TWEETING **TESLA**

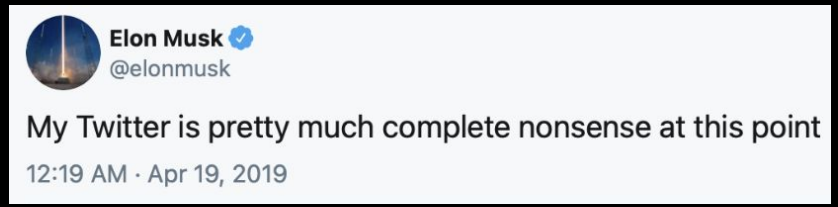
COLUMBIA UNIVERSITY
OF NEW YORK
FU FOUNDATION
SCHOOL OF ENGINEERING
FINTECH BOOTCAMP

| **JANUARY 2021**

CONTRIBUTORS |
TAMARI BACHULASHVILI
AIDAN DOWDLE
LISA ESBERGER
MINYEONG HAN
CARLOS TACCHI

INTRODUCTION

Our team has used natural language processing, sentiment analysis and machine learning in order to determine whether Elon Musk's tweets have an impact on Tesla's stock price. We used this analysis to create a trading algorithm to trade Tesla stock.



KEY QUESTIONS

IS IT POSSIBLE TO PREDICT THE SHORT-TERM CHANGES IN TESLA'S STOCK USING ELON MUSK'S TWEETS?

TESLA RETURNS

Can we predict the change in daily and weekly returns using machine learning?

PROFITABILITY

Which(buy, sell, hold) strategy could be the most profitable?

POWER OF TWEETS

Is there a correlation between Elon Musk's Tweets and returns?

TRADING

Could investors and day traders benefit from an algorithm that tracks Tweets?

01

Data Cleaning

Cleaning our data and sorting through potential features.

02

Data Exploration

Gaining a deeper understanding of the story told by the data.

03

Model Training & Evaluation

Testing our hypothesis through random forest regressor

04

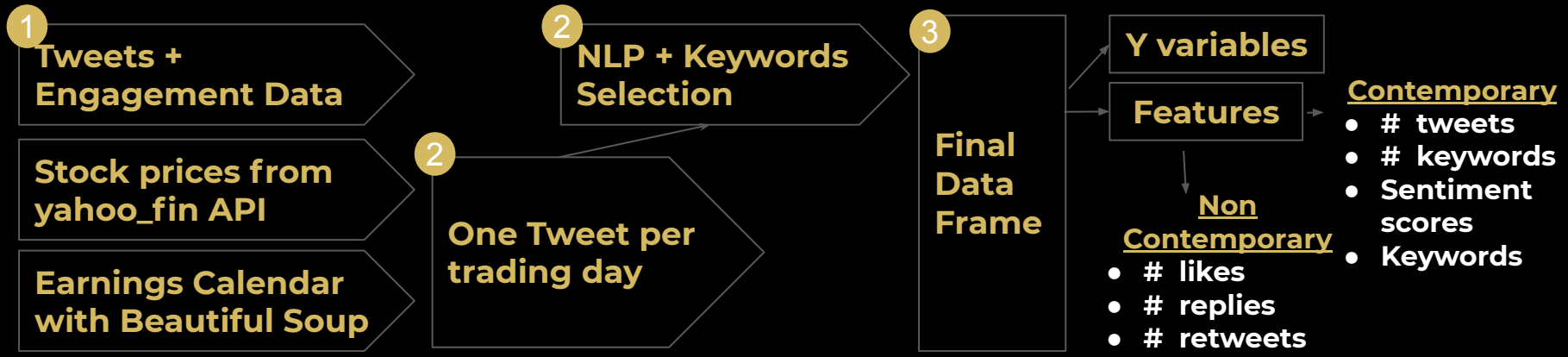
Project Conclusions

Evaluating our conclusions on Tesla stock for profit

01

DATA CLEANING

DATA CLEANING



1

• Different date formats:	Mainly use of to_datetime to have indexes in same format for joins
• Stock splits:	Shifted away from Alpaca API to one that adjusted prices for stock splits
• Web scraping:	Used Beautiful Soup to extract data
• Twitter data:	Used twint to scrape twitter data that could not pull via twitter API

2

• Adjusting dates of Tweets:	Tweets received after 16:00:00 are pushed to t+1
• Grouping Tweets per day:	Count tweets, combine strings + sum engagement data
• Pushing Tweets to trading days	Iterative process to push tweets to next immediate trading day
• Keywords count	lower() + for loop + apply() to use with data frame

3

• Y variable generation	Price change difference between TSLA and QQQ Drop of NAs
• Features	Sentiment on no tweet days is shown as neutral
• Time horizon	Dropped first six months since TSLA IPO due to lack of tweet activity




SENTIMENT ANALYSIS

VADER Model

- ➔ Classified tweets as negative, positive, and neutral.
  - ◆ Extracted polarity scores.
  - ◆ Created pos/neg/neu labels derived from the compound score.

