

Stock Change Prediction Utilizing Social Media Pools

Michael Petrillo

Colorado State University – Global Campus

ABSTRACT

Online interaction has transformed the world's communications over the last decade (Grange, 2018). Users write comments, posts, and replies creating massive data repository-creating user-generated content (UGC). Social pools knowledge bases that create unique opportunities (Grange, 2018). Pooled UGC around the New York Stock Exchange (NYSE) a unique collection of information into the insights of trading. A subset of the NYSE indexes and stocks will include, Amazon, Walmart, Standards and Poor's Index, Apple, Netflix, Disney, Nvidia, Boeing, the Volatility Index and Telsa. NYSE is an extremely difficult and moving target showing the potential of using UGC for machine learning (ML) techniques such as support vector machine (SVM). Classification represented by integers from -2 to 2 standing for changes in the market corresponding to decline, minor decline, flat, minor growth or growth. Prediction is based on the change in delta percentage of the close price and the open over the open price from month to month (Akhter T. & Yong O,2019). Utilizing SVM requires a large amount of data, therefore the prediction is the predicted change forecasted in the next hour instead of month to month. The overall findings based on the scoring of accuracy of the SVM model had improved scores for many the experimental groups over the control showing the hypothesis that UGC is helpful for NYSE SVM predictions.

RESEARCH HYPOTHESIS

The hypothesis is focused on determining if an SVM ML model can utilize stock market pooled UGC to show an improved accuracy for a specific index or stock. The control group is a dataset with the same financial information of the stock market but without the UGC. Utilizing UGC the hypothesis is that at least one of the combinations of subreddits, will have a higher accuracy score for the test dataset compared to a control group score of the test dataset. Using no subreddit data as a control group, the SVM will either change or see no benefit from using the UGC. If the

UGC is able to score higher in the test vs train dataset then can be value to incorporating the data to advance models.

The paper's focuses could be qualitative research on how to set up an ML around UGC, but the goal is quantitative research to show the overall impact of the UGC. The overall product of the analysis will be a matrix of each subreddit and subreddit combination showing the score of the test data sets performance in relation to each other combination.

Stock Change Prediction Utilizing Social Media Pools

Veteran traders have admitted to underestimating the 900,000-user analysis on stock trading on Reddit (Kawa, 2020). All the insights are documented records in Reddit that can be turned into wangled data points for machine learning. Since all of the information and insights are freely available (Boe, 2020). The challenge is understanding how to take unstructured data to allow for predictions to emerge from the analysis. Being able to create an accurate model utilizing data pooled together focused on understanding an advanced dataset as the NYSE would show the power in data analytics. Pooled social media UGC is a valuable tool for research objectives (Grange, 2018).

OBJECTIVES

Understanding that UGC is as valuable as the metrics found within the data are important to improving how data sets are developed. Data mining and parameter tuning are major components for developing ML around NYSE datasets (Göçken et al, 2019). Creating strong relational metrics for the input data is key. Support vector machine (SVM) was chosen based on the scikit-learn "choose the right estimator" (scikit-learn developers, 2020). There was not enough data to try a regression model and there was no text data. Data mining will create a

dataset to be used but configuring the dataset to be useful for the SVM and controlling the interpretation is required for a successful mode. While there is a wealth of information around quantifiable values for correlating to the NYSE, opening up pool information from UGC will present key insights (Grange,2018). By approaching the meta-analysis for quantitative research, the reader can see different avenues to invest in a qualitative exploration for further perfecting an ML technique around UGC.

OVERVIEW OF STUDY

The dataset for the NYSE is from January 30th, 2020 to April 23rd of 2020. The UGC from Reddit are captured from January 1st, 2019 to April 24th of 2020 but only used within an hour of the NYSE dataset. The dataset is collected on a 30-minute interval over the last 60 days which is the largest amount of granular and range of time allowed by google. Scikit-learn supplies a test and training scoring based on the performance of two randomized data sets (scikit-learn developers, 2020).

LITERATURE REVIEW

Social Media Pools. Reading for understanding UGC is based “The Generativity of Social Media: Opportunities, Challenges, and Guidelines for Conducting Experimental Research” to understand how to account for different pools of information (Grange, 2018). Defining social media data as a valuable tool for research was key to picking the information for data analytics. Overall, the research is around using research pools for stock information so knowing how to conduct the experimental research is key.

