2019 DATATHON

TEAM 22 REPORT

Chutong Xiao Wanxing Dai

Yutong Liu Qinyang Fang

CONTENTS

1 Introduction

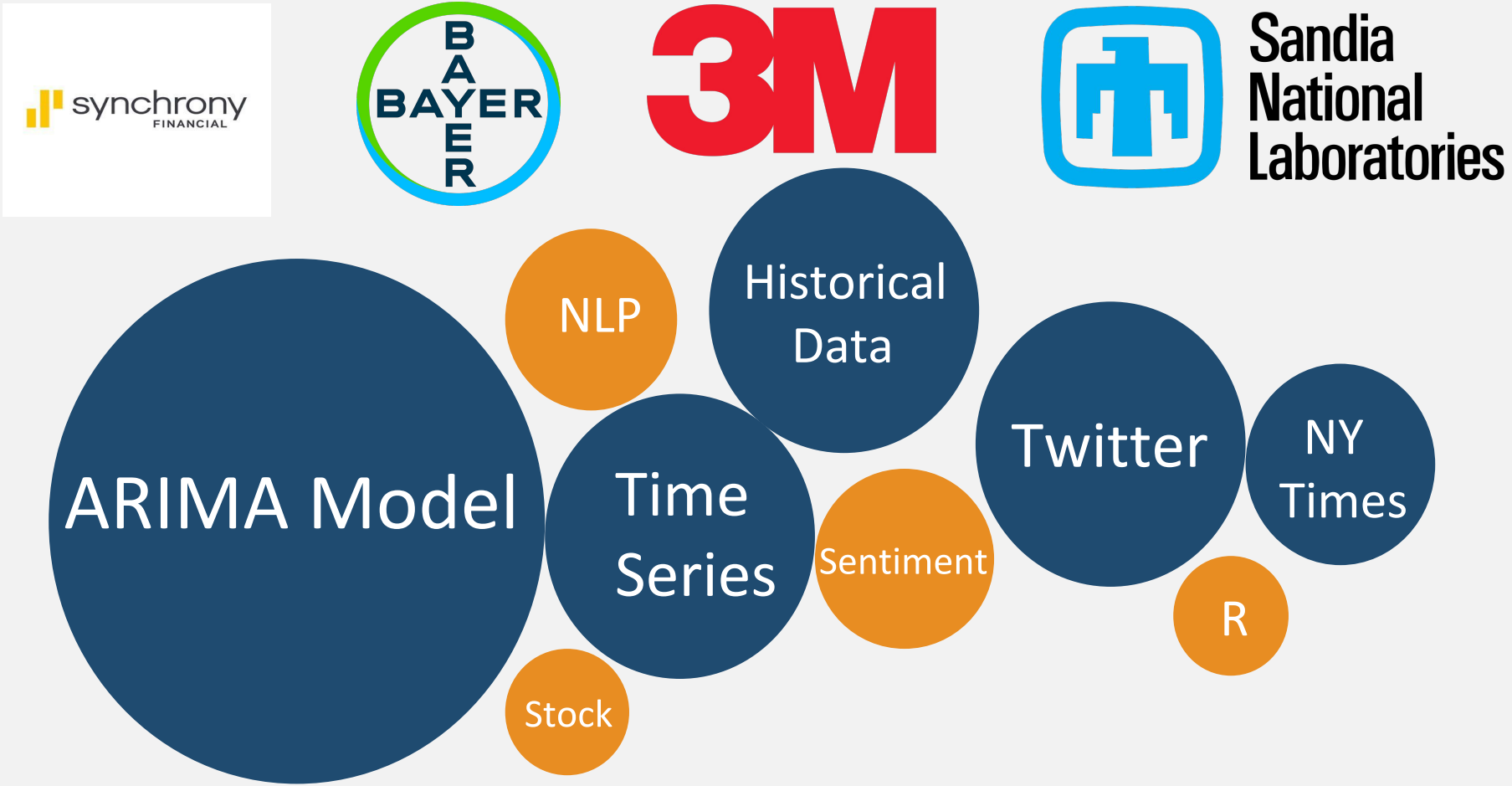
