Twitter Sentiment Analysis

EXECUTIVE SUMMARY

Why on earth did we do this?

Both ATVI and CDPR offer unique case-studies for developing a model to determine trading strategies.

- Both companies have an historically involved following on social media, especially Twitter.
- Both companies have a particularly rabid fanbase.
- Both companies have recently experienced high-profile PR nightmares with ensuing Twitter and social media backlash.

Project Concept

- Create a machine learning model that determines the public twitter sentiment towards companies such as Activision Blizzard (ATVI) after their issue regarding their misogynistic workplace and CD Projekt Red (CDPR) with their Cyberpunk 2077 game flopping in the market for releasing an incomplete game.

- We looked into a number of tweets to check how it affected the stock prices of ATVI, CDPR and NASDAQ.

PROJECT ANALYSIS

- Machine Learning Models
  - Base - Sequential model
  - RNN - Bidirectional LSTM (BiLSTM)
  - RNN 2 - Bidirectional LSTM (BiLSTM)

Models

Model Summaries


```
# View summary
model_basic.summary()
```

```
Model: "sequential_4"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding (Embedding)        (None, None, 16)          160000
_________________________________________________________________
global_average_pooling1d_4 ( (None, 16)                0
_________________________________________________________________
dense_8 (Dense)              (None, 16)                272
_________________________________________________________________
dense_9 (Dense)              (None, 1)                 17
=================================================================
Total params: 160,289
Trainable params: 160,289
Non-trainable params: 0
_________________________________________________________________
```

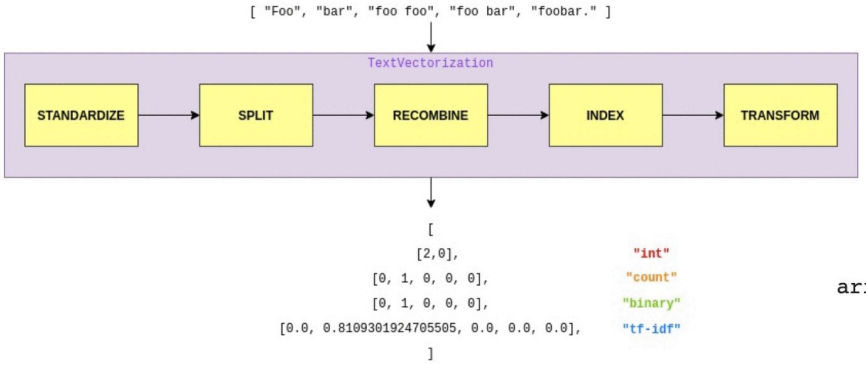
```
# View summary
model_rnn.summary()
```

```
Model: "sequential_5"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding (Embedding)        (None, None, 64)          640000
_________________________________________________________________
bidirectional (Bidirectional (None, 128)               66048
_________________________________________________________________
dense_10 (Dense)             (None, 64)                8256
_________________________________________________________________
dense_11 (Dense)             (None, 1)                 65
=================================================================
Total params: 714,369
Trainable params: 714,369
Non-trainable params: 0
_________________________________________________________________
```

```
# View summary
model_rnn_2_layer.summary()
```

```
Model: "sequential_6"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding (Embedding)        (None, None, 64)          640000
_________________________________________________________________
bidirectional_1 (Bidirection (None, None, 128)         66048
_________________________________________________________________
bidirectional_2 (Bidirection (None, 64)                41216
_________________________________________________________________
dense_12 (Dense)             (None, 64)                4160
_________________________________________________________________
dropout (Dropout)            (None, 64)                0
_________________________________________________________________
dense_13 (Dense)             (None, 1)                 65
=================================================================
Total params: 751,489
Trainable params: 751,489
Non-trainable params: 0
_________________________________________________________________
```

Text Vectorization

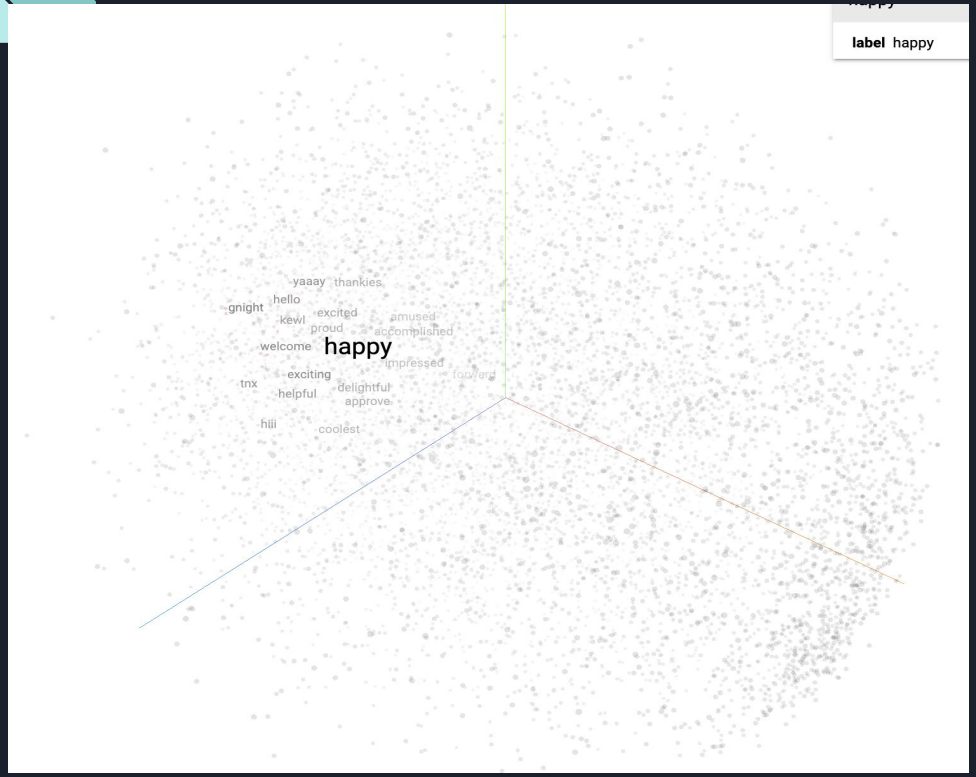



```
# This would give you.
Vocabulary:     [b'foo', b'bar', b'foobar']
Prediction data:    ['foo', 'bar', 'foobar', 'foo foo', 'OOV']
Encoded prediction data:
[[2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
 [3, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
 [4, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
 [2, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
 [1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]]
```

```
array(['', '[UNK]', 'i', 'to', 'the', 'a', 'my', 'and', 'you', 'is', 'it',
       'for', 'in', 'of', 'im', 'on', 'me', 'so', 'have', 'that', 'but',
       'just', 'with', 'be', 'at', 'its', 'not', 'was', 'this', 'now',
       'good', 'up', 'day', 'get', 'all', 'out', 'are', 'like', 'no',
       'go', 'dont', 'do', 'your', 'today', 'too', 'going', 'love',
       'work', 'cant', 'got', 'time', 'from', 'back', 'lol', 'u', 'one',
       'what', 'will', 'know', 'we', 'about', 'really', 'am', 'had',
       'can', 'see', 'some', 'well', 'if', 'still', 'new', 'want',
       'night', 'think', 'how', 'amp', 'thanks', 'home', 'as', 'when',
       'there', 'oh', '2', 'more', 'much', 'off', 'miss', 'here', 'need',
       'they', 'last', 'an', 'then', 'been', 'hope', 'morning', 'great',
       'has', 'tomorrow', 'ill'], dtype='<U27')
```

```
Original:       [b"@NiGhT_RaVeN13  ANO! but i don't have any money atm "]
Round-trip:     [UNK] ano but i dont have any money atm

Original:       [b'@jamespcho one of my favorite shows ']
Round-trip:     [UNK] one of my favorite shows

Original:       [b'@AnnieD1 do u no why nk have cancelled ausie ? carley is well pissed ']
Round-trip:     [UNK] do u no why nk have cancelled [UNK] [UNK] is well pissed

Original:       [b'@DJ_AM but then again a lot of your 22k+ followers prolly already told you ']
Round-trip:     [UNK] but then again a lot of your [UNK] followers prolly already told you

Original:       [b'want some cereal hardly any milk ']
Round-trip:     want some cereal hardly any milk
```