Reddit Praw. Pulling the data from Reddit to use the comments, replies and titles as an open source requires creating an extraction, transform and load (ETL) process. Understanding the python-based programming library Praw allows turning Reddit into an open-source (Beo,

2020). Praw utilizes AWS and the methods for pulling the data are not intuitive, the data source is not structured like SQL. By reading the Praw documentation, setting up the ETL based on the functions was possible.

Big Data Ethics. Ethical decisions around storing and associating the user are based on the “Ethics of Big Data” (David & Patterson, 2012). Keeping in mind that the data is still personal and important even though it may seem massive each comment is important to the user. By understanding the nature of data is based on usage. “Using Ethical Reasoning to Amplify the Reach and Resonance of Professional Codes of Conduct in Training Big Data Scientists” has insights into structuring research projects such that the information collected cannot be used for other purposes (R. E. Tractenberg et al, 2014). Keeping the overall design focused on research and not trying to leverage data points such as usernames or use unknown software is key.

Stock Analytical Trends. The choices and set up for machine learning experimentation is from "Stock price prediction using hybrid soft computing models incorporating parameter tuning and input variable selection" to help set up a base for the meta-analysis and choose values that are well defined for parameter tuning (Göçken et al, 2019). The machine learning stock price prediction had issues with limitations in the data. Additional parameters and recommendations for machine learning for asset management using Pandas and Numpy (FirmAL, 2020). Learning how to understand the metrics for the financial data were based on the library Yfinance (Aroussi, 2019). By focusing on what could be gathered and looking at previously known limitations, the reading help create a base for the setup.

The basis for the target variable for prediction similar to “Adaptive market hypothesis and momentum effect: Evidence from Dhaka Stock Exchange” but using day to day instead of month to month (Akhter T. & Yong O ,2019). A base script for the machine learning techniques

is from the documentation of scikit-learn (scikit-learn developers, 2020). The machine learning programing and test scoring from the trained data is using coding from scikit-learn developers (scikit-learn developers, 2020). The specific SVM of Support Vector Classification was chosen based on the reading for SVM (scikit-learn developers, 2020).

By collecting reading on social media, looking at the gaps with current machine learning attempts in data available and metrics the goal became creating a method around UGC for the NYSE. Making an applied method required creating scripting and data processes around Praw and Yfinance that help develop the model.

RESEARCH DESIGN

Methodology

Financial Dataset. The stock market data is available from Yahoo's finances. The data can be pulled in real-time from an application programming interface (API) but is also available as a daily report out (Aroussi, 2019). The values pulled as Date, Open, High, Close, Volume and Adj Close (Aroussi, 2019). The NYSE dataset are pulled in increments of a half an hour. The data used to predict the next hour will be based on the on a time range of current time to as far back as an hour and a half.

Overall Descriptive Statistics of Each Subreddit

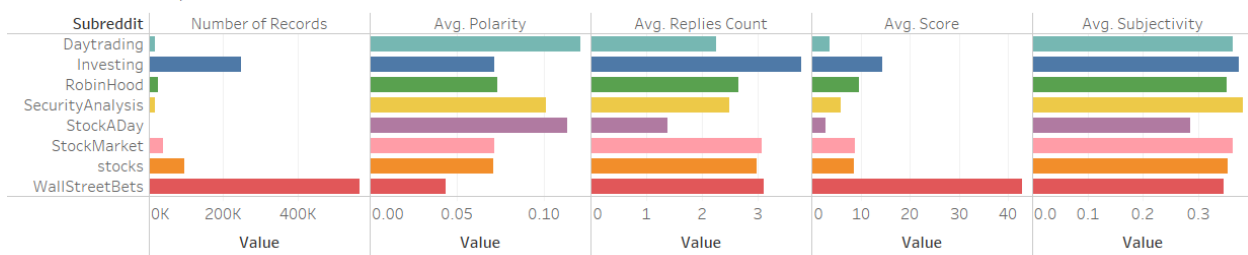


Figure 1. The descriptive statistics around the content created in each subreddit showing their diversity in subjectivity, polarity, scoring and data collected.