2 API

3 Prediction Model

4 Possible Improvements

01

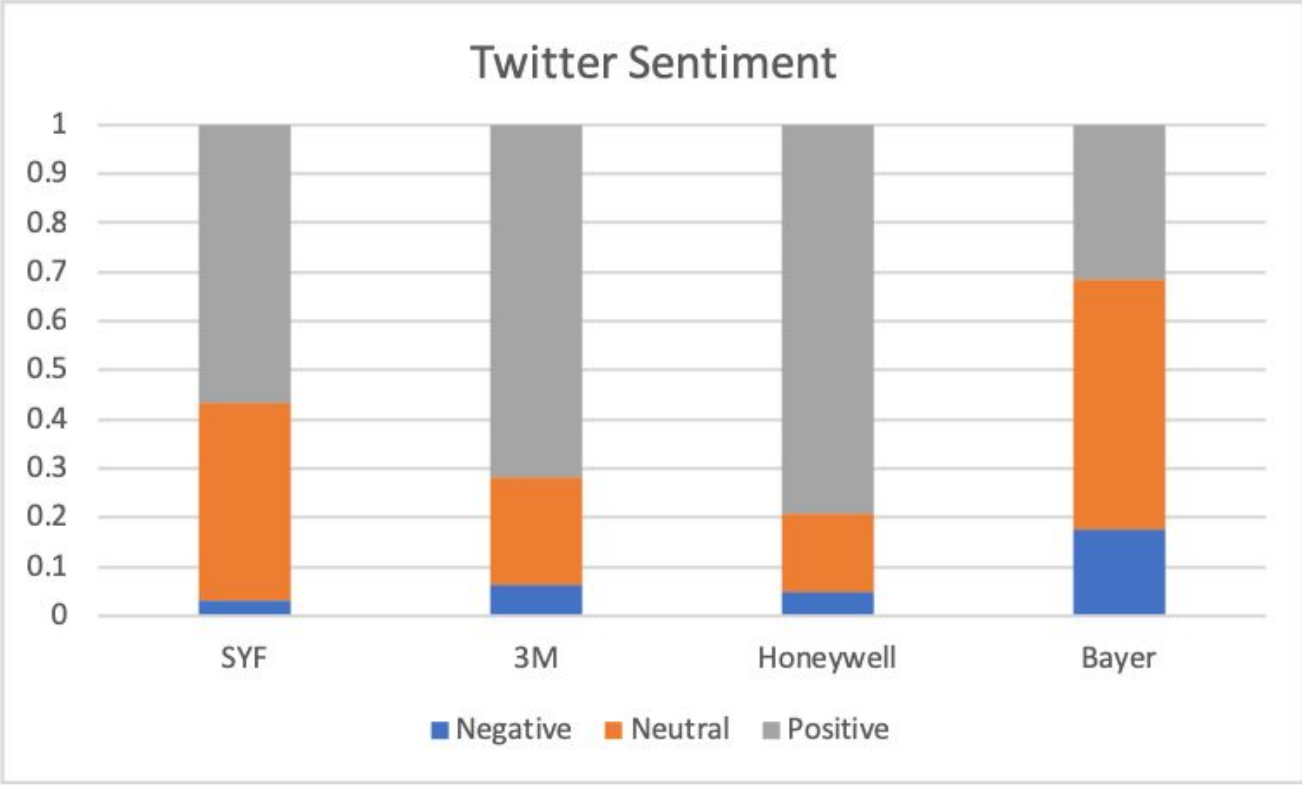
Introduction

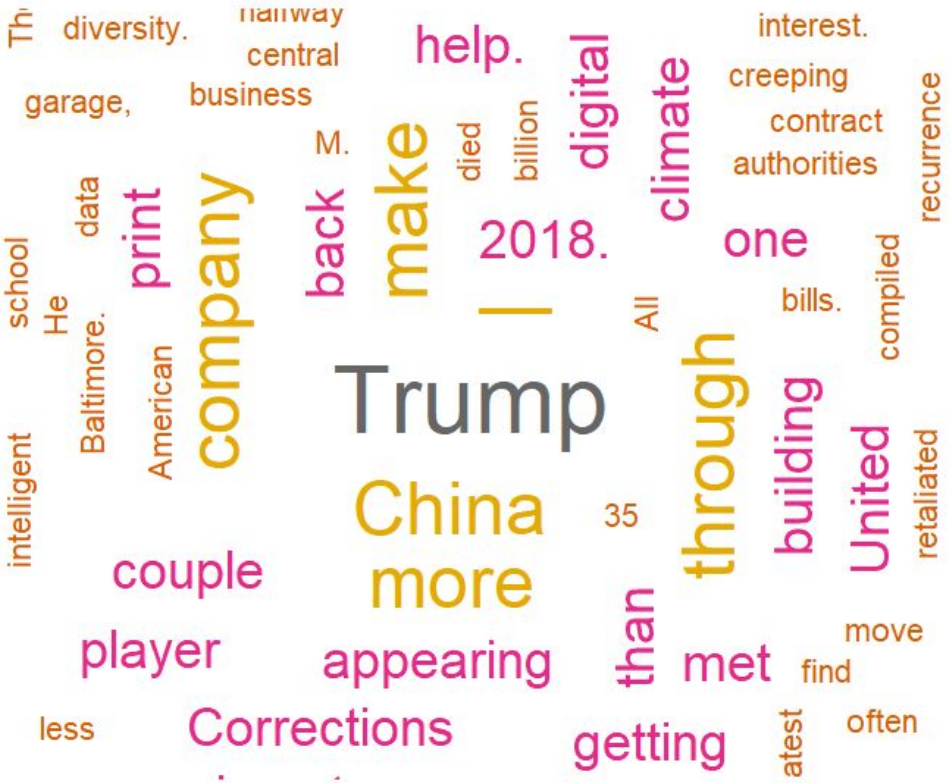

synchrony FINANCIAL
BAYER
3M
Sandia National Laboratories
ARIMA Model
NLP
Historical Data
Time Series
Sentiment
Twitter
NY Times
R
Stock

