RGR Stock Price Forecasting Project

By Jack Wang

PROBLEM STATEMENT

Stock prices are hard to predict because they are not only affected by the performance of the underlying companies but also the expectations from the general public. As known, the stock price of firearm companies are highly correlated to the public opinions toward gun control.

My model intends to predict the stock price of one of the largest firearm company in the states, RGR (Sturm, Ruger & Co., firearm company), by using its historical stock price, public opinions toward gun control, and its financial reports to SEC.

EXECUTIVE SUMMARY

The goal of my project is to build a time series regression model that predicts the stock price of RGR.

The data I am using would be historical stock price from **Yahoo Finance**, twitter posts scraped from **Twitter**, **subreddit** posts mentioned about gun control, and also the financial reports to **SEC**. I will do sentiment analysis on the text data and time series modeling on the historical stock price data. The model will be evaluated using MSE.

DATA COLLECTION

HISTORICAL STOCK PRICE DATA

Use **Yahoo Finance** to collect historical stock price.

- Open_price: the stock price when market opens

TWITTER DATA

Use **twitterscraper** to collect historical tweets. Keyword is very specific - gun control.

- tweet_word_count_sum
- tweet_compound_score_sum
- tweets_sum
- tweet_word_count_mean
- tweet_compound_score_mean

SUBREDDIT DATA

/politics

- redd_pol_score_mean
- redd_pol_comment_mean
- redd_pol_compound_mean
- redd_pol_score_sum
- redd_pol_comment_sum
- redd_pol_post_count

/guns

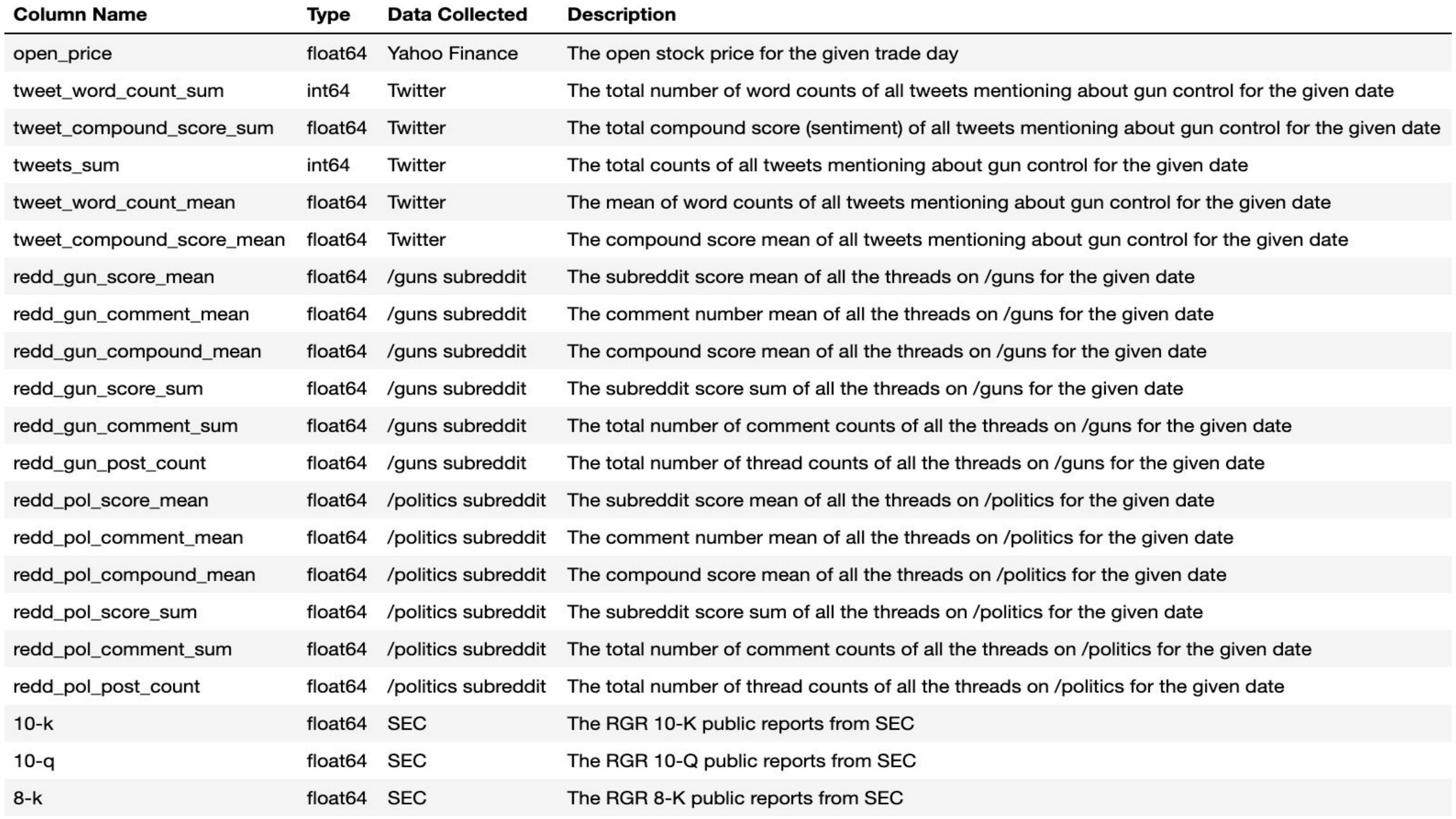
- redd_gun_score_mean
- redd_gun_comment_mean
- redd_gun_compound_mean
- redd_gun_score_sum
- redd_gun_comment_sum
- redd_gun_post_count

SEC DATA

Use **SEC** website to collect public reports.

