Bitcoin Evolution Analytics: Twitter Sentiments to Predict Price Change as Bearish or Bullish

Naman Srivastava
Indian Institute of Information Technology Dharwad
srinaman2@gmail.com

Omkar Gowda
Indian Institute of Information Technology Dharwad
omkarkgowda@gmail.com

Shreyas Bulbule
Indian Institute of Information Technology Dharwad
shreyasbulbule007@gmail.com

Siddharth Bhandari
Indian Institute of Information Technology Dharwad
siddharthrbhandari2002@gmail.com

Dr. Animesh Chaturvedi
Indian Institute of Information Technology Dharwad
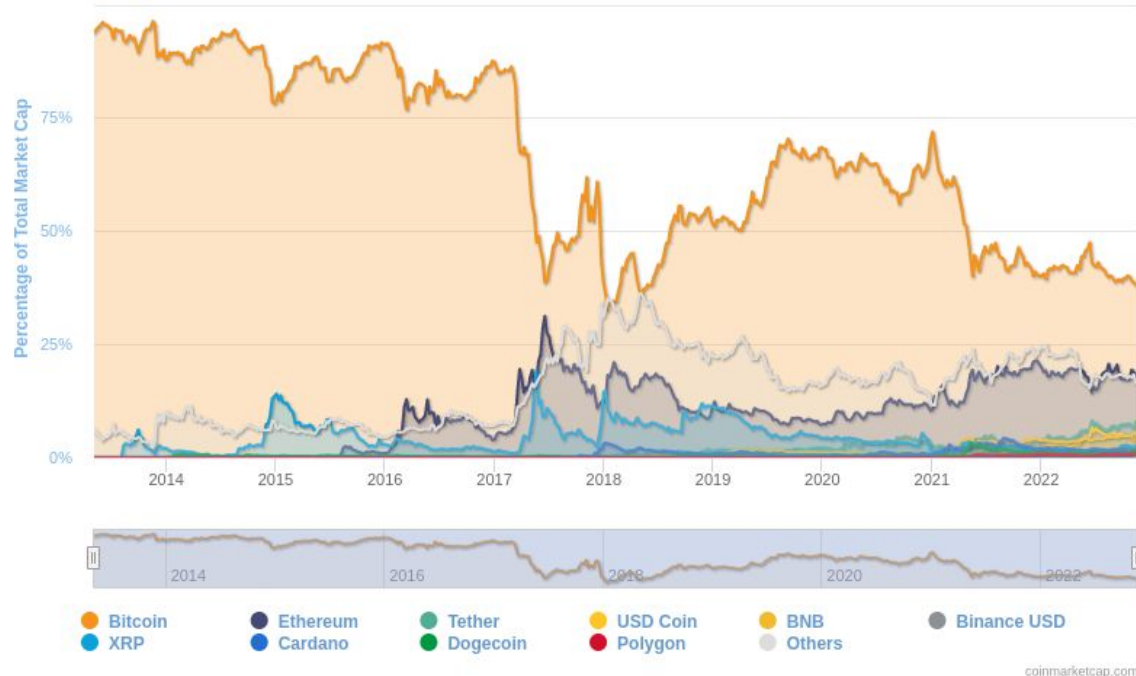animesh.chaturvedi88@gmail.com

**2022 IEEE International Conference on Big Data (IEEE BigData 2022)
December 17-20, 2022 @ Osaka, Japan**

Motivation to study Bitcoin

Bitcoin still holds the highest market capitalization of more than 300 Billion USD*.

Over 100 million people own bitcoin

The Bitcoin market is extremely volatile.

Percentage Market Cap For Cryptocurrencies

Image Credit: Global Cryptocurrency Market Charts | CoinMarketCap

* As of November 2022

How does Twitter Affect Bitcoin

Social Media platforms like Twitter can provide us information about what people are saying.

For entities like Bitcoin, that are difficult to assign a fixed value are heavily influenced by sentiments of investors.

However, all tweets have different effect on the price.

Image Credit: Buyucoin

Research Questions (RQ)

We aim to perform the Prediction of Bitcoin Big Data as either Bearish or Bullish

RQ1: How to predict Bitcoin price change with binary classes (Bearish or Bullish) using sentiment in Tweets?

