

Deep dive into predicting stock prices using technical indicators as well as sentiment analysis.



CZ1115 Mini - Project by Neel, Sashank and Wei Ling

Problem Definition

To explore the extent to which Amazon stock price predictions can be characterized by public sentiments on Twitter using multivariate regression and sentiment analysis.



Datasets used

- Twitter API
- Yahoo! Finance
- Alpha Vantage Stock API (Amazon Stocks - AMZN)
 - Stock Time Series - Intraday
 - Technical Indicators - Small Moving Average (SMA) and Exponential Moving Average (EMA)

Data collection, curation and preparation

- Coded functions to acquire intraday, EMA and SMA data using company stock ticker.
- Data was cleaned and optimized to concatenate data frames to match in terms of date.
- The required time frame was from 2001 - 2021, and the last 7 days were used to compare closing prices with predicted prices and polarity scores.
- Ticks were concatenated in order to sort data and prepare for exploratory data analysis.



Utility functions used to curate data

```
def get_tech():
    key = '84E88MB3ZLGNJO2H'
    symbol = input('Ticker : ')
    outputsize = 'compact'
    interval = input('Interval- 1min,5min,15min,30min,60min,daily,weekly,monthly : ')
    time = input('Time Period : ')
    tech_indi = input('Technical Indicator- SMA,EMA,VWAP,MACD,Stochastic Oscillator')

    ti = TechIndicators(key,output_format='pandas')

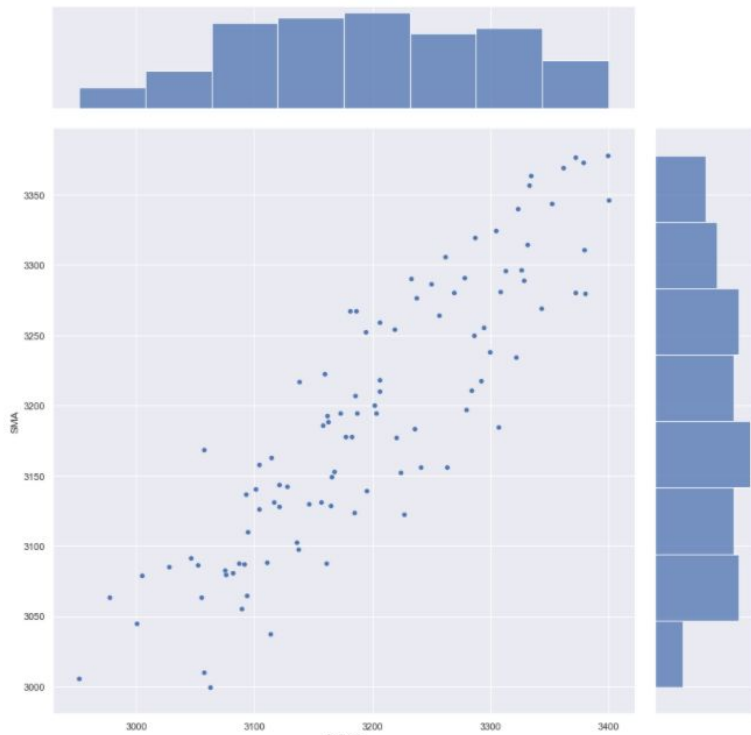
    if tech_indi == 'SMA':
        state = ti.get_sma(symbol, interval=interval, time_period=time, series_type='close')
    elif tech_indi == 'EMA':
        state = ti.get_ema(symbol, interval=interval, time_period=time, series_type='close')
    elif tech_indi == 'VWAP':
        state = ti.get_vwap(symbol, interval=interval)[0]
    elif tech_indi == 'MACD':
        state = ti.get_macd(symbol, interval=interval, series_type='close')[0]
    elif tech_indi == 'Stochastic Oscillator':
        state = ti.get_stoch(symbol, interval=interval)[0]
    elif tech_indi == 'RSI':
        state = ti.get_rsi(symbol, interval=interval, time_period=time, series_type='close')
    elif tech_indi == 'Bollinger bands':
        state = ti.get_bbands(symbol, interval=interval, time_period=time, series_type='close')
    else:
        print('Wrong Entry')
    return state
```