- 10-K: company’s annual report including financial statement
- 10-Q: company’s quarter report including financial statement
- 8-K: company’s unscheduled events that are important

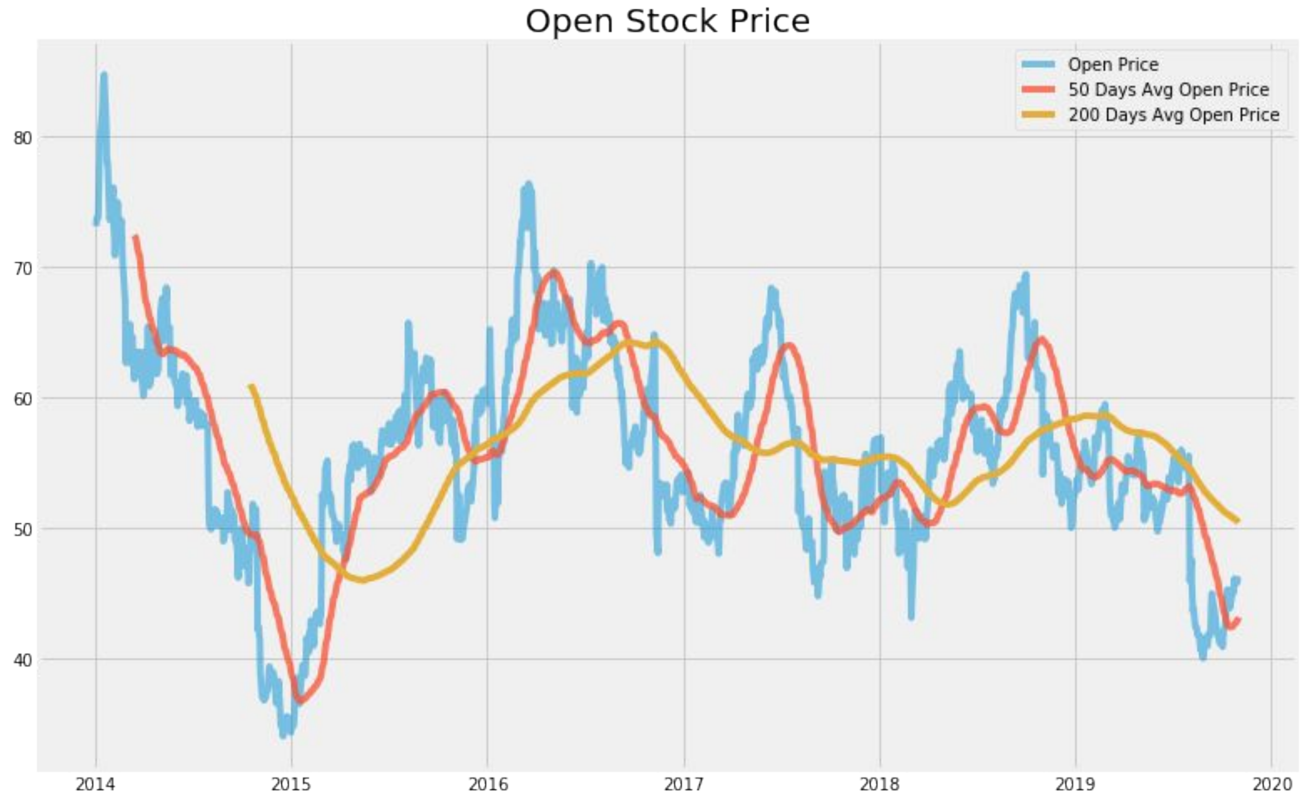
DATA DICTIONARY



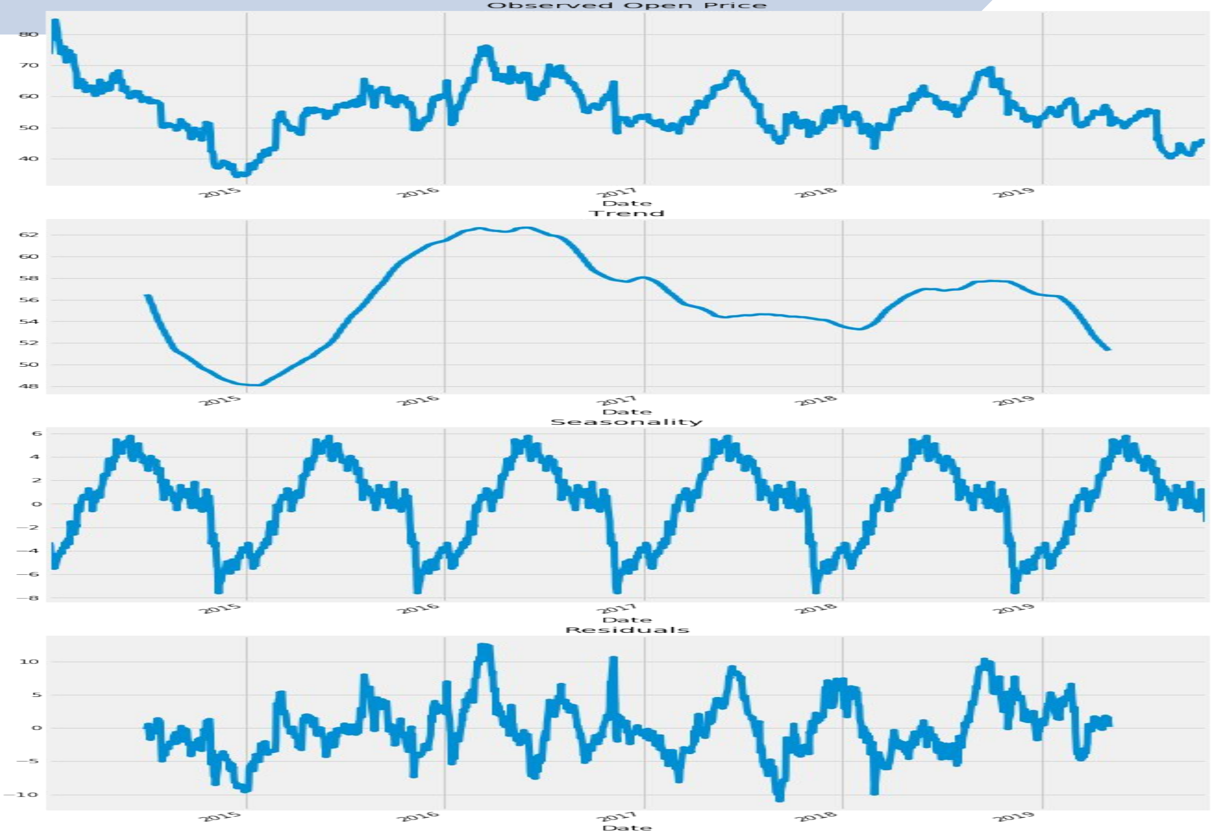
Column Name	Type	Data Collected	Description
open_price	float64	Yahoo Finance	The open stock price for the given trade day
tweet_word_count_sum	int64	Twitter	The total number of word counts of all tweets mentioning about gun control for the given date
tweet_compound_score_sum	float64	Twitter	The total compound score (sentiment) of all tweets mentioning about gun control for the given date
tweets_sum	int64	Twitter	The total counts of all tweets mentioning about gun control for the given date
tweet_word_count_mean	float64	Twitter	The mean of word counts of all tweets mentioning about gun control for the given date
tweet_compound_score_mean	float64	Twitter	The compound score mean of all tweets mentioning about gun control for the given date
redd_gun_score_mean	float64	/guns subreddit	The subreddit score mean of all the threads on /guns for the given date
redd_gun_comment_mean	float64	/guns subreddit	The comment number mean of all the threads on /guns for the given date
redd_gun_compound_mean	float64	/guns subreddit	The compound score mean of all the threads on /guns for the given date
redd_gun_score_sum	float64	/guns subreddit	The subreddit score sum of all the threads on /guns for the given date
redd_gun_comment_sum	float64	/guns subreddit	The total number of comment counts of all the threads on /guns for the given date
redd_gun_post_count	float64	/guns subreddit	The total number of thread counts of all the threads on /guns for the given date
redd_pol_score_mean	float64	/politics subreddit	The subreddit score mean of all the threads on /politics for the given date
redd_pol_comment_mean	float64	/politics subreddit	The comment number mean of all the threads on /politics for the given date
redd_pol_compound_mean	float64	/politics subreddit	The compound score mean of all the threads on /politics for the given date
redd_pol_score_sum	float64	/politics subreddit	The subreddit score sum of all the threads on /politics for the given date
redd_pol_comment_sum	float64	/politics subreddit	The total number of comment counts of all the threads on /politics for the given date
redd_pol_post_count	float64	/politics subreddit	The total number of thread counts of all the threads on /politics for the given date
10-k	float64	SEC	The RGR 10-K public reports from SEC
10-q	float64	SEC	The RGR 10-Q public reports from SEC
8-k	float64	SEC	The RGR 8-K public reports from SEC