FREQUENCY ANALYSIS



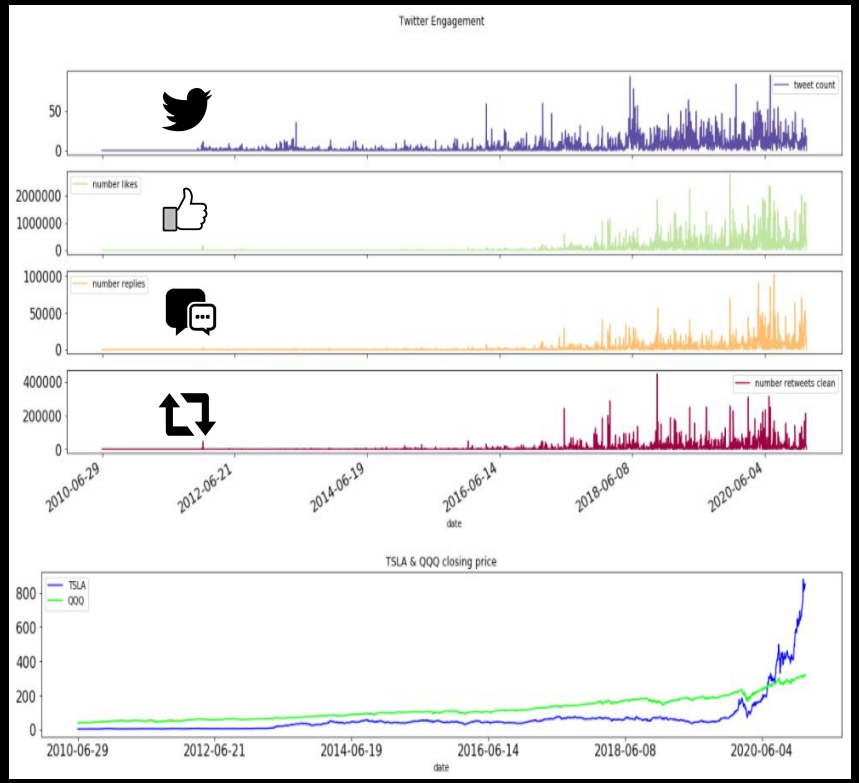

02

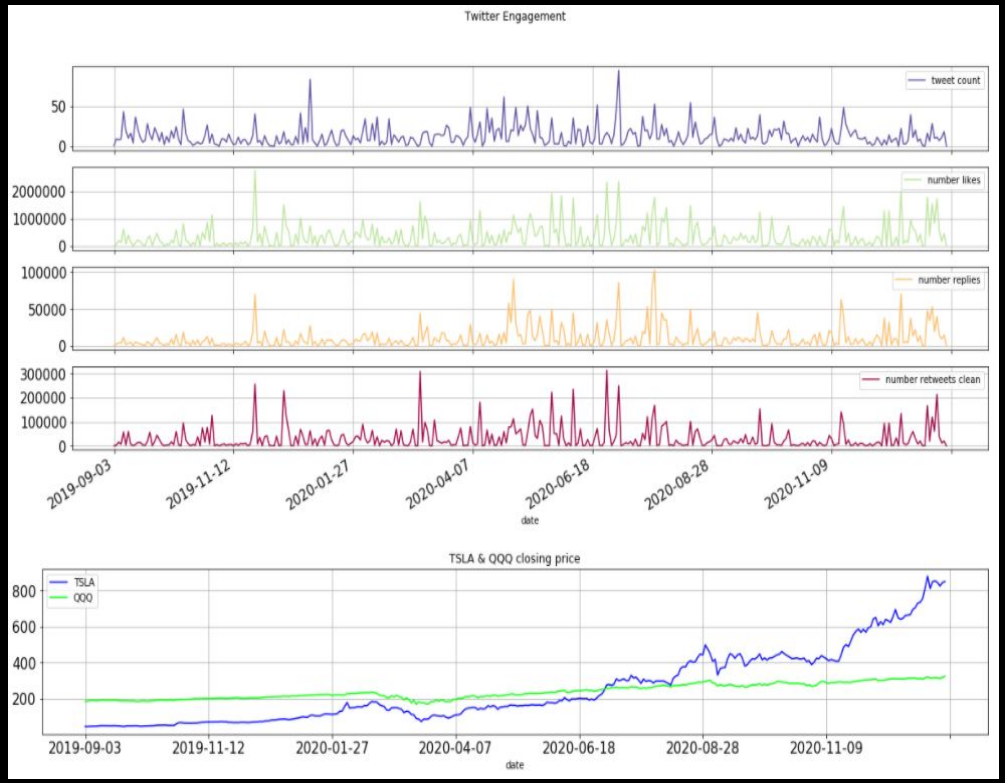
DATA EXPLORATION

DATA EXPLORATION

TWITTER ENGAGEMENT FROM 2010 - 2020

On June 29, 2010, Tesla made its debut as a public company - the first initial public offering (IPO) of a domestic automaker in a half century. The IPO price (pre stock split) had been $17 a share.