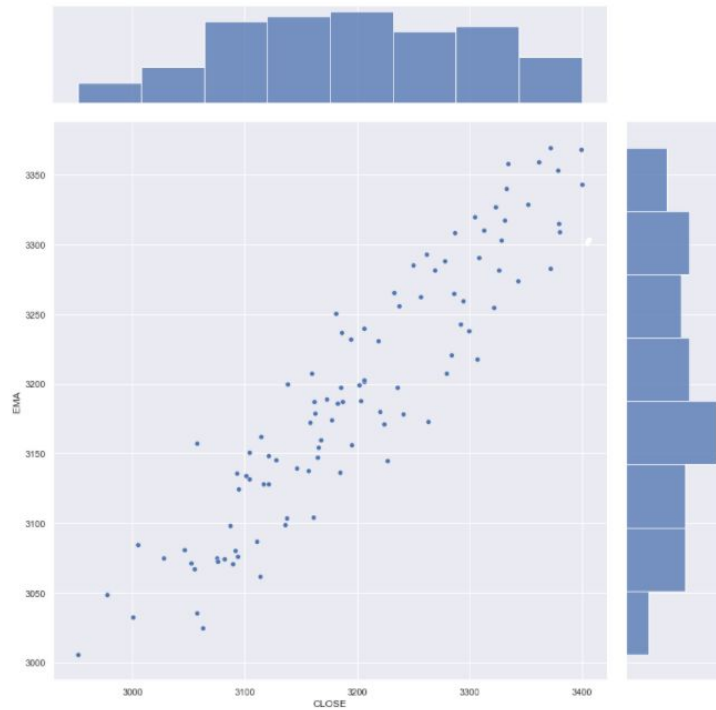
Exploratory Analysis

- Collected **statistical information** on the variables obtained.
- Visualised the **uni-variate distributions** of each variable (EMA, SMA, CLOSE)
- Visualised each variable against closing price using **jointplot**
- Visualised correlation matrix using **heatmap**
- Drew pairs of variables against each other using **pairplot**

