

W U T I S

Price Movement Prediction using Natural Language Processing

Systematic Trading Division

Yiqing Rosaline Chen

Isaac Manopla

Florian Kollarczik

Thomas Ziller

Vienna, 06/12/2020

Team Overview

Systematic Trading



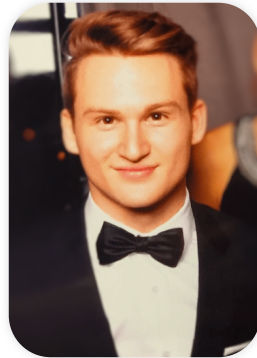
Yiqing Rosaline
Chen

Analyst

- Backtesting
- Trading Strategy



- B.S. – 6th Sem. – (IB)
- B.S. – 3rd Sem. – (CS)



Florian
Kollarczik

Analyst

- Natural Language Processing
- Signal Prediction



- B.S. (WU) – 5th Sem.



Isaac
Manopla

Analyst

- Backtesting
- Trading Strategy



- B.A. – 4th Sem. – (IB)



Thomas
Ziller

Analyst

- Natural Language Processing
- Signal Prediction



- B.S. (WU) – 5th Sem.

Natural Language Processing

Systematic Trading

Tokenization

Today, I was swimming in the sea
with my girlfriend.



[Today] [I] [was] [swimming]
[in] [the] [sea]
[with] [my] [girlfriend]

[Today] [was] [swimming]
[sea] [girlfriend]



[Today – NN] [was – VBD]
[swimming – VBG]
[sea – NN] [girlfriend – NN]

Position of Speech

Filtering

[Today] [I] [was] [swimming]
[in] [the] [sea]
[with] [my] [girlfriend]



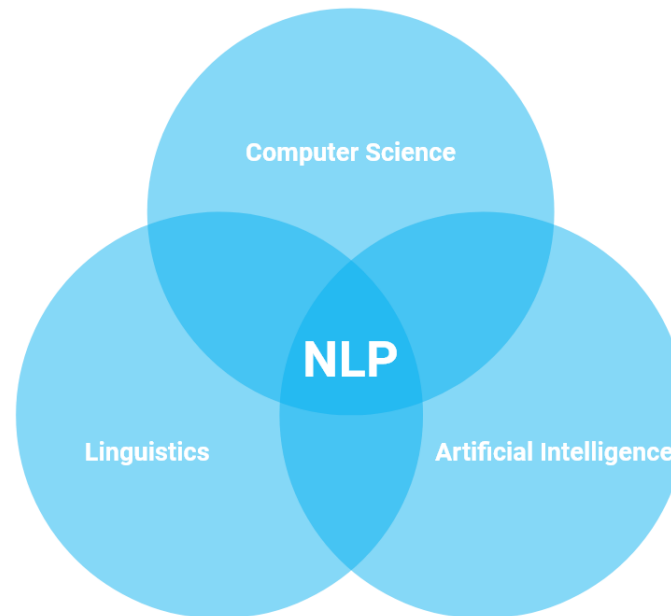
[Today] [was] [swimming]
[sea] [girlfriend]

[Today – NN] [was – VBD]
[swimming – VBG]
[sea – NN] [girlfriend – NN]



[Today] [be] [swim]
[sea] [girlfriend]