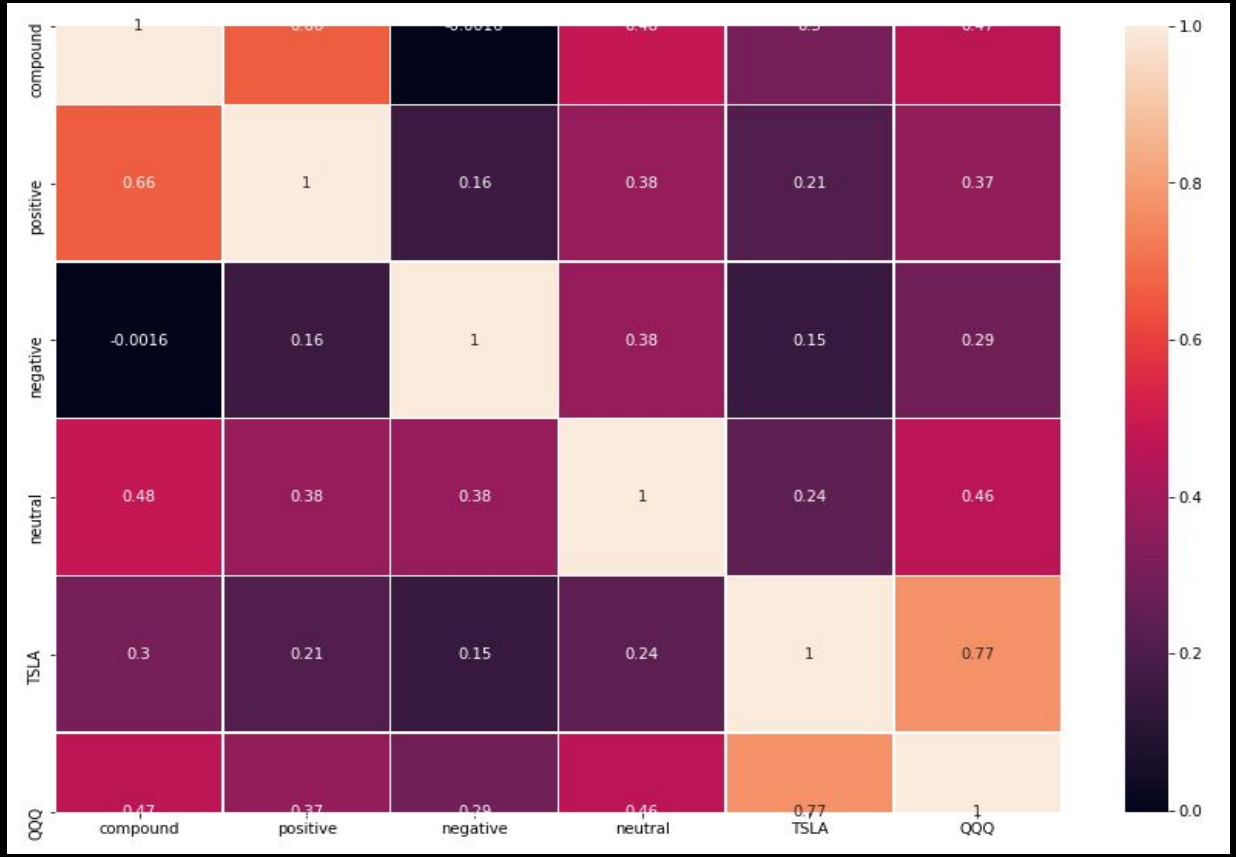




DATA EXPLORATION

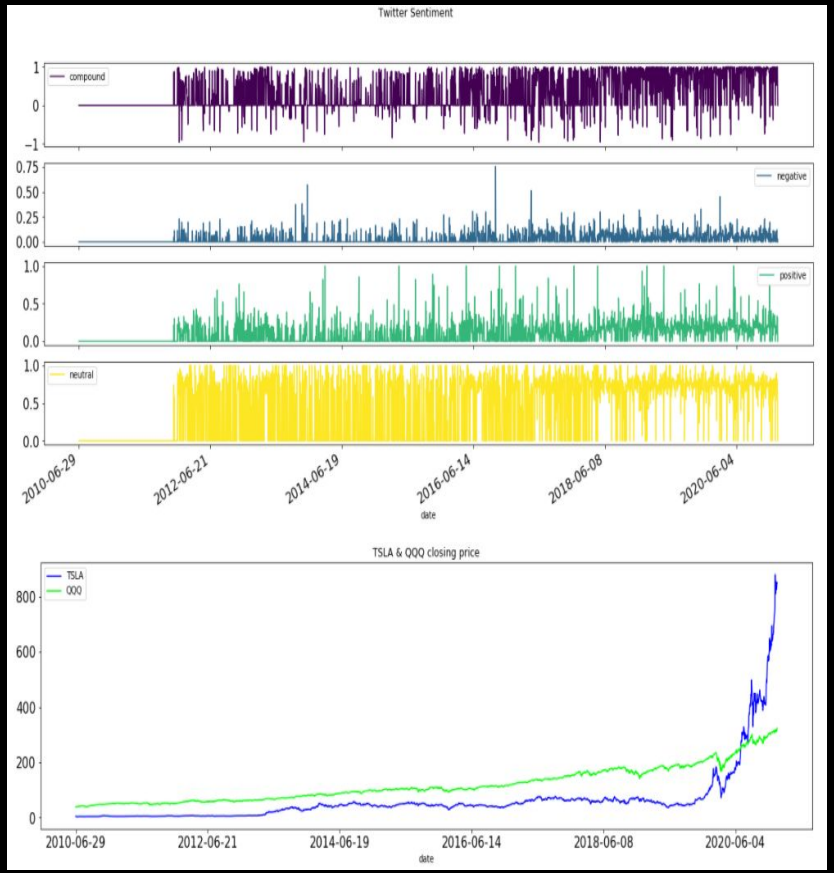
TWEET SENTIMENT CORRELATION & TESLA STOCK GROWTH

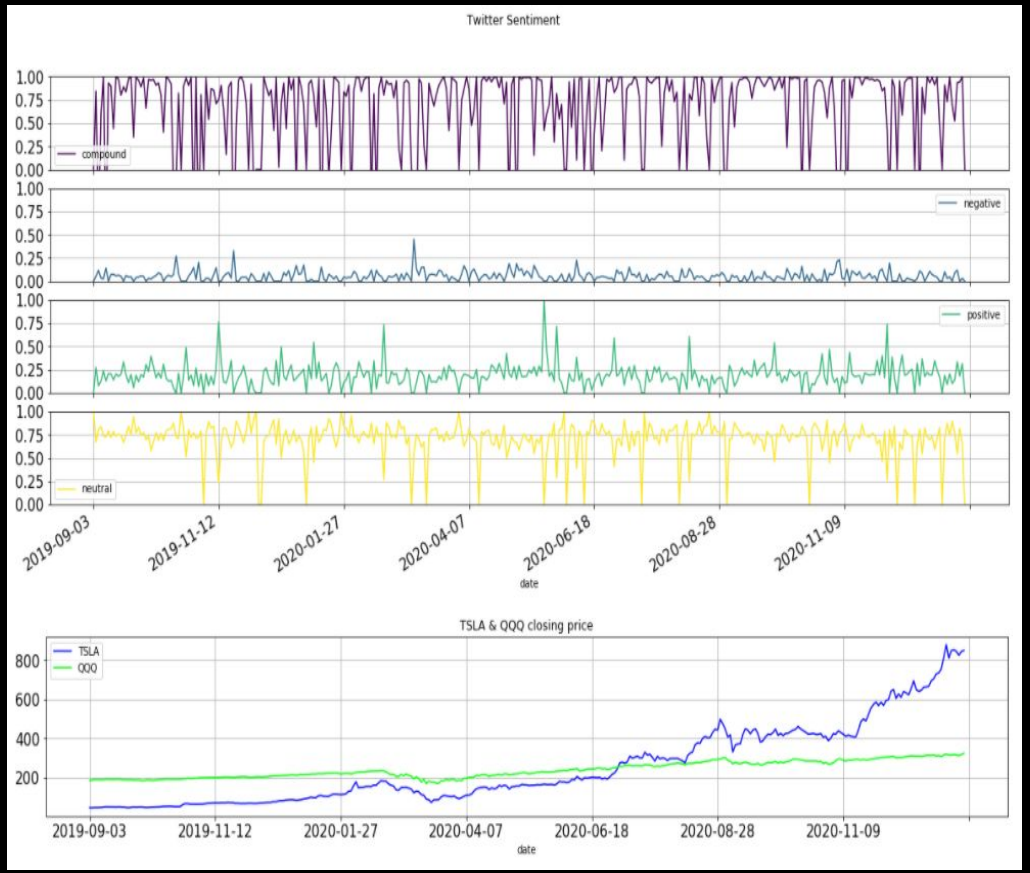
‘Compound’ score and ‘Neutral’ sentiment appear to have a relatively higher correlation with TSLA stock closing price, except QQQ closing price used to offset the market effect in our machine learning model.



DATA EXPLORATION

SENTIMENT SCORES VS. TESLA STOCK PRICE





03
MODEL TRAINING & EVALUATION

OUR PREDICTIVE MODEL

RANDOM FOREST REGRESSION

