TEAM 093: Utilizing Social listening to identify Cryptocurrency trends

Motivation

- Traditional money market volatility is controlled by **central authorities**.
- Cryptocurrency market is however more driven by **public perception** and **confidence** in currency
- **Our solution reduces this uncertainty** by modelling external factors with investor sentiment.
- Our short-term goal is to help investors gain better return on crypto asset investment and a long-term goal is to increase acceptance and utilization of crypto asset class, making transactions more secure and fast.

Data

3 years of post and comment data was extracted from 2 most common sub-reddits **r/Cryptocurrencies** and **r/Cryptomarket** with **5.5M** total and 4.2k hourly active members. Collections of **446K posts** are collected at a rate of 12,000 top posts per month, extracted at day level with **~260k words per day**. Using Pushshift API, it took **21 hours** of crawling. Additional post parameters were collated from **172.8 M comments** associated.

An important factor to estimate future demand for a cryptocurrency, data on **retail sales, consumer sentiment** and **unemployment** was fetched from **Alphavantage.co**

Market price values for 100 crypto currencies were sourced from an existing dataset in data.world which is refreshed with an active connection to **coinmarketcap.com** with **~150K records** for 3 years.

AMD and **NVDIA** stock price along with global stock indices **DJI, GSPC, IXIC** and **NYA** and **Gold** were fetched from **Yahoo Financial**

Online search frequency data was gathered using **Pytrends** library for all currencies. A spearman correlation of 0.78 was observed with 60 days in bitcoin future price prediction

Novel Approach & Design

Our approach **educates users of ever-changing social cues** and understand the impact of association between them. It will also open a window to better understand general investors sentiments, dependencies of price. There are two major parts:

1. **Extracting and processing social media data** to identify conversation topics, finding association between them and quantifying popularity, trendiness and hotness of these topic. Along with correct identifications of investor sentiments, with models trained on the twitter conversations of past 3 years, understanding crypto jargons, associated slangs and emotions, compared with historically known approaches. **Testing and Improvement**: We explored relevance of "popular" tokens generated and their association to understand effectiveness and updated our logic to populate "total popularity score" which is a function of posts, comments likes and rewards, for rolling last 30 days.
2. **Quantifying the impact of external, derived financial factors along with quantified social media data** on crypto price and to gauge the prediction improvement. **Testing and Improvement:** We tuned XGBoost hyperparameters and filtered features based on GMIC values to maximize our prediction accuracy.

COMPARISON WITH OTHER METHODS: While some studies have factored social influence most studies performed till date focus only on **price prediction accuracies**[1][2][3]. **None has worked on answering the question of crypto trends illiteracy and causal sentiment analysis.** Our tool fills these gaps by educating users of ever changing social cues and understand the **impact of association** between them. It opens a window to better understand general investors sentiments, dependencies of price trend on specific sentiments and finds the most probable explanation of market movements, therefore result in more informed decisions.

Prediction Model using **Fundamental** and **Technical** Factors

Reddit data **Topic Modelling**

Post Topic identifier — 1

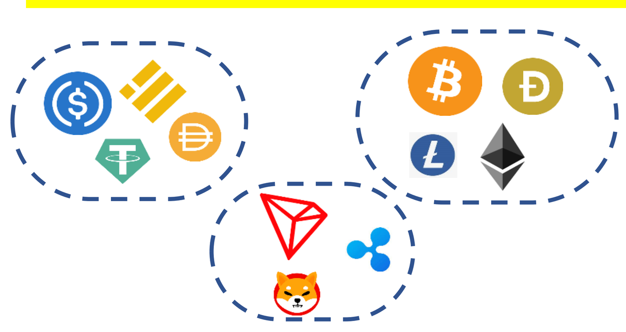
Topic importance scoring: Popularity, Hotness, Trend — 2

Post Topic Prioritization and Associations — 3

Topic associated Sentiment Identification and Aggregations — 4

Feature Engineering

Similarity and Cluster Analysis



XG Boost based Price Forecasting Model (Social + Technical + Fundamental Factors)