EXPLORATORY DATA ANALYSIS

GOLDEN CROSS & DEATH CROSS

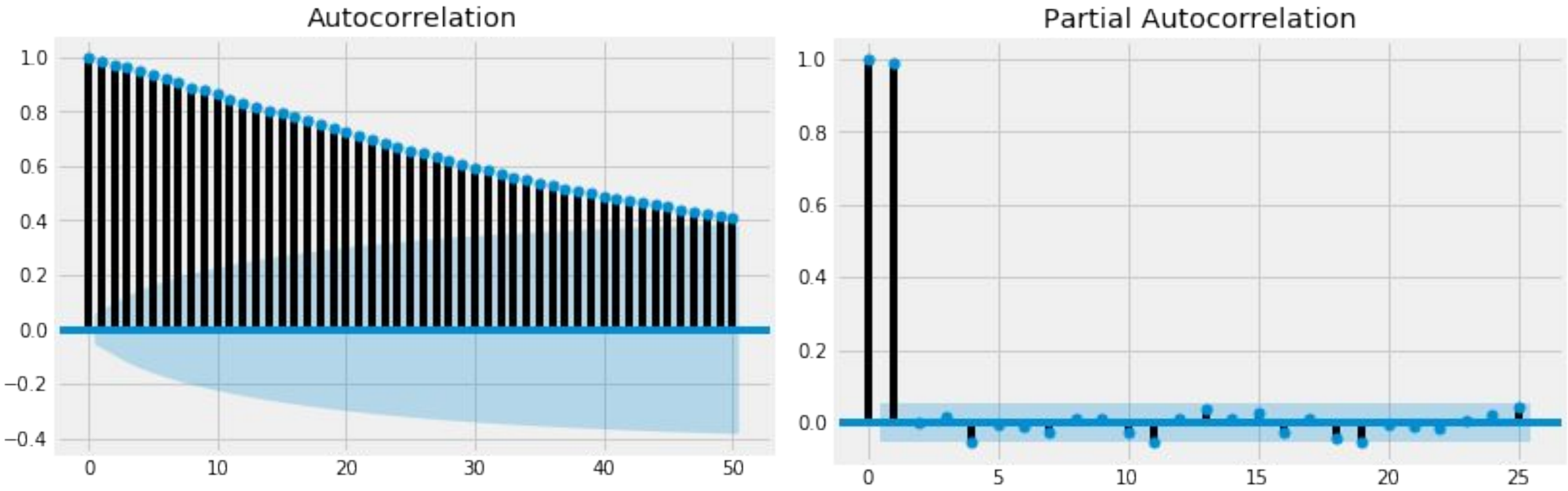


TIME SERIES DECOMPOSE

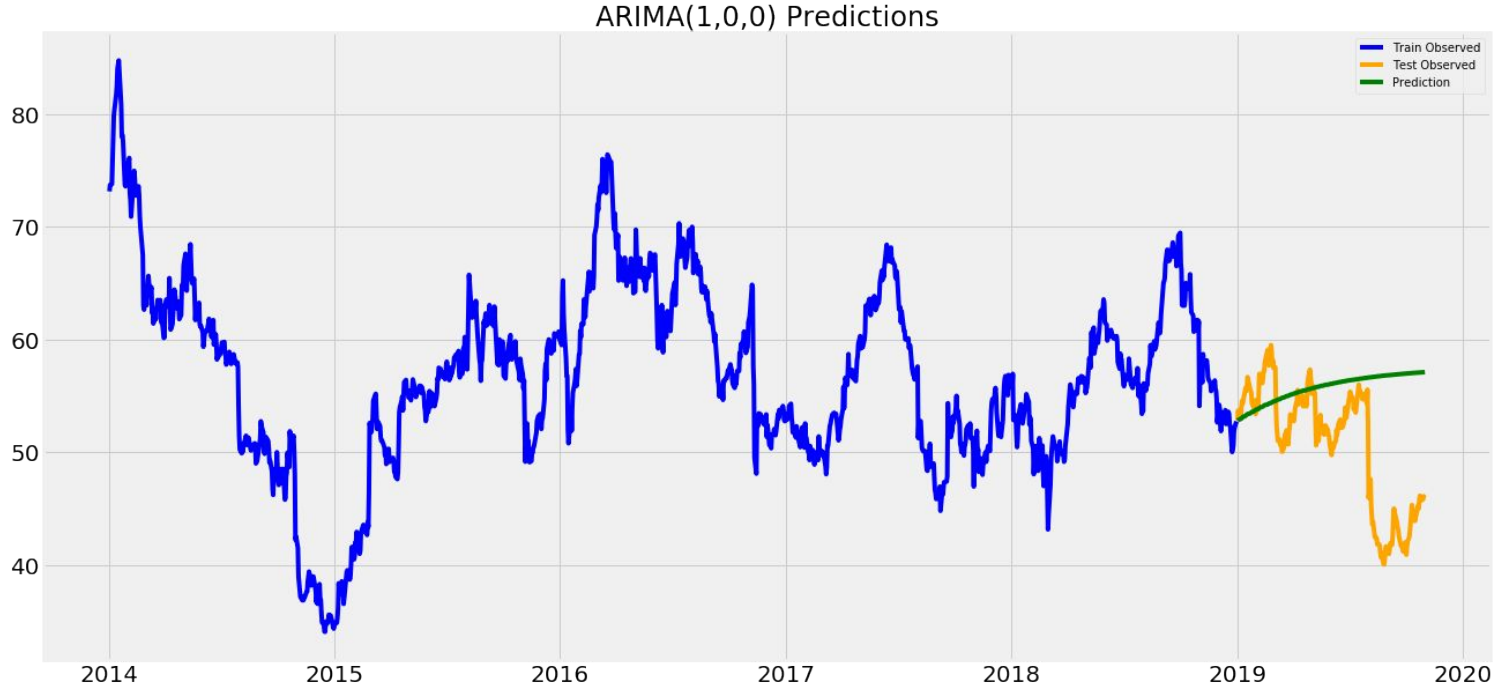


MODELING

ARIMA (1, 0, 0)


Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
0
10
