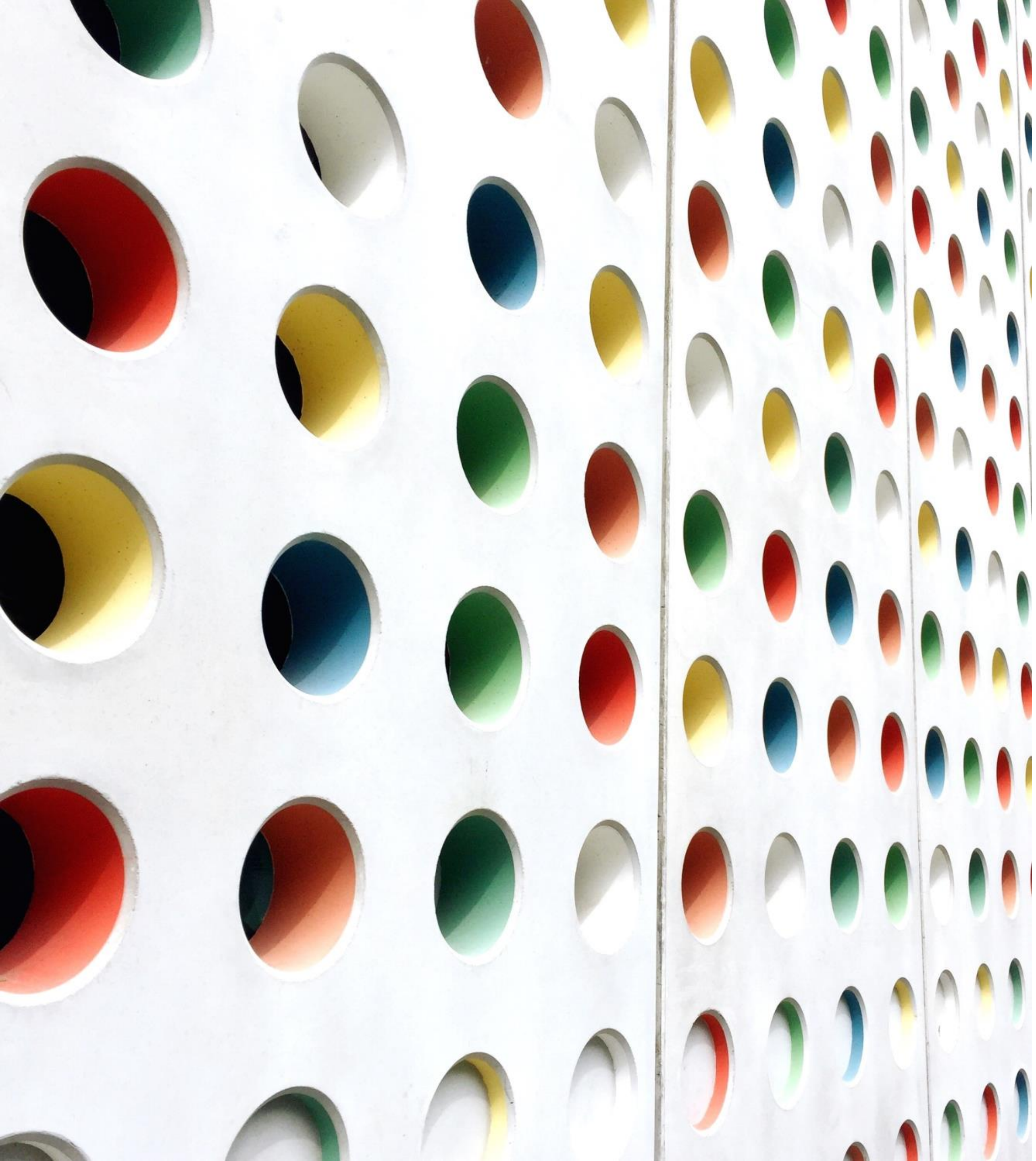


How News affects stock prices

BY YOHANNES ABEBE

DATA SCIENCE: CAPSTONE PROJECT

Content

- ❑Introduction
- ❑Data Collection
- ❑EDA
- ❑Modeling
- ❑Conclusion

Datasets Collection

- Problem Statement

  To learn specific news, contents, and posts that affect a company's stock prices and create a model to predict the closing price.

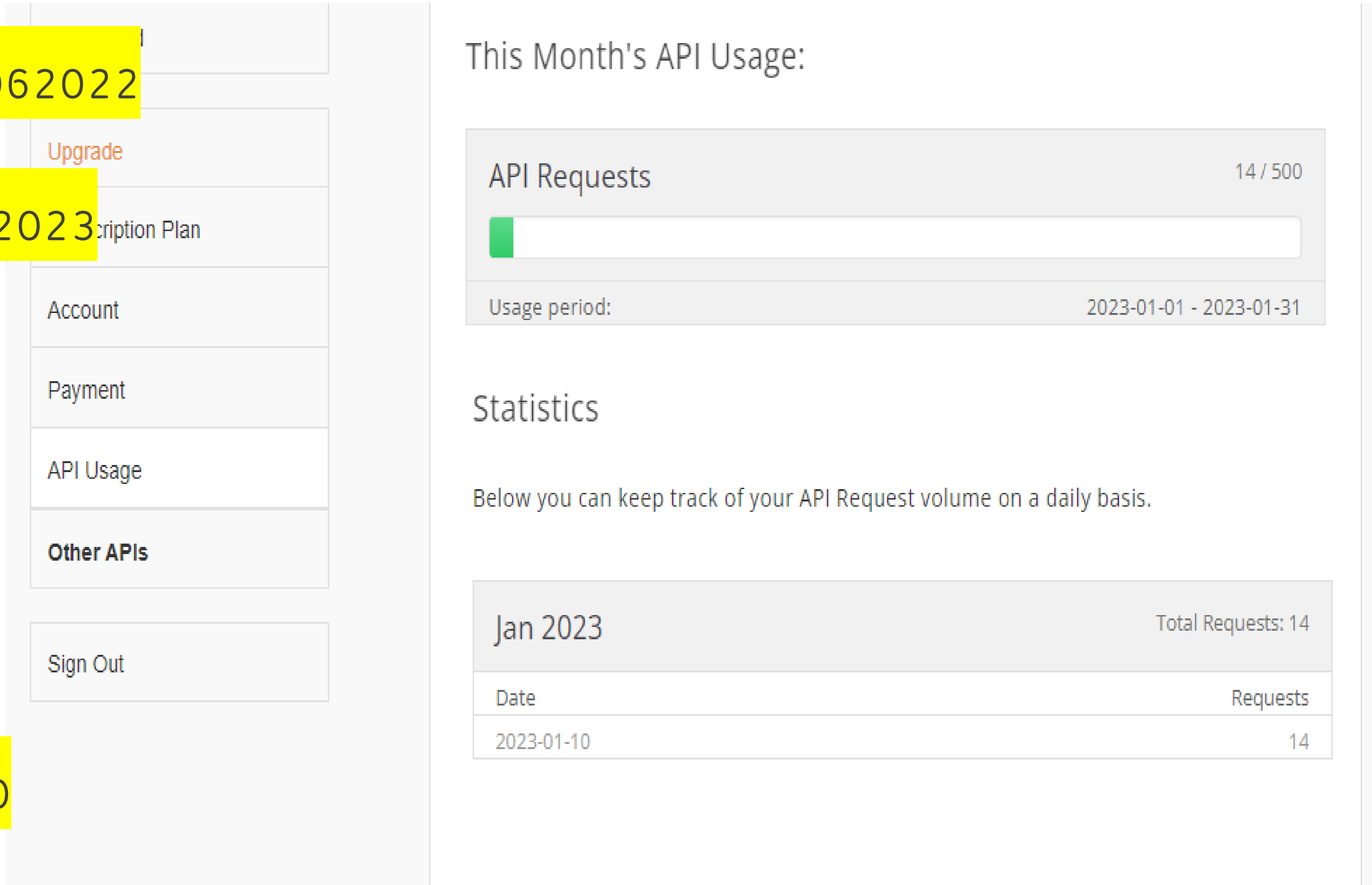
- Why: Interested in modeling as a result of my past experience.