02

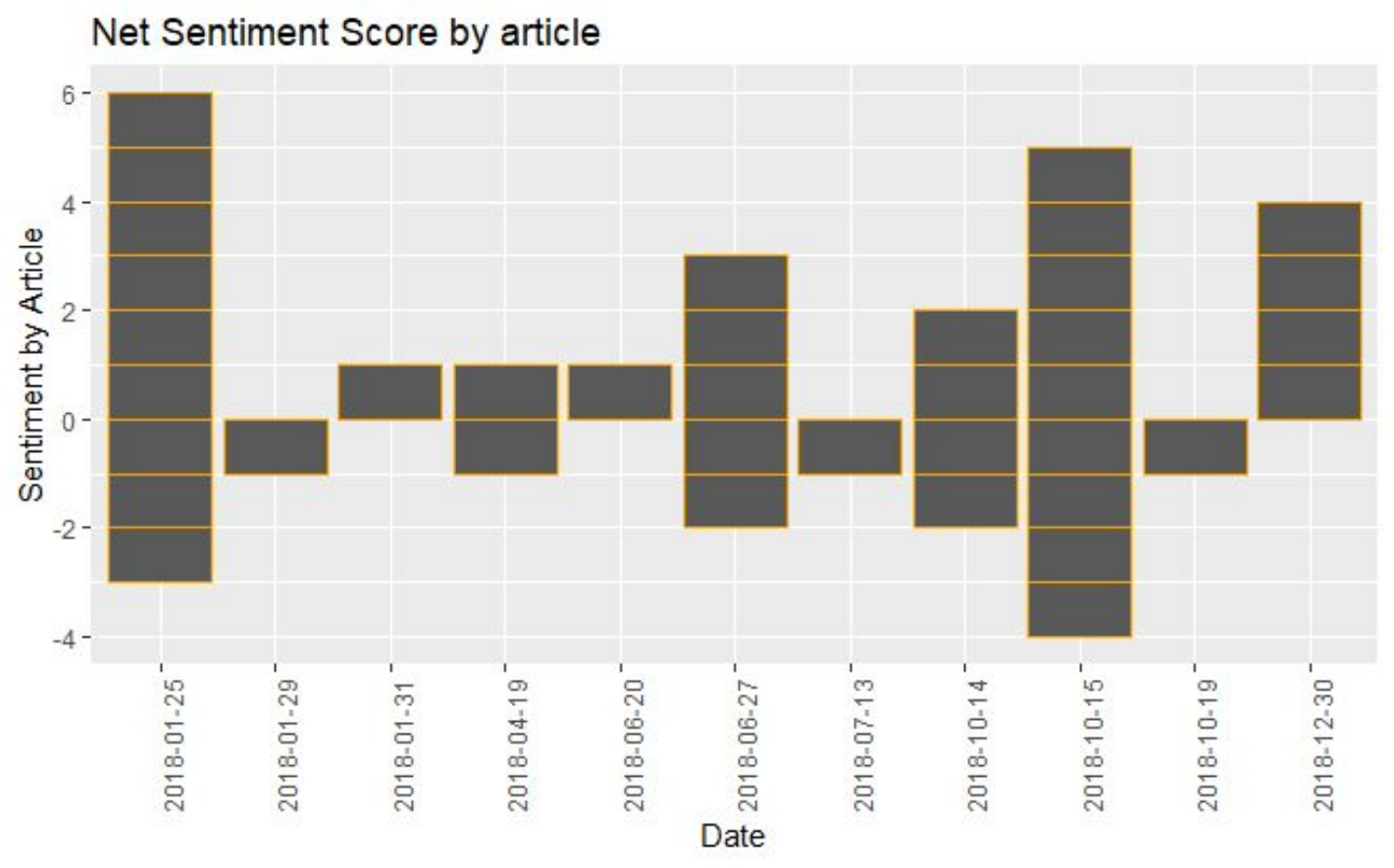
NLP on APIs

Twitter API


Twitter Sentiment
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
SYF
3M
Honeywell
Bayer
Negative
Neutral
Positive


Th
diversity.
hallway
central
business
garage,
help.
digital
interest.
creeping
contract
authorities
climate
recurrence
died
billion
M.
make
back
company
print
data
school
He
Baltimore.
American
intelligent
2018.
one
bills.
compiled
All
Trump
through
building
United
retaliated
China
35
couple
more
player
appearing
than
met
move
find
Corrections
getting
less
often
atest