RQ2: Is the "*UserSentiment*" score a better measure for predicting Bitcoin price change as either Bearish or Bullish?

RQ3: Which machine learning models give the highest F1 score and accuracy for predicting price change when considering Sentiments, Followers, and Technical Indicators?

System Evolution Analytics

We define (S_i, ER_i, t_i) for an evolving system represented as a **State Series** (**SS**),
a collection of states (or data points)

$$SS = \{S_1, S_2 \dots S_N\} \text{ at various time points } \{t_1, t_2, t_3 \dots t_N\}.$$

Analogous, to the works of Animesh Chaturvedi et al. published in IEEE Computer Society Conferences and IEEE Transactions/Journals

A. Chaturvedi and A. Tiwari, "System evolution analytics: Deep evolution and change learning of inter-connected entities," in 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 3075–3080, IEEE, 2018.

A. Chaturvedi and A. Tiwari, "System evolution analytics: Evolution and change pattern mining of inter-connected entities," in 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 3877–3882, IEEE, 2018.

A. Chaturvedi, A. Tiwari, D. Binkley, and S. Chaturvedi, "Service evolution analytics: change and evolution mining of a distributed system," IEEE Transactions on Engineering Management, vol. 68, no. 1, pp. 137–148, 2020

A. Chaturvedi and A. Tiwari, "System network complexity: Network evolution subgraphs of system state series," IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 4, no. 2, pp. 130–139, 2018.

A. Chaturvedi, A. Tiwari, and N. Spyratos, "minstab: Stable network evolution rule mining for system changeability analysis," IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 5, no. 2, pp. 274–283, 2019.

A. Chaturvedi and A. Tiwari, "Sysevorecomd: Graph evolution and change learning based system evolution recommender," in 2018 IEEE International Conference on Data Mining Workshops (ICDMW), pp. 1499–1500, IEEE, 2018.

A. Chaturvedi, A. Tiwari, and S. Chaturvedi, "Sysevorecomd: Network reconstruction by graph evolution and change learning," IEEE Systems Journal, vol. 14, no. 3, pp. 4007–4014, 2020.

A. Chaturvedi, A. Tiwari, S. Chaturvedi, and P. Li`o, "System neural network: Evolution and change based structure learning," IEEE Transactions on Artificial Intelligence, vol. 3, no. 3, pp. 426–435, 2022.

Bitcoin State Series

Bitcoin State Series (Bitcoin-SS) is a collection of states (or bitcoin data points), such that :

$$SS = \{S_1, S_2 \dots S_N, S_{N+1}\} \text{ at (N+1) time points } \{t_1, t_2 \dots t_N, t_{N+1}\}$$

The bitcoin state series can be pre-processed to make an evolving BiTcoin Data

$$BTDs = \{BTD_1, BTD_2, BTD_3 \dots BTD_N, BTD_{N+1}\}$$

There exist a relationship between (S_i, BTD_i, t_i), such that the bitcoin state S_i and the evolving BiTcoin Data BTD_i are representing bitcoin big data at the ith time point t_i, where 'i' varies from 1 to $N+1$.

Bitcoin State Series from the BiTcoin Data

Time Stamp	User Sentiment Score of VADER Positive	User Sentiment Score of VADER Negative	User Sentiment Score of VADER Neutral	User Sentiment Score of VADER Compound	User Sentiment Score of BERT	User Sentiment Score of Polarity	User Sentiment Score of Subjectivity	Aroon Down	Aroon Up	AD Line	Label
2021-02-05 12:00:00	11.08	0	34.91	31.7768	32.8066	25.3	36.8	16	44	3291559810	Bullish
2021-02-05 13:00:00	102.37	0	8244.62	334.86	7675.87	204.62	546.49	16	44	3290077562	Bearish
2021-02-05 14:00:00	65.792	0	9898.20	139.11	9285.61	2272.22	4229.49	16	44	3279059002	Bearish
2021-02-05 15:00:00	0	36.036	359.96	-108.19	236.64	0	184.8	16	44	3282308951	Bearish
2021-02-05 16:00:00	1589.42	222.30	12193.28	4070.4	10751.2	1862.41	3887.97	16	44	3289743917	Bullish
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
2022-02-18 23:00:00	3589.18	1720.63	1622.19	7736.58	1367.40	1124.30	7852.39	68	12	-284511124	Bullish

Bitcoin State Series from the BiTcoin Data

Each time point is per hour basis.

the state endpoint is Closing price of the market per hour.