- Future Development: Create a framework for future development of additional features.

Datasets Collection

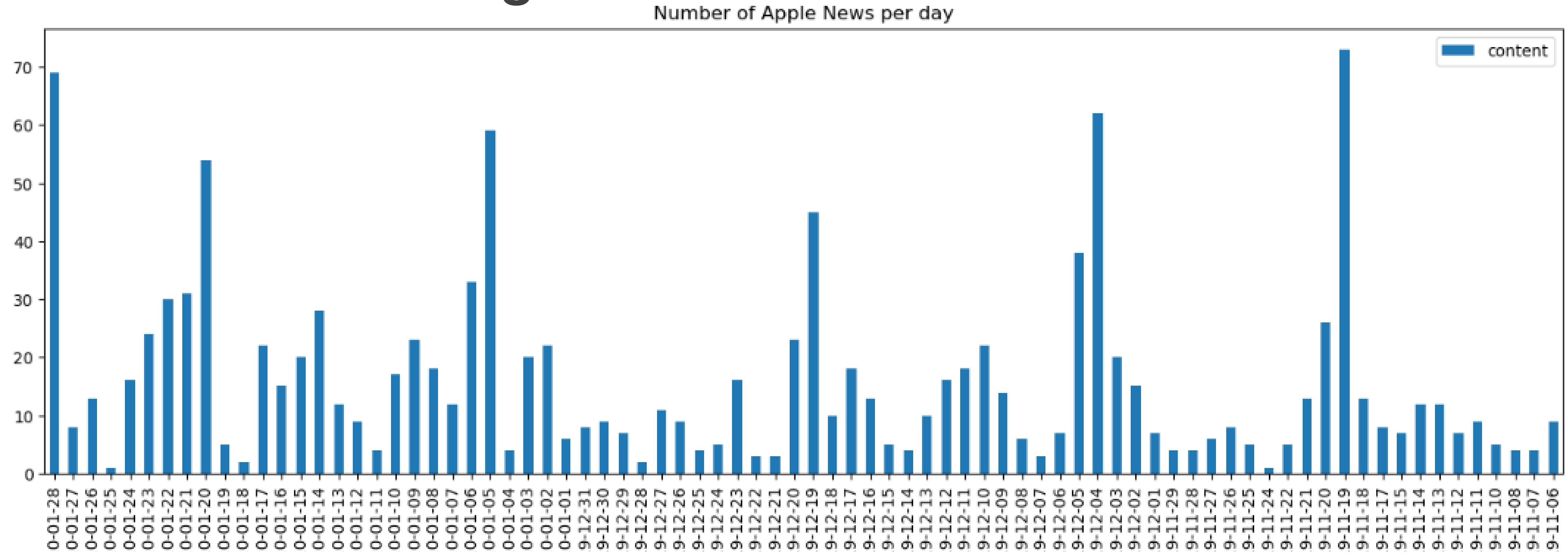
Financial data, News, Posts providers

- Tweeter – 3 Stocks from 012021- 062022
- Finnhub – Stock news 082022 - 012023
- GoogleNews
- Yahoo finance
- CNBC
- Kaggle – 221513 072012 - 012020