Honeywell



Honeywell

Positive = 1

Negative = -1

03 Prediction Model

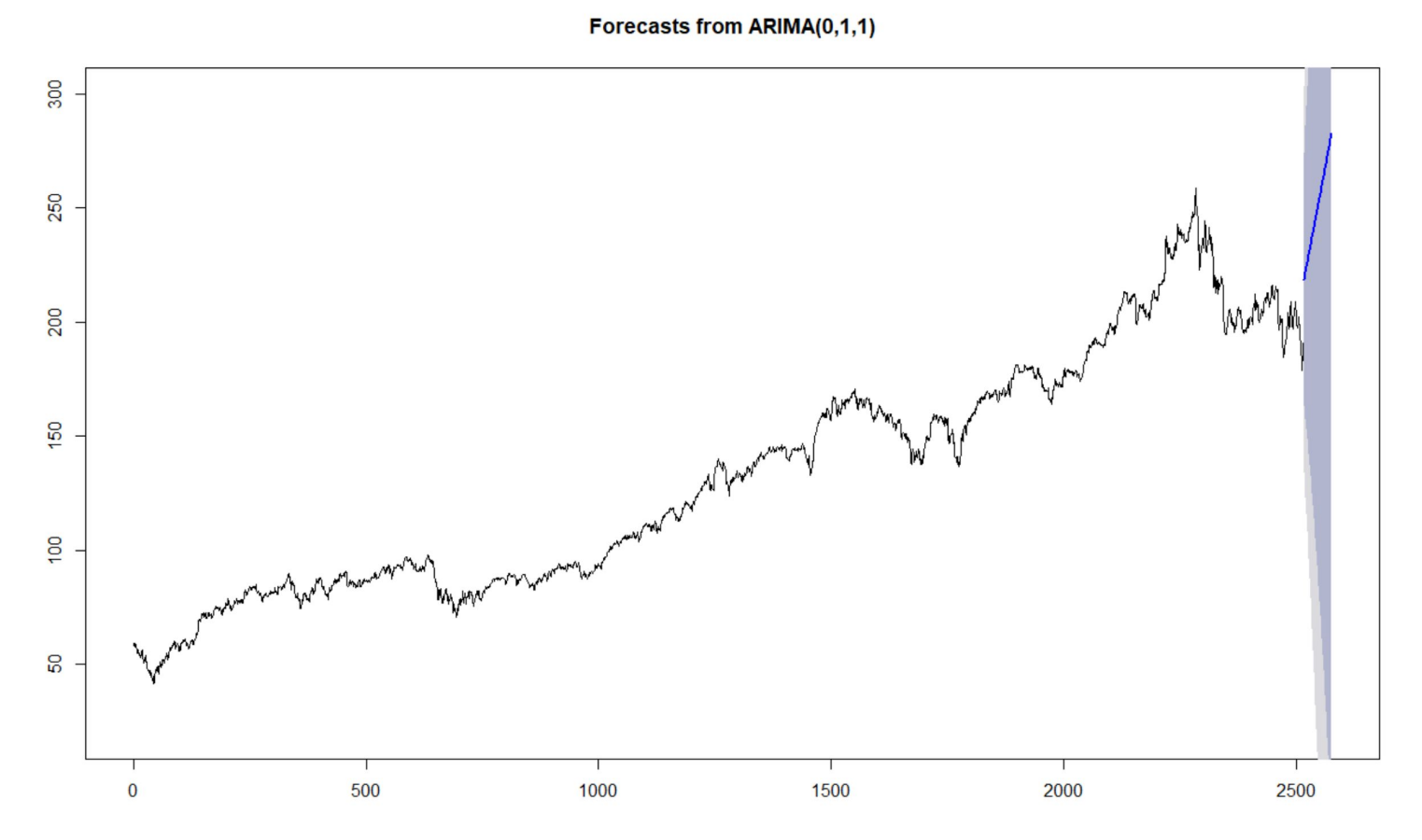