In a day of *24* hours, it will be *24* time points and *24 BTDs.*

In a week of 7 days, it will be *(7 X 24)* time points and (7 X 24) *BTDs.*

Bitcoin Evolution Learning Models

The change at a given TimeStamp is to be determined as a function of the Bitcoin State Series created using the BiTcoin Data (BTDs)

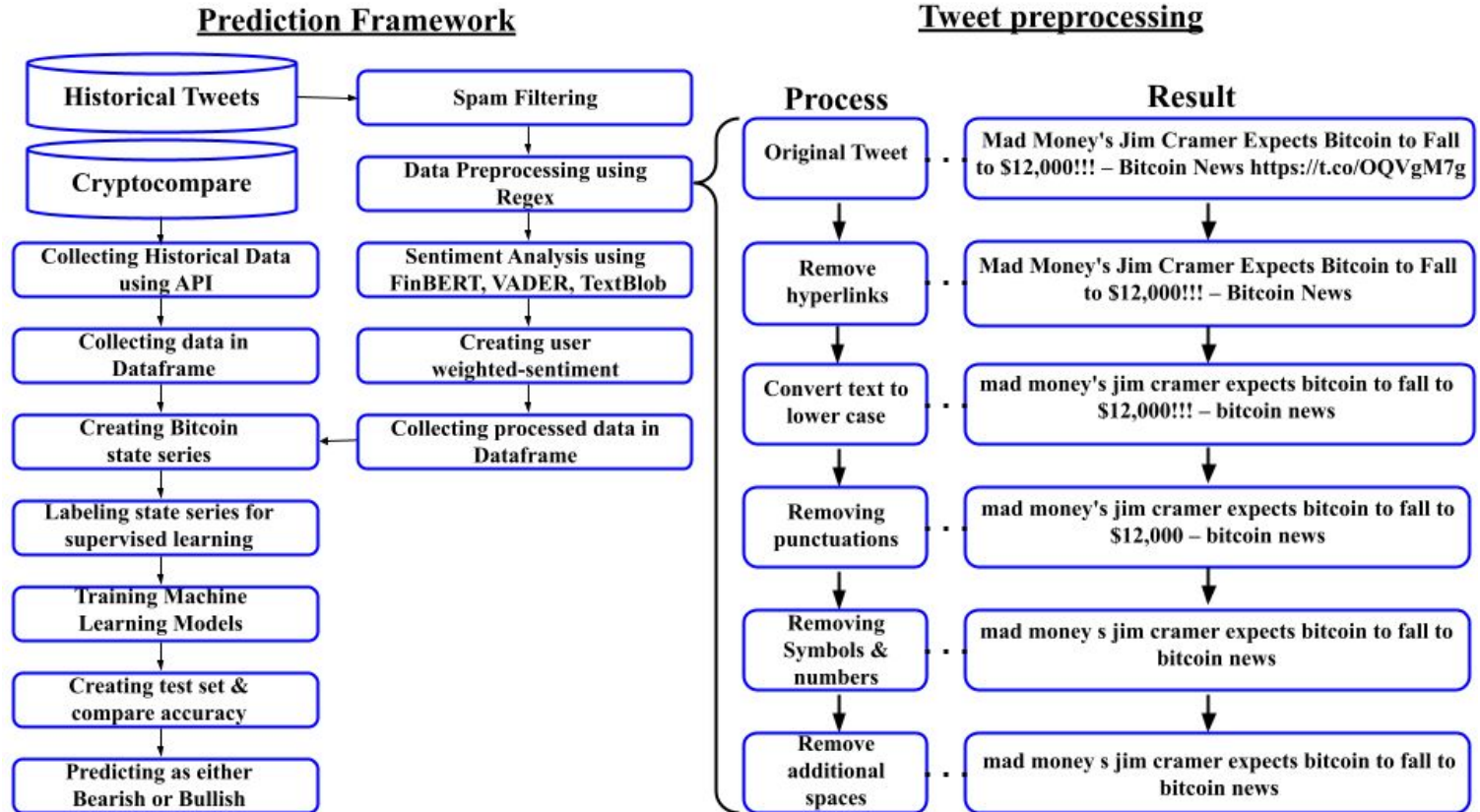To study the relation between price change and the features generated.

