TradeSpade: Price Signal Forecast for Financial Assets

Anurag Bejju , Manan Parasher , Nikitha Ravi , Rishabh Singh

Objective

- Our product provides day traders assistance with intra day trading by predicting **Buy** and **Sell** signals in order to maximize profits and make optimized decisions.
- Depict the influence of social media and everyday news on market fluctuations.

Motivation

- Can we design a model that helps with Stock and Crypto Currency forecast based on features other than just OHLCV ?
- Can we find the impact of global factors on market volatility and derive the correlation between them?
- Which technical indicators are most important for market direction analysis ?

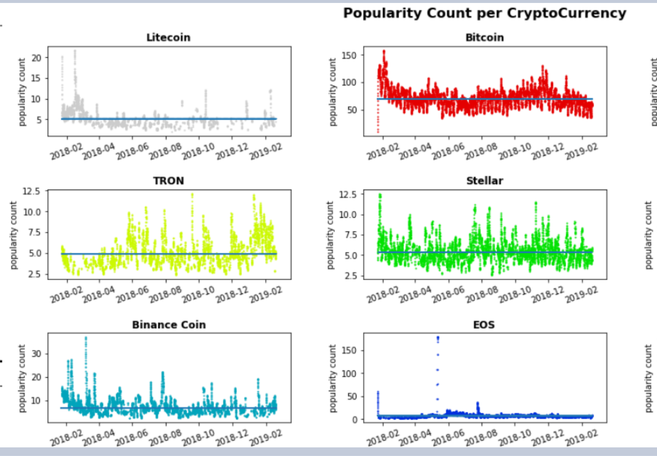
Data Collection (1st March 2018 – 19th Feb 2019) (Size > 50 GB)

Collected Twitter and Reddit data for social media quantitative metrics and financial as well as general news for qualitative metrics.

Category	Stocks			Crypto Currency			
Assets Used	Binance Coin, Bitcoin , Bitcoin Cash ,EOS, Ethereum , Litecoin, Stellar, TRON, XRP			Walmart Inc., Alphabet Apple Inc., Chevron, Exxon Mobil, Microsoft, Coca-Cola, Home Depot, Wells Fargo			
Data Type	Financial	General News	Financial News	Reddit		Twitter	Financial
Data Collected	OHLCV	News, Title, Publisher, Links	Title, Sub Title, Date, Publisher	Score, Subreddit, Title, Comments		Tweet, Comments, Re-Tweets, Likes	OHLCV
Collection Source	finam.ru	Custom News API - (Modified as per requirement)	Financial Times (Customized)	Pushshift API Metrics		Twitter API + Custom API for older Tweets.	Crypto-compare
Data Count	Hourly Data	6,875,000 News Articles	25,000 Financial News Articles	1,284,023 Reddits (Crypto)	557,391 Reddits (Stocks)	468,888 Tweets (Stocks)	328,704 Tweets (Crypto)	Hourly Data

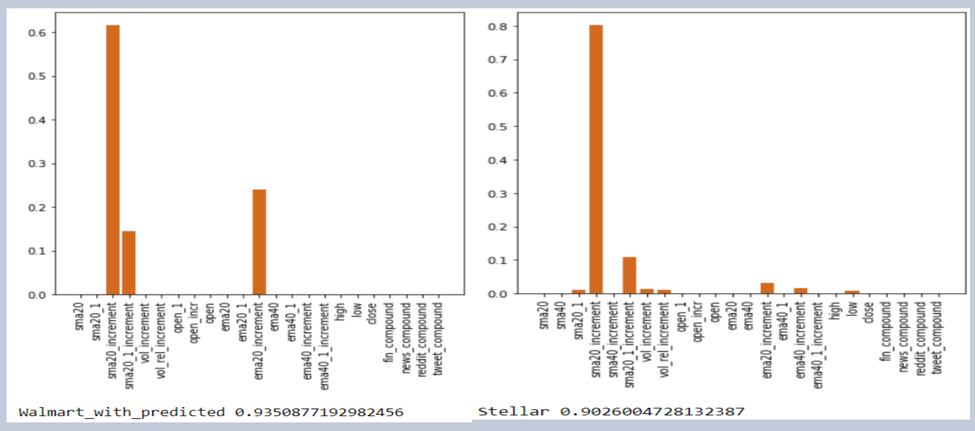
Exploratory Data Analysis

- Analyzed data collected from six different external sources.
- Performed NLP to extract sentiment scores and establish relationship between social media and news with stock and crypto assets.
- Performed in-depth data analysis to test reliability and authenticity of data being considered.
- Visualize trends in asset price fluctuations and the impact of sentiments scores of the various data sources.



Model Training

- Used different models to compute the price signals and obtained the best results with *XGBoost*.
- Used basic features like *OHCLV* as well as *simple moving averages, exponential moving averages* and *increments*, in addition to the news features from four different sources.



Visualization

- Developed and designed an interactive web application with multiple widgets and dynamic graphs.
- **Tools**: *Flask, JQuery, HTML, ChartJS, Apexcharts* ; **Hosted on:** *SFU Cloud*

Model Inference

- Out of all models, **XGBoost** provided the best results with **90%** accuracy for most stocks and close to **90%** for most cryptocurrencies.
- The most important feature was difference between the current and the previous SMA (Simple moving avg)20
- SMA40 proved to be more important for stocks than crypto.
- Including the news features improved the accuracy, but only marginally.

Conclusions and Future Work

- As future work, we can include other features such as Support and resistance, relative strength indicator and other breadth indicators to train our model.
- Also enable current implementation to work on live streaming data as well as provide real-time predictions.
- Predicting actual prices rather than just the market direction and see how good the models fare.