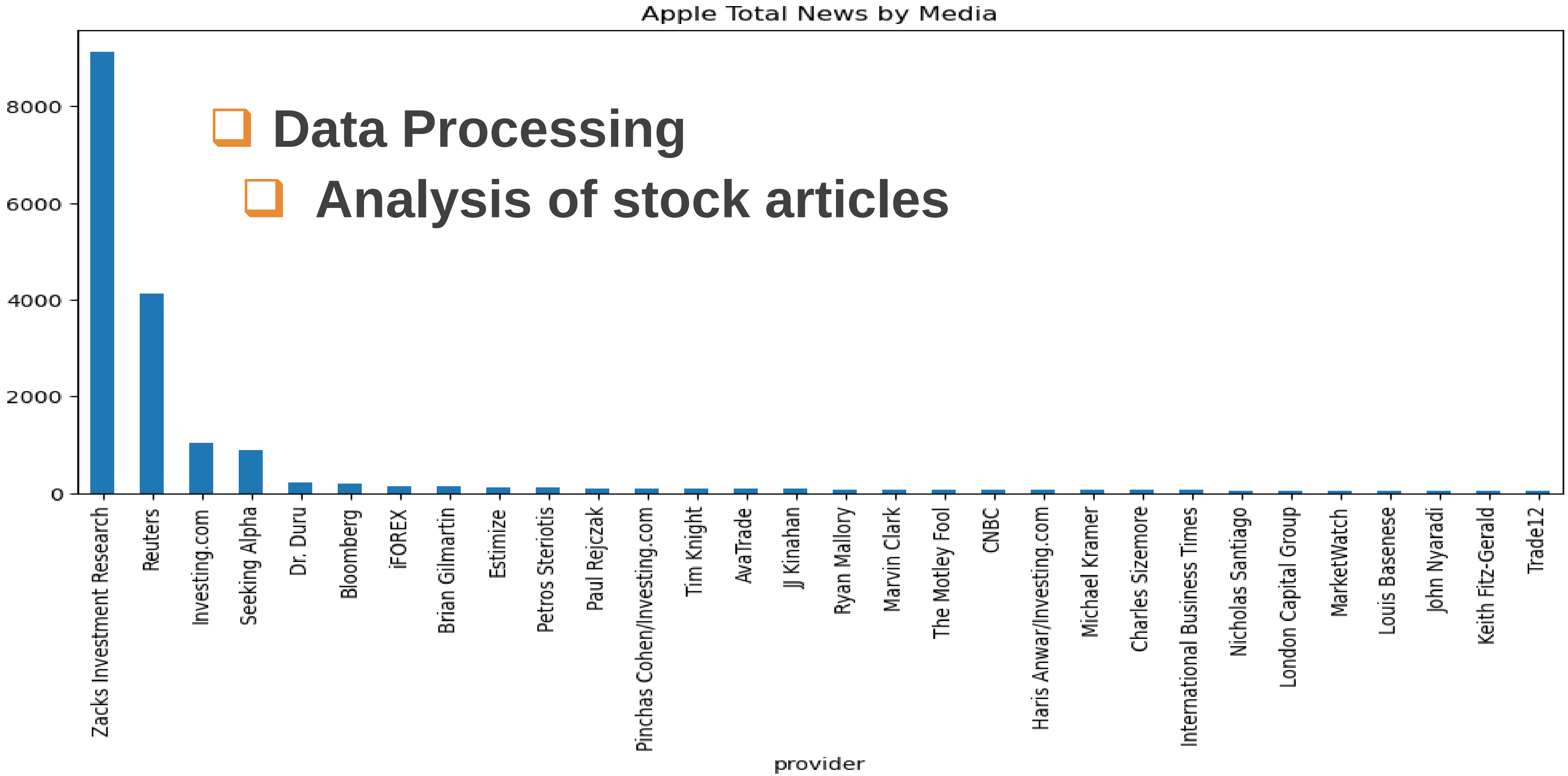
Explanatory Data Analysis

❑ Data Processing



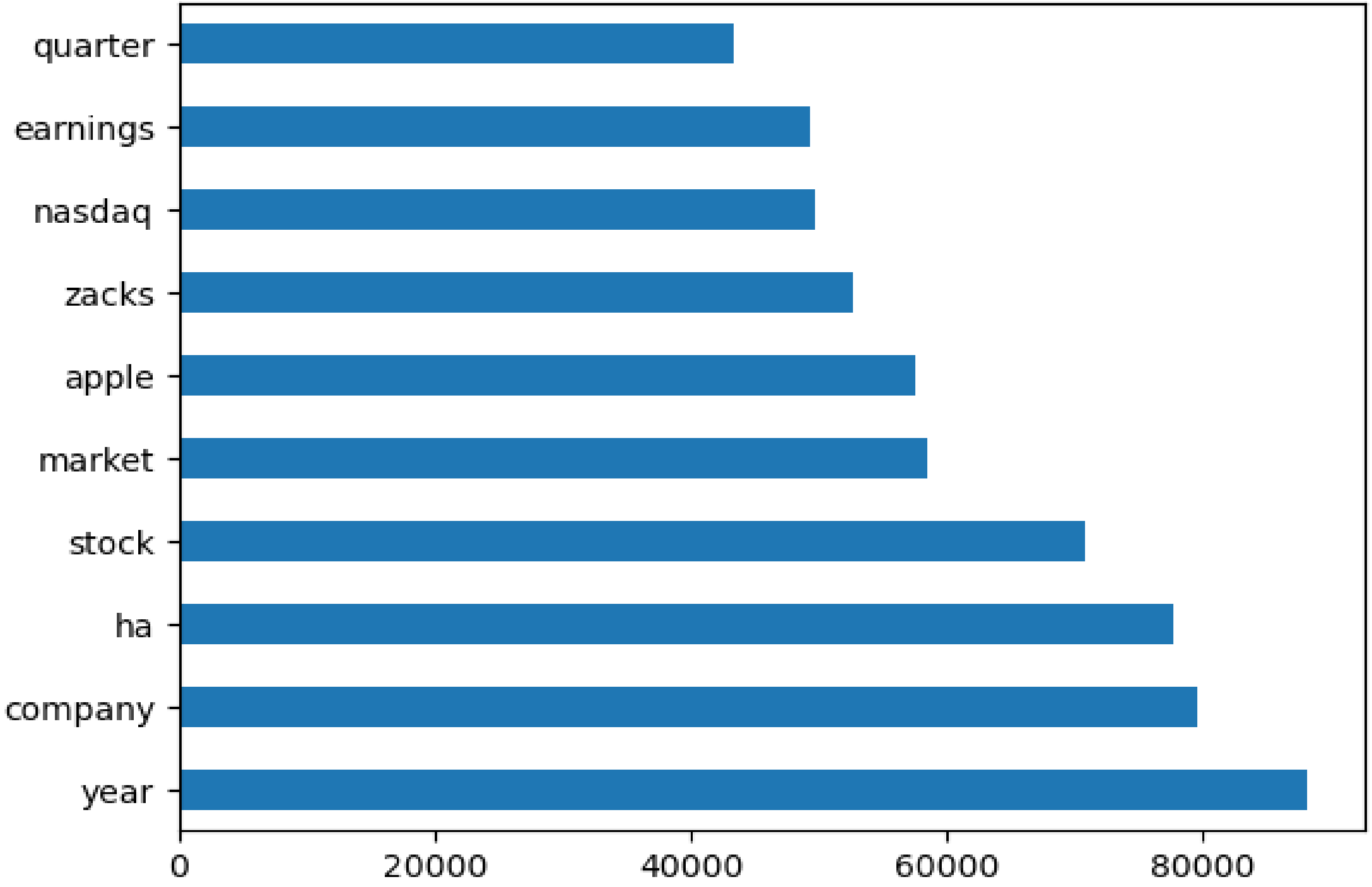
Explanatory Data Analysis

- Data Processing
  - Analysis of stock articles



Explanatory Data Analysis

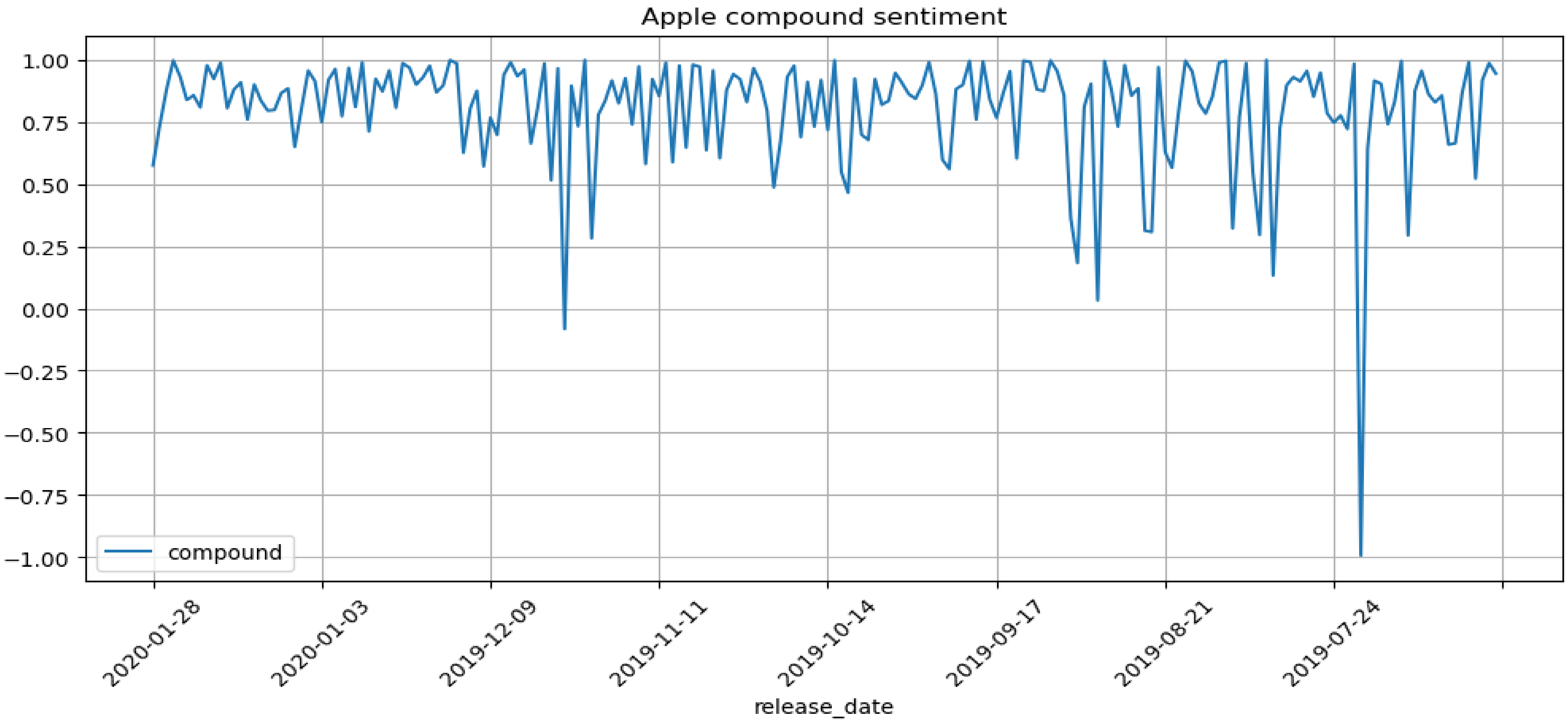
- **Data Processing**
  - **NLP**
    - **tokenizer & Lemmatize**



Explanatory Data Analysis

Sentiment score

	compound
release_date	
2019-11-27	-0.082350
2019-07-20	-0.993300
2019-06-02	-0.026750
2019-03-30	-0.691500
2019-03-23	-0.690800
...	...
2012-09-25	-0.817600
2012-09-12	-0.492333
2012-09-06	-0.980300
2012-08-10	-0.539900
2012-07-30	-0.263200



Explanatory Data Analysis

Sentiment

score

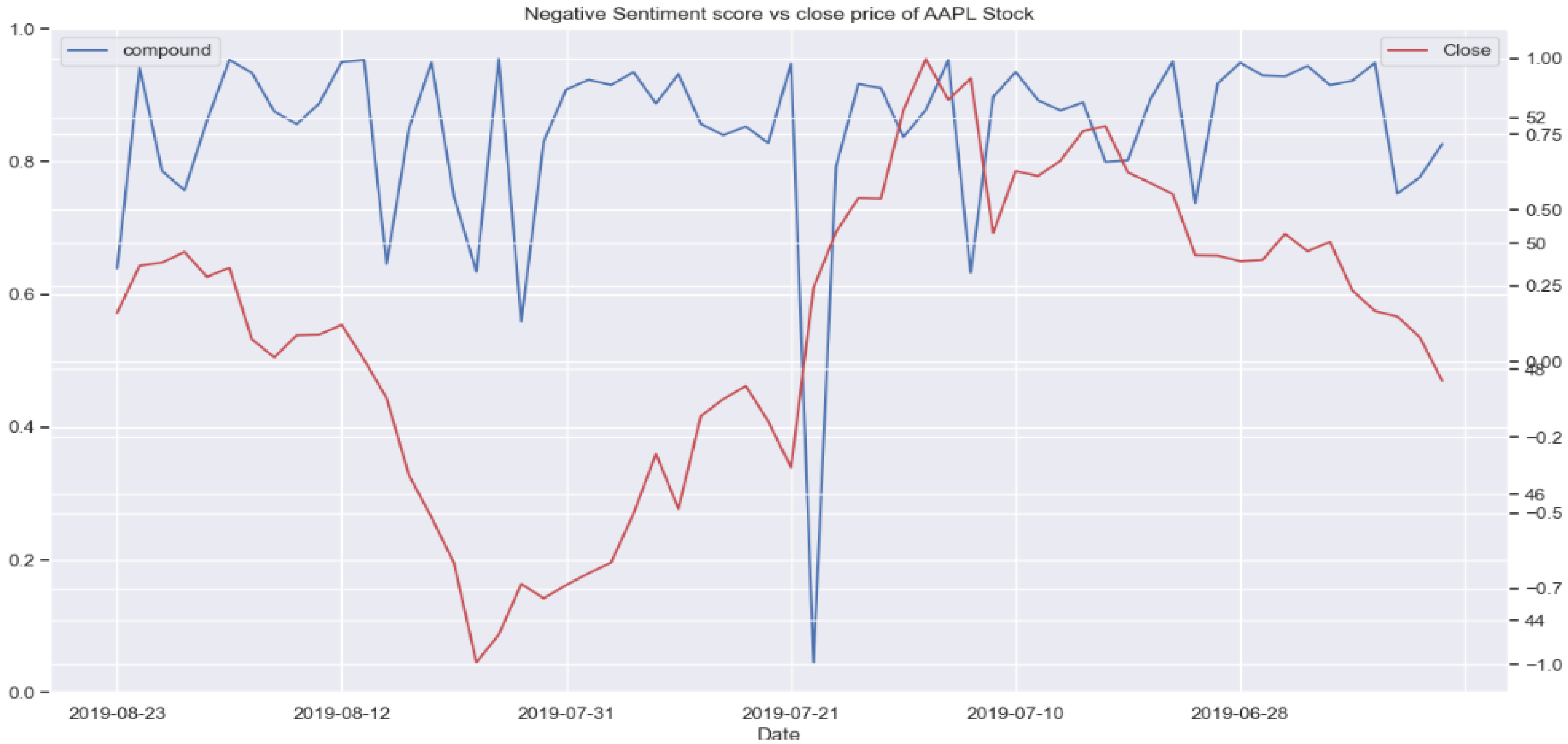
release_date	compound
2019-11-27	-0.082350
2019-07-20	-0.993300
2019-06-02	-0.026750
2019-03-30	-0.691500
2019-03-23	-0.690800
...	...
2012-09-25	-0.817600
2012-09-12	-0.492333
2012-09-06	-0.980300
2012-08-10	-0.539900
2012-07-30	-0.263200

```
'tim kelly sam nussey kyoto japan reuters many victim arson attack japanese animation studio young bright future joining april shaken company presic
ent said saturday death toll climbed 34 thursday attack kyoto animation well known television series movie wa japan worst mass killing two decade wa
poignant youth victim country population among world oldest many victim attack ancient capital kyoto young woman company president hideaki hatta sai
d joined u april eighth july gave small first bonus said people promising future lost life know say rather feeling anger word hatta said fifteen vic
tim 20 11 30 public broadcaster nhk said six 40 one wa least 60 age latest victim man died hospital wa known name victim disclosed yet wait definiti
ve identification police confirmed identity suspect shinji aoba ex convict shop robbery formally arrested treated heavy burn unusual police release