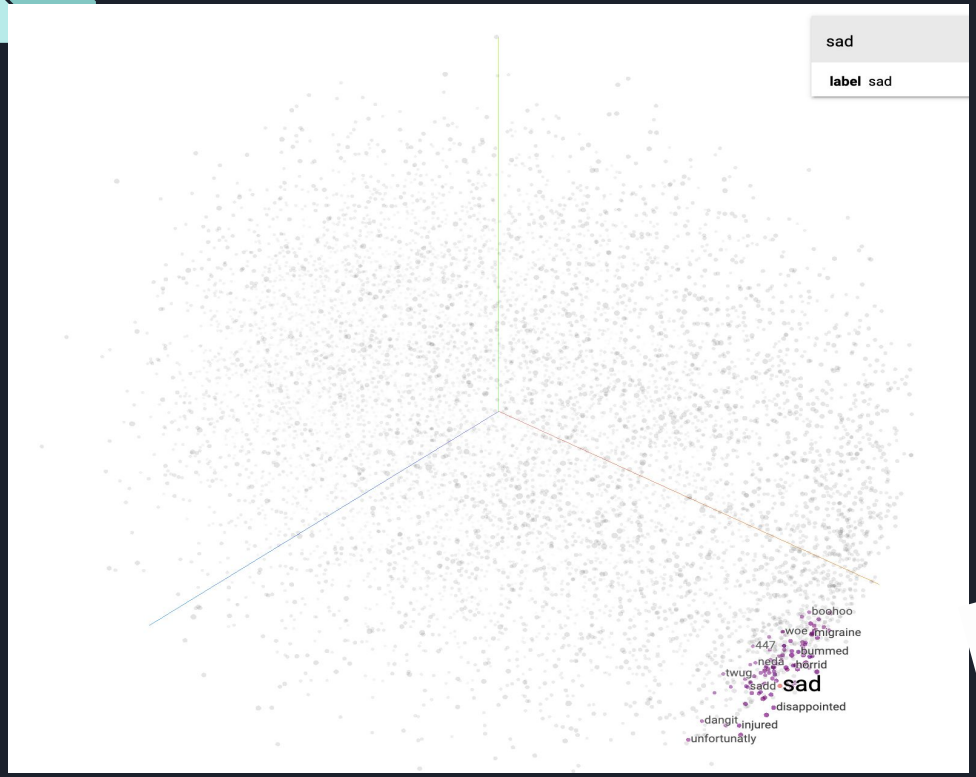
Model Embeddings - Happy



Nearest points in the original space:

excited	0.103
proud	0.123
loving	0.124
exciting	0.129
optimistic	0.131
grateful	0.135
pleased	0.136
cool	0.137
accomplished	0.138
heheh	0.143
np	0.143
gracias	0.146
pleasure	0.150
psyched	0.151
introducing	0.151
blessing	0.152
thankies	0.154
sweeet	0.154
yaaaay	0.155
smiley	0.157

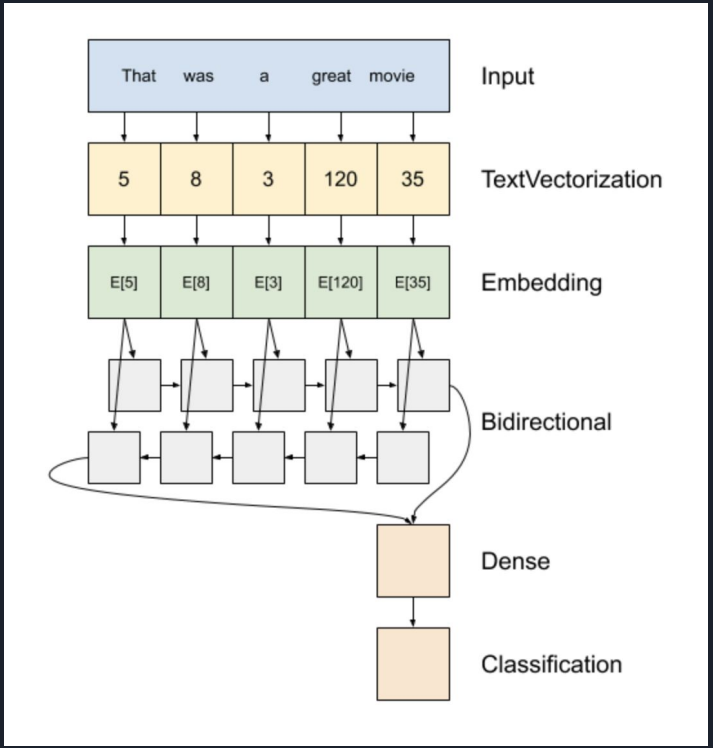
Model Embeddings - Sad



Nearest points in the original space:

bummed	0.039
dissapointed	0.041
depressing	0.041
unhappy	0.045
horrid	0.046
depressed	0.048
rainin	0.050
sadd	0.054
ouchie	0.054
upsetting	0.055
sadly	0.056
poisoning	0.057
disappointment	0.058
saddens	0.058
disapointed	0.059
condolences	0.059
damaged	0.059
injured	0.059
heartbroken	0.061
carradine	0.063

Our RNN and RNN2 model structure - BiLTSM



Regular LSTM Explained

I jumped into the ...

