CRYPTO SENTIMENT ANALYSIS USING MACHINE LEARNING

Ketan Bonde[1] , Vina M. Lomte[2] , Prathamesh Bhalerao[3] , Pratik Chavan[4] , Vrushabh Bhandalkar[5]
[1][3][4][5]RMD School College of Engineering, Pune, India.
[2]Professor, RMD School College of Engineering, Pune, India

Abstract :

The rapid evolution of the cryptocurrency market, mirroring traditional currency systems but operating digitally and decentralized, marks a significant paradigm shift. Despite cryptographic safeguards, the industry remains under intense scrutiny due to its nascent nature. Cryptocurrencies, emerging as a distinct asset class driven by financial technology, present a rich landscape for research. Understanding public sentiment towards cryptocurrencies is pivotal for a comprehensive understanding of this ecosystem.

This study employs a Long Short-Term Memory (LSTM) neural network to predict Bitcoin's price in US dollars, utilizing the Keras library in Python. Renowned for its accuracy and ability to capture long-term dependencies, LSTM models offer promising avenues for price forecasting. Drawing from an extensive literature review, we validate LSTM's effectiveness in predicting cryptocurrency prices. Our goal is to devise a robust solution applicable to various cryptocurrencies, ensuring the highest possible accuracy. This research contributes to the advancement of cryptocurrency markets, empowering informed decision-making in this dynamic financial realm. Overall, this study represents a significant advancement in the prediction of cryptocurrency sentiment, offering valuable insights for investors, traders, and policymakers alike. By harnessing the power of advanced analytics

Keywords: Cryptocurrency, Sentiment analysis, Machine learning, Deep learning, LSTM(Long short-term memory), Bitcoin.

I. INTRODUCTION

Challenges and uncertainties in the cryptocurrency trading market can be overcame by developing a sophisticated recommendation system. This system leverages sentiment analysis and Long Short- Term Memory (LSTM) neural networks to provide recommendation to users whether to buy, sell, or hold their investments. While the primary focus is on Bitcoin, the project is designed to be adaptable for various cryptocurrencies, providing users with accurate, data-driven insights to guide their trading decisions.

Cryptocurrency market has been developed at an exceptional pace since its emergence. Cryptocurrency is a digital currency however it is not controlled by any central authority to make online payments. It uses system ledger entries called 'tokens' to make online payments for goods and services. Elliptical curve encryption and public-private key pairs are used as cryptographic algorithms. Similarly, hashing functions are utilized to protect online payments and ensure legitimate and unique transactions. Bitcoin was the first block chain-based cryptocurrency introduced in 2009 and itremains important and leading the market today . Such cryptocurrencies include Bitcoin clones, as well as, entirely new currencies with additional features [1].

II. LITERATURE SURVEY

People show emotions for everyday communication. Emotions are identified by facial expressions, behavior, writing, speaking, gestures and physical actions. Emotion plays a vital role in the interaction between two people. The detection of emotions through text is a challenge for researchers. Enabling machines with the ability to recognize emotions in a particular kind of text such as twitter's tweet has important applications in sentiment analysis and affective computing. We have worked on the newly published gold dataset (AIT-2018) and propose a model consisting of lexical based using WordNet-Affect and Emo Sentic Net with supervised classifiers for detecting emotions in a tweet text.[1] Humans have the power to feel different types of emotions because human life is filled with many emotions.Most of the study is using a machine learning technique. In this paper, we classified 7 emotions such as anger, fear, joy, love, sadness, surprise, and thankfulness using deep learning technique that is Long Short-Term Memory (LSTM) and Nested Long Short-Term Memory (Nested LSTM). We have compared our results with Support Vector Machine (SVM) Nested LSTM gets the best accuracy of 99.167%, while LSTM gets the best performance in term of average precision at 99.22%, average recall at 98.86%, and f1- score at 99.04%.[2]