suspect name making arrest plagiarism claim aoba life modest two floor apartment building 500 km 310 mile kyoto rural suburb outside omiya commuter
hub north tokyo 27 year old neighbor said aoba grabbed yelled noise dispute started yelling face shut grabbed collar started pulling hair wa terrify
ing neighbor declined identified said aoba suspected carrying arson attack believed novel plagiarized local medium say hatta said idea plagiarism cl
aim adding seen correspondence suspect people living near studio said saw man fitting aoba description park day attack police suspect may spent day
area prepare resident rie bannai said nephew saw man sleeping bench park rui yamaguchi work factory nearby said saw suspect wa detained looked like
mannequin hair blackened one pant leg gone around calf looked like burnt said beautiful warm near blackened studio building smell burning timber lir
gered neighborhood animation fan many nationality local resident queued add growing pile flower drink offering bing xie 25 chinese student kyoto uni
versity said could forgive arsonist criminal doe seems mentally disturbed forgive young people kyoto animation beautiful warm hard accept gone polic
e guarded site investigator roof near many died connecting stairwell examined torched three story building hatta said building needed torn wa badly
damaged tribute victim lit social medium world leader apple inc aapl chief executive offering condolence hashtag prayforkyoani studio known among fa
n ha become popular kyoto animation produce popular anime series sound euphonium also known violet evergarden ha shown netflix nasdaq nflx fumiko sh
imizu 91 life close building watched fire balcony home said regularly greeted staff met street used see going convenience store group six buy lunch
box lot young people twenty thirty died young would alright somebody like die anytime said'
```