20
30
40
50
Partial Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25

ARIMA (1, 0, 0)

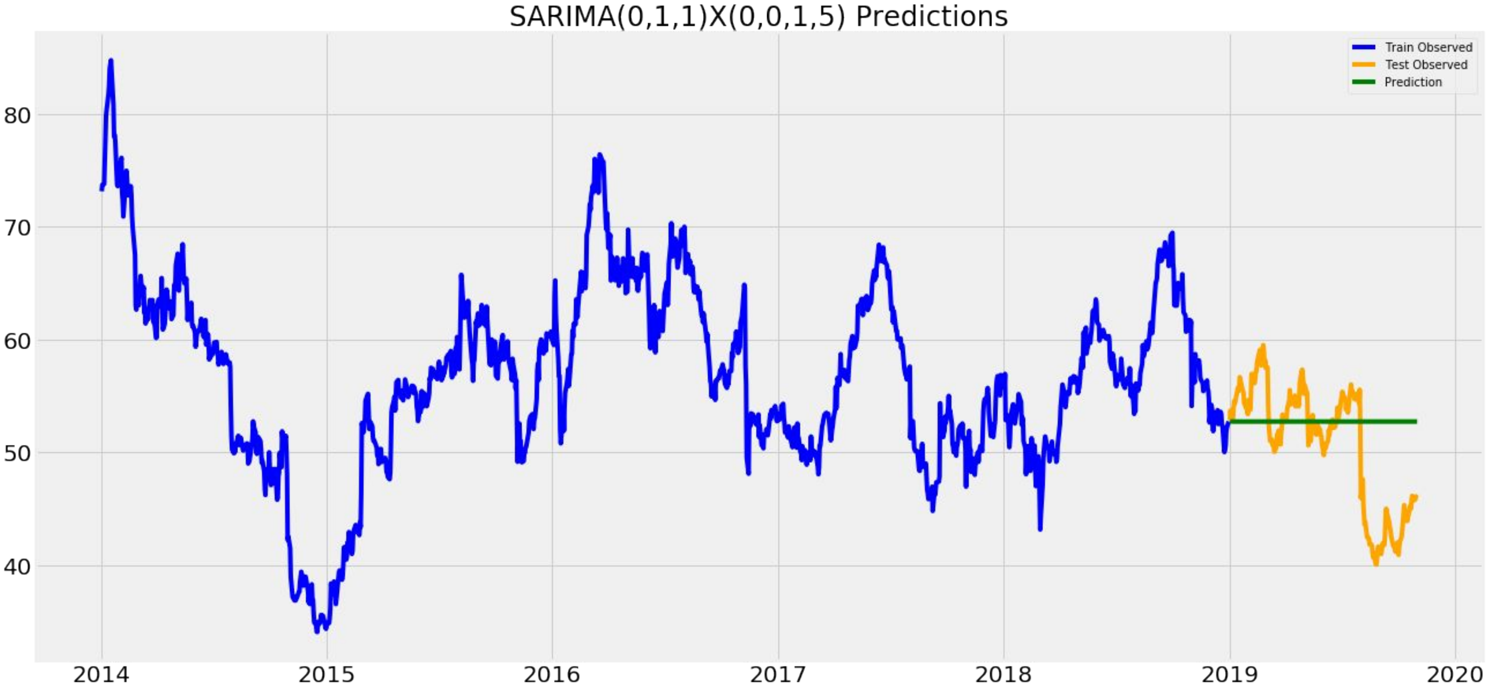

ARIMA(1,0,0) Predictions
Train Observed
Test Observed
Prediction
80
70
60
50
40
2014
2015
2016
2017
2018
2019
2020