Exploratory Analysis - JointPlot



SMA vs CLOSE



EMA vs CLOSE

Visualization



SMA, EMA and Closing Prices against dates

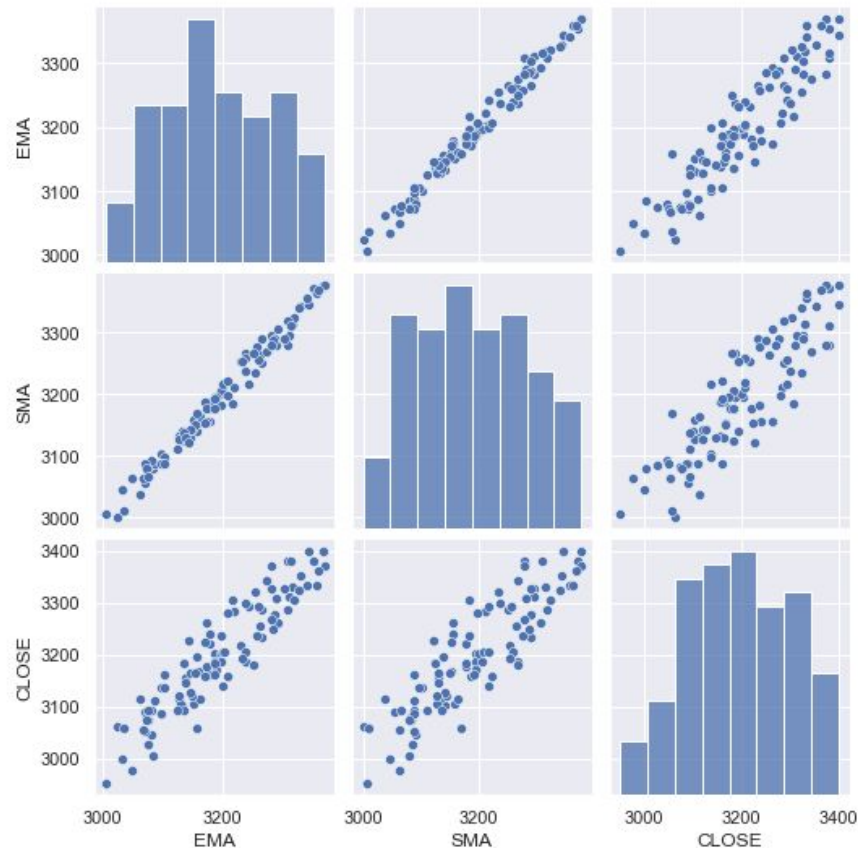


Exploratory Analysis - Heatmap



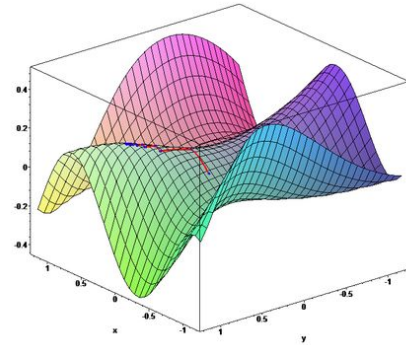


Exploratory Analysis - Pairplots



Multivariate Linear Regression

- We decided to use multivariate linear regression for analysing the trends in the stock price and how good indicators such as SMA and EMA are in predicting the stock price.
- Since our objective is to predict the future stock price and compare it with the public's sentiment analysis, linear regression has to be used in order to predict the future stock price. Any other ML model such as classification will not work as this is a prediction problem.



Methodology

- We used the data that we extracted from alpha vantage as our independent variables. We extracted technical indicators(SMA,EMA) and used these to predict the closing price(dependent variable) of the Amazon stock.
- To achieve this,, we created a train test split to get a train dataset and a test dataset(test size=0.2) . After fitting the data, we obtained the coefficients of regression for the dependent variables.
- We then proceeded to obtain the R squared values and the MSE values for the train and test dataset respectively.

```
Intercept of Regression      : b = [-406.88546829]  
Coefficients of Regression   : a = [[-2.00766693  3.13658434]]
```

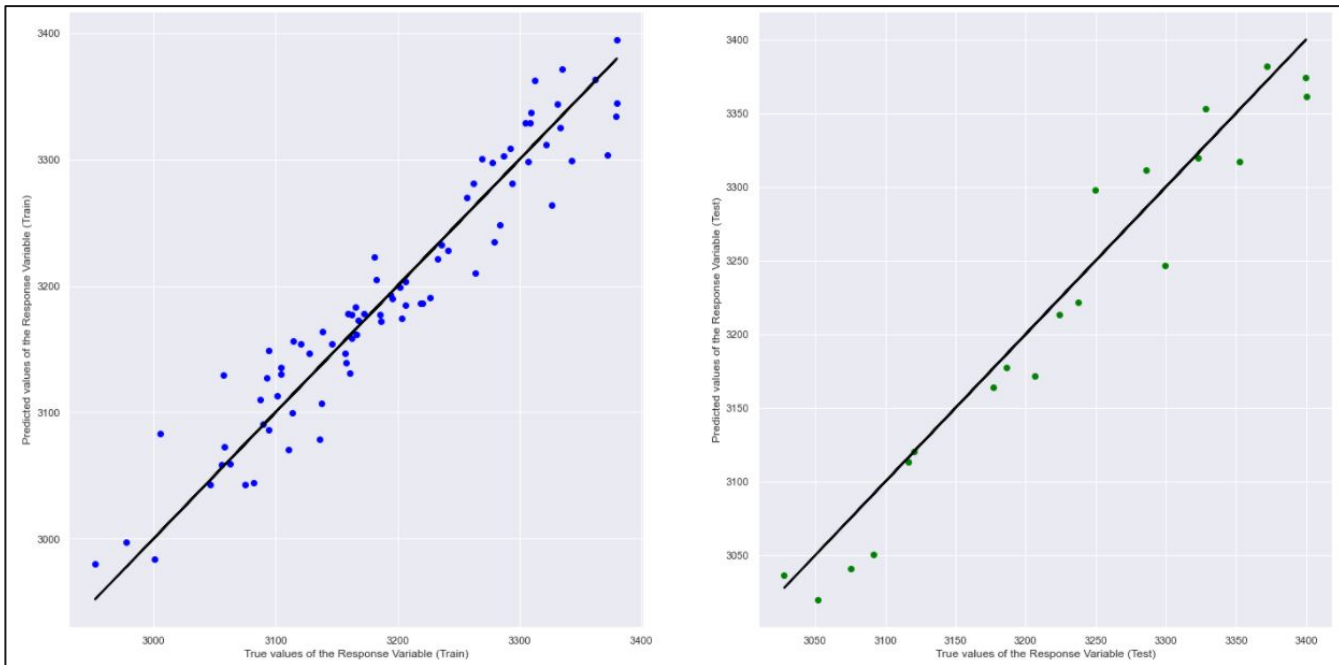
	Predictors	Coefficients
0	SMA	-2.007667
1	EMA	3.136584