Explanatory Data Analysis

Sentiment score

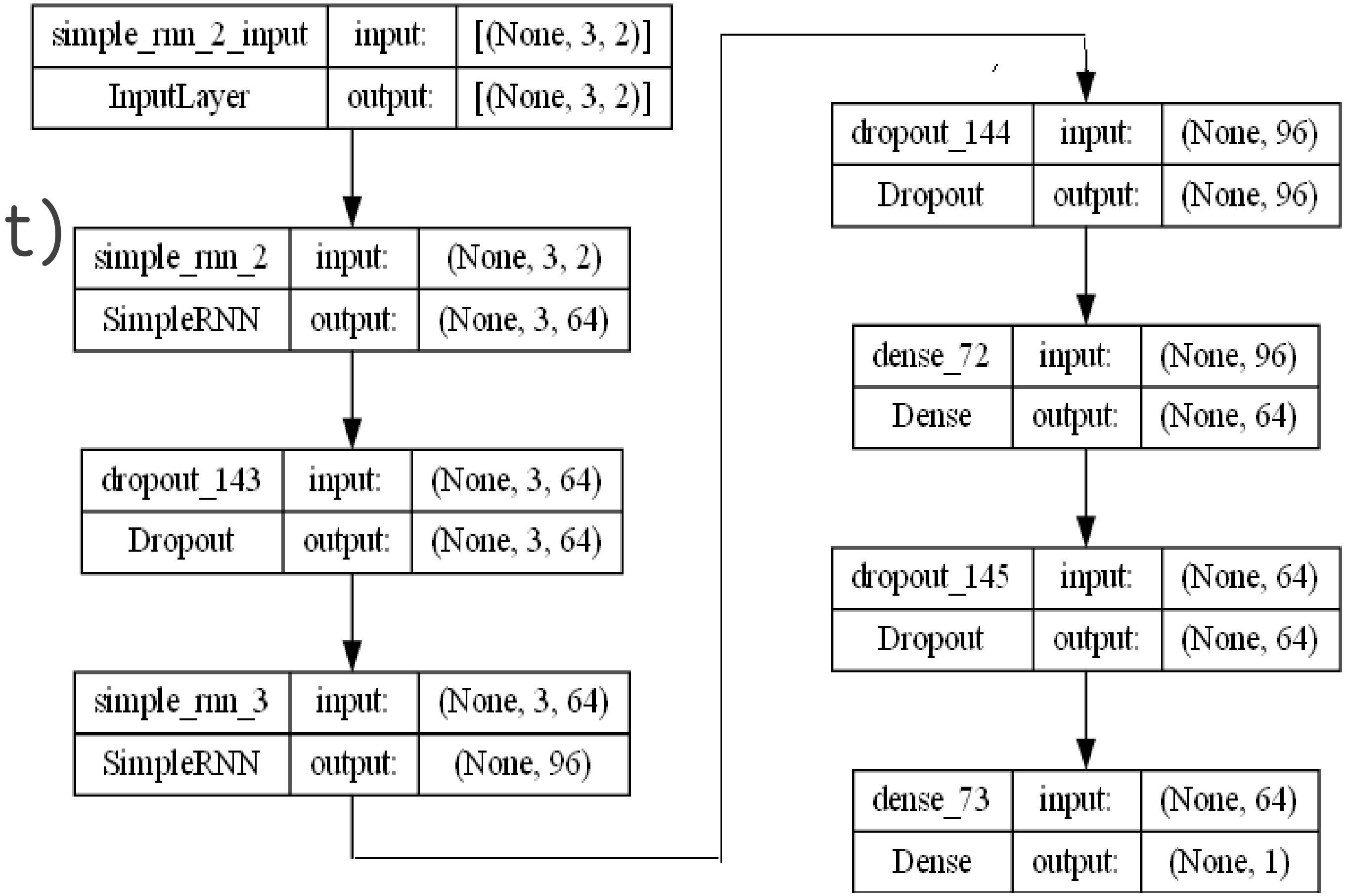
	compound
release_date	
2019-11-27	-0.082350
2019-07-20	-0.993300
2019-06-02	-0.026750
2019-03-30	-0.691500
2019-03-23	-0.690800
...	...
2012-09-25	-0.817600
2012-09-12	-0.492333
2012-09-06	-0.980300
2012-08-10	-0.539900
2012-07-30	-0.263200