Reddit Dataset. Reddit allows users to pull the data from their website. Utilizing the package Praw, the Reddit development team allows for querying the data from specific subreddits (Boe, 2020). Looking at Figure 1, the Reddit dataset shows the demographical differences between the user pools based on the UGC. Praw utilize an API system that allows users to have a secret key to track the amount of data pulled so that they do not abuse the system (Boe, 2020).

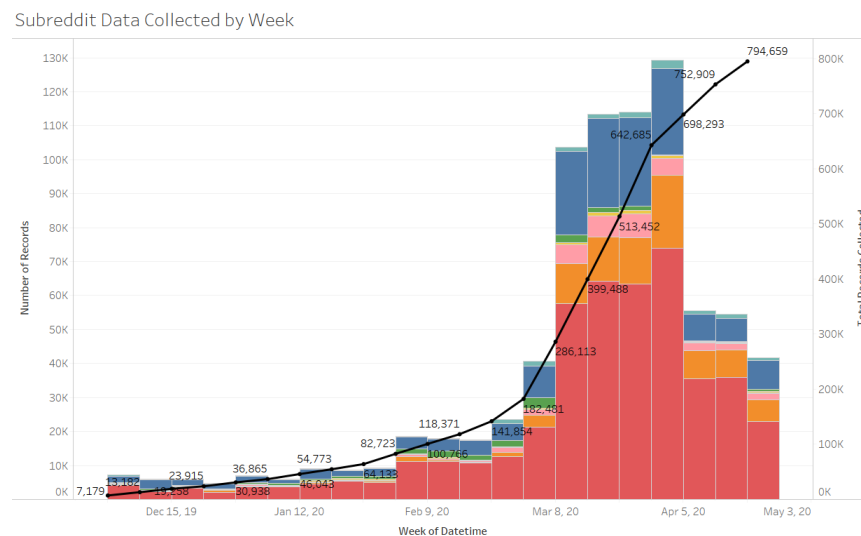


Figure 2. Number of Subreddit comments, replies and titles collected over the 8 weeks research project.

Reddit data is gathered based on the top or hot posts meaning that the top or hot posts will change throughout the day and can change drastically based on what happens in the stock market as seen in Figure 2. There is also no real way to ensure every post is gathered but a significant sample size of posts will give an indication of the overall group. To ensure that the data is not duplicated Praw has a unique key that can be used to only add values to the MySQL DB if they have not been previously added. Praw required at least python 3.5 and a knowledge of how Reddit works to understand the tree structure of the data. Each post can have a subset of comments and each comment can have nested replies (Boe, 2020). Reddit collects metadata

around each element that can be retrieved from Praw (Boe, 2020). Therefore, an open data source of UGC is available for research.

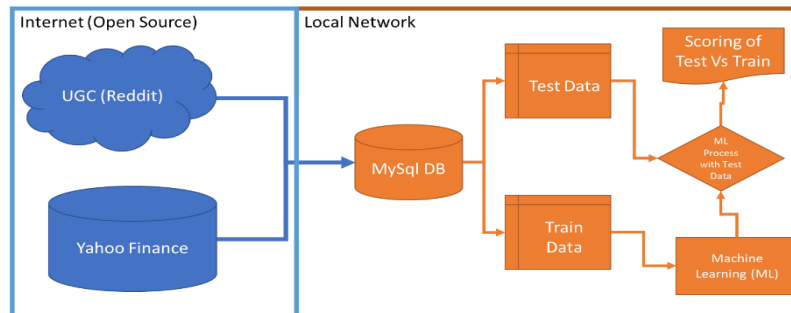


Figure 3. Data Process to ensure data security utilizing localized network and MySQL Security.

The tech stack for the extract, transform and load (ETL) process is Python 3.7 and MySQL as shown in Figure 3. Python will manage the API process from yfinance and Praw that can be automatically triggered daily utilizing windows task scheduler. As python pulls in the data and is stored in memory, the script will transform the data to better suit the analysis. Pandas, a python data framework, will collect the data from the yahoo finances and Praw to transform the values before sending the data to MySQL (pandas, 2020). Data exploration and visualization will be done in Tableau utilizing open database connectivity to MySQL.