BiLSTM Explained

I jumped into the … with only a small blade

I jumped into the … and swam to the other shore

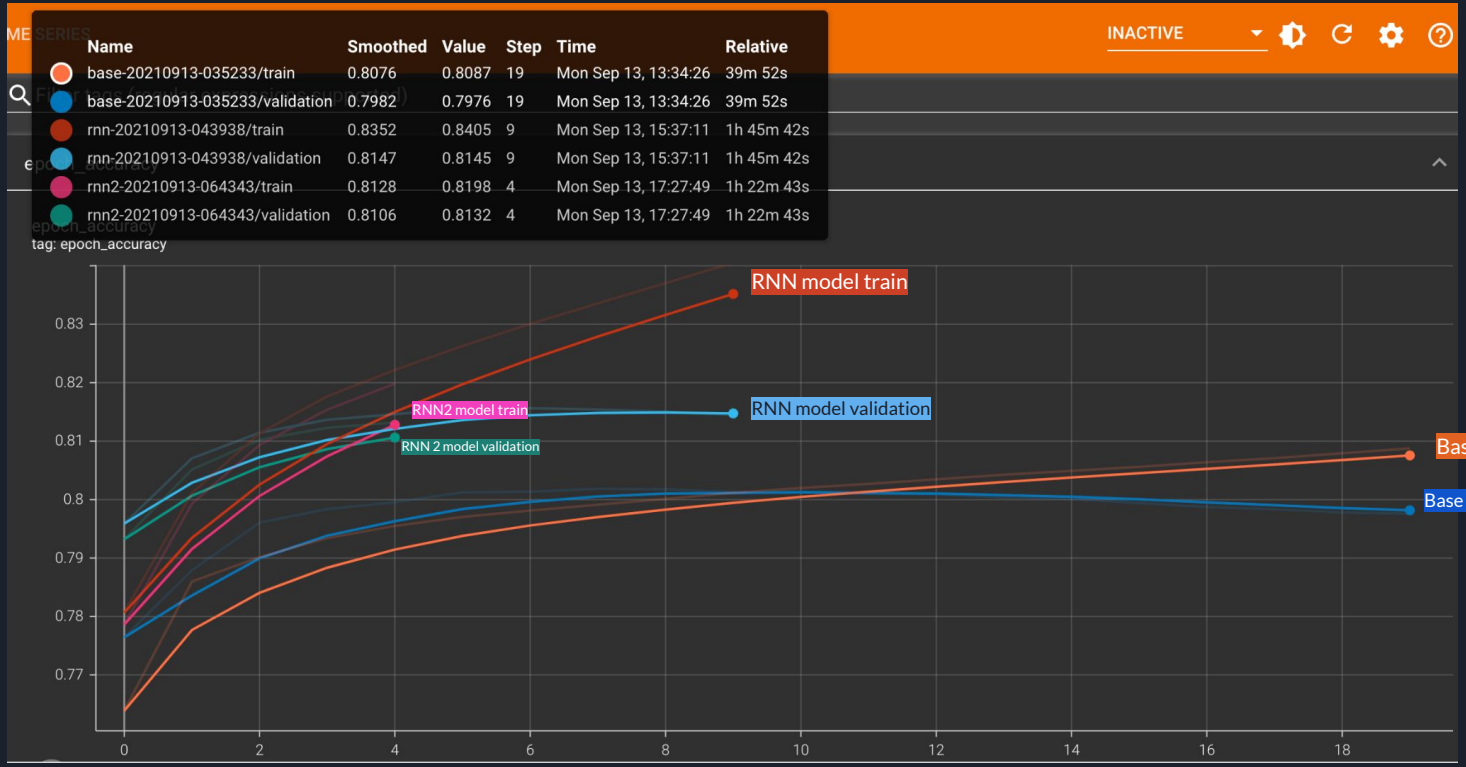
I jumped into the … from the 10m diving board

Quick Model Prediction Example

```
input: @umeshunni yeah iPhone does have more going for it.. now thats decided.. all thats left is wait for 3G here..
score: 0.808093

input: @BenjaminHouston oh I wish you guys were coming here  I wanna see savvy!!! ( and you guys of course) ha
score: 0.744839

input: @CaliSmiles06 hmm. ;) yuh, up &amp;amp; out with the #rat boys &amp;amp; girls.
score: 0.182670

input: never thought my daughter would want to borrow my clothes... even less that she would suggest that i borrow hers... luckily it didnt fit
score: 0.423570
```

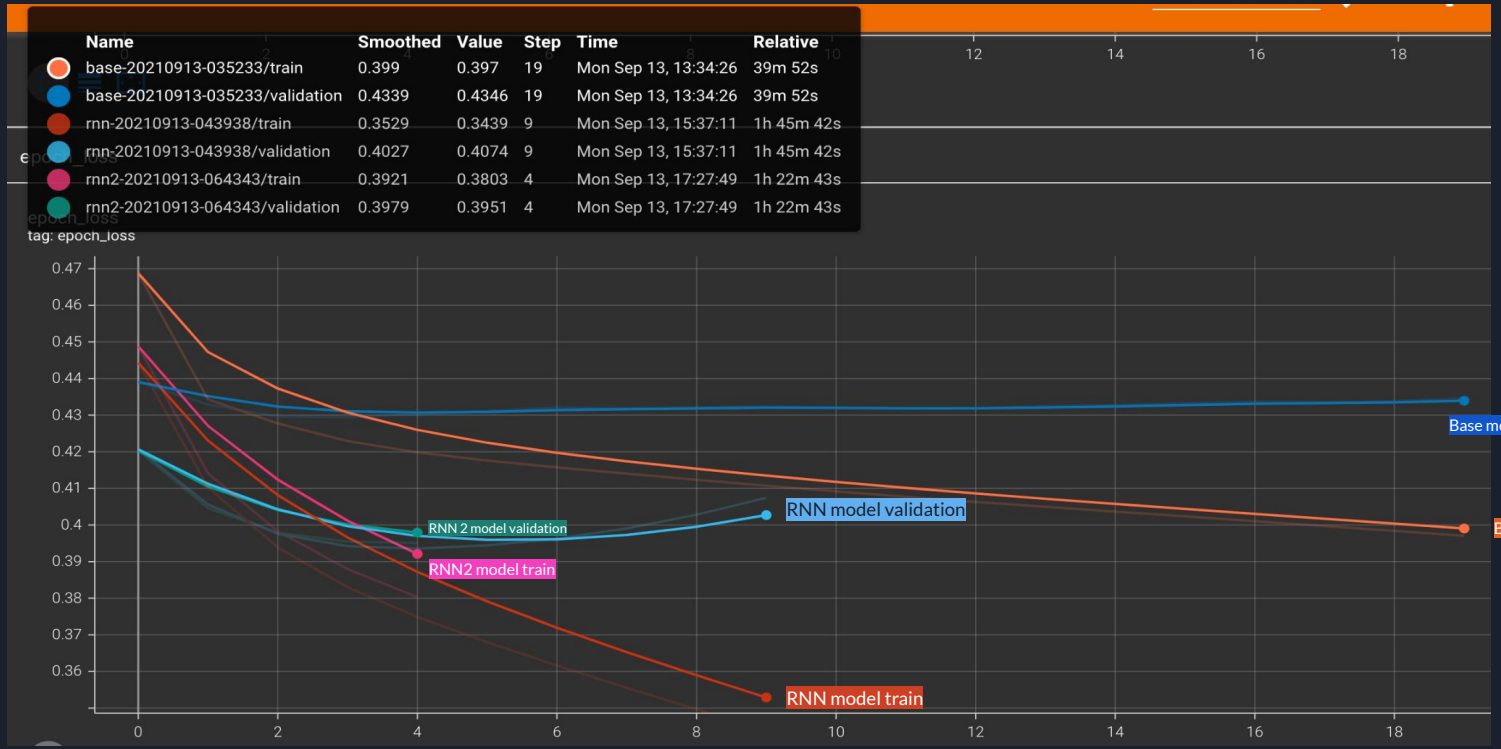
All Models Accuracy/Loss Summary

All Models Accuracy



- Base model- starts to overfit around epoch 14
- RNN - overfits at epoch 7
- RNN 2- is still pretty good

All Models Loss



- Base - validation model - starts to overfit around epoch 14 since the loss is going up.
- RNN- validation starts to overfit at epoch 6
- RNN2- is still doing good

Model Summary

		Accuracy	Loss
1	Base Model	79.8%	43.5%
2	RNN Model	81.4%	40.7%
3	RNN Model 2	81.3%	39.5%