Build Models

Model Used

- Linear Regression
- GRU (Gated Recurrent Unit)
- RNN (Recursive Neural Network)
- LSTM (Long Short-Term Memory)

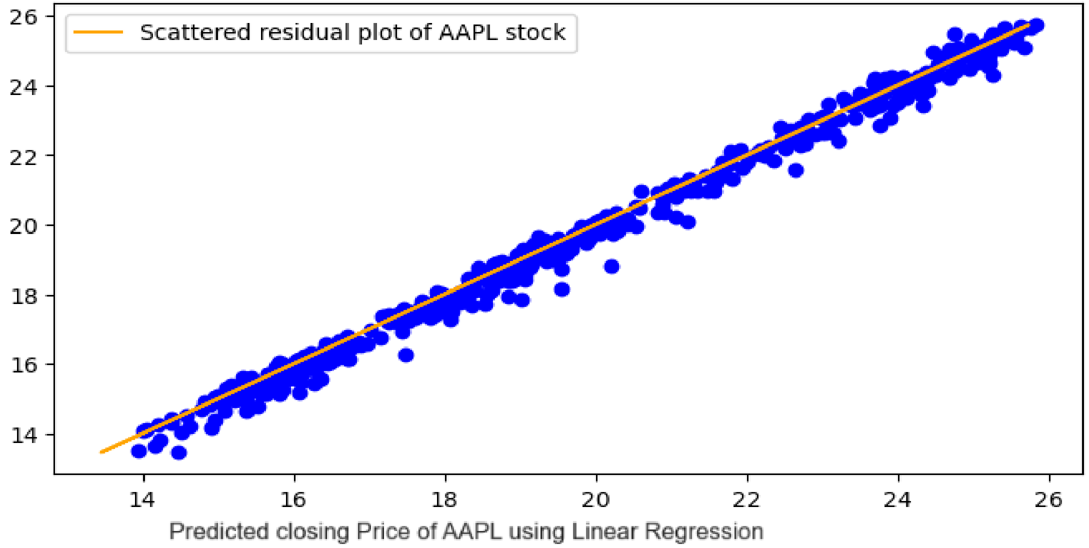


Build Models …

Linear Regression

Linear Regression:	
Train score	0.998467
Test score	0.990316
rmse	0.31766

Linear Regression	
Actual	Predicted
25.74749947	25.62897801
25.61750031	25.60230001
24.94000053	24.46949041
24.41749954	24.28663658
24.06500053	23.7838852
24.38500023	24.30603232
24.6875	25.02405547
24.95249939	25.29328952
25.18250084	25.10461924
25.25499916	25.27499134

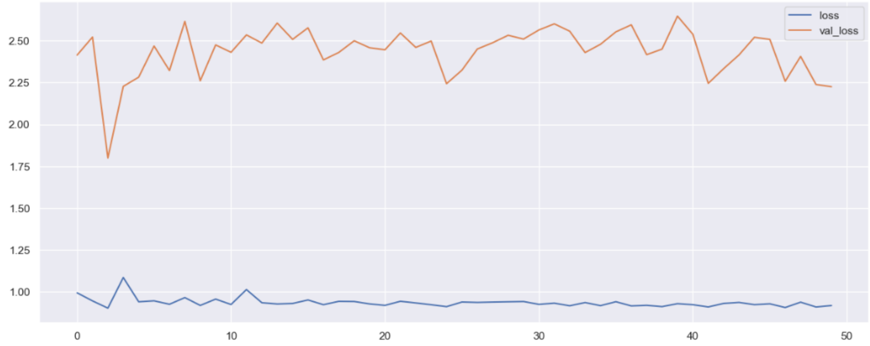


Build Models …

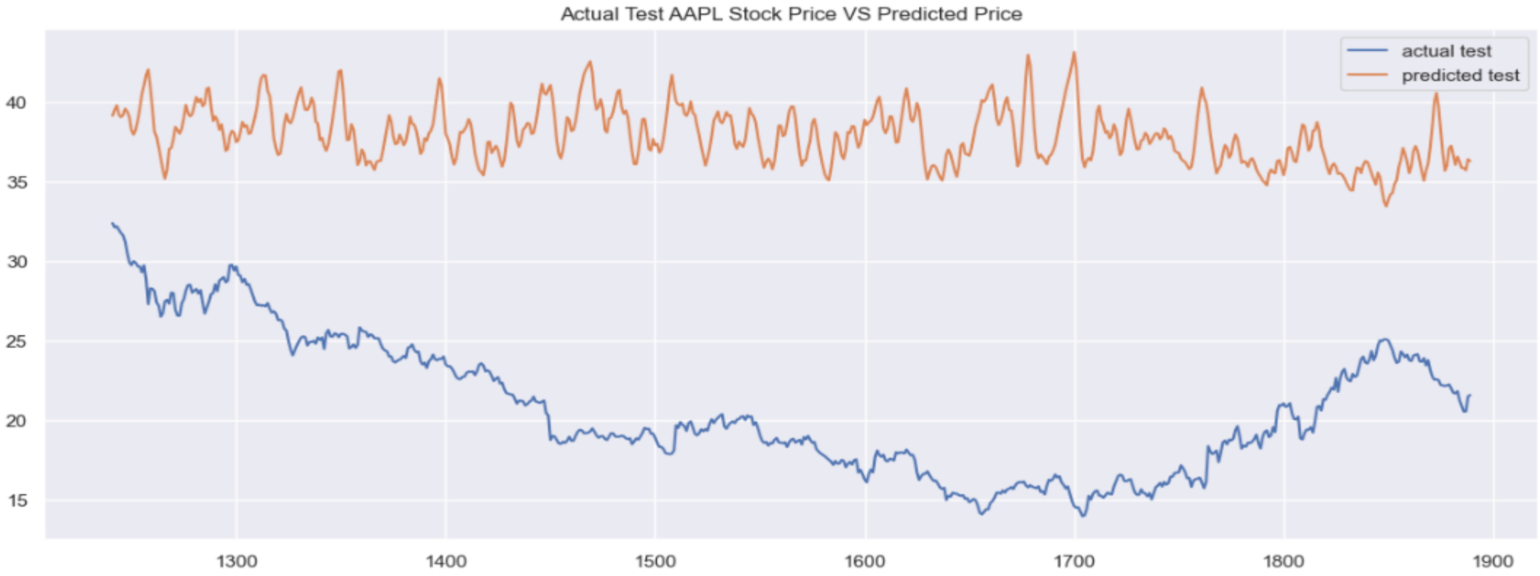
GRU with tweets sentiment only

Parameters used:	
Length	[3,11,50]
batch	[16,32, 64]
ecpoch	[50,100, 200]