Each record in the training data is labelled (as Bearish or Bullish) we can use them to train supervised Bitcoin Evolution Learning Model.
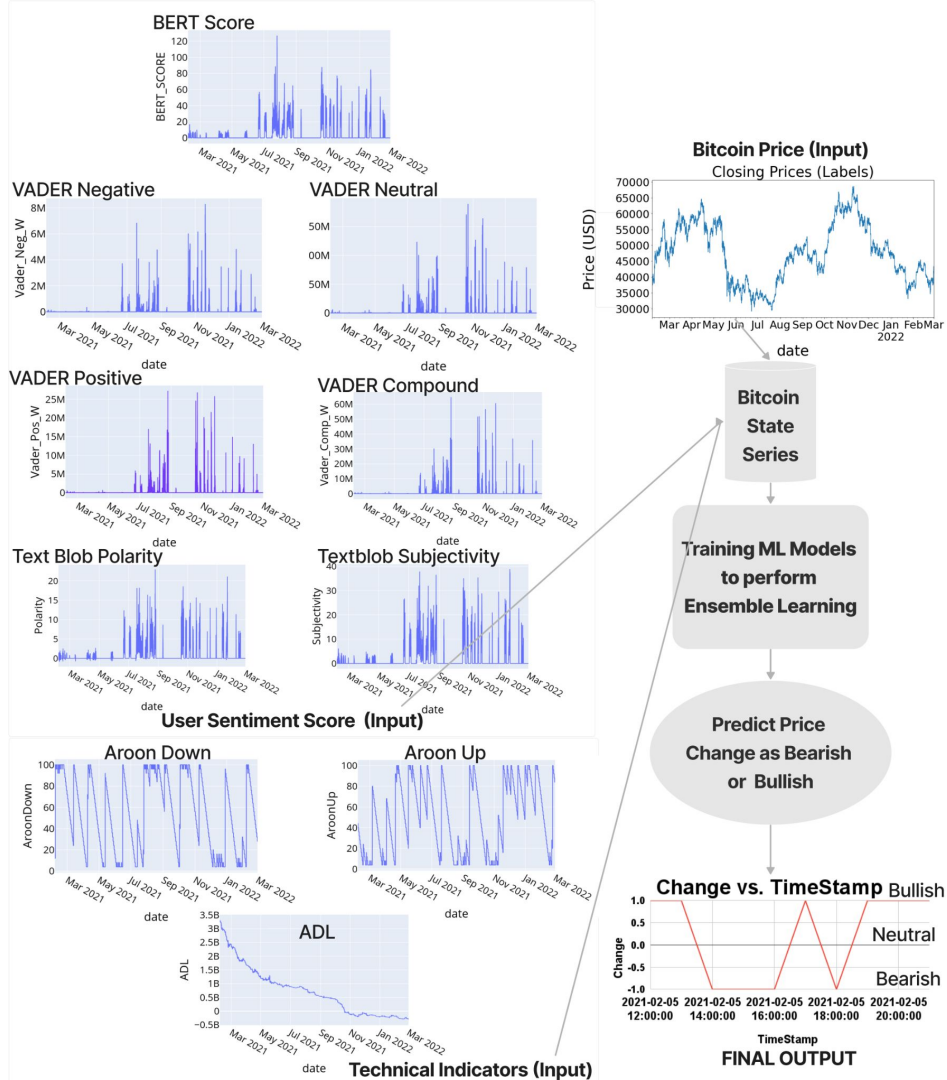
Good features contribute positively in the overall accuracy and f-measure of the 27 Machine Learning models.

$$Avg. \quad Accuracy = \frac{1}{27} \sum_{i=1}^{27} Accuracy\_i \qquad Avg. \quad F{-}Measure = \frac{1}{27} \sum_{i=1}^{27} F{-}Measure\_i$$