Decision trees are sensitive to the specific data on which they are trained.

If the training data is changed, the resulting decision tree can be quite different and in turn the predictions can be quite different.

They carry a big risk of overfitting, and tend to find local optima because they can't go back after they have made a split.

To address these weaknesses, we turn to Random Forest Regression to obtain mean prediction of the individual trees as output.

MODEL TRAINING & EVALUATION

RANDOM FOREST REGRESSION

TRAINING PROCESS

- Time window: Started on June 29, 2010 when Tesla Motors launched its IPO on NASDAQ.
- Estimators: 500
- Train / test split: Random split with shuffling (test size 25%)
- Modeled with difference in daily return and weekly return, respectively.

MODEL EVALUATION

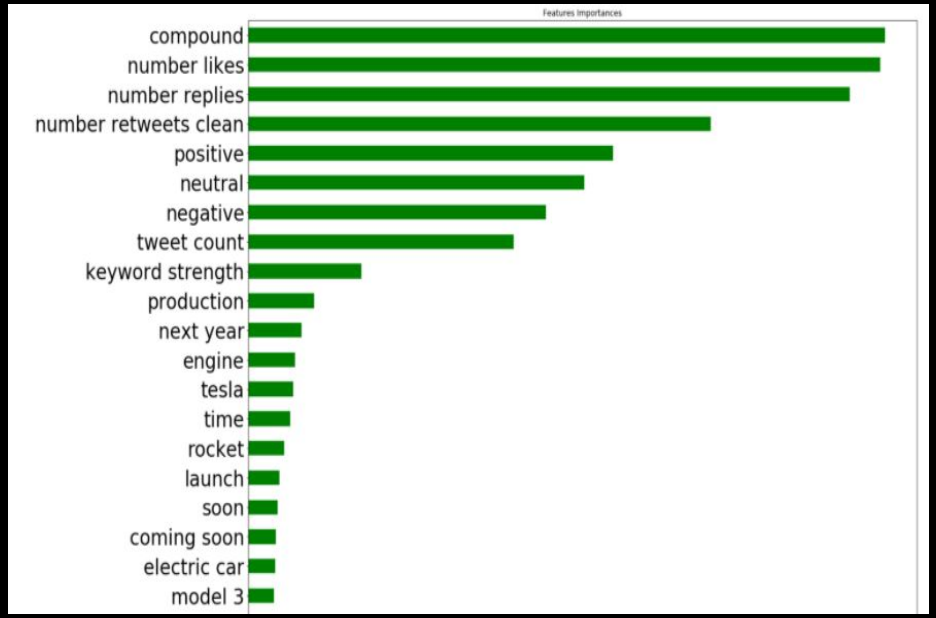
- Mean Squared Error (MSE)
  - Closer to 0 indicates that the model is closest to the regression line.
  - Daily return MSE for train data: 0.0005
  - Daily return MSE for test data: 0.001
  - Weekly return MSE for train data: 0.002
  - Weekly return MSE for test data: 0.006