Gather Twitter Data

Twitter Sentiment140 Dataset

created_at	tweet	prediction
2020-08-19 19:35:09	Cyberpunk 2077 Historia de Night City #Cyberpunk2077 #Cyberpunk #NightCity https://t.co/Ti5fK6dN1F	0.724559
2020-08-19 19:28:03	Might have to start streaming so that I can justify spending hundreds of hours playing #Cyberpunk2077 when it releases in November 💪 👀 https://t....	0.677329
2020-08-19 19:02:52	the cursed DAY 91 we have arrived for the 3rd time 👀 #Cyberpunk2077 https://t.co/UXaOrbD0so	0.684503
2020-08-19 18:59:40	👀 #Cyberpunk2077 Collector's box Was gifted to me & it has a 🔥 Samurai Pin to add to my collection 😃 https://t.co/cl4dUEQ0Ov	0.910197
2020-08-19 18:55:37	I can't believe they did it to us again. #Cyberpunk2077 https://t.co/o3IbLKXhDX	0.533198
...	...	...
2020-08-17 12:35:01	Choose your weapon! #rtxon #geforcertx #rtx2080ti #nvidiageforce #cdprojektred #cyberpunk2077 #gamingsetup #yellowandblack #rgblights https://t....	0.594399
2020-08-17 12:26:59	Why am I playing Nomad? There's something about losing all you know is routine and comfortable in the world, being thrust into something so out of...	0.222436
2020-08-17 12:05:11	Concept Art - Xbox Series_Cyberpunk2077_Collector — - . #XboxSeriesX #Cyberpunk2077 #collector #conceptart __> @XboxFR @xbox @CyberpunkGame @...	0.590995
2020-08-17 12:03:10	Saw these today in @GAMEdigital and was very tempted.. #Cyberpunk2077 https://t.co/xLsfhUw5lI	0.459139
2020-08-17 11:56:30	I am so ready for this game 😃 Night City and The Badlands are going to be an incredible Open World to explore #Cyberpunk2077 https://t.co/UrF12a...	0.985989

<< Sigmoid function

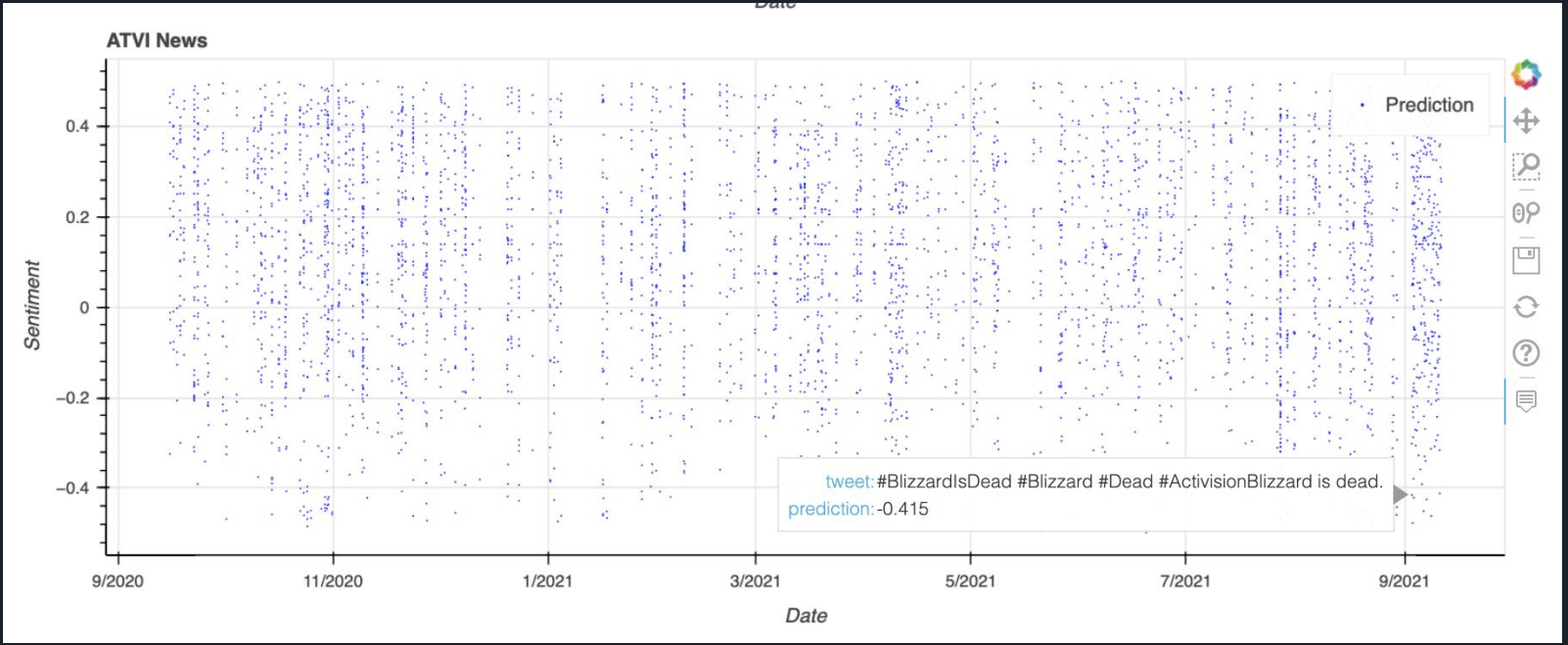
Note: The reason why we adjusted it so that we can easily add it and create a cumulative sum of good/bad predictions of 'x' days easier

created_at	tweet	prediction
2020-08-19 19:35:09	Cyberpunk 2077 Historia de Night City #Cyberpunk2077 #Cyberpunk #NightCity https://t.co/Ti5fK6dN1F	0.224559
2020-08-19 19:28:03	Might have to start streaming so that I can justify spending hundreds of hours playing #Cyberpunk2077 when it releases in November 💪 👀 https://t....	0.177329
2020-08-19 19:02:52	the cursed DAY 91 we have arrived for the 3rd time 👀 #Cyberpunk2077 https://t.co/UXaOrbD0so	0.184503
2020-08-19 18:59:40	👀 #Cyberpunk2077 Collector's box Was gifted to me & it has a 🔥 Samurai Pin to add to my collection 😃 https://t.co/cl4dUEQ0Ov	0.410197
2020-08-19 18:55:37	I can't believe they did it to us again. #Cyberpunk2077 https://t.co/o3IbLKXhDX	0.033198
...	...	...
2020-08-17 12:35:01	Choose your weapon! #rtxon #geforcertx #rtx2080ti #nvidiageforce #cdprojektred #cyberpunk2077 #gamingsetup #yellowandblack #rgblights https://t....	0.094399
2020-08-17 12:26:59	Why am I playing Nomad? There's something about losing all you know is routine and comfortable in the world, being thrust into something so out of...	-0.277564
2020-08-17 12:05:11	Concept Art - Xbox Series_Cyberpunk2077_Collector — - . #XboxSeriesX #Cyberpunk2077 #collector #conceptart __> @XboxFR @xbox @CyberpunkGame @...	0.090995
[illegible]	Saw these today in @GAMEdigital and was very tempted.. #Cyberpunk2077 https://t.co/xLsfhUw5lI	-0.040861
[illegible]	[illegible]ready for this game 😃 Night City and The Badlands are going to be an incredible Open World to explore #Cyberpunk2077 https://t.co/UrF12a...	0.485989