Goodness of Fit of Model
Explained Variance (R^2)
Mean Squared Error (MSE)

Train Dataset
: 0.9267354969023606
: 911.6118027577589

Goodness of Fit of Model
Explained Variance (R^2)
Mean Squared Error (MSE)

Test Dataset
: 0.9064929411715916
: 642.6551740994557





Sentiment Analysis

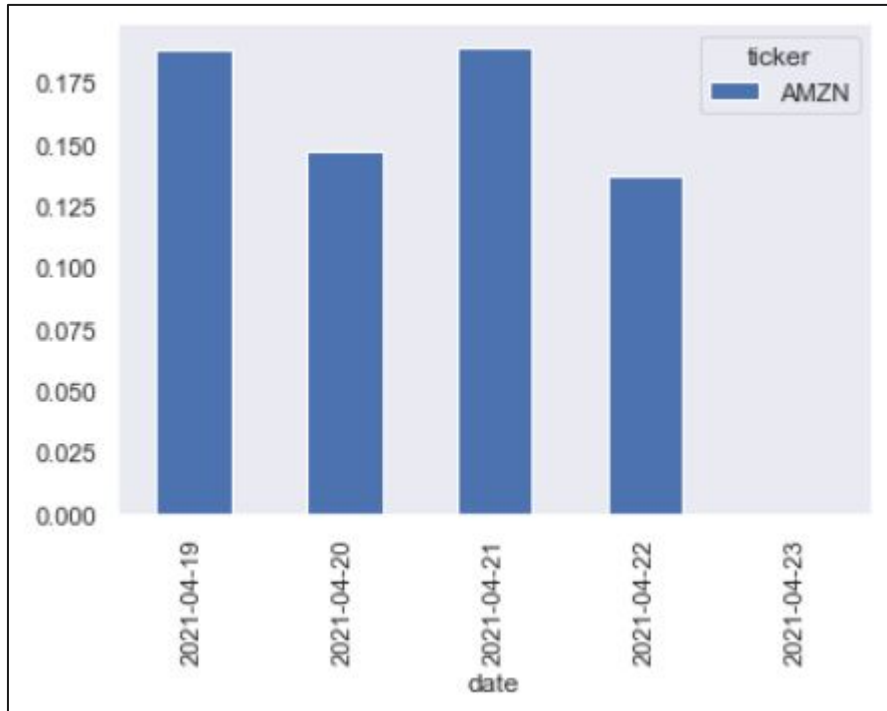
- We decided to use sentiment analysis to compare and contrast our stock price prediction with the public's opinion of Amazon to see a correlation.
- Data was scraped from Twitter's API using nltk and newspaper3k to get AMZN specific data.
- Use of vader sentiment library to assess subjectivity and polarity to assign scores to data.
Used sentiment intensity analyzer to see a connect sentiment scores with stock prices.