CHANGE IN FEATURE IMPORTANCE

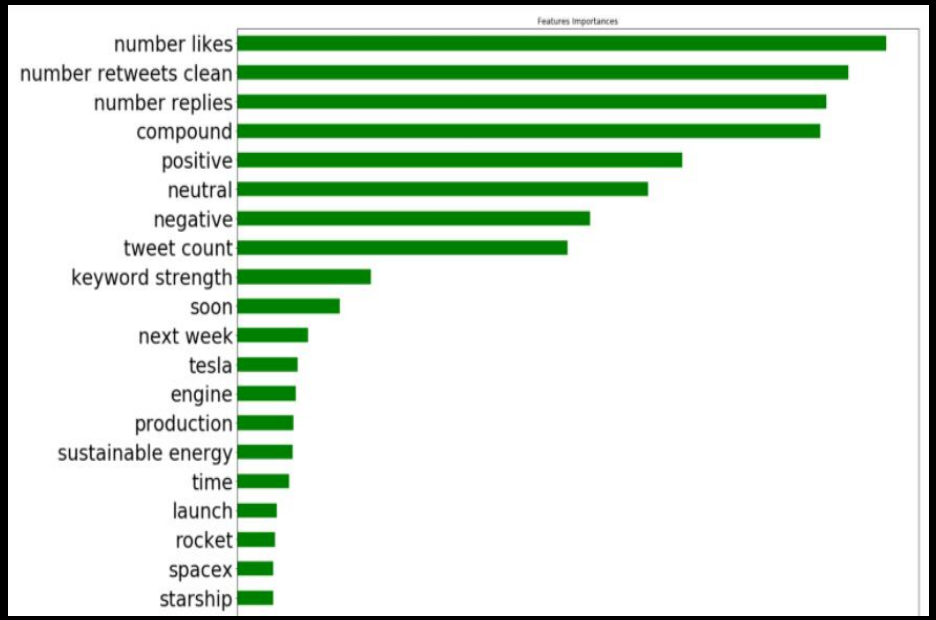
IMPACT ON DAILY & WEEKLY RETURNS

- Both Twitter sentiment & engagement features affect returns the most.
- Compound sentiment score has a slightly greater impact on daily returns.
- Engagement features have a slightly greater impact on weekly returns.

DAILY RETURNS



WEEKLY RETURNS



04

PROJECT CONCLUSIONS

DISCLAIMER

- NOT INVESTMENT ADVICE -

THIS IS A STUDENT PROJECT THROUGH TRILOGY / COLUMBIA UNIVERSITY OF NEW YORK

The content contained in the FinTech Boot Camp (the "Course") is for informational purposes only. You should not construe any such information or other material provided pursuant to the Course (the "Materials") as investment, financial, tax, legal or other advice. The Materials are not investment advice and any observations concerning any security, trading algorithm or investment strategy provided in the Course is not a recommendation to buy, sell or hold such investment or security or to make any other investment decisions. Trilogy Education Services, LLC, its parent, subsidiaries, and affiliates ("Trilogy") do not provide any advice regarding the nature, potential value, risk or suitability of any particular investment strategy, trading algorithm, transaction, security or investment. Any use of the Materials, and any decisions made in reliance thereon, including any trading or investment decisions or strategies, are made at your own risk. You should seek the advice of a competent, licensed professional if you require investment, trading or other advice. No employee, agent or representative of Trilogy is authorized to provide any professional advice in connection with the Course and any such advice, if given, is in violation of Trilogy's policies, is unauthorized and may not be relied upon. Nothing contained in the Materials constitutes a solicitation, recommendation, endorsement, or offer by Trilogy or any third party service provider to buy or sell any securities or other financial instruments in this or in in any other jurisdiction whether or not such solicitation or offer would be unlawful under the securities laws of such jurisdiction.

