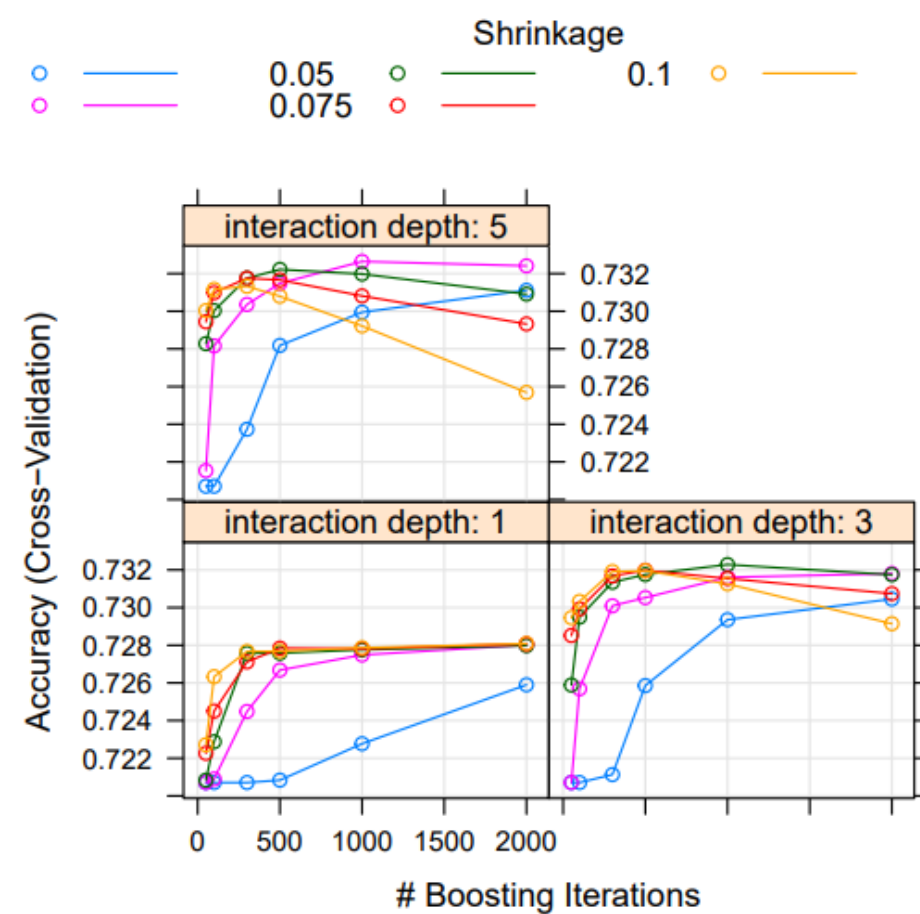
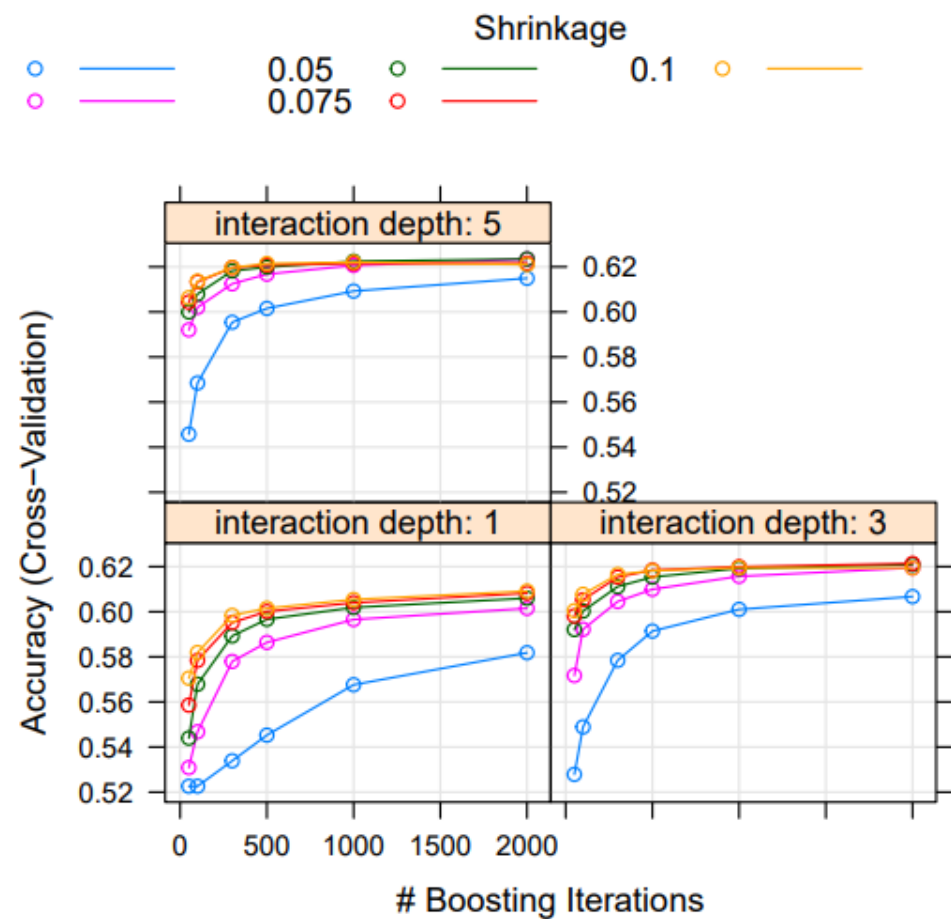