Lemmatization



Data – Reddit World News / S&P 500

Systematic Trading



389,750

Observations

9

Years

Sample

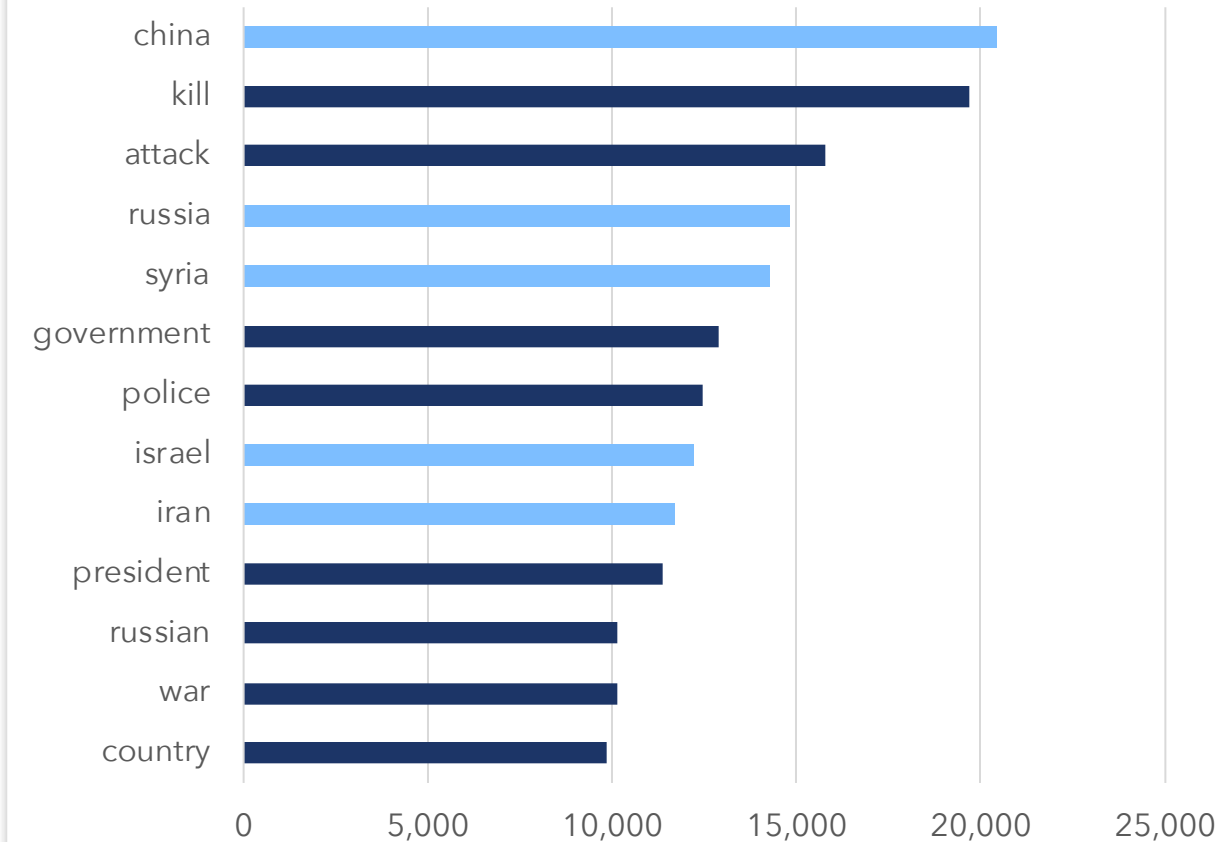
Sources: Yahoo Finance API / Kaggle

Date	Daily Return	Signal	Headlines
2014-03-24	0.004404	Sell	scores killed in pakistan clashes...
2014-03-25	-0.007000	Sell	the un mismanagement program...
2014-03-26	-0.001900	Buy	us plans for missile shields...
2014-03-27	0.004640	Buy	archbishop of canterbury...
2014-03-28	0.007924	Buy	migrant workers told to abandon...
2014-03-29	0.007039	Sell	four injured in firebomb attack...
2014-03-30	0.002853	Buy	japan oks additional role for...