For each crypto

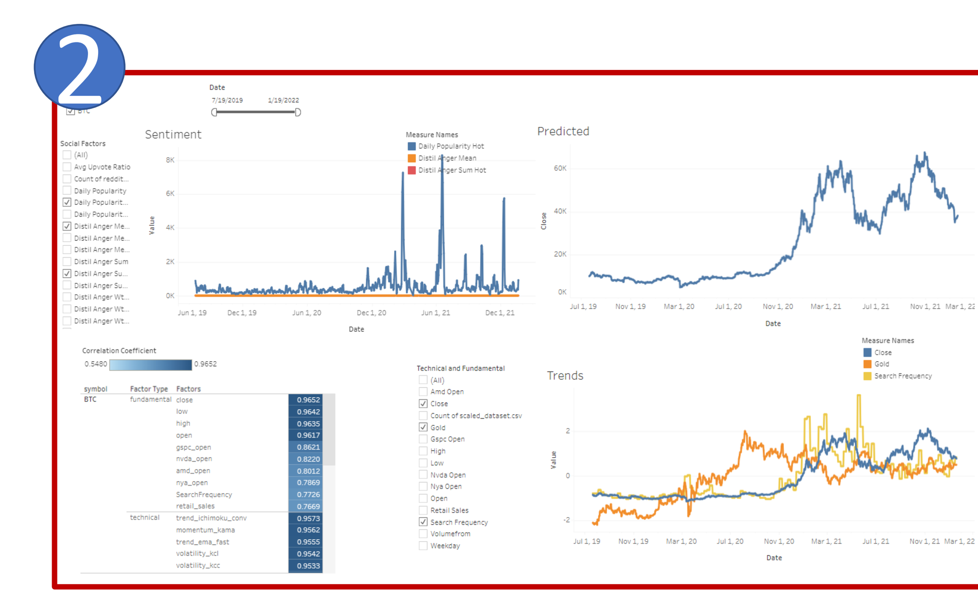
User Experience

Testing and Improvement: Ease of visualization was tested by outside user ratings and comments.

Screen 1 helps user develop contextual understanding of new crypto currencies what it is about, other associated currencies and trending comments on Reddit.

Screen 2 helps user gauge the price movement and prediction of different cryptocurrencies and further helps user make investment decision by quantifying the stakes.

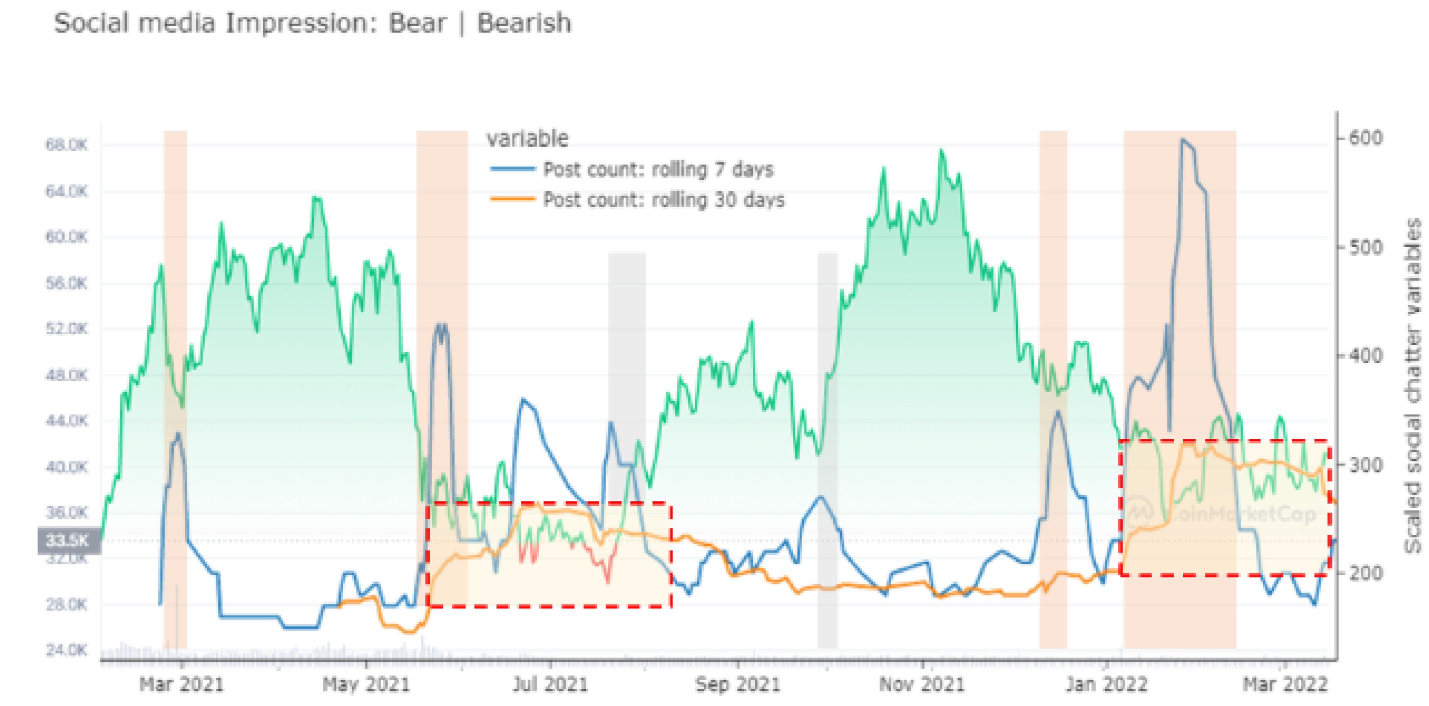
Overall, we received positive feedback for our visualization and **8/10** people felt that they can now make a more informed investment decision