# RANDOM FOREST



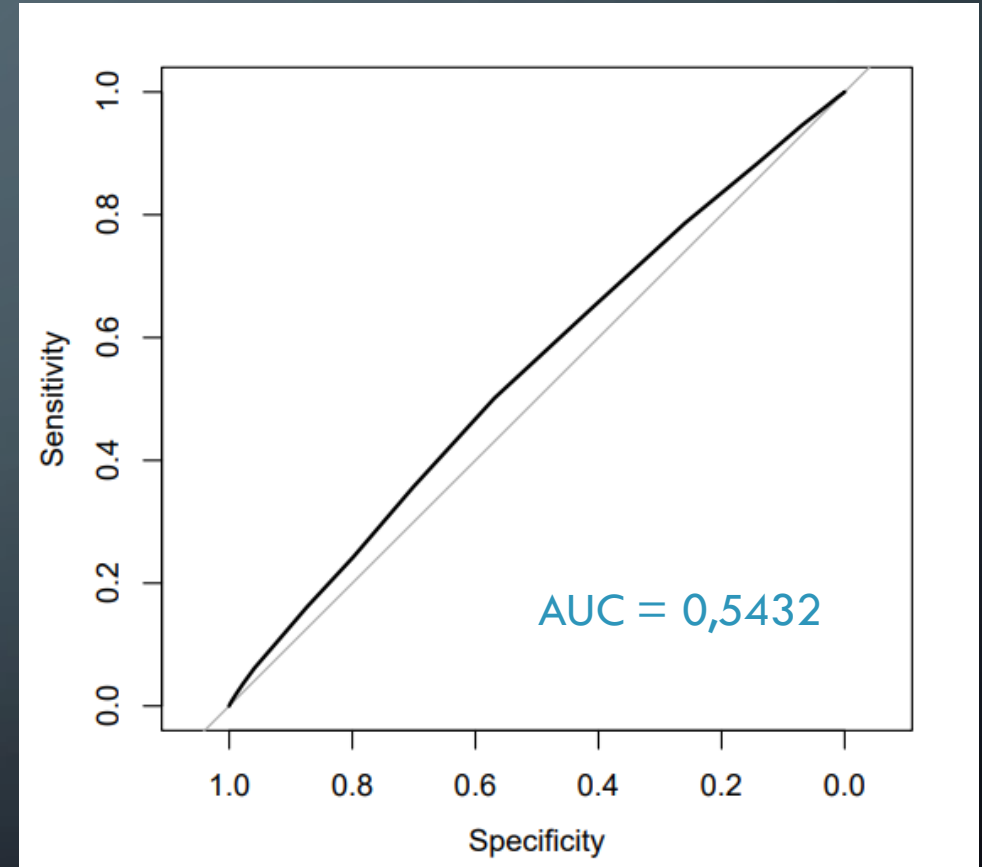
# GRADIENT BOOSTING MACHINE



# PRELIMINARY RESULTS

PREDICT: RETURNS POSITIVE OR NEGATIVE

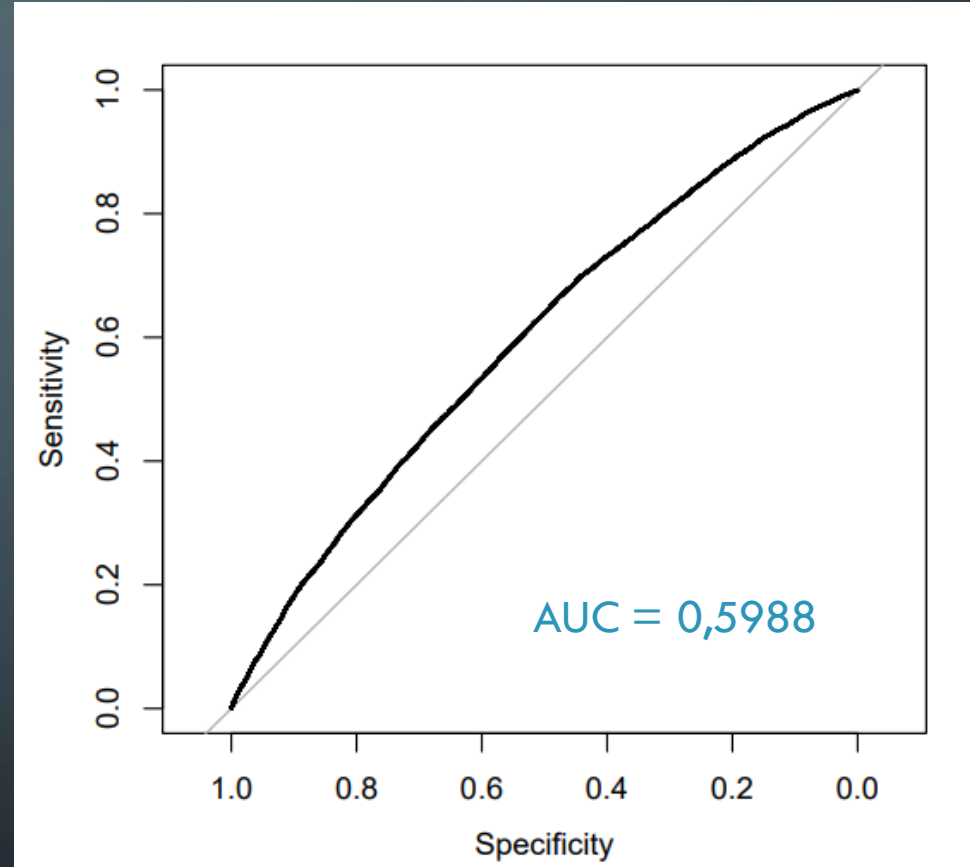
Algorithm	Accuracy	Kappa
<b>KNN</b>	<b>0,5320</b>	<b>0,0693</b>
Neural Network	0,5193	0,0690
Random Forest	0,5185	0,0392
Gradient Boosting Machine	0,5262	0,0544



# PRELIMINARY RESULTS

PREDICT: MONTHLY RETURNS<sup>2</sup> > 0,01

Algorithm	Accuracy	Kappa
KNN	0,6106	0,0348
Neural Network	0,6108	0,0454
Random Forest	0,6116	0,0407
<b>Gradient Boosting Machine</b>	<b>0,6127</b>	<b>0,0250</b>



# COMPARISON PRELIMINARY RESULTS

Algorithm	Accuracy	Kappa
<b>KNN</b>	<b>0,5320</b>	0,0693
Neural Network	0,5193	0,0690
Random Forest	0,5185	0,0392
Gradient Boosting Machine	0,5262	0,0544

Algorithm	Accuracy	Kappa
KNN	0,6106	0,0348
Neural Network	0,6108	0,0454
Random Forest	0,6116	0,0407
<b>Gradient Boosting Machine</b>	<b>0,6127</b>	0,0250

# CONCLUSION - LOOKING FORWARD

- Results are not satisfying enough to be confident in using these models as a selling or buying indicator
  - Need for improvement
- Data side: find additional variables that explain return and volatility
- Algorithm side: - try to improve tuning
  - use XGB boosting instead of GBM
  - use Keras or MXNet instead of nnet





ANY  
QUESTIONS?