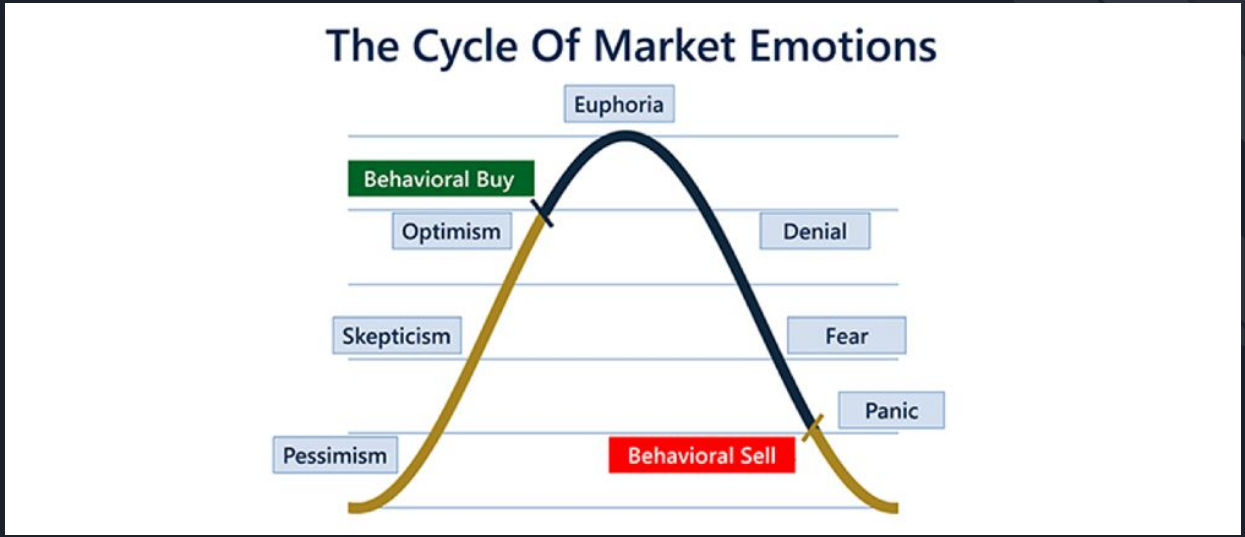
<< Centered around 0

+ 0.5 good tweet
- 0.1 bad tweet
- 0.5 worst tweet

ATVI Sample Sentiment/Prediction Signal



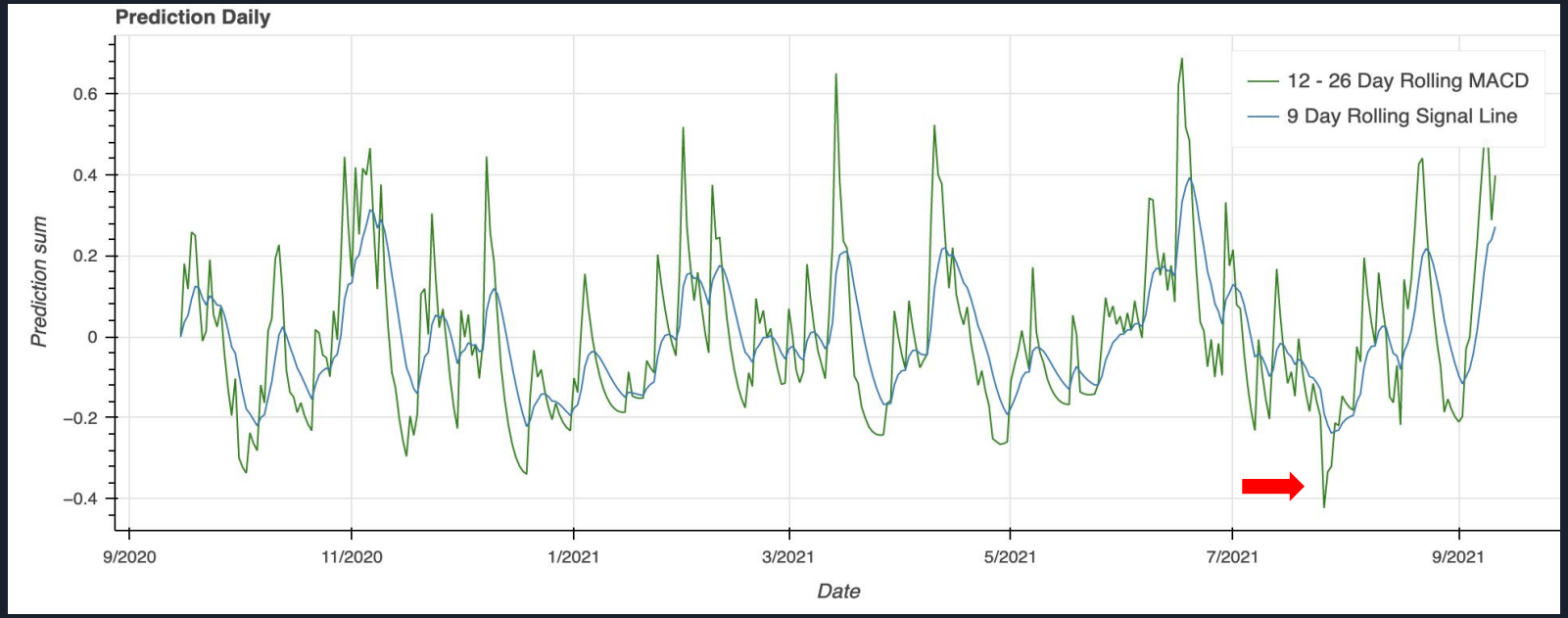
Stock Price Analysis



Activision Blizzard (ATVI)

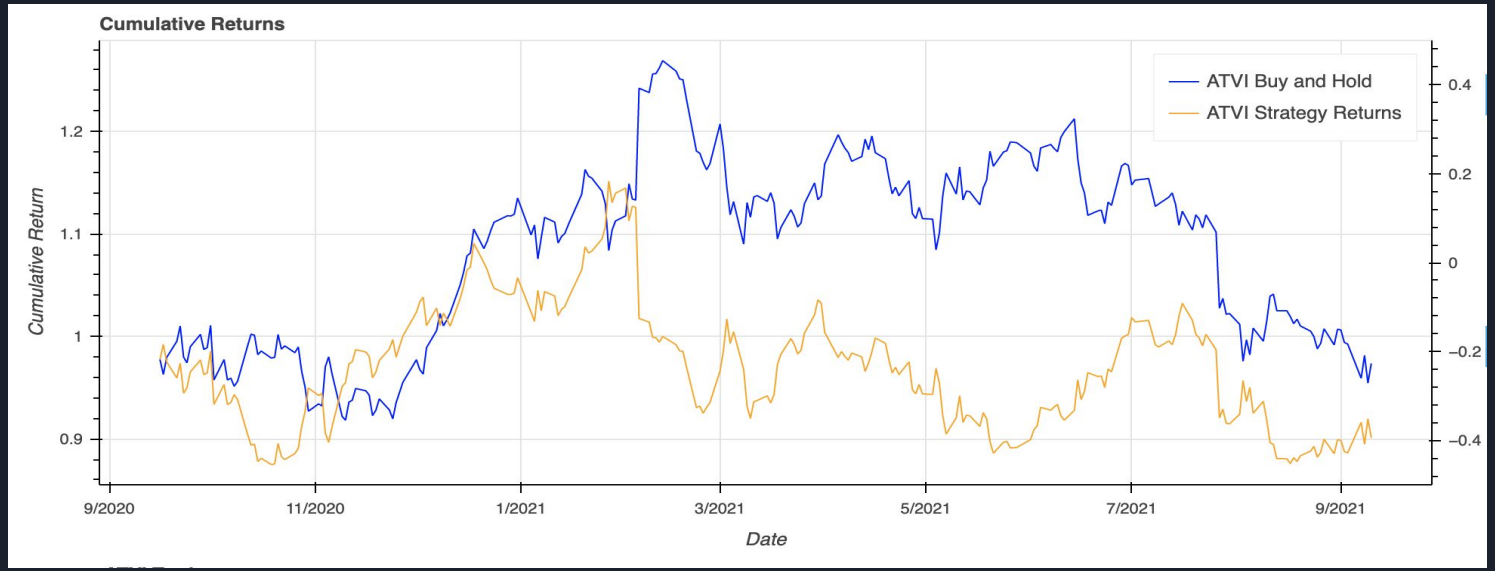
>> Around July 20th, ATVI Blizzard was sued for its workplace culture. As shown on July 22th- 28th, its stock price was down.