Log Returns Plot

Time Series

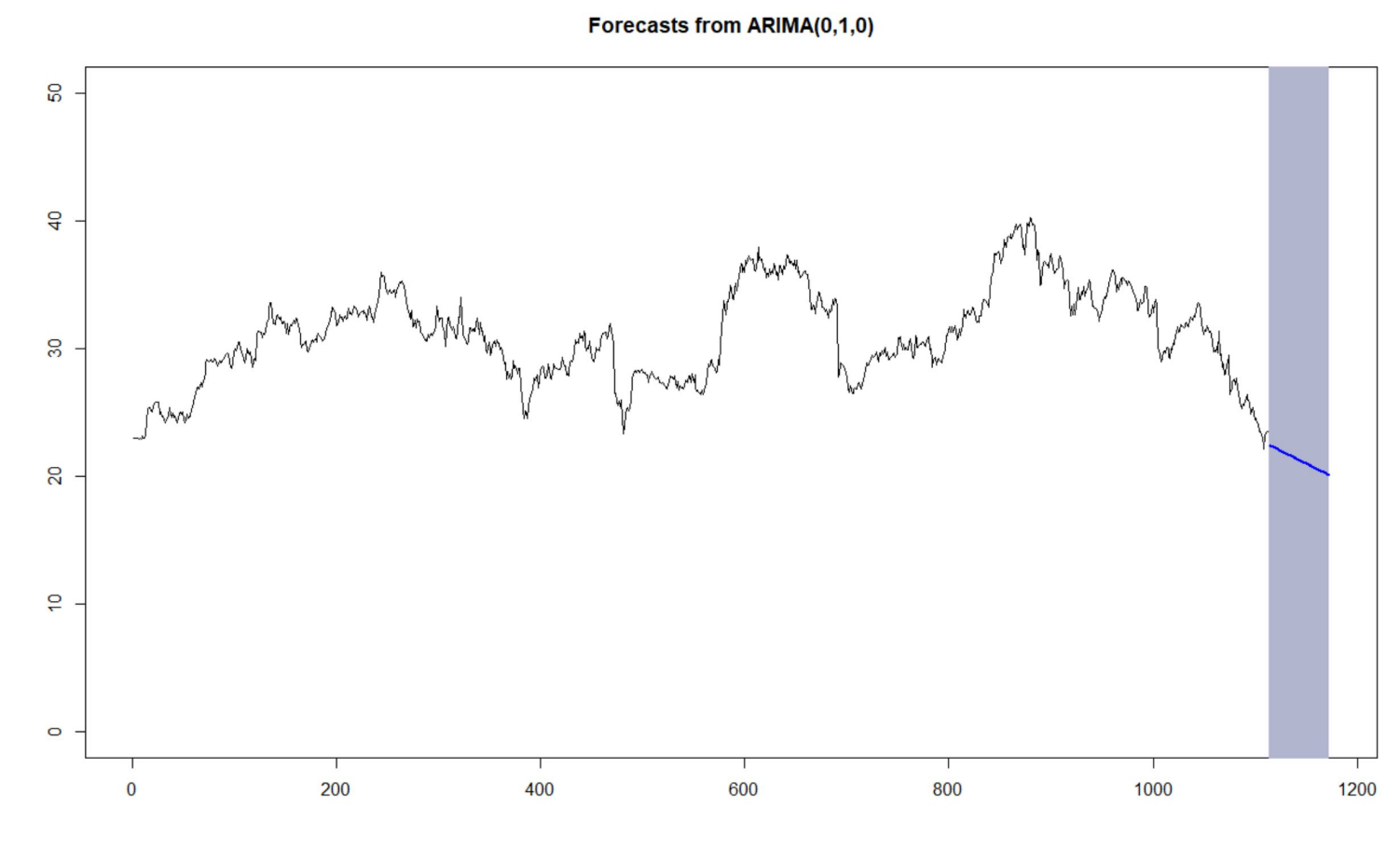
ARIMA MODEL