SARIMA MODEL

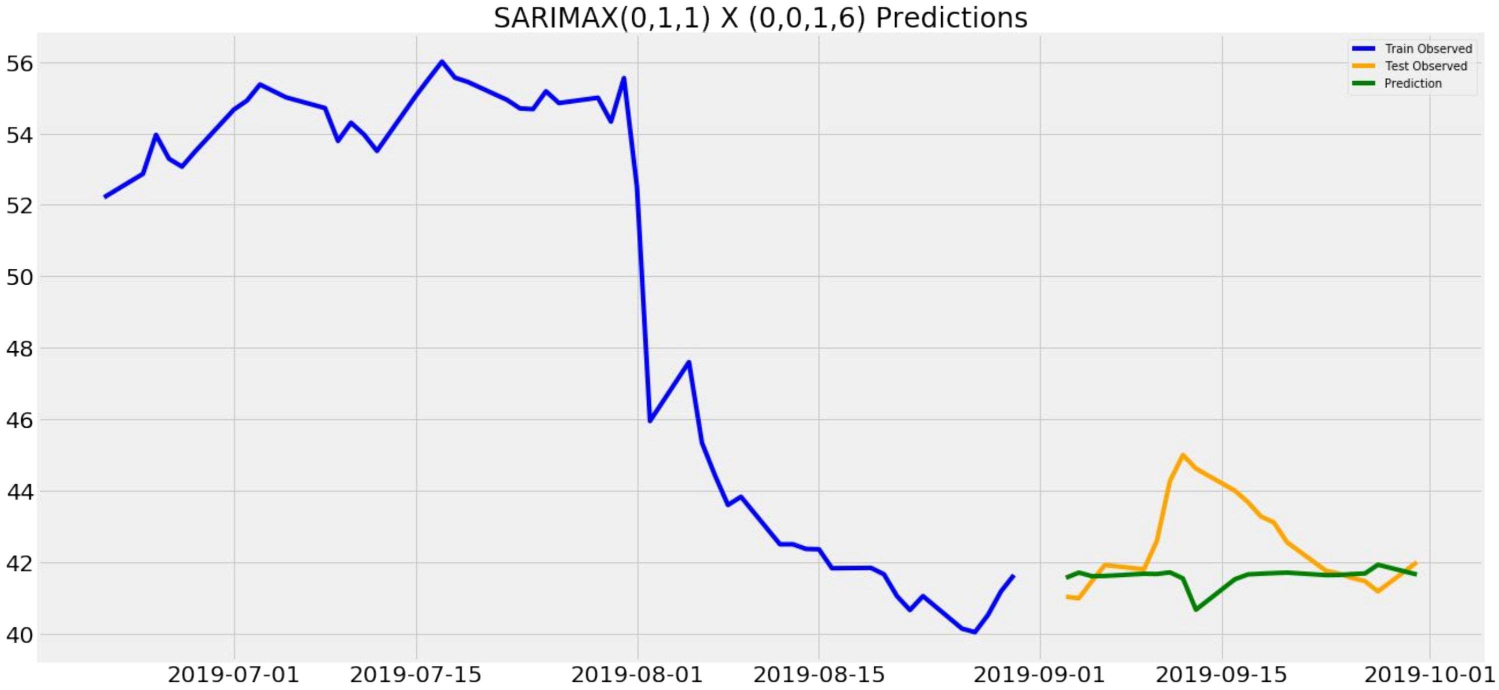
I picked the PDQS using gridsearch over AIC score (Akaike Information Criteria).

It measures the simplicity & goodness of fit. The lower the **AIC score** the better.

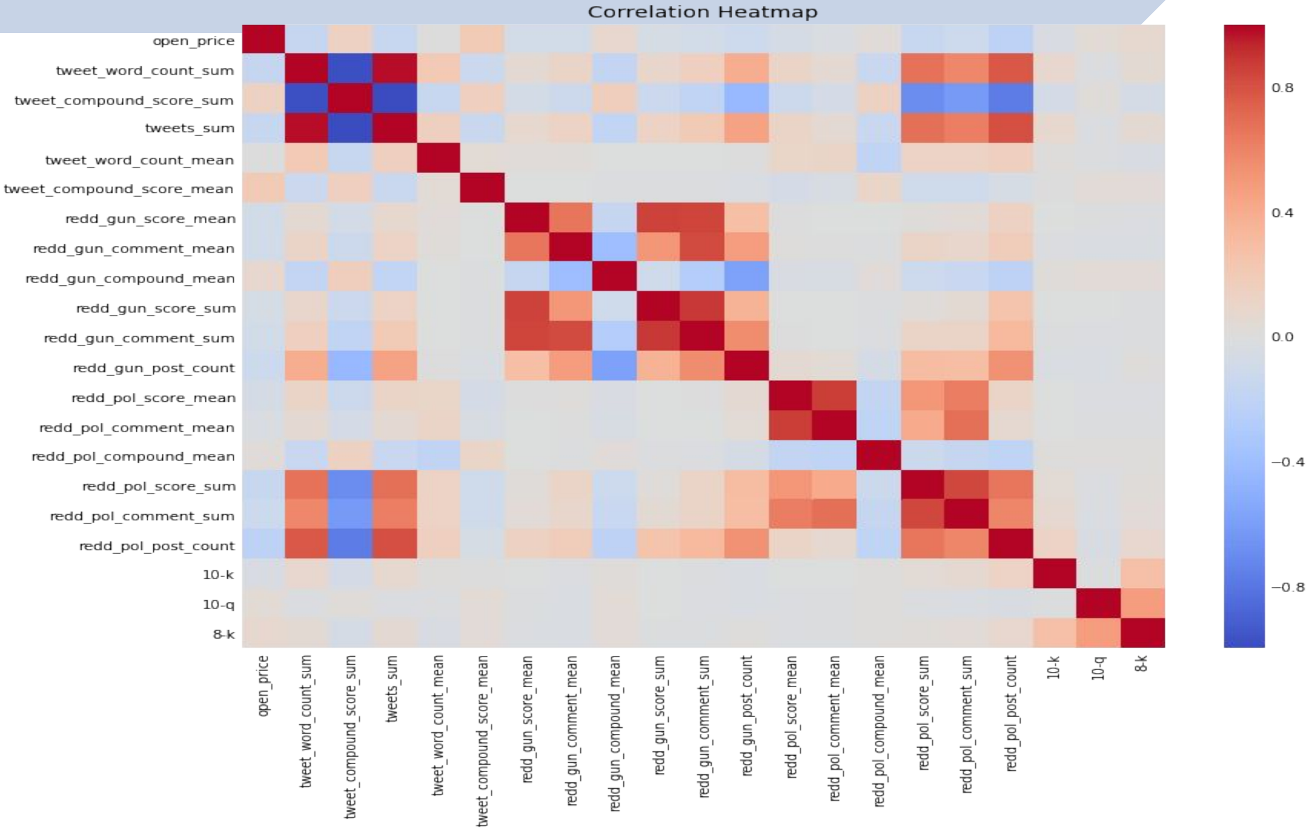
SARIMA (0,1, 1) X (0, 0, 1, 5)