Experiments and Results:

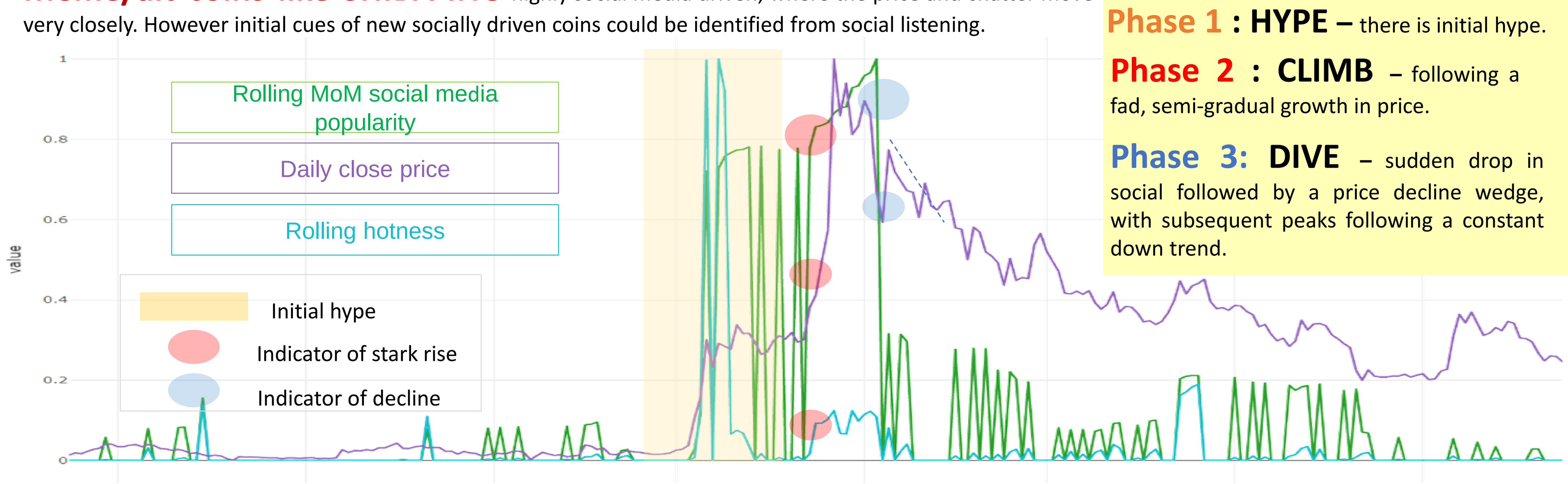
1 Investor conversations associated with specific subtopics, contain information about their views, news or applications.

The figure below shows user conversation involving "bear" an "bearish" effectively shows the realization of fear in investors, where the **conversations containing these keywords spike after the start of dip cycle and its largest initial drop.**
Width of the spike is proportional to impact of the drop. While the conversation of these keyword remain higher than the 1.96 σ level of the conversations before the dip, the price value was observed to stay down or worsened, gauging the width of possible down period.



2 Impact of Social Sentiment factors differs for different coins, GMIC of ~0.73 with price is observed for SHIBA INU whereas for Ethereum it is ~0.41

- **Meme/alt coins like SHIBA INU** highly social media driven, where the price and chatter move very closely. However initial cues of new socially driven coins could be identified from social listening.



Phase 1 : HYPE – there is initial hype.