III. PROPOSED SYSTEM

Cryptocurrency markets are notoriously volatile, driven by a complex interplay of factors ranging from technological advancements to market sentiment. In this proposed system, we aim to leverage sentiment analysis techniques to forecast cryptocurrency prices. By analyzing public sentiment expressed through social media, news articles, and other sources, we intend to build predictive models that can anticipate market movements and trends.The primary objective of this research is to design and implement a robust sentiment analysis framework capable of extracting sentiment from diverse textual sources relevant to cryptocurrency markets. This framework will leverage state-of-the-art NLP techniques to analyze sentiment patterns and trends. Finally, the research aims to explore practical applications and implications of the proposed system for various stakeholders, including investors, traders, financial analysts, and policy makers.The research holds significant importance for investors and traders in cryptocurrency markets by providing more accurate and timely price predictions. Incorporating sentiment analysis with LSTM models can offer valuable insights into market sentiment trends, helping stakeholders make informed decisions regarding buying, selling, or holding cryptocurrencies.

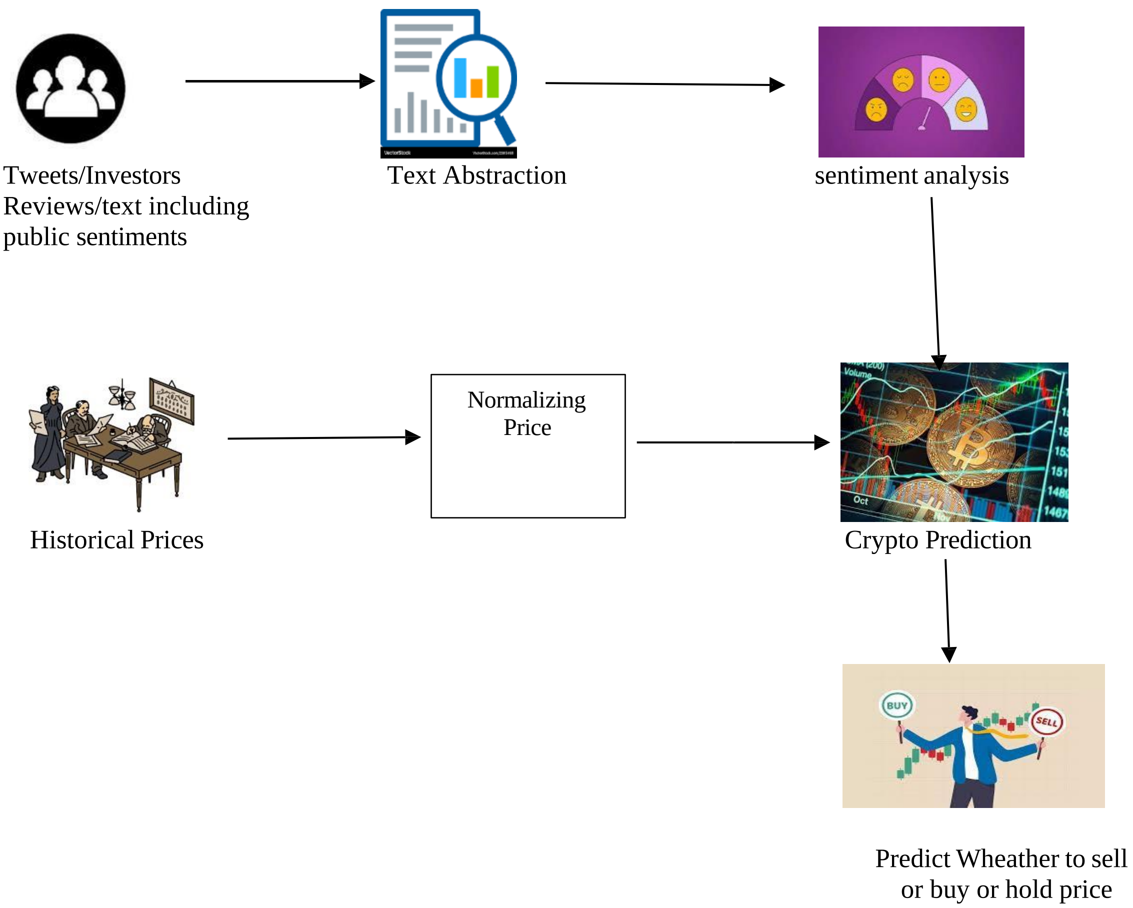

Figure No 01 : System Architecture

IV. Algorithm

Step 1. Processing the Data

i. Gather historical cryptocurrency price data for the target cryptocurrency(s) of interest.
ii. Obtain textual data from various sources such as social media platforms, news articles, and blogs related to cryptocurrencies.
iii. Clean the price data by handling missing values, outliers, and inconsistencies.