Methods

NYSE Metrics. While the data from yahoo finance is structured the data, the data will need to be processed to be able to relate to social media data. Looking at the open, close, and high will give a sign of the changes in the market throughout the day. Delta of the open and close will show the change throughout the day. The delta of the open and close over the open will show the percentage of change. Calculating the daily return as the adjusted closing value/price minute the previous closing value/price over the previous closing value/price has been shown to measure the momentum anomaly on a monthly basis but will be altered increments of 30 minutes

for this analysis (Akhter T. & Yong O,2019). Values are collected and merged with the previous hour of trading and the previous hour and a half before that using the equations using return series (1)(2)(3)(4). The intent is to allow for metrics to provide data for the target value and baseline of the stock performance. Predicting the stock price purely on UGC would be impossible. The goal is to allow the SVM to utilize the comments and stock dataset to predict a change.

$$\Delta_{future} = (Close_{future} - Open_{current}) / Open_{current} \quad (1)$$

$$\Delta_{current} = (Close_{current} - Open_{previous}) / Open_{previous} \quad (2)$$

$$\Delta_{previous} = (Close_{previous} - Open_{previous}) / Open_{previous} \quad (3)$$

$$Volume_Delta_{current} = (Volume_{current} - Volume_{previous}) / Volume_{previous} \quad (4)$$

Only the target value will use a value from the future dataset. Both the experimental and control group will use the same metrics created from the stock dataset. The dataset using UGC will have a different distribution of the data changing the overall importance of the values based on the number of comments as the aggregation of data is focused on a by comment. The control group of only stock data will have the same metrics, but nothing from the Reddit comments.

Reddit Metrics. The unstructured text data from Reddit requires manipulation. Each title, comment, and reply will be data mined for the title of the post, overall score, author, created date, number of replies, subreddit, type, polarity and subjectivity of the comment. The package TexBlob which using natural language processing tasks such as part-of-speech to figure out polarity and subjectivity (Loria, 2018). Additional all ten stock indicators will have a separate column to see if they are called out directly by the comment and a feeling if the comment is indicating to buy or sell. A unique comment, reply, and title value is added to prevent duplicate data when the script is rerun.

Field Name	Data Type	Description	Example
comment	string	The text from a reddit comment.	Buying Spy Puts.
score	bigint	The reddit score based on users.	4
author	string	comment author.	somedude_2020
created	datetime	date and time created.	43919.67014
replies_count	bigint	Number of nested responses.	10
subreddit	string	where did the comment come from in Reddit.	Investing
type	string	Is the data from a comment, title or response.	comment
polarity	double	Text Bolb's comment assessment of polarity	0
subjectivity	double	Text Bolb's comment assessment of subjectivity	0
gspc	double	Was the stock or index mentioned? Will the user buy calls or puts? 0 for only mentioned, 1 if the user is buying calls or stock and -1 if they are buying puts or selling.	0
vix	double		
aapl	double		
dis	double		
tlsa	double		
nflx	double		
ba	double		
wmt	double		
amzn	double		
nvda	double		
movement	double	Does the user mention buying,selling or both? 1 for buying, -1 for selling and 0 for both.	0

Table 1. The metrics created from the unstructured data from Reddit in Python.

By collecting the social metrics and turning the unstructured text into structured metrics the analysis can start on looking for correlations as shown in Table 1. Keeping the data focused on the latest increments of data allows for the Reddit data to be tied to events happening in the market. Reddit's culture of trying to be the most liked comment or first comment for an event helps create UGC that may help react or predict change (Aroussi, 2019). Each comment is assigned values based on their score, number of replies to be used in an aggregation of performance for that 30-minute window.

SVM model. Looking at the results of the data will require machine learning to analyze the large volume of data. The goal will be to figure out what metrics can be created from the social media data and the stock market results to show a correlation for the target variable (1). Utilizing Scikit-learn in python will help create machine learning process to evaluate the data

(Hao& Ho,2019). Scikit-learn provides comprehensive machine learning as a convenient toolkit for education (Hao& Ho,2019).

$$\text{accuracy}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} 1(\hat{y}_i = y_i) \quad (5)$$

The SVM model with the overall best accuracy score, as seen by equation (5), was kernel type poly, scale used for game and power transformation for preprocessing for the y or y_{true} compared to the $\hat{y}$ or y_{pred} (scikit-learn developers, 2020).