Best parameters	
optimizer	rmsprop
Test score	2.2254
length	11
epochs	50



	actual test	predicted test
1241	32.375000	39.141136
1242	32.112499	39.511097
1243	32.180000	39.787495
1244	31.957500	39.212704
1245	31.770000	39.043598

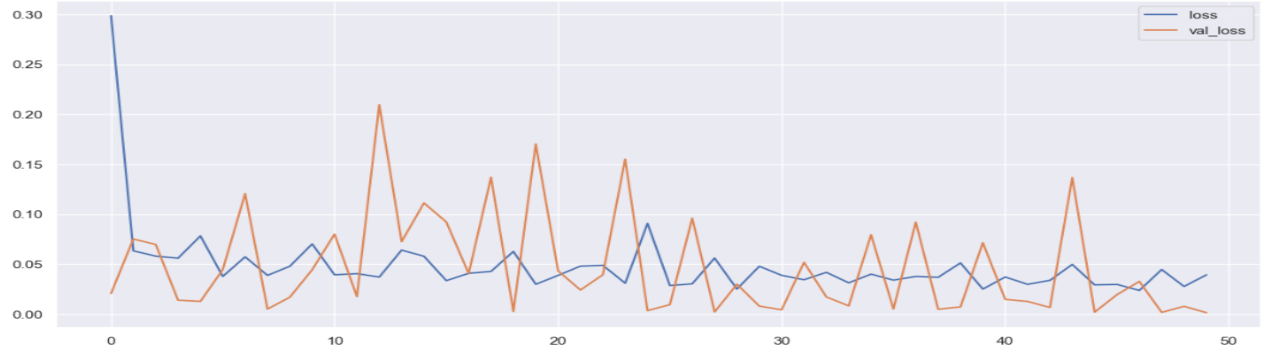


Build Models …

GRU with tweets/open price of stock

Parameters used:	
Length	[3,11,50]
batch	[16,32, 64]
ecpoch	[50,100, 200]

Best parameters	
optimizer	rmsprop
Test score	0.00156
length	11
epochs	50



	actual test	predicted test
1241	32.375000	32.486126
1242	32.112499	32.396526
1243	32.180000	32.255733
1244	31.957500	32.112583
1245	31.770000	32.006588

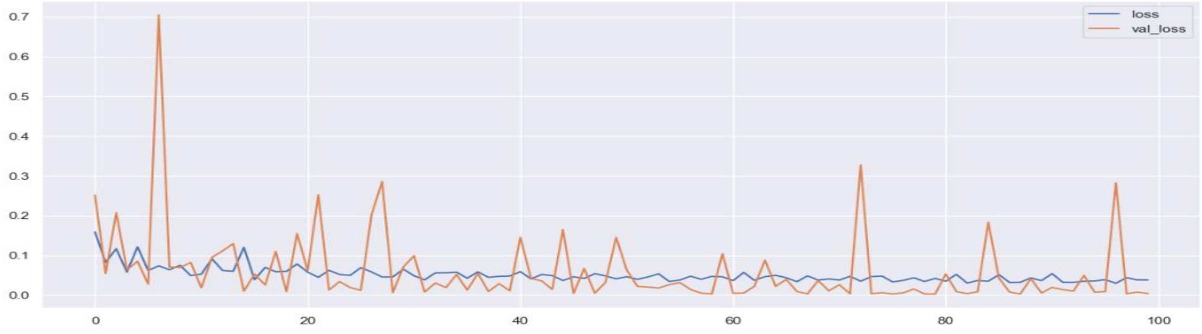
Actual Test AAPL Stock Price VS Predicted Price

Build Models …

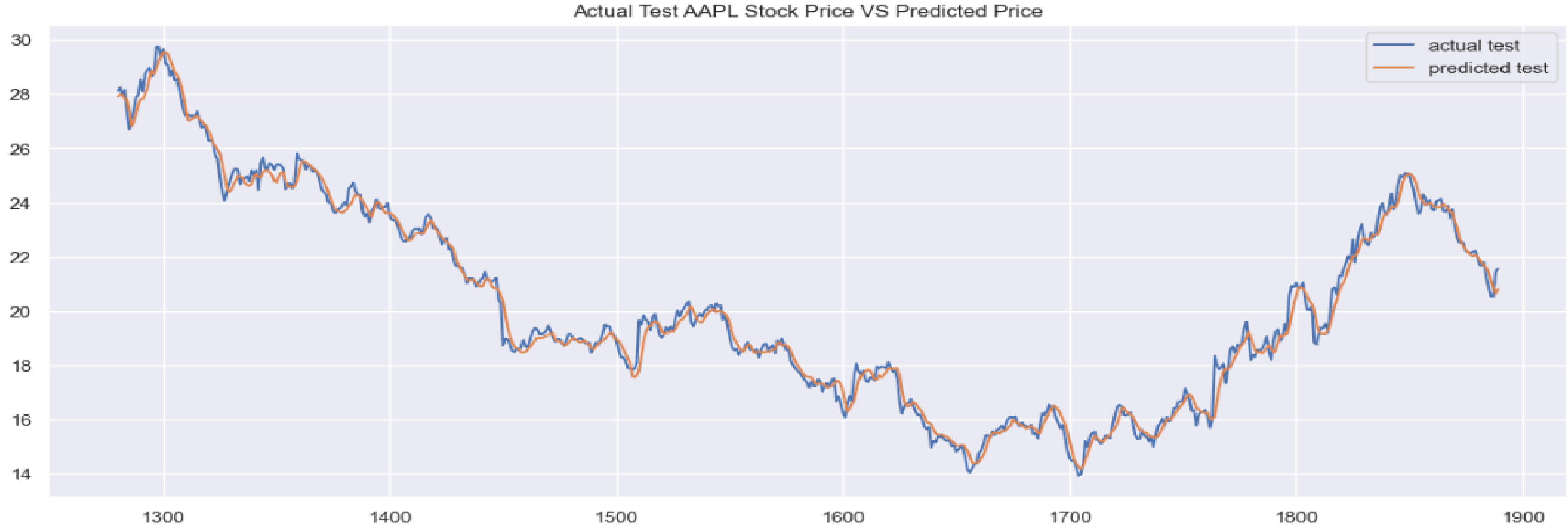
LSTM with tweets/open price of stock

Parameters used:	
Length	[3,11,50]
batch	[16,32, 64]
ecpoch	[50,100, 200]