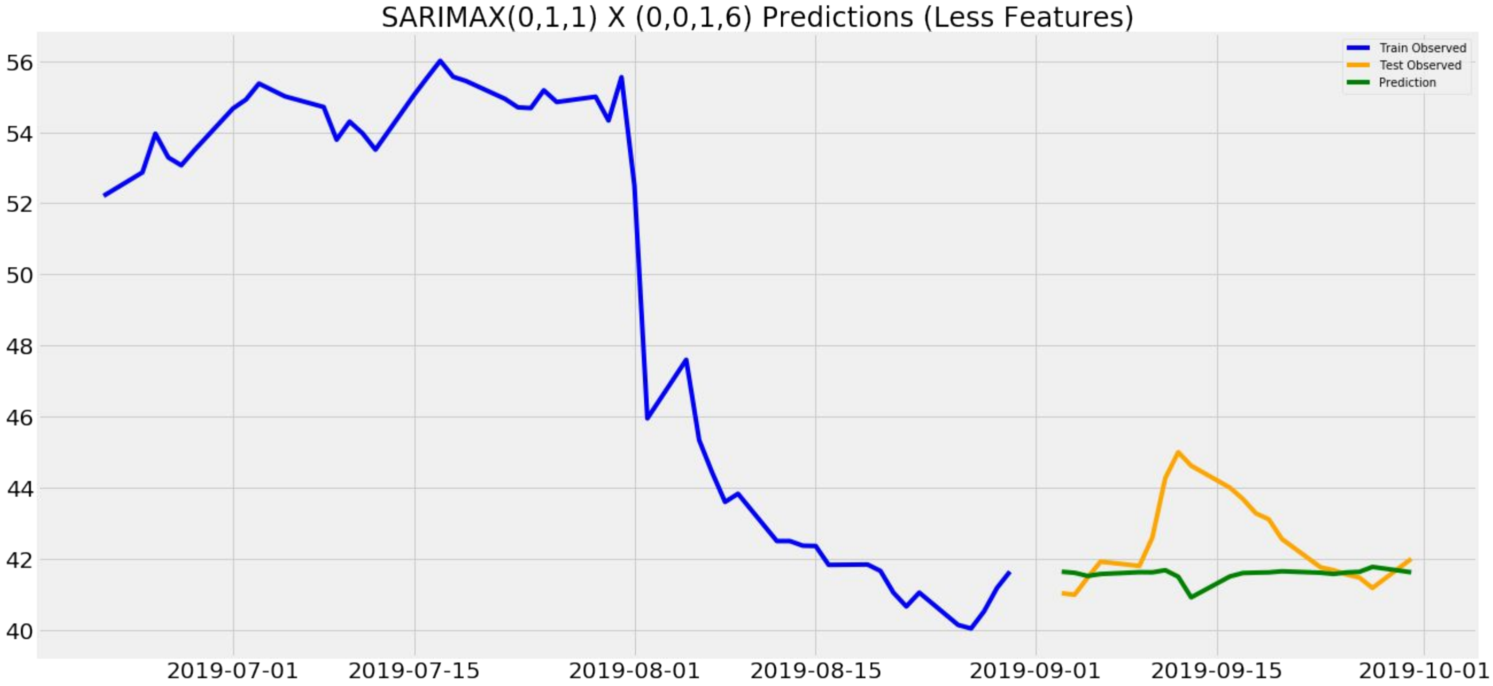
SARIMAX


SARIMAX(0,1,1) X (0,0,1,6) Predictions
Train Observed
Test Observed
Prediction
56
54
52
50
48
46
44
42
40
2019-07-01
2019-07-15
2019-08-01
2019-08-15
2019-09-01
2019-09-15
2019-10-01