ATVI Signals



- The tweets regarding Blizzard lawsuit, clearly shows that their stocks will drop based on their signals. (July is the month that the lawsuit was made public.)

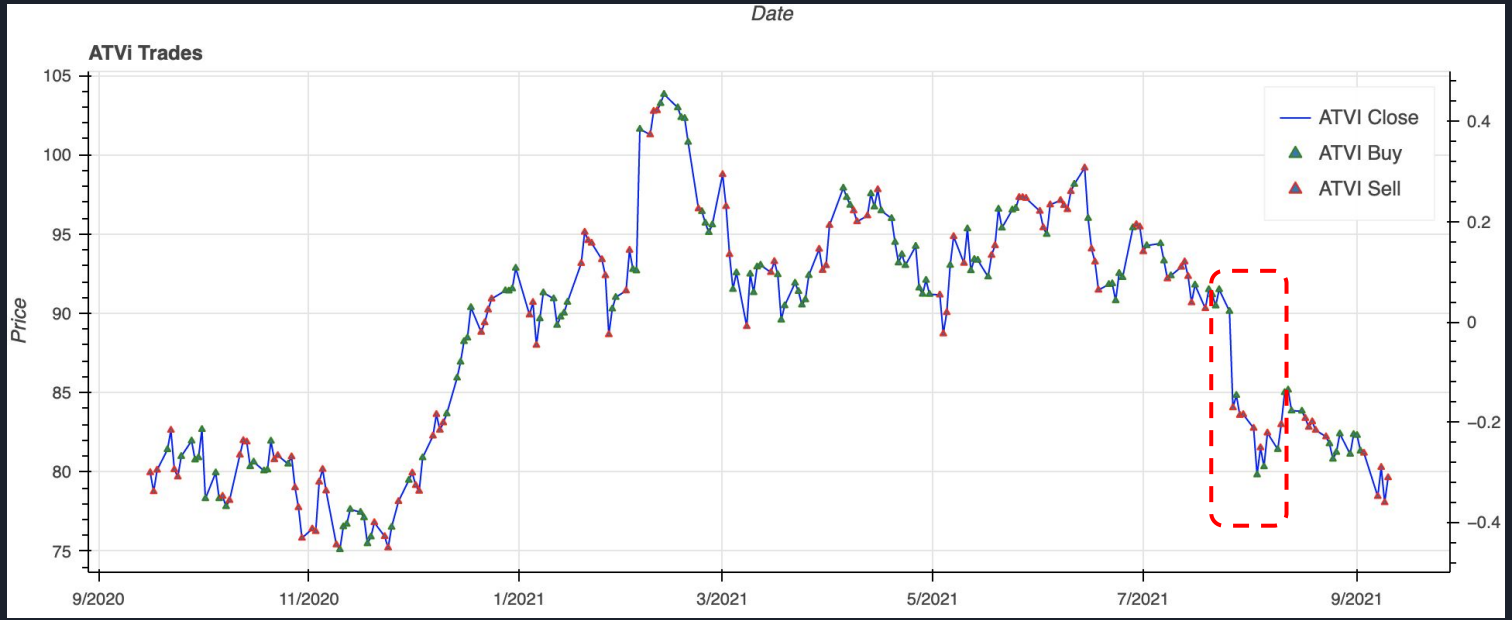
ATVI Strategy Returns



	ATVI Buy and Hold	ATVI Strategy Returns
CAGR	-0.00403395	-0.0788749
Max Drawdown	0.247953	0.239486
Annual Volatility	0.282273	0.282269
MAR Ratio	-0.016269	-0.32935
Sharpe Ratio	0.125876	-0.151096

- Around end of July, we can see the drop in returns for buy and hold while our strategy had a surge upwards.

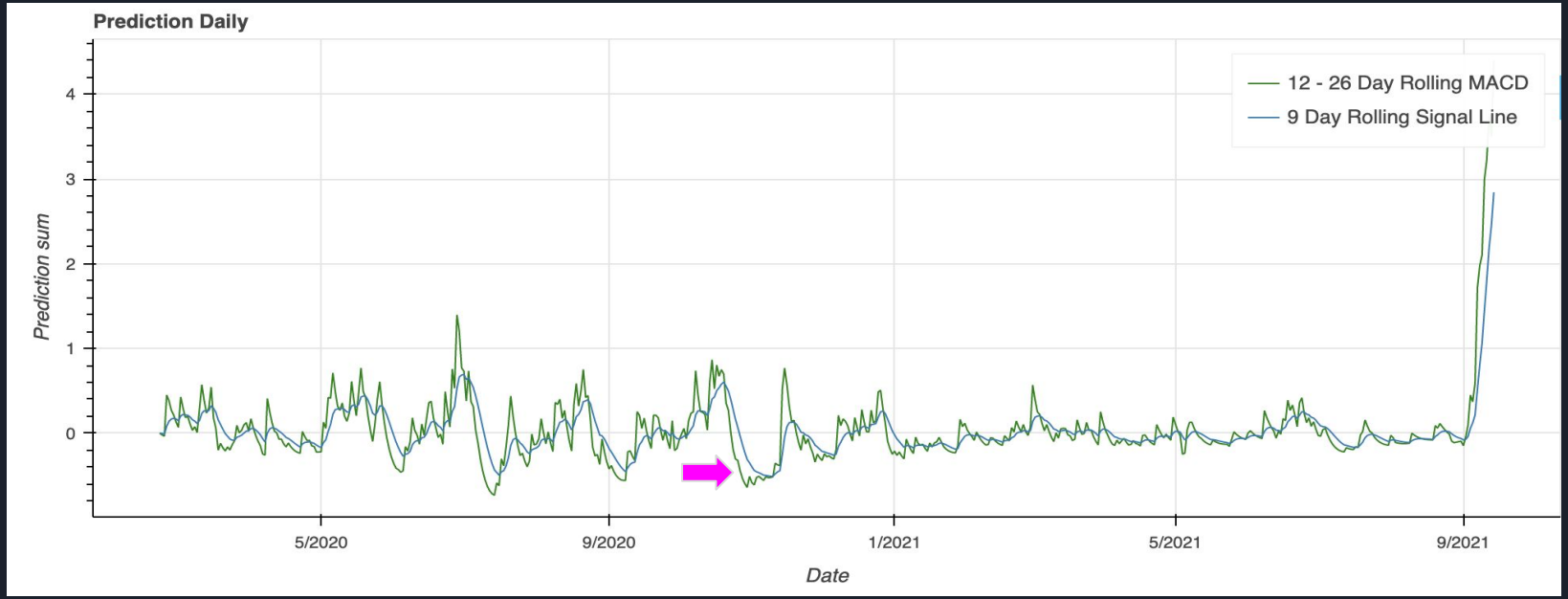
ATVI Trades


Date
ATVi Trades
Price
105
100
95
90
85
80
75
9/2020
11/2020
1/2021
3/2021
5/2021
7/2021
9/2021
0.4
0.2
0
−0.2
−0.4
ATVI Close
ATVI Buy
ATVI Sell