SVM True Vs Predicted - ^GSPC

Subreddit	Y	Pred Y				
		-1.5 or Less Delta %	-1.5 to -0.75 Delta %	-0.75 to 0.75 Delta %	0.75 to 1.5 Delta %	1.5 or Greater Delta %
No Subreddit Data	-1.5 or Less Delta %			9		
	-1.5 to -0.75 Delta %			36		
	-0.75 to 0.75 Delta %			492		
	0.75 to 1.5 Delta %			46	2	
	1.5 or Greater Delta %			4		1
StockMarket	-1.5 or Less Delta %	1		8		
	-1.5 to -0.75 Delta %		4	29		
	-0.75 to 0.75 Delta %			382		
	0.75 to 1.5 Delta %			38	8	
	1.5 or Greater Delta %			3		1

Figure 4. The classifications are a wide range due current market variance. The model utilizing StockMarket subreddit data was able to predict accurately negative changes which the pure stock dataset was not.

Limitations

Classifications. The flat classification of -0.75 to 0.75 is a very wide range as shown in Figure 4. If there was less variance in the market there might be more opportunity to home in on the range. If the overall goal was a qualitative approach the range would be a major concern. There are no false positives for the predictions outside of zero so the model could be useful for acting or not acting but overall the design was based on figuring out how to get the Reddit dataset positioned more than designing an applied model.

Data Source. Reddit does not have a way to ensure all comments and replies are collected as data collected are based off different groupings of the top, hot or controversial of the last day, week, month or year (Boe, 2020). Metrics for likes and replies are also based off a when the comment are pulled meaning that the information is a snapshot. As the comments, replies and titles are only pulled once they cannot be updated. A modification of the script could be made to update values but if the comment is deleted then the update would remove valuable data. Ideally, the number of comments in the 30-minute period could be added. As such, there is currently there is no assurance each comment, title and reply is recorded but this might be the best given that the snapshot might be a better understanding of the comment then if other changes are made.

Timing. Prediction of change of the market will also be based around the delta of openings meaning that there could be a large amount of fluctuation before the day that are not captured. The model would require looking at the update every 30 minutes for the next hour and trending over time. The data source would also need to ensure to wait until all of the reddit comments possible were captured within the 30 minutes, so the prediction of an hour is the hour minus the ETL, two and a half minutes. Creating a more applied model would have the amount of time pushed out further, the overall goal would need to encompass ensuring all of the data is collected properly for the time interval.

Ethical considerations

Research Goal. Teaching goals have a ethical framework for Big Data (R. E. Tractenberg et al, 2014). By having identified a goal as a hypothesis for the educational research the structure of the question has a finite end. The overall intent is designing a process to test and train ingested comments for utilization by machine learning techniques. Looking at the final

matrix product of tested and trained scores keeps the scope of the project from leaving the control of the research. The overall project is within a vacuum such that the data is not used for other projects. By keeping the data in a controlled environment like a localized MySQL server the intent of the collected data does not leave the research environment. MySQL server access is required per internet protocol which is only local access as the information is a closed loop on a laptop (Oracle, 2020). Additionally, the database (DB) is password protected.

User Generated Content. The information that will be shared outside of the project is a compiled result of the research of the tested and trained materials. An extract of the MySQL DB can be provided with encryption if required (Oracle, 2020). This will allow for showing the quantitative results without the risk of providing access to the collected information to other entities. By aggregating the data there is no ability to backtrack the source data without having the source data itself. Furthermore, the ML processes are able to be validated by an independent process collecting the same data and storing the information on an independent network. Knowing the ML process is also well-defined utilizing premade and validated methodologies (scikit-learn developers, 2020). Following the security and ensuring the privacy of the users is key to creating a research project for data analytics.