Predict price change as Bearish or Bullish using Tweet pre-processing

Prediction Framework

- Historical Tweets → Spam Filtering
- Cryptocompare
- Spam Filtering → Data Preprocessing using Regex
- Collecting Historical Data using API
- Data Preprocessing using Regex → Sentiment Analysis using FinBERT, VADER, TextBlob
- Collecting data in Dataframe
- Sentiment Analysis using FinBERT, VADER, TextBlob → Creating user weighted-sentiment
- Creating Bitcoin state series ← Collecting processed data in Dataframe
- Creating user weighted-sentiment → Collecting processed data in Dataframe
- Labeling state series for supervised learning
- Training Machine Learning Models
- Creating test set & compare accuracy
- Predicting as either Bearish or Bullish

Tweet preprocessing

| Process	Result
Original Tweet	Mad Money's Jim Cramer Expects Bitcoin to Fall to $12,000!!! – Bitcoin News https://t.co/OQVgM7g
Remove hyperlinks	Mad Money's Jim Cramer Expects Bitcoin to Fall to $12,000!!! – Bitcoin News
Convert text to lower case	mad money's jim cramer expects bitcoin to fall to $12,000!!! – bitcoin news
Removing punctuations	mad money's jim cramer expects bitcoin to fall to $12,000 – bitcoin news
Removing Symbols & numbers	mad money s jim cramer expects bitcoin to fall to bitcoin news
Remove additional spaces	mad money s jim cramer expects bitcoin to fall to bitcoin news

Bitcoin State Series to train Machine Learning (ML) models for predicting Bitcoin price change

5 V's of Big Data in our BiTcoin Data

- **Volume:** 9000+ states in the Bitcoin state series fusing over a million tweets

- **Velocity:** Twitter generates an enormous amount of data to understand the market sentiment

- **Variety:** FinBERT, VADER, TextBlob, Aroon, ADL

- **Veracity:** Our data is raw and collected from Twitter API

- **Value:** Bitcoin's Market Value (investment)

5 V's of Big Data

Volume

Velocity

Variety

Veracity

Value
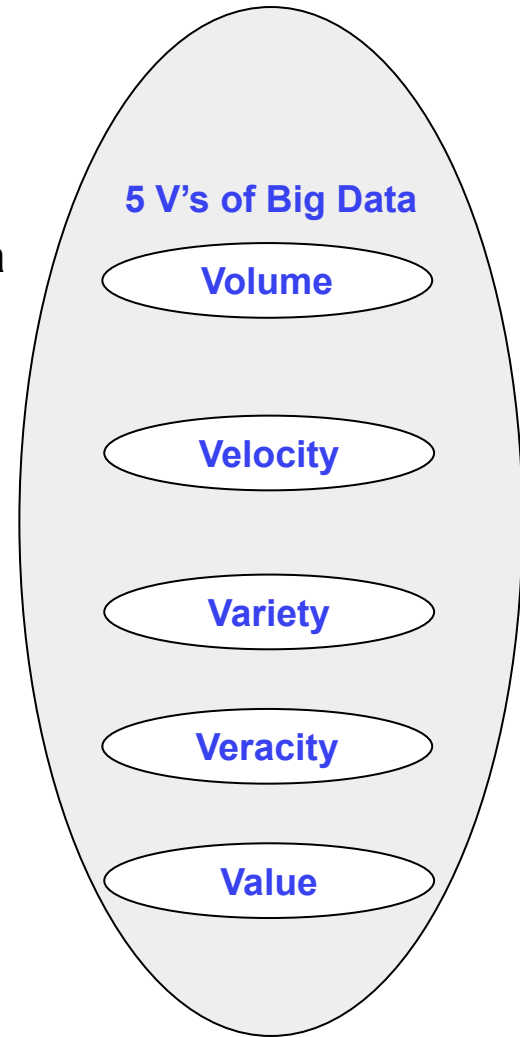
Sentiment Features

FinBERT → BERT_label, BERT_score

Vader → neg_val, pos_val, neu_val, comp_val

UserSentiment : S x number of followers the user has who tweeted.

Here S is any numerical score of Sentiment extraction algorithms.