Descriptive Statistics

Systematic Trading

Most frequent words



5,589,154

Total Words

74,325

Unique Words

1,638,153

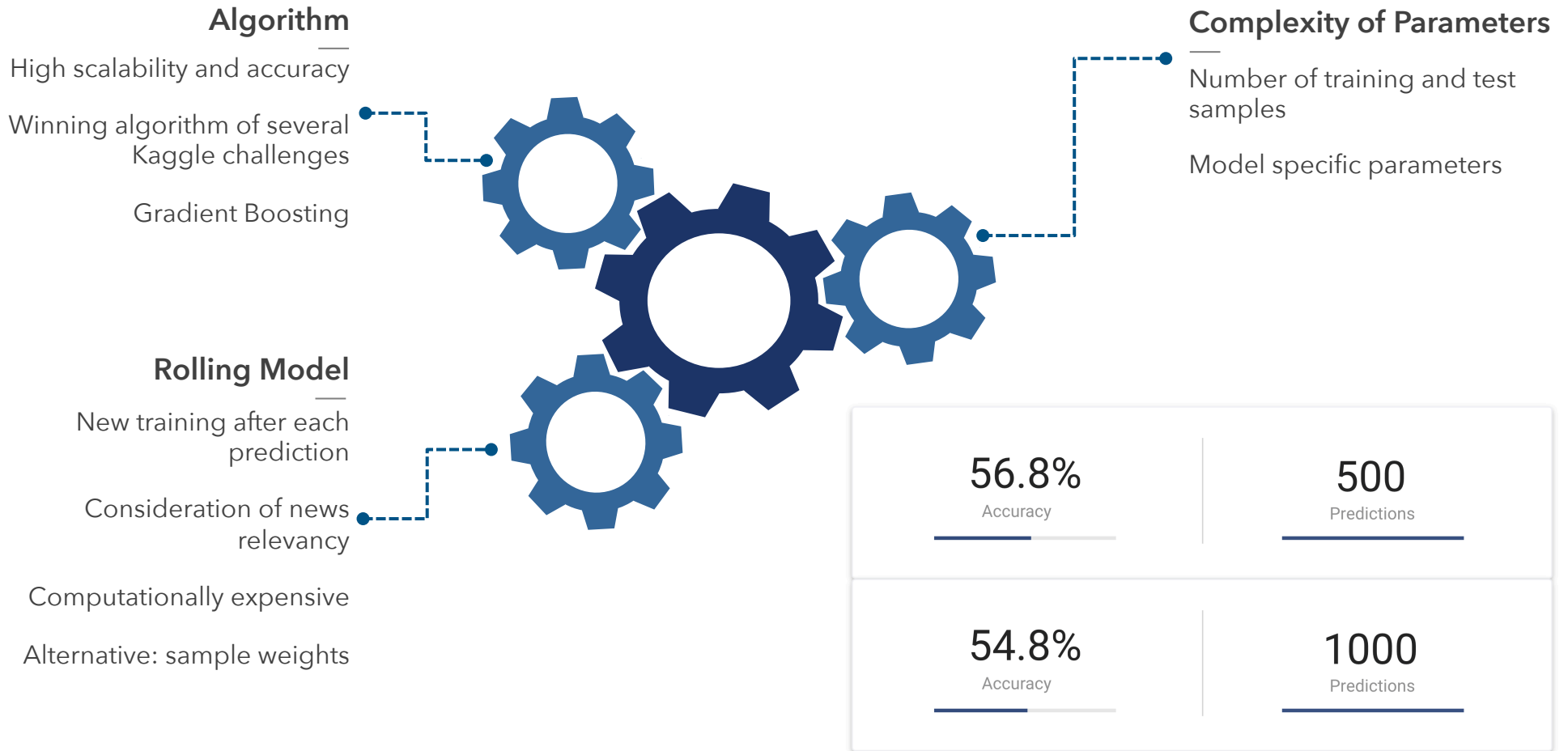
Total After Processing

1,478

Unique After Processing

Trading Signal Prediction with XGBoost

Systematic Trading



Implementation

- Determined by NLP signals (Long/Short)
- Model will go Long or Short once every signal appears
- Hold position when same consecutive signal
- Positions are executed at closing price

Asset Allocation

- SPDR S&P 500 ETF Trust (SPY)
- S&P 500 (^GSPC)
- Initial account balance of \$100,000
- Every trade will allocate 100% of the current balance
- Time horizon: 2 years

Risks & Constraints

- Volatile markets
- Accuracy of the signals
- Calculation of the model for a certain date range and asset

Limitations

- Dataset timeframe 2008 – 2016
- Limited to world news on Reddit
- Lack of technical indicators
- Limited usage for today's stock market
- Supplementary investing strategy
- Best performance in volatile markets