Best parameters	
optimizer	rmsprop
Test score	0.003704
length	11
epochs	100



	actual test	predicted test
1241	32.375000	32.621872
1242	32.112499	32.533440
1243	32.180000	32.404701
1244	31.957500	32.267345
1245	31.770000	32.122486

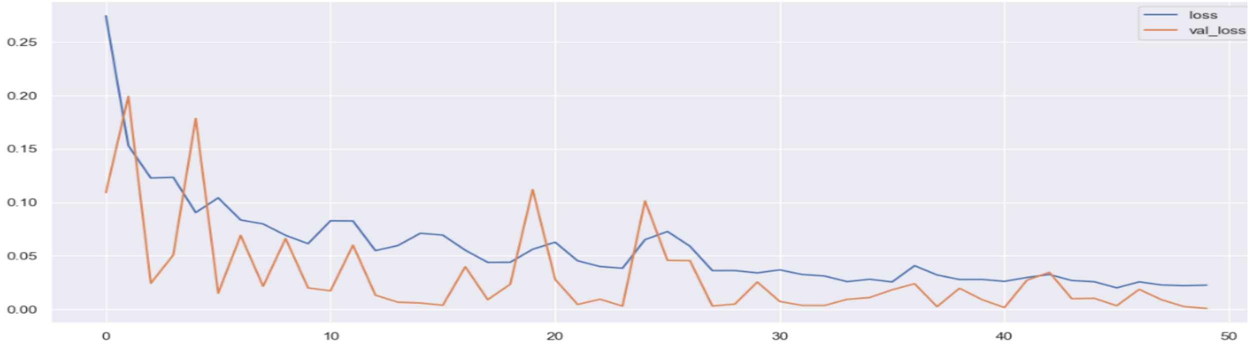


Build Models …

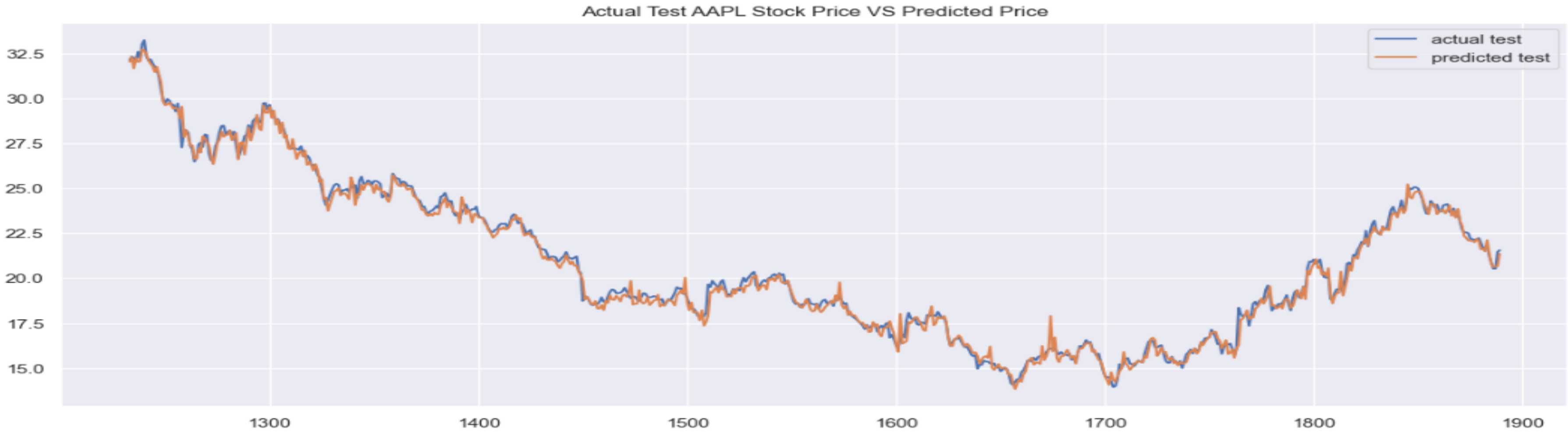
RNN with tweets/open price of stock

Parameters used:	
Length	[3,11,50]
batch	[16,32, 64]
ecpoch	[50,100, 200]

sentiment score & stock opening price	
RNN	
optimizer	adam
Test score	0.0008744
length	11
epochs	200



	actual test	predicted test
1233	32.134998	32.050903
1234	32.340000	32.329674
1235	32.272499	31.681116
1236	32.115002	32.260738
1237	32.605000	32.038532



Build Models …

GRU with tweets/open price of stock - Loss /val-loss

```
Train Error score: loss: 0.006313977763056755
Test Error score: acc: 0.005312368739396334
20/20 [==============================] - 1s 8
11/11 [==============================] - 0s 8
Train Error score: loss: 0.029037179425358772
Test Error score: acc: 0.08606790006160736
Train Error score: loss: 0.003029243787750601
Test Error score: acc: 0.00957771297544241
Train Error score: loss: 0.02036929503083229
Test Error score: acc: 0.013182457536458969
Train Error score: loss: 0.006194890011101961
Test Error score: acc: 0.00370423449203372
20/20 [==============================] - 1s 1
11/11 [==============================] - 0s 1
Train Error score: loss: 0.0317319892346859
Test Error score: acc: 0.10949279367923737
Train Error score: loss: 0.005554272327572107
Test Error score: acc: 0.010423670522868633
Train Error score: loss: 0.04360645264387131
Test Error score: acc: 0.05949308723211288 5
Train Error score: loss: 0.012417167425155 64
Test Error score: acc: 0.026213442906737328
========================
Best test score : 0.00370423449203372
<IPython.core.display.Image object>
```

Conclusion

- RNN model predicted better than others
- Very difficult to predict with only sentiment score.
- Attempt to execute all models with combination of news, tweets, and other social media.

Just stock opening price	
optimizer	rmsprop
Test score	0.00139811
length	50
epochs	50

sentiment score & stock opening price	
LSTM	
optimizer	rmsprop
Test score	0.0037043
length	11
epochs	100

sentiment score & stock opening price	
GRU	
optimizer	rmsprop
Test score	0.000984
length	50
epochs	50

sentiment score & stock opening price	
RNN	
optimizer	adam
Test score	0.0008744
length	11
epochs	200

sentiment score only	
GRU	
optimizer	rmsprop
Test score	2.22546
length	11
epochs	50

sentiment score & stock opening price	
RNN	
optimizer	rmsprop
Test score	0.000898
length	11
epochs	200

Questions?

Suggestion