Estimate
& ForeCast

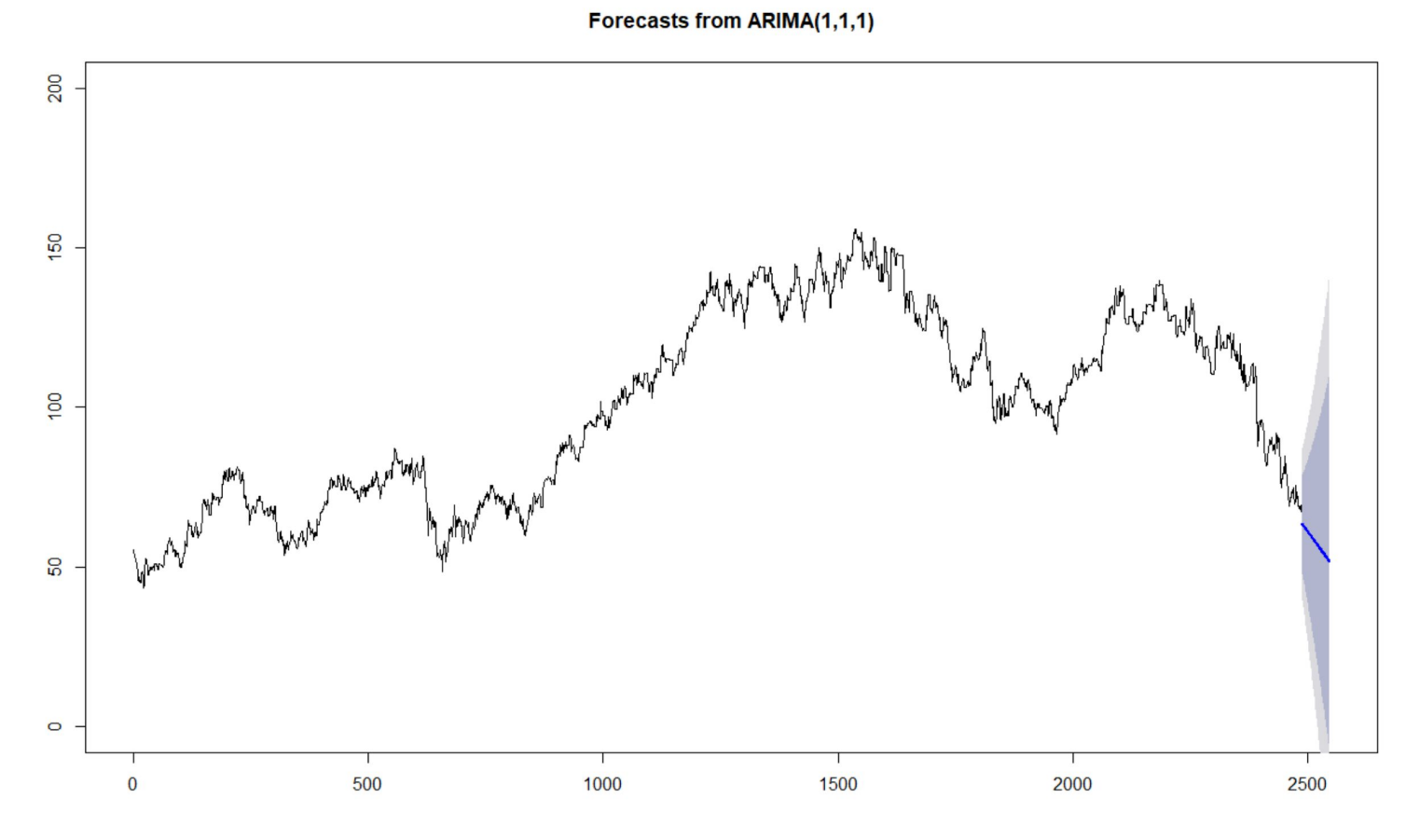
Autoregressive Integrated
Moving Average