https://columbia.bootcampcontent.com/columbia-bootcamp/cu-nyc-fin-pt-09-2020-u-c/-/blob/master/15-Algorithmic-Trading/1/FinTech-15-1_v1.0.2.pdf

BACKTESTED TRADING RESULTS

03 Jan 2017 - 31 Dec 2020

VALUE OF 3,000 TESLA SHARES ON 31 DEC 2020 = $2,117,009.95

BUY DAY 1 & HOLD

COST: $231,691.14

Final Purchase: 14 JUL ‘20

6 Shares per Trigger

$1,885,318.81

BUY DAY 5 & HOLD

COST: $252,695.93

Final Purchase: 20 AUG ‘20

5 Shares per Trigger

$1,864,314.02

BUY ON WEEKLY INTERVAL & HOLD

COST: $295,472.10

Final Purchase: 30 OCT ‘20

15 Shares per Trigger

$1,821,537.85

BUY ALL ON 03 JAN 2017 & HOLD

COST: $130,193.00

Final Purchase: 03 JAN ‘17

3,000 Shares per Trigger

$1,986,815.96

CONCLUSIONS

IS IT POSSIBLE TO PREDICT THE SHORT-TERM CHANGES IN TESLA'S STOCK USING ELON MUSK'S TWEETS?

Is there a correlation between Elon Musk's Tweets and returns?

- Yes, we have discovered a correlation between Elon Musk's Tweets and changes in Tesla's returns using sentiment analysis.

Can we predict the change in daily and weekly returns using machine learning?

- Yes, our model Random Forest Regressor appeared to predict daily and weekly returns well as it achieved a mean squared error close to zero.

Could day traders and investors benefit from an algorithm that tracks Tweets?

- Yes and no, day traders could benefit from receiving an alert when the machine learning algorithm indicates correlated activity.
- However, a trading algorithm would not benefit intraday traders in this case.
- Based on our analysis, a buy and hold investor would benefit from waiting until the hot news cools off before buying additional shares.

What (buy, sell, hold) strategy could be the most profitable?

- In Tesla's case, a buy and hold at the time of IPO or a systematic investment strategy would likely be the most profitable than trading following news.

PROJECT LIMITATIONS

TIME DELAYS

- The popularity of the Tweets cannot be gauged accurately in real time.
- Day Traders rely on trading algorithms that report data within milliseconds of events to make buy/sell decisions in rapidly moving markets.

EXPANDING ALGORITHMIC TRADING

- Use additional features to determine if a better model exists.
- Moving averages, bollinger bands, flagging for press releases, regulatory changes, fundamental data, guidance reporting on pricing expectations, etc.

VOLATILITY

- Further exploring Tweets impact on volatility may benefit traders.
- If there is a correlation, traders could use options to benefit from bursts of Tesla trading activity.

Q&A

Questions?

Elon Musk ✓
@elonmusk

Turns out there's more ways to use Autopilot than we imagined

4:46 PM · May 9, 2019

♡ 168.3K 24.7K people are Tweeting about this

Elon Musk ✓
@elonmusk

I love Twitter

12:50 PM · Dec 21, 2017

♡ 170.8K 37.5K people are Tweeting about this

Elon Musk ✓
@elonmusk

Will those who write the algorithms ever realize their negativity bias?

7:52 PM · Jul 17, 2020 · Twitter for iPhone

10.4K Retweets **898** Quote Tweets **138K** Likes

Elon Musk ✓
@elonmusk

Replying to @elonmusk

I was always crazy on Twitter fyi

12:11 AM · Apr 15, 2019

♡ 79.6K 5.6K people are Tweeting about this

Elon Musk ✓
@elonmusk

Am considering taking Tesla private at $420. Funding secured.

12:48 PM · Aug 7, 2018

♡ 86.1K 21.2K people are Tweeting about this

Elon Musk ✓
@elonmusk

Tesla stock price is too high imo

11:11 AM · May 1, 2020

♡ 191.8K 34.5K people are Tweeting about this

Elon Musk ✓
@elonmusk

You're an idiot

9:48 PM · Dec 14, 2017

♡ 37.5K 5.4K people are Tweeting about this

Elon Musk ✓
@elonmusk

Who controls the memes,
controls the Universe

1:05 PM · Jun 26, 2020 · Twitter for iPhone

31.2K Retweets **168.9K** Likes

- We greatly appreciate your time and attention -

Thank you