**Abstract:**

*Paragraph synopsis of research*

This paper aims to evaluate the impact of social media sentiment on Stock Prices.

The research centers its analysis on public company Tesla, one of retail investors favorite stocks [1], and whose CEO, Elon Musk, is active on social media platforms. Both these facts create an interesting and extensive pool of data points that the paper exploits to extract sentiment analysis and further study its correlation with stock market performance.

Data aggregation combines (1) YTD Tweets (from January- October 2020) containing the world ‘Tesla’. These were extracted crawling the web for the tweet URLS (snscrapein call in Python), which granted us access to over 340,000 tweets. (2) YTD Tesla EOD (End of Day) stock prices extracted from Nasdaq API.

The paper concludes twitter stock sentiment on Tesla is correlated with its stock price performance. This judgement is made after contrasting different sentiment analysis libraries and stock prediction models, deploying various parameters to minimize error rates by contrasting results.

**Introduction:**

*Description of your project*

The focus of the project is to leverage the data preprocessing and modeling techniques learnt during the semester to research whether stock sentiment from Twitter is a statistically significant parameter in predicting Tesla stock performance.

The project consists of 3 phases: (1) Preprocess training data and train sentiment analysis extraction model, (2) Sentiment Extraction from our Tesla Stock dataset, (3) Stock return prediction.

1. After contrasting different possible training data sets, Sentiment140’s, consisting of over 500,000 tweets from 2009 was deemed adequate to train our model given its size, and format; It presents the same data preprocessing challenges as our test data, such as special characters.

The data was transformed with NLTK tokenizers and stemmers. TF-IDF later created a vector matrix that set the grounds to perform our research.

NLTK’s VADER lexicon was used to generate stock sentiment on this dataset, which we subdivided earlier on between training and testing. Accuracy score obtained was 80% and classifier comprised a Positive, Negative or Neutral tag.

1. Sentiment extraction from our Tesla tweets:

First we generated stock sentiment by replicating homework 3: we used the positive and negative word list provided to calculate the sentiment of the Tesla tweets as either positive, or negative, but 11% of the dataset didn’t include either positive or negative words. In order to improve the model we applied VADER, resulting in all tweets being assigned a sentiment tag.

1. Stock return prediction was motivated by not only the team member’s will to apply the material covered in class but also our interests in finance, financial modeling and machine learning techniques for estimation. The models applied are Linear Regression and Random Forest.

[1]https://www.investopedia.com/retail-faves-are-beating-hedge-fund-mutual-fund-picks-5024920