For eg. BERT_score, neg_val, pos_val, neu_val, comp_val

Technical Indicators

We also explore the effect of combining Sentiment Analysis with Technical Indicators:

Aroon up and Aroon down:

$$AroonDown = \frac{25 - \#PeriodsSince25PeriodLow}{25} \times 100$$

$$AroonUp = \frac{25 - \#PeriodsSince25PeriodHigh}{25} \times 100$$

Accumulation/Distribution Line:

$$AD_t = AD_{t-1} + V_t \left(\frac{(C_t - L_t) - (H_t - C_t)}{H_t - C_t} \right)$$

here H_t, L_t, C_t and V_t denote High, Low, Close and Volume respectively for a timestep t.

Ensemble Learning models used

- *Hard voting* - Hard voting entails picking the prediction with the highest number of votes

- *Random Forest* - a classification algorithm consisting of many decisions trees

- *Bagging Ensemble* (Bootstrap Aggregation) - Ensemble learning method that is commonly used to reduce variance within a noisy dataset
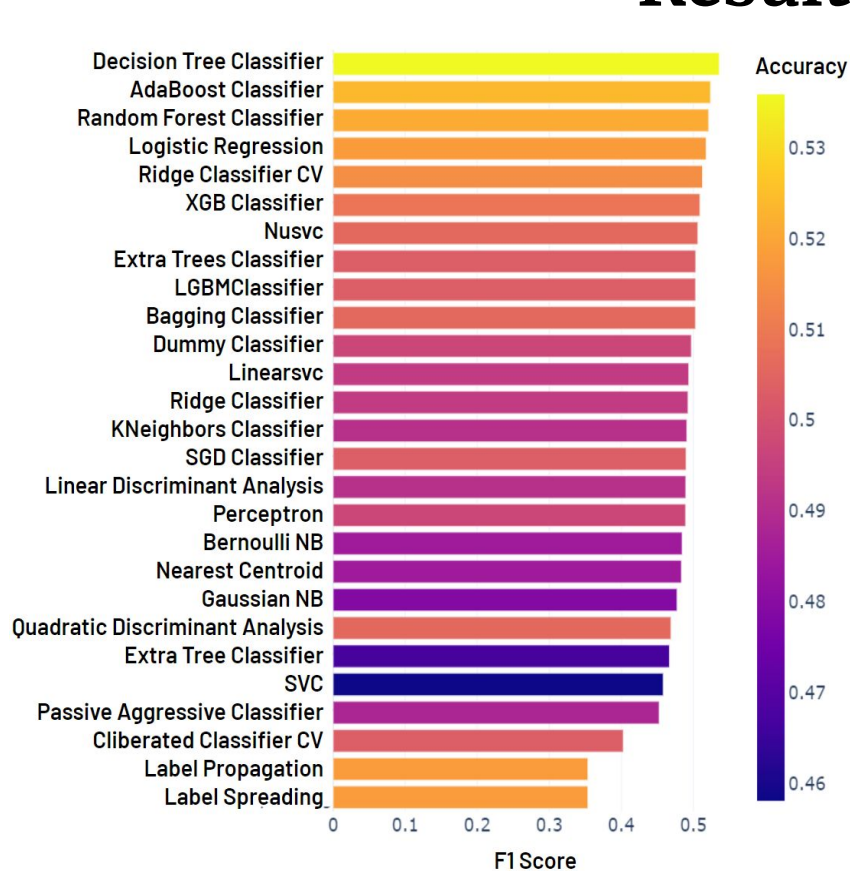
Ensemble Learning models used