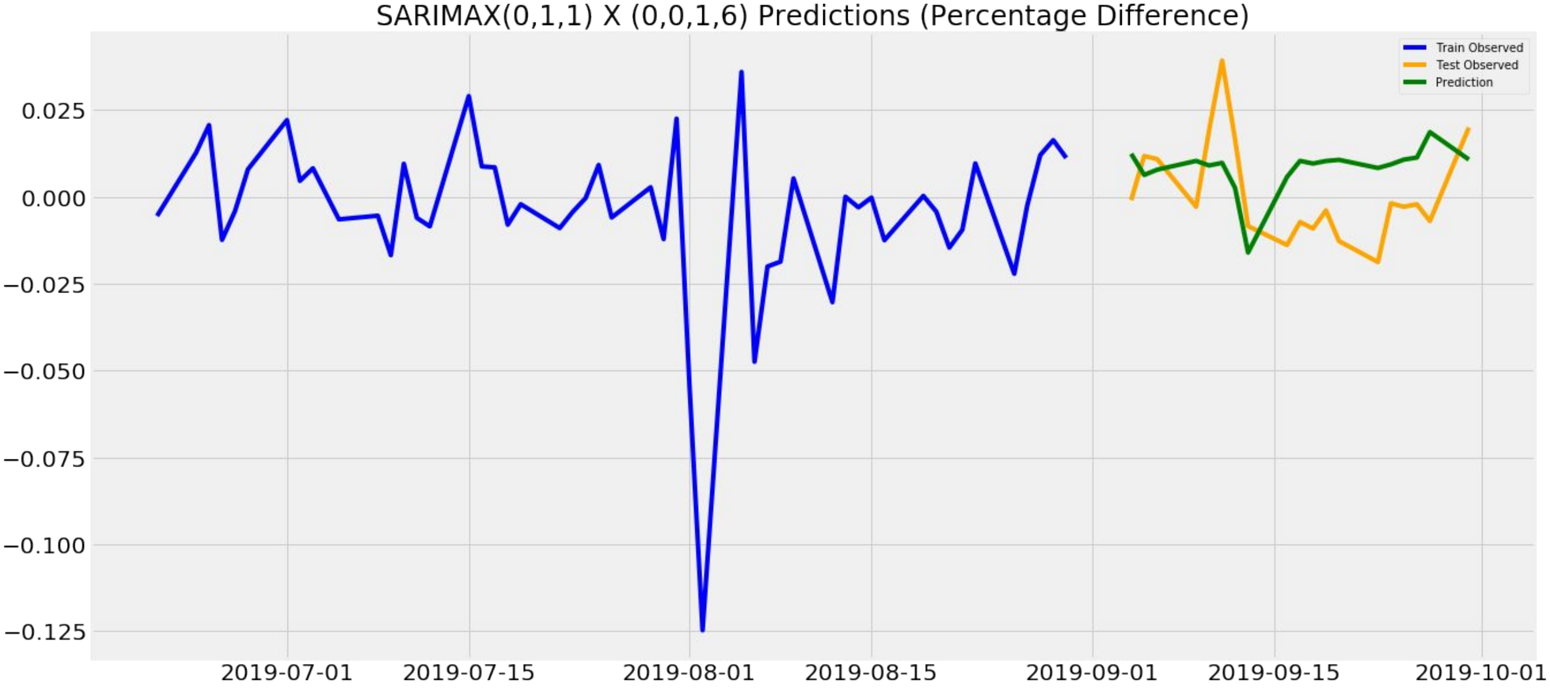
SARIMAX - LESS FEATURES



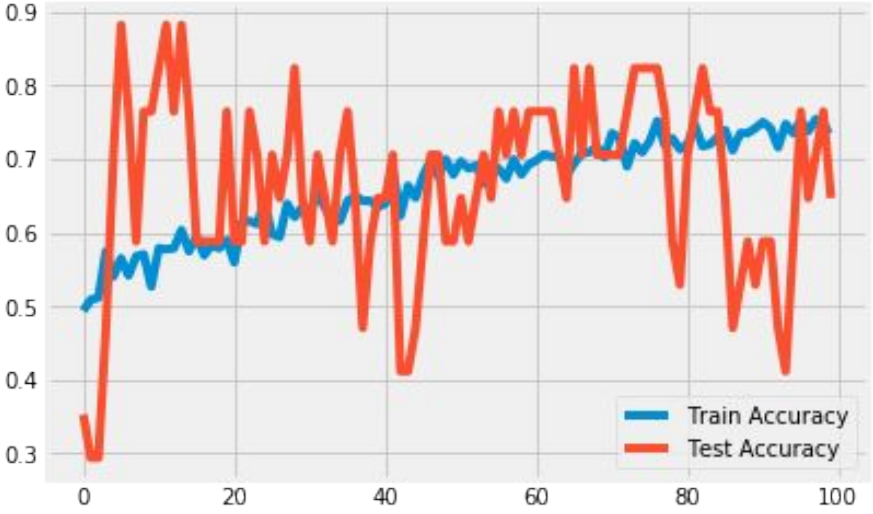
SARIMAX - LESS FEATURES

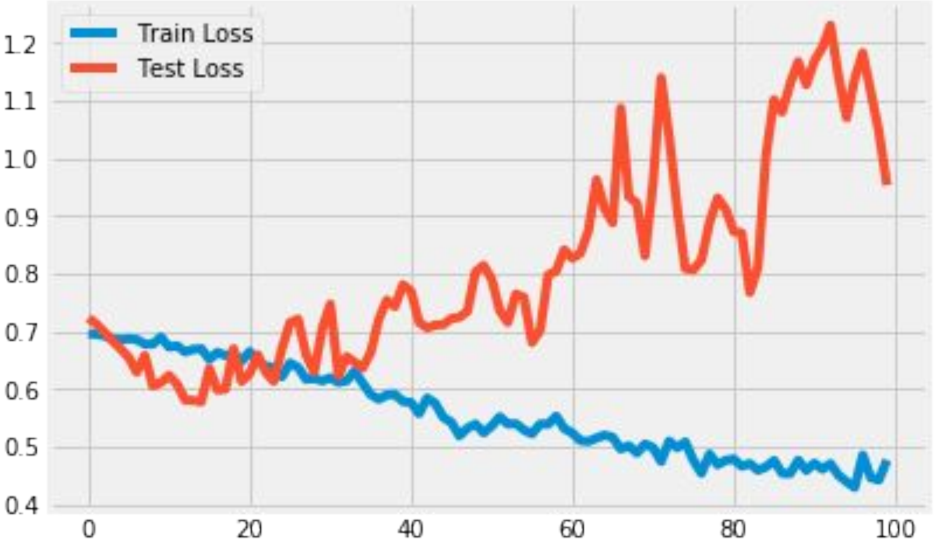


SARIMAX - PERCENTAGE DIFFERENCE



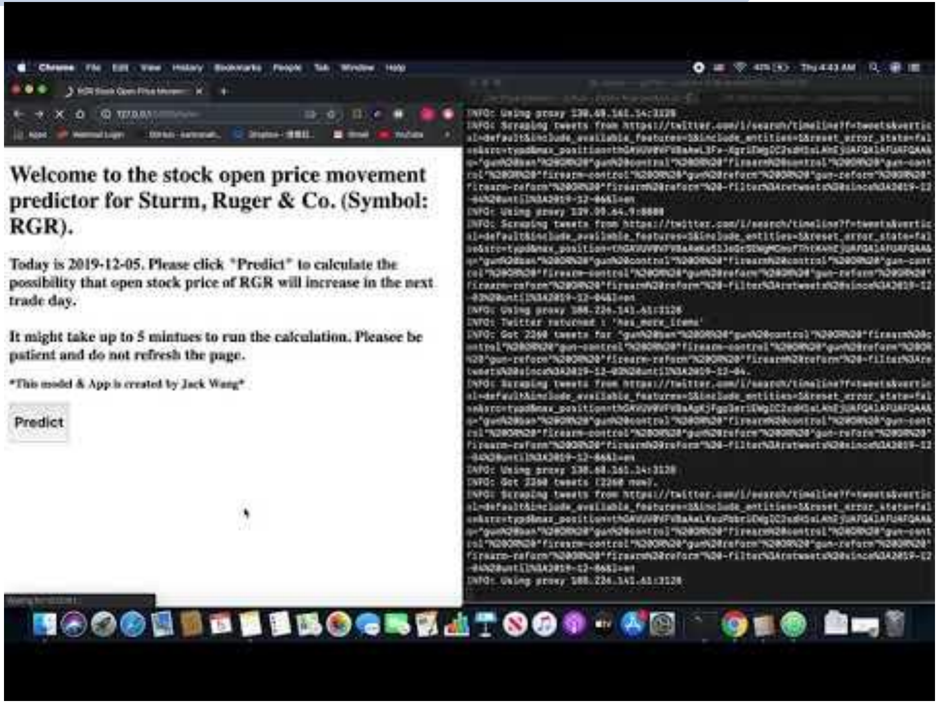
NEURAL NETWORK - RNN CLASSIFIER





FLASK DEMO

FLASK VIDEO DEMO



Welcome to the stock open price movement predictor for Sturm, Ruger & Co. (Symbol: RGR).

Today is 2019-12-05. Please click "Predict" to calculate the possibility that open stock price of RGR will increase in the next trade day.

It might take up to 5 mintues to run the calculation. Pleasee be patient and do not refresh the page.

This model & App is created by Jack Wang

```
INFO: Using proxy 138.68.161.14:3128
INFO: Scraping tweets from https://twitter.com/i/search/timeline?f=tweets&vertic
al=default&include_available_features=1&include_entities=1&reset_error_state=fal
se&src=typd&max_position=[illegible]&
q="gun%20ban"%20OR%20"gun%20control"%20OR%20"firearm%20control"%20OR%20"gun-cont
rol"%20OR%20"firearm-control"%20OR%20"gun%20reform"%20OR%20"gun-reform"%20OR%20"
firearm-reform"%20OR%20"firearm%20reform"%20-filter%3Aretweets%20since%3A2019-12
-04%20until%3A2019-12-06&l=en
INFO: Using proxy 139.99.64.9:8080
INFO: Scraping tweets from https://twitter.com/i/search/timeline?f=tweets&vertic
al=default&include_available_features=1&include_entities=1&reset_error_state=fal
se&src=typd&max_position=[illegible]&
q="gun%20ban"%20OR%20"gun%20control"%20OR%20"firearm%20control"%20OR%20"gun-cont
rol"%20OR%20"firearm-control"%20OR%20"gun%20reform"%20OR%20"gun-reform"%20OR%20"
firearm-reform"%20OR%20"firearm%20reform"%20-filter%3Aretweets%20since%3A2019-12
-03%20until%3A2019-12-04&l=en
INFO: Using proxy 188.226.141.61:3128
INFO: Twitter returned : 'has_more_items'
INFO: Got 2200 tweets for "gun%20ban"%20OR%20"gun%20control"%20OR%20"firearm%20c
ontrol"%20OR%20"gun-control"%20OR%20"firearm-control"%20OR%20"gun%20reform"%20OR
%20"gun-reform"%20OR%20"firearm-reform"%20OR%20"firearm%20reform"%20-filter%3Are
tweets%20since%3A2019-12-03%20until%3A2019-12-04.