Phase 2 : CLIMB – following a fad, semi-gradual growth in price.

Phase 3: DIVE – sudden drop in social followed by a price decline wedge, with subsequent peaks following a constant down trend.

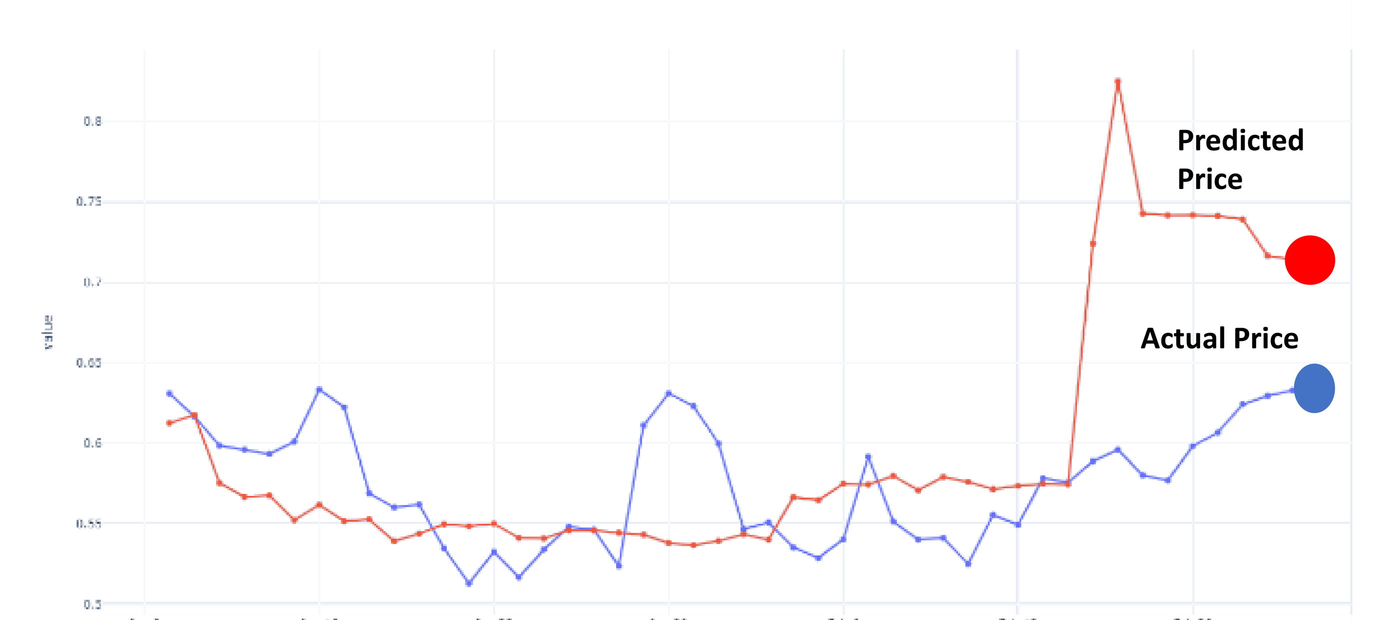
Whereas relatively **stable coins** like **ETHEREUM** follow a different trend. This information was vital for us to tune parameters for each currency model and make accurate predictions.

Phase 1 : HOPE – avg weighted positive sentiment increases

Phase 2 : BEGINNING initialization of short term high growth

Phase 3 : CLIMB – the increase in price Is generally steep,

3 Our XGBoost model achieved an MSE 0.005 of for a 30 day horizon for Bitcoin using learning rate: 0.05, L1: 0.17, L2: 0.01. Most studies till now have focused on a much smaller horizon and achieved relatively less accuracy over longer period [4][5].



REFERENCES: [1] How Does Social Media Impact Bitcoin Value? A Test of the Silent Majority Hypothesis,2018, Journal of Management Information Systems 35(1):19-52 [2] Global cryptocurrency trend prediction using social media, 2021, Information Processing and Management 58(6):102708 [3] Crypto Currency Prediction Model using ARIMA, Turkish Journal of Computer and Mathematics Education Vol.11 No.03 (2020), 1654-1660 [4] Time Series Analysis of Cryptocurrency returns and volatilities, Journal of Economics and Finance (2021) 45:75-94 [5] Forecasting cryptocurrency prices time series using Machine Learning,2019, SHS Web of Conferences 65(1):02001