Adaboost (Adaptive Boosting) Ensemble - It learns from the mistakes by increasing the weight of misclassified data points.
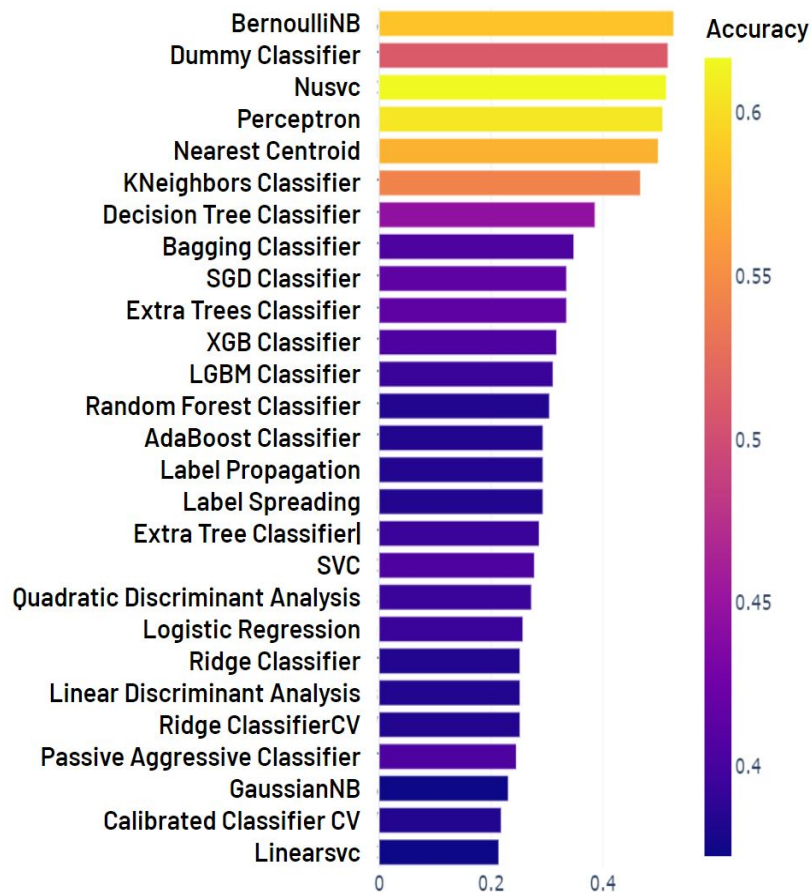
Gradient Boosting Ensemble - It learns from the mistake — residual error directly, rather than update the weights of data points.