FINDINGS

Subreddit And Stock/Index Compared to SVM Skicit Accuracy Score

Subreddit	Symbol										Grand Total
	^GSPC	WMT	AAPL	AMZN	NFLX	DIS	NVDA	BA	^VIX	TLSA	
StockMarket	0.9684	0.9579	0.9579	0.8211	0.8000	0.8316	0.8000	0.6211	0.5158	0.4571	0.7731
Investing	0.9492	0.9576	0.9407	0.8475	0.8305	0.7881	0.7712	0.6017	0.5169	0.4474	0.7651
stocks	0.9684	0.9474	0.9474	0.8211	0.8105	0.8105	0.8000	0.5895	0.4947	0.4521	0.7642
No Subreddit Data	0.9492	0.9492	0.9407	0.8475	0.8305	0.7966	0.7627	0.6102	0.5169	0.4286	0.7632
ALL	0.9492	0.9492	0.9407	0.8390	0.8305	0.7966	0.7797	0.5932	0.5169	0.4286	0.7623
WallStreetBets	0.9407	0.9492	0.9322	0.8305	0.8220	0.7881	0.7627	0.5847	0.5085	0.4286	0.7547
Grand Total	0.9542	0.9517	0.9433	0.8344	0.8207	0.8019	0.7794	0.6001	0.5116	0.4404	0.7638

Table 2. Shows the overall accuracy scoring based on the test vs train data set for the SVM. The smaller subreddits had greater potential then the largest subreddit.

The results of the accuracy score are based on the target value or y_{true} matching the predicted value or y_{pred} as seen in equation (1) and the results in Table 2 (scikit-learn developers, 2020). Overall, three of the subreddits performed on average better than the control group of pure stock dataset. StockMarket had the highest accuracy average rating for the NYSE stocks and indexes but was one of the least active subreddits. WallStreetBets which was the most active was the worst performing.

CONCLUSION

The UGC was able to give a slightly better score in comparison of the test verse the training dataset based off the conditions for the SVM utilizing a classification of integers. The less active stock based subreddits had the better accuracy scores. This may be due to the overall active being pooled directly on stocks instead of other activities. The purpose of the hypothesis is to determine if there is a value of the Reddit pooled user generated content to give a stronger accuracy in comparison to the control group of no Reddit data. The Reddit data was able to push the SVM model to try predictions outside of the flat, which was the safe bet for the model's

accuracy. The investing subreddit did also have a larger overall score. None of the models performed well in stocks or index with higher volatility such as Telsa, VIX or Boeing which have had historical fluctuations.

RECOMMENDATIONS

Increasing the granularity of classification of the dataset reduces the accuracy of the overall model. As the SVM is a classification and not a regression there is less ability to look for trends. Looking at Table 3, there are key areas where the merged dataset of UGC and NYSE helps but would need to be further developed.

Experiment Vs Control Delta of Subreddit and Stock/Index Score based on SVM Accuracy

Subreddit	Symbol										Grand ..
StockMarket	0.019	0.009	0.017	-0.026	-0.031	0.035	0.037	0.011	-0.001	0.029	0.010
stocks	0.019	-0.002	0.007	-0.026	-0.020	0.014	0.037	-0.021	-0.022	0.023	0.001
Investing	0.000	0.008	0.000	0.000	0.000	-0.008	0.009	-0.009	0.000	0.019	0.002
No Subreddit Data	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ALL	0.000	0.000	0.000	-0.009	0.000	0.000	0.017	-0.017	0.000	0.000	-0.001
WallStreetBets	-0.009	0.000	-0.008	-0.017	-0.008	-0.008	0.000	-0.025	-0.008	0.000	-0.008
Grand Total	0.005	0.003	0.003	-0.013	-0.010	0.005	0.017	-0.010	-0.005	0.012	0.001

Table 3. Shows the overall scoring of the test vs train data set for the SVM with an increased granularity. Overall Investing had the greatest potential.

Getting the 100k data points recommend require for a regression model around 15.24 years' worth of data considering the data collect was 1078 data points over 60 days. More data points and better metrics created from data mining would help the overall performance to improve this. A model could be made with 5-minute increments instead of 30 minutes but that may affect the reddit data collected.