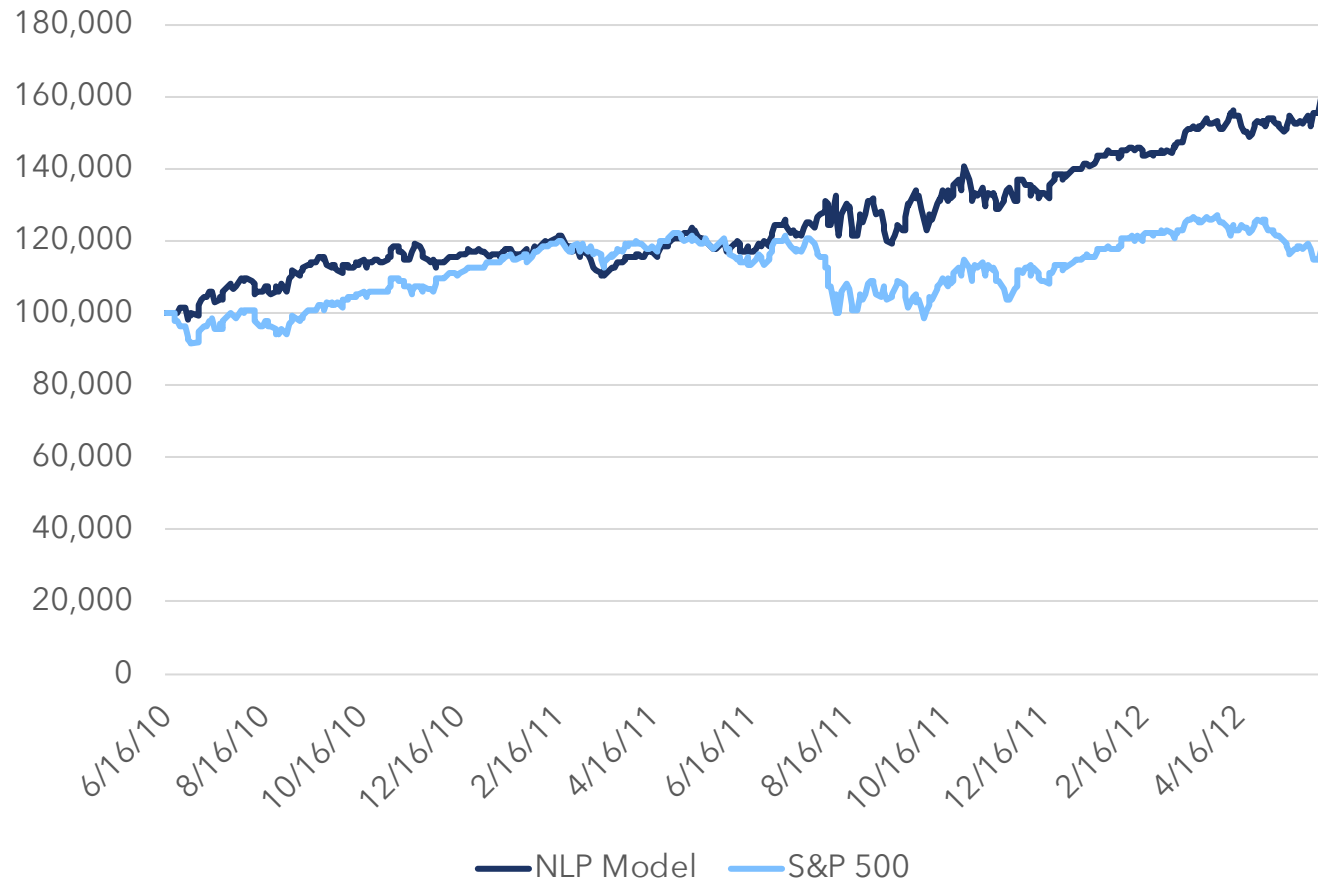
Future Work

- Provided model is limited to world news and only gathers data from one news source
- Re-train model and add extensions (Hyper Parameter Tuning)
- Sentiment Analysis
- Implementation of tree-based algorithms to the model: 50% random forest & 50% (extreme) gradient boosting