XGBoost Ensemble (Extreme Gradient Boosting) - A scalable, distributed gradient-boosted decision tree (GBDT) machine learning library. It provides parallel tree boosting and is the leading machine learning library for regression, classification, and ranking problems.
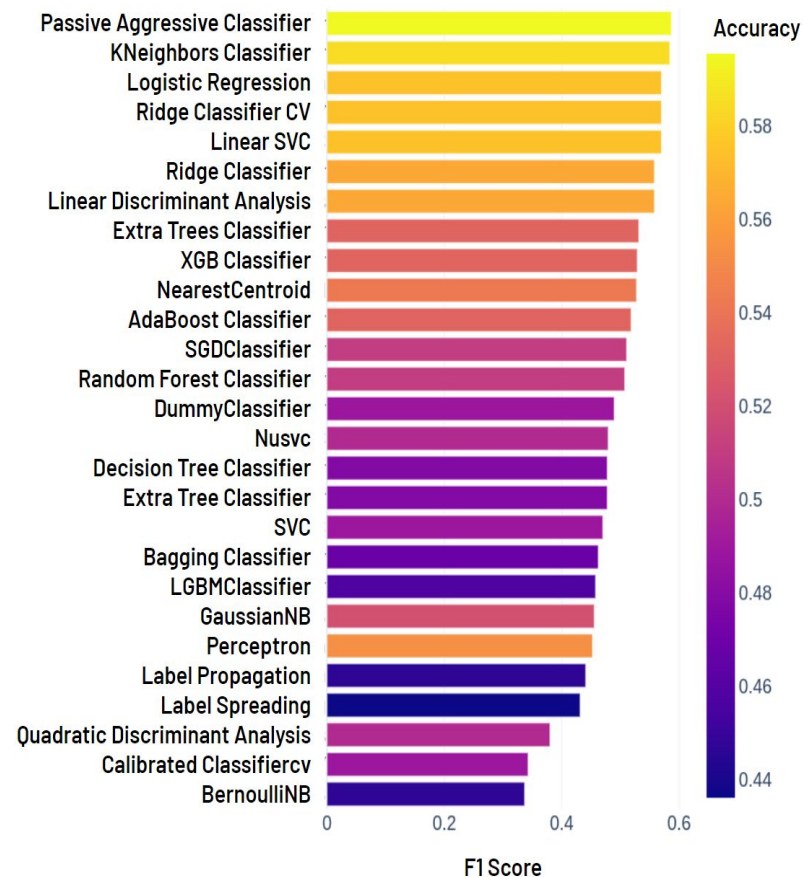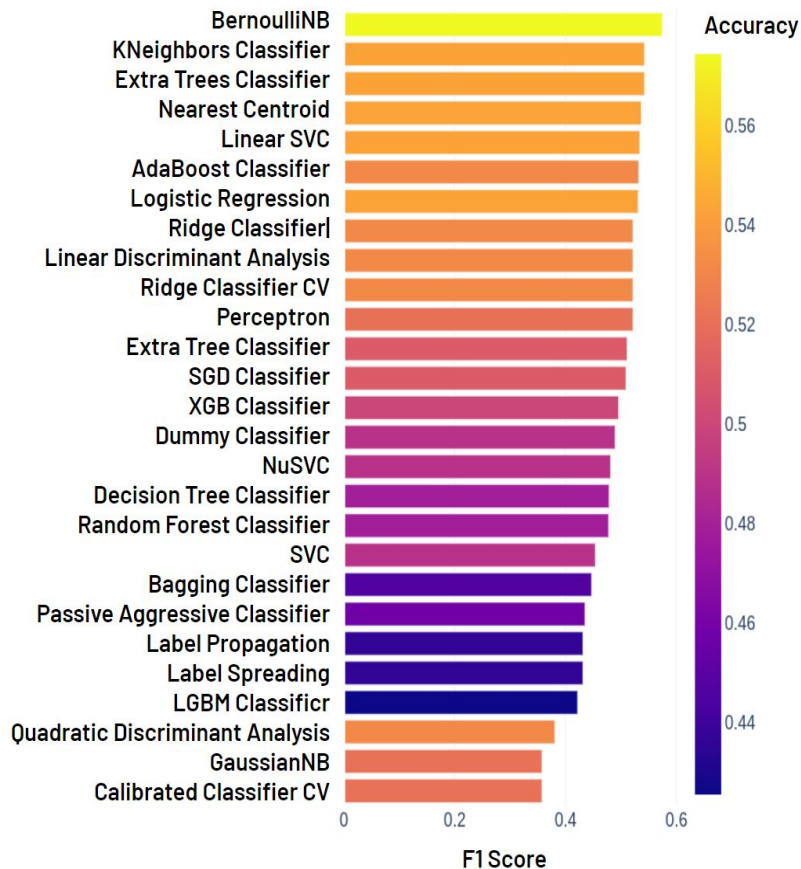
Results

Accuracy and F-Measure Using only *UserSentiment*

Accuracy and F-Measure Using only Sentiment

Results

Accuracy and F-Measure Using *UserSentiment* and Technical Indicators

Accuracy and F-Measure Using Sentiment and Technical Indicators

Ensemble Learning models results

COMPARATIVE ANALYSIS OF ENSEMBLE LEARNING SCORES.

Bitcoin and Tweet Features	Average		Maximum among the 27 models	
	Avg. Accuracy	Avg. F-Measure	Max. Accuracy	Max. F-Measure
Sentiment Scores	43.36%	33%	**62% (NuSVC)**	52%
Sentiments Scores and Technical Indicators	51.3%	49%	53%	53%
UserSentiment Scores	52.5%	48%	54.6%	65.4%
UserSentiment Scores and Technical Indicators	**52.60%**	**51%**	60.2%	**69.5% (Gradient Boosting)**

Conclusion

- We introduced *Bitcoin Evolution Analytics* which uses *Bitcoin State Series* to predict Bitcoin Price Change in the next hour as Bearish or Bullish

- We found that *UserSentiment Scores* is a better predictor than Sentiment Scores.

- We found that *Gradient Boosting* model performed the best. Combination Technical and Sentiments Indicators perform better than only Sentiments.

Future Work

- In the future, we plan to extend *Bitcoin Evolution Analytics* to retrieve various system properties like *System Network Complexity*, *System stability*, and changeability.

- The Bitcoin Evolution Analytics would also be helpful in evolution recommendation by constructing a *System Neural Network*.

Thank You

Welcome for
Questions & Answers