INFO: Scraping tweets from https://twitter.com/i/search/timeline?f=tweets&vertic
al=default&include_available_features=1&include_entities=1&reset_error_state=fal
se&src=typd&max_position=[illegible]&
q="gun%20ban"%20OR%20"gun%20control"%20OR%20"firearm%20control"%20OR%20"gun-cont
rol"%20OR%20"firearm-control"%20OR%20"gun%20reform"%20OR%20"gun-reform"%20OR%20"
firearm-reform"%20OR%20"firearm%20reform"%20-filter%3Aretweets%20since%3A2019-12
-04%20until%3A2019-12-06&l=en
INFO: Using proxy 138.68.161.14:3128
INFO: Got 2200 tweets (2200 new).
INFO: Scraping tweets from https://twitter.com/i/search/timeline?f=tweets&vertic
al=default&include_available_features=1&include_entities=1&reset_error_state=fal
se&src=typd&max_position=[illegible]&
q="gun%20ban"%20OR%20"gun%20control"%20OR%20"firearm%20control"%20OR%20"gun-cont
rol"%20OR%20"firearm-control"%20OR%20"gun%20reform"%20OR%20"gun-reform"%20OR%20"
firearm-reform"%20OR%20"firearm%20reform"%20-filter%3Aretweets%20since%3A2019-12
-04%20until%3A2019-12-06&l=en
INFO: Using proxy 188.226.141.61:3128
```

CONCLUSION & IMPROVEMENT

CONCLUSION

Stock prices are hard to predict.

IMPROVEMENT

1. More data may help. I built the model on data from 2016 October - 2019 October because scraping Twitter data took a lot of time. So I would like to see how much the models will improve if we have more data.
2. More useful exogenous variables. I have included 20 features already, but some of them are not helping the model. It seems like the financial reports 10Q, 10k, and 8K don't help much.
3. Better models. I did not dig into the popular neural network models due to time limitation. I will definitely check out [LSTM] and some other models.

REFERENCE

REFERENCE

- Yahoo Finance
- SEC
- twitter
- twitterscraper
- A Guide to Time Series Forecasting with ARIMA in Python 3
- How to Scrap Reddit using pushshift.io via Python
- List of mass shootings in the United States
- Bug in ARIMA predict(): ValueError: Must provide freq argument if no data is supplied #3534
- Using AIC to Test ARIMA Models

QUESTIONS