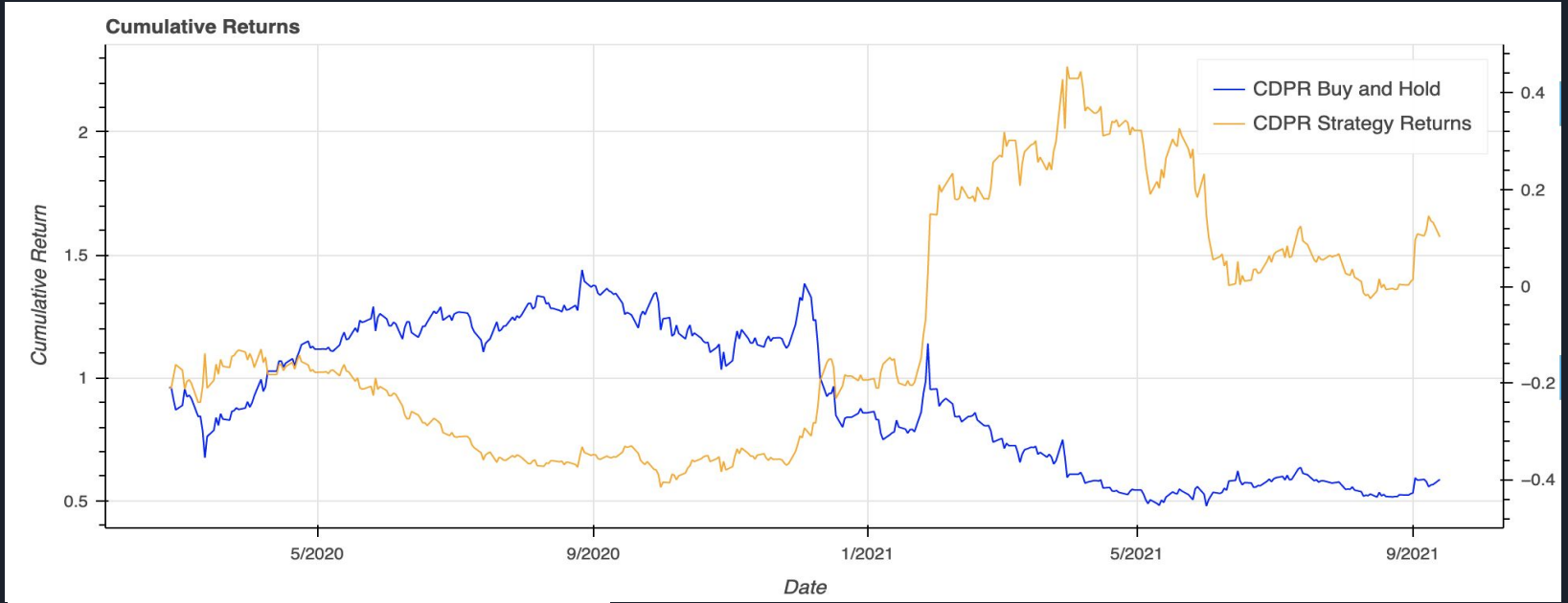
CD Projekt Red (CDPR)

>> CDPR released a highly anticipated game titled “Cyberpunk 2077” on December 10th, 2020 - the title flopped magnificently in the wave of their heavy marketing and consumer hype.

CDPR Signals

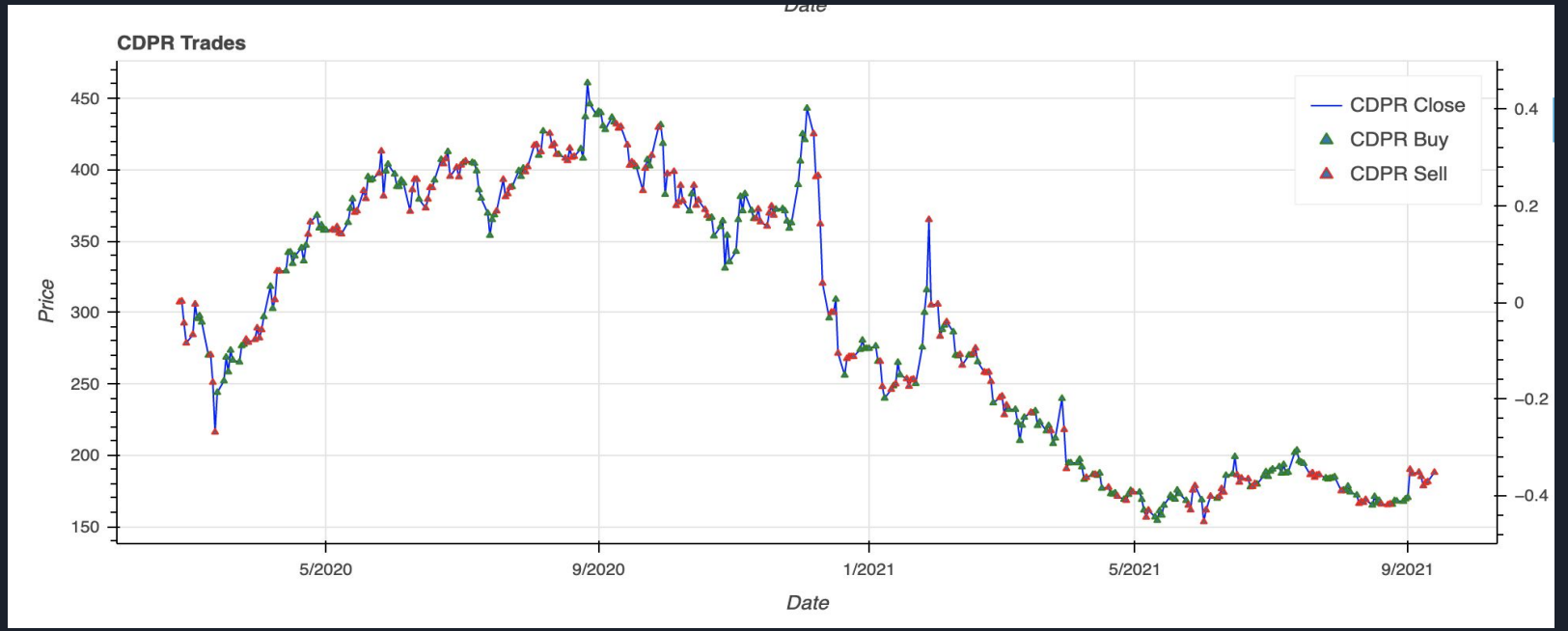


CDPR Strategy Returns



	CDPR Buy and Hold	CDPR Strategy Returns
CAGR	-0.263575	0.360009
Max Drawdown	0.667144	0.503295
Annual Volatility	0.582063	0.581354
MAR Ratio	-0.395079	0.715303
Sharpe Ratio	-0.235334	0.817747