Compiled sentiment data frame using NLP

	ticker	date	time	headline	neg	neu	pos	compound
0	AMZN	2021-04-23	12:24AM	Xiaomi Said to Mull Investing in AI Chipmaker ...	0.000	1.000	0.000	0.0000
1	AMZN	2021-04-22	07:06PM	3 Great Growth Stocks to Buy Now at a Discount	0.000	0.472	0.528	0.7717
2	AMZN	2021-04-22	06:35PM	Is Kroger Stock A Buy Right Now? Here's What E...	0.000	1.000	0.000	0.0000
3	AMZN	2021-04-22	05:08PM	Microsofts Big Deal and Coinbases Big Debut	0.000	1.000	0.000	0.0000
4	AMZN	2021-04-22	03:21PM	Amazon Targets Automated Checkout In Regular S...	0.000	0.805	0.195	0.1779
5	AMZN	2021-04-22	02:37PM	AeroFarms Serves Up High-Growth Greens Investo...	0.000	0.829	0.171	0.1139
6	AMZN	2021-04-22	02:30PM	Top mistakes graduating college students make ...	0.188	0.677	0.135	-0.1779
7	AMZN	2021-04-22	02:00PM	Amazon to Open London Hair Salon: Another Step...	0.171	0.698	0.132	-0.1280
8	AMZN	2021-04-22	01:06PM	DISH Stock Extends Rally on Amazon Deal. Heres...	0.000	0.789	0.211	0.2960
9	AMZN	2021-04-22	12:33PM	Amazon (AMZN) Reports Next Week: Wall Street E...	0.000	0.650	0.350	0.5106
10	AMZN	2021-04-22	11:50AM	Why Clean Energy Stock Moved Higher Today and...	0.139	0.511	0.350	0.4404
11	AMZN	2021-04-22	11:20AM	President Biden's Tax Proposal Targets Big Tec...	0.000	0.775	0.225	0.4404
12	AMZN	2021-04-22	10:20AM	Influencers with Andy Serwer: Carolyn Everson	0.000	1.000	0.000	0.0000
13	AMZN	2021-04-22	10:00AM	Chinese Billionaire Chen Tianqiaos Top 10 Stoc...	0.000	0.795	0.205	0.2023
14	AMZN	2021-04-22	09:51AM	Azure, Teams & Xbox Adoption to Aid Microsoft ...	0.000	1.000	0.000	0.0000
15	AMZN	2021-04-22	08:49AM	Amazons AWS Partners With DISH Network	0.180	0.820	0.000	-0.0258
16	AMZN	2021-04-22	08:45AM	3 Lessons From Jeff Bezos' Final Letter to Sha...	0.000	0.855	0.145	0.1779
17	AMZN	2021-04-22	08:30AM	4 Top Retail Picks for 2021	0.000	0.690	0.310	0.2023
18	AMZN	2021-04-22	08:05AM	Thursdays Pre-Market: Heres What You Need To K...	0.000	1.000	0.000	0.0000
19	AMZN	2021-04-22	07:45AM	3 Reasons to Buy and Hold This E-Commerce Stock	0.000	1.000	0.000	0.0000

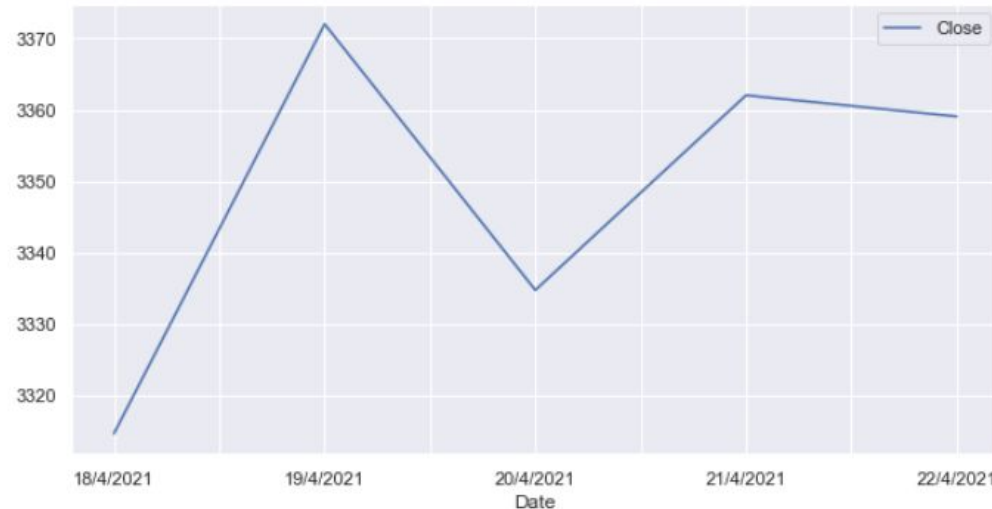
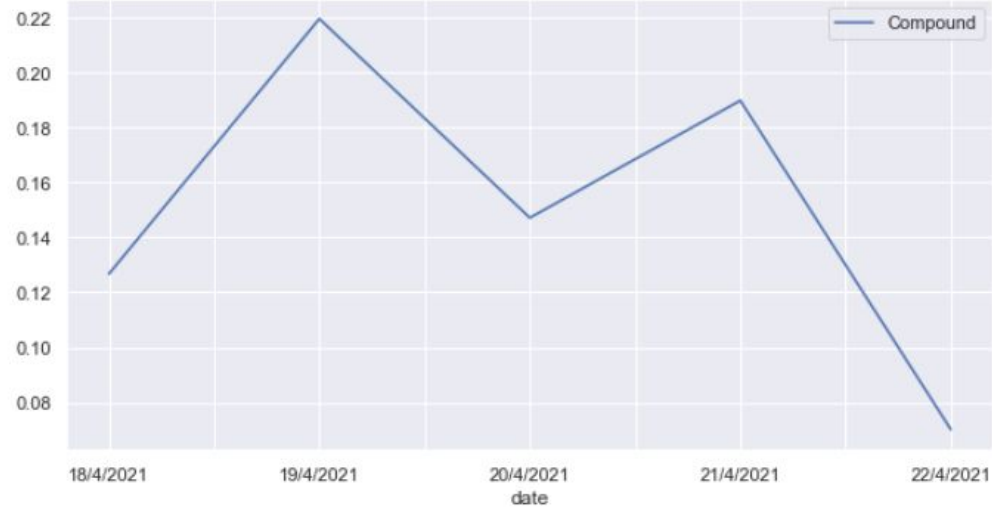
Histogram based on compound values