Price Performance

Systematic Trading

W U T I S



Performance – S&P 500



8.62%
Annualized Return

17.98%
2-yr Return

Performance – NLP Model

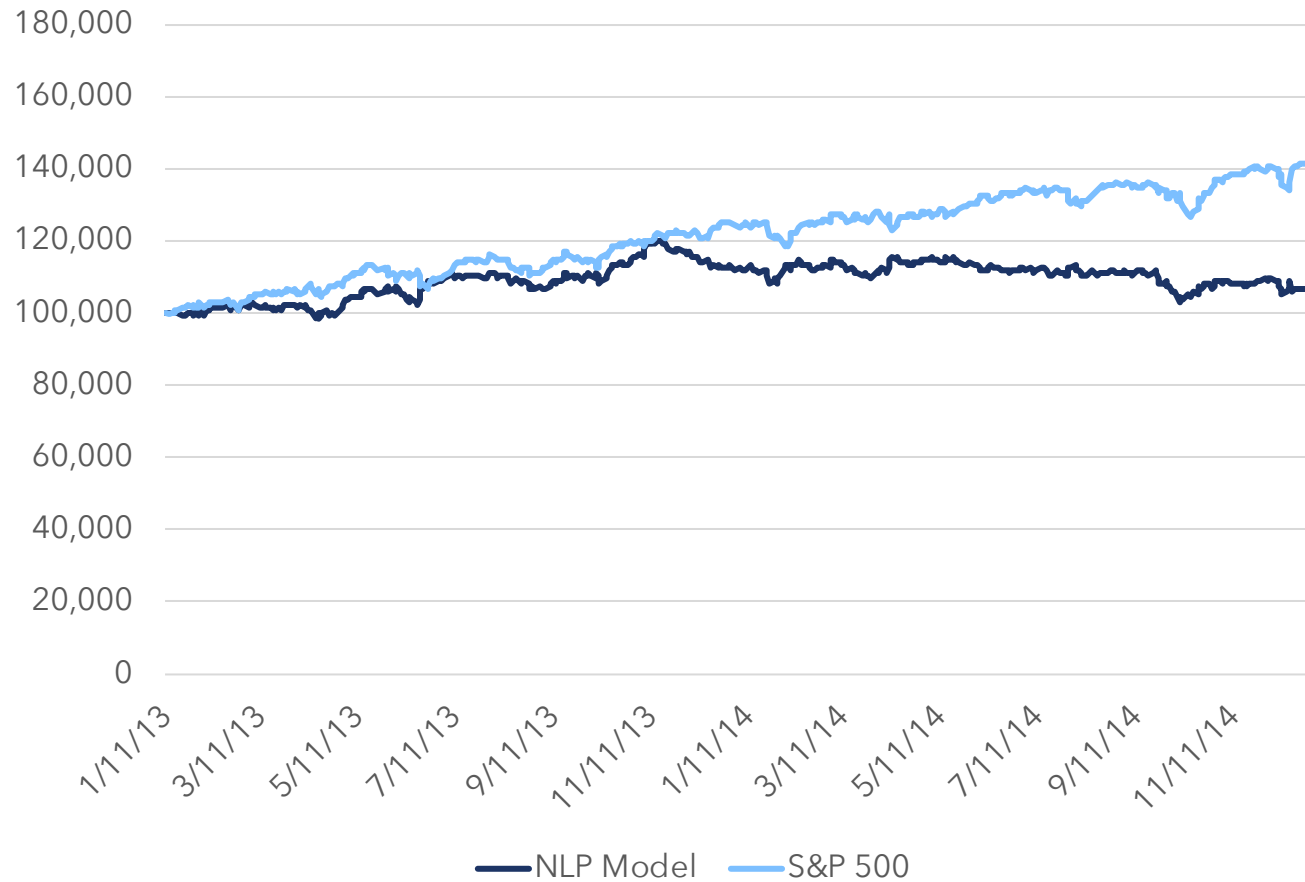


25.56%
Annualized Return

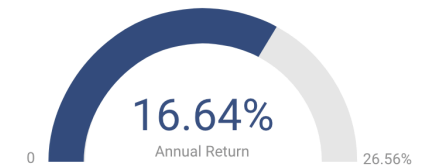
60.16%
2-yr Return

Price Performance

Systematic Trading



Performance – S&P 500



16.64%
Annualized Return

36.04%
2-yr Return

Performance – NLP Model



2.15%
Annualized Return

4.35%
2-yr Return