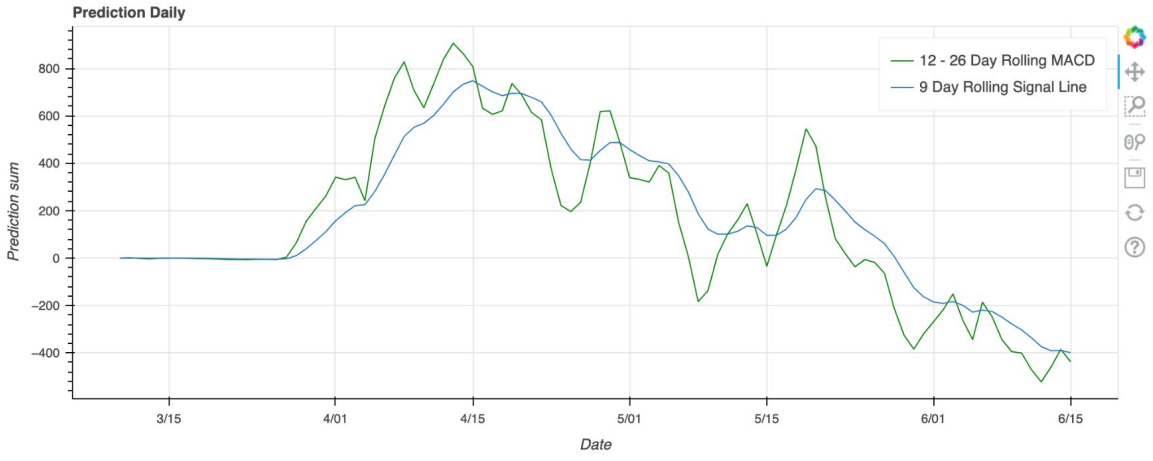
CDPR Trades


CDPR Trades
Price
Date
450
400
350
300
250
200
150
5/2020
9/2020
1/2021
5/2021
9/2021
0.4
0.2
0
−0.2
−0.4
CDPR Close
CDPR Buy
CDPR Sell

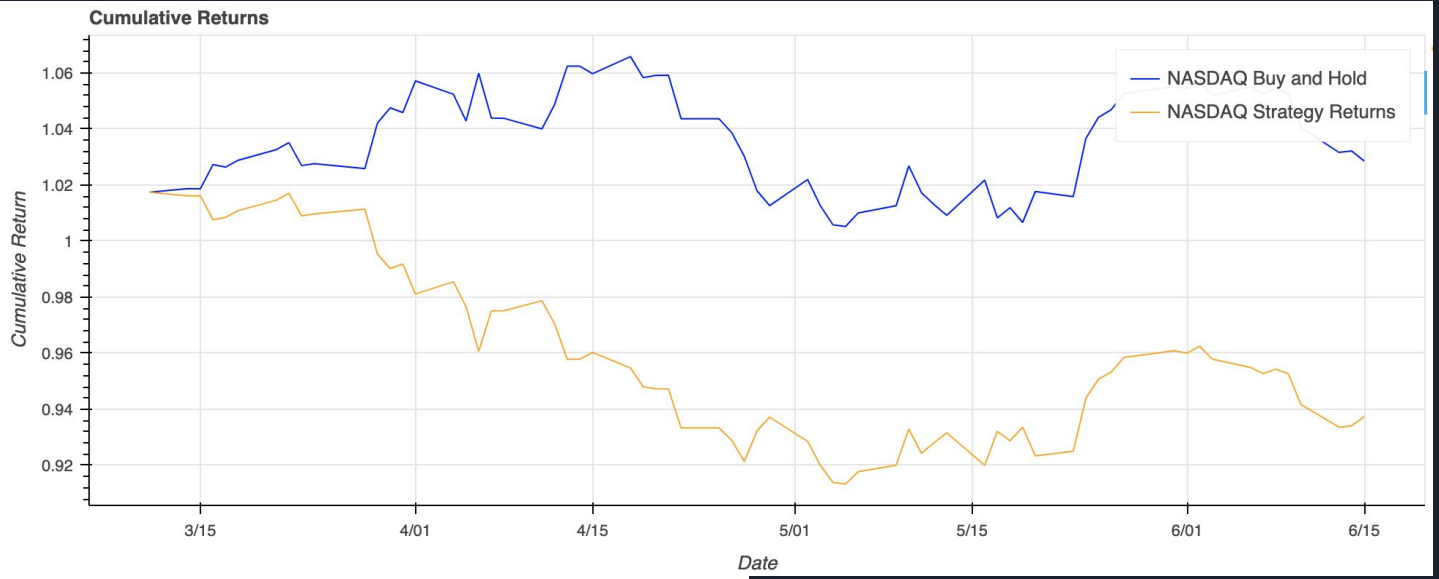
NASDAQ

NASDAQ Signals

```
# Display Nasdaq trade signal
p = figure(
    title="Prediction Daily",
    x_axis_label='Date',
    x_axis_type='datetime',
    y_axis_label='Prediction sum',
    plot_width=1000,
    plot_height=400)
p.line(nasdaq_signals.index, nasdaq_signals['MACD'], legend_label=f"{short_window} - {long_window} Day Rolling MACD", line_color='green')
p.line(nasdaq_signals.index, nasdaq_signals['signal'], legend_label=f"{signal_window} Day Rolling Signal Line")
show(p)
```



NASDAQ Strategy Returns



	NASDAQ Buy and Hold	NASDAQ Strategy Returns
CAGR	0.0411123	-0.262054
Max Drawdown	0.0568642	0.102341
Annual Volatility	0.121492	0.119971
MAR Ratio	0.722991	-2.56058
Sharpe Ratio	0.401418	-2.54905

Conclusion

- With the two most prominent companies, ATVI and CDPR having a huge twitter backlash and tweets regarding their respective issues. This model seemed to predict their respective stock prices.

- This model doesn’t work great with NASDAQ and likely would work similarly on other indexes. This is most likely because the NASDAQ is more diversified which dilutes and somewhat mitigates the “business risk” our strategy utilizes.

Sourced Data

- Data gathered from
  - Twitter- through Twint
  - FinViz
  - Alpaca API
- Datasets
  - Sentiment140
  - Activision Blizzard (ATVI)
  - CDProjektRed (CDPR)
  - Nasdaq100

Technology Used

- TensorFlow
- NLP and Sentiment Analysis
- Google Colab
- Twint

Team Members and Their Tasks

01	**Colin Benjamin** in/colinbenjamin	• Created performance metrics for our model and the buy and hold model. • Created a trading bot that takes both long and short positions based on twitter sentiment. • Charted Model Performance
02	**Justine Cho** in/justinecho	• Created a dataframe from the Alpaca API and pulled relevant stock tickers • Pulled the closing prices and calculated percent change of these prices. • Compiled the README.md file.
03	**Christopher Henderson** in/chris-henderson123/	• Helped manage central project goals and scope of project • Suggested companies to provide analysis for • Composed CDPR twitter pull notebook portion
04	**Nathan Patterson** in/natepatterson/	• Data clean up and save it to csv files • Used Twint to pull relevant tweets for the project. • Used machine learning model to train, fit and predict the accuracy of the tweets.

Successes

- Managed to pull tweets using a Twitter API and call specific hashtags relevant to ATVI, CDPR, Nasdaq 100.
- Managed to pull real time closing prices for the tickers used in the project.
- Decided on a machine learning sentiment analysis model.

Challenges

- Running the machine learning models took a while.
- The limitations on Google Colaboratory with regard to using its GPU and some libraries were hard to work with.
- The fickleness of running Twint on Google Colab has been challenging but not without resolution.
- With 3rd party API, we noticed that we weren't collecting all the tweets, and that losing those data messed up the model.

Lessons Learned

- Using a new machine learning model, Recurrent Neural Network.
- To save a particular data to csv so one wouldn't reach the limits on its respective API calls.
- Some imports may or may not work well with Google Colab.

Next Steps

- Additional topics to research:
  - Expanding the tickers pulled to other relevant stock tickers to see correlation with the public sentiment using tweets.
  - To use a different machine learning model that can further improve its accuracy.
- Plan for future development:
  - Use different stock tickers and add new performance metrics.

Links

- GitHub Repo: https://github.com/Best-Brain-Gang/twitter_sentiment_analysis.git

THANK YOU FROM YOUR BEST BRAIN GANG!

ANY QUESTIONS?