- Y axis represents the compound scores
- X axis is the last 7 days worth of data on AMZN
- This correlation was further evidenced by similar movements in AMZN stock price graphed next:



Trend analysis

- First chart shows the variation of sentiments with respect to time in the last week.
- Second chart shows the variation of closing price with respect to time in the last week,
- We see a quite a strong correlation between sentiments and predicted price.




Data Driven Insights



- While using public sentiments and opinions as a predictor of stock price was quite fun to play around with, we see that technical indicators such as SMA and EMA are obviously much more accurate to predict closing prices for AMZN, with a correlation score of more than 0.9
- Our multivariate regression model had an R^2 value of 0.896 and a MSE value of 1419.193.
- The coefficients of regression are 2.963 and -1.866 for EMA and SMA respectively. This indicates that if we are observing the closing price by using EMA and hold all the other independent variables constant, the closing price will change by a factor of 2.963.
- These values show that SMA and EMA are extremely good variables for prediction of the closing price of the stock.
- Moreover, positive and negative outlooks of the public on amazon affect Amazon's trading stock price as seen by the plots and a correlation score of 0.64

Improvements and Sources of Errors



Access to a larger time frame of twitter data could have lead to a more vast comparison of sentiments and their relationship with stock prices for Amazon. More knowledge on natural language processing syntax could have also yielded a more accurate result.



Split up of Work

Neel

- Utility functions for extracting data from ticker
- Web scraping and extracting data from twitter API
- Sentiment Analysis and Trend Analysis

- Importation of relevant libraries need for codes
- Tidying up of code lines for better readability
- Exploratory Analysis

Sashank

- Data extraction into variables
- Codes for combining the data
- Plots of stock prices vs dates, technical indicators vs dates, Rolling mean vs dates.
- Multivariate linear regression

Wei Ling



References and Citations

- (n.d) 'Sentiment Analysis using Vader', article retrieved from:
[https://towardsdatascience.com/sentimental-analysis-using-vader-a3415fef7664\](https://towardsdatascience.com/sentimental-analysis-using-vader-a3415fef7664)
- (n.d) 'Multivariate Linear Regression' article retrieved from
<https://www.hackerearth.com/practice/machine-learning/linear-regression/multivariate-linear-regression-1/tutorial/#:~:text=This%20is%20quite%20similar%20to,due%20to%20the%20added%20variables,\>
- (n.d) Visualizations, article retrieved from
https://pandas.pydata.org/pandas-docs/stable/user_guide/visualization.html
- (n.d) 'NLP for beginners', article and source files retrieved from
<https://www.analyticsvidhya.com/blog/2018/02/natural-language-processing-for-beginners-using-textblob/>



Thank you.