The Reddit user generated comments does empower the support vector machine to take more chances at predicting when a flaw will change, this could be perfected in a qualitative

analysis but as far as showing a quantitative different there is value in the relation to the pooled user data. If developed around a qualitative approach to the specific metrics, the test and training score should be able to improve for stocks with low variances.

REFERENCES

- Akhter, T., & Yong, O. (2019). Adaptive market hypothesis and momentum effect: Evidence from Dhaka Stock Exchange. *Cogent Economics & Finance*, 1. Retrieved from <https://doi.org/10.1080/23322039.2019.1650441>
- Aroussi, R. (2019). yfinance 0.1.54. Retrieved from <https://pypi.org/project/yfinance/>
- Boe, B. (2020). PRAW: The Python Reddit API Wrapper. Retrieved from <https://praw.readthedocs.io/en/latest/>
- David, K., & Patterson, D. (2012). *Ethics of Big Data*.
- FirmAI. (2020). <https://github.com/firmai/machine-learning-asset-management>. Retrieved from Machine Learning in Asset Management.
- Göçken, M. Ö. (2019). Stock price prediction using hybrid soft computing models incorporating parameter tuning and input variable selection. *Neural Computing & Applications*, 577–592. Retrieved from <https://doi.org/10.1007/s00521-017-3089-2>
- Grange, C. (2018). The Generativity of Social Media: Opportunities, Challenges, and Guidelines for Conducting Experimental Research. . *International Journal of Human-Computer Interaction*, 943–959. Retrieved from <https://doi.org/10.1080/10447318.2018.1471573>
- Hao, J. &. (2019). Machine Learning Made Easy: A Review of “Scikit-learn” Package in Python Programming Language. *Journal of Educational and Behavioral Statistics*, 348–361.
- Kawa, L. (2020, February). *Reddit’s Profane, Greedy Traders Are Shaking Up the Stock Market*. Retrieved from Bloomberg: <https://www.bloomberg.com/news/articles/2020-02-26/reddit-s-profane-greedy-traders-are-shaking-up-the-stock-market>
- Loria, S. (n.d.). *TextBlob*. Retrieved from textblob.readthedocs.io: <https://textblob.readthedocs.io/en/dev/>
- NumPy Developers. (2020). *NumPy*. Retrieved from <https://numpy.org/>
- Oracle. (2020). <https://dev.mysql.com/doc/mysql-security-excerpt/5.7/en/security.html>.
- Ramazan Gençay, M. D. (2001). An Introduction to High-Frequency Finance. In M. D. Ramazan Gençay.
- Roelants, P. (2019). *How to implement a simple RNN*. Retrieved from https://github.com/peterroelants/peterroelants.github.io/blob/master/notebooks/RNN_implementation/rnn-implementation-part01.ipynb
- scikit-learn developers . (2019). 1.10. *Decision Trees*. Retrieved from <https://scikit-learn.org/stable/modules/tree.html#tree>
- scikit-learn developers. (2011). *Generalized Linear Models*. Retrieved from https://ogrisel.github.io/scikit-learn.org/sklearn-tutorial/modules/linear_model.html

scikit-learn developers. (2019). *1.7. Gaussian Processes*. Retrieved from https://scikit-learn.org/stable/modules/gaussian_process.html

scikit-learn developers. (2020). *scikit-learn*. Retrieved from <https://scikit-learn.org/stable/>

scikit-learn developers. (2020). *sklearn.model_selection.train_test_split*. Retrieved from https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html

scikit-learn developers. (n.d.). *Choosing the right estimator*. Retrieved from https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html.

Tractenberg, R. E., Russell, A. J., Morgan, G. J., FitzGerald, K. T., Collmann, J., Vinsel, L., . . . Dolling, L. M. (2014). *Using Ethical Reasoning to Amplify the Reach and Resonance of Professional Codes of Conduct in Training Big Data Scientists*. *Sci Eng Ethics*.

Programming Code

Title: Reddit Monitoring Capstone

Author: Michael Petrillo

Date: May 1st, 2020

Code version: 1.0

Availability: https://github.com/mryagerr/reddit_monitoring_capstone?fbclid=IwAR2_2trwh000Ild3dH61JSkyQABFcKI3ZN3Y0itep0OGFMlnhAILHF8baow