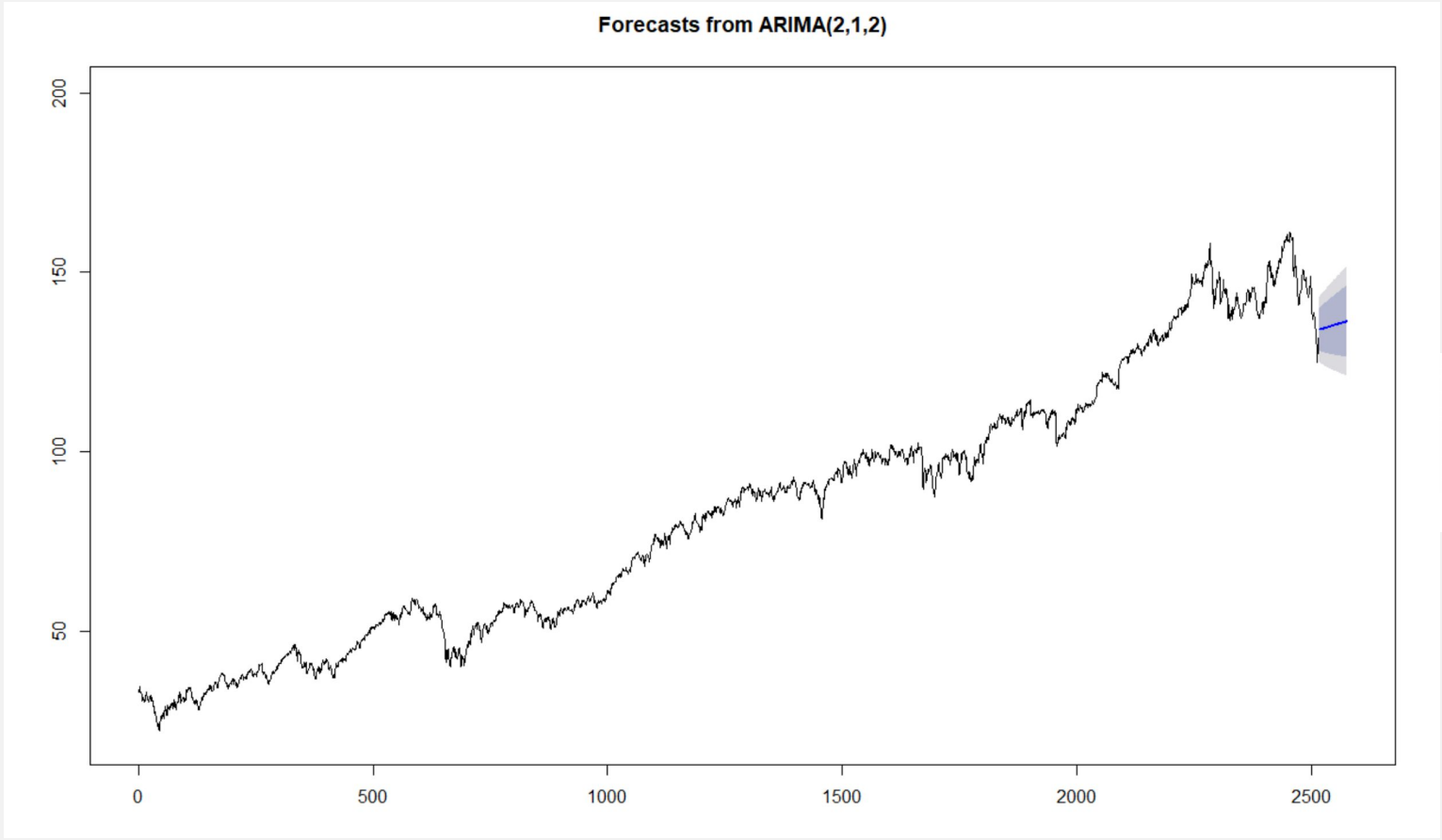

Forecasts from ARIMA(0,1,1)
300
250
200
150
100
50
0
500
1000
1500
2000
2500

3M



synchrony FINANCIAL


Forecasts from ARIMA(1,1,1)
200
150
100
50
0
0
500
1000
1500
2000
2500


Forecasts from ARIMA(2,1,2)
200
150
100
50
0
500
1000
1500
2000
2500

Honeywell

Winner:

04 Possible Improvements

For the sentiment of NYTimes articles, we could have generated more data but due to time limits we only selected data in 2018.

Time

Twitter only allows access to data at most 7 days prior to the present date, which leads to insufficient tweets collected

Limits

We tried to change the legend color of historical data to red and blue, but we encountered some technical difficulties.

Color

We only used one model to make prediction. Our analysis could be more precise with the combination of EGARCH model.

Model

Thank you!

References:

https://rpubs.com/leenam/FitchDSC1

http://dataaspirant.com/2018/03/22/twitter-sentiment-analysis-using-r/

https://www.cnblogs.com/bicoffee/p/3838049.html