Step 2. Splitting the Data

Splitting the data for cryptocurrency price prediction using sentiment analysis with LSTM involves dividing the dataset into appropriate subsets for training, validation, and testing.

Step 3. Training the LSTM Model

Training an LSTM model for cryptocurrency price prediction using sentiment analysis involves several steps, from defining the model architecture to optimizing hyperparameters and monitoring training progress.

Step 4. Evaluating the LSTM Model

Evaluating an LSTM model for cryptocurrency price prediction using sentiment analysis involves assessing its performance, accuracy, and generalization ability.

Step 5. Fine-Tuning the LSTM Model

Fine-tuning an LSTM model for cryptocurrency price prediction using sentiment analysis involves optimizing various hyperparameters, model architecture choices, and training strategies to improve performance, convergence speed, and generalization ability.

Step 6. Making Predictions

Making predictions for cryptocurrency price using sentiment analysis with LSTM involves feeding new or unseen data into the trained LSTM model to generate predictions for future cryptocurrency prices.

Embedding Layer

An embedding layer is a fundamental component in deep learning models, especially in natural language processing (NLP) tasks, where it plays a crucial role in representing categorical data such as words or tokens as continuous vectors in a high-dimensional space.

$X^{(e)}=W^{(e)} \cdot X$ ....(1)

LSTM Layers

LSTM (Long Short-Term Memory) layers are a type of recurrent neural network (RNN) layer that is designed to handle long-term dependencies and capture temporal patterns in sequential data.

$h^{(t)},c^{(t)}=\mathrm{LSTM}(X^{(e,t)},h^{(t-1)},c^{(t-1)})$ ....(2)

Output Layer

The output layer in a neural network is the final layer that produces the model's predictions or outputs based on the processed input data. Its design and activation function depend on the type of task the neural network is performing, such as classification, regression, or sequence generation.

$P(Y)= \mathrm{softmax}(W^{(o)} \cdot h^{(N)}+b^{(o)})$ .... (3)

Loss Function

A loss function, also known as a cost function or objective function, is a crucial component in machine learning and deep learning algorithms. Its role is to quantify the difference between the predicted outputs of a model and the actual target values during the training process. The goal of training a model is to minimize this loss function, which reflects how well the model is performing relative to the desired outcome.

$L=-\sum iYilog(P(Y)i)$..(4)

Notation

- X - Input text data.
- Y - Sentiment label (positive, negative, neutral).
- W^(e) - Word embedding matrix.
- X^(e) - Word embeddings for the input text.
- H^(t) - LSTM hidden state at time step t.
- C^(t) - LSTM cell state at time step t.
- L - Number of LSTM layers.
- N - Sequence length.
- E - Embedding dimension.
- H - Number of LSTM units.
- W(o) - Output layer weight matrix.
- b(o) - Output layer bias.

V. RESULTS AND DISCUSSION

The results of our cryptocurrency sentiment analysis using machine learning techniques demonstrate the effectiveness of our approach in capturing and analyzing sentiment within the cryptocurrency domain. Using state-of-the-art machine learning algorithms, including natural language processing (NLP) techniques, sentiment analysis models were trained on this dataset to classify sentiments as positive, negative, or neutral. The performance of these models was evaluated using standard metrics such as accuracy, precision, recall, and F1-score.

Our experiments yielded promising results, with our sentiment analysis models achieving high levels of accuracy in classifying sentiments across different cryptocurrencies. Additionally, we observed that our models were able to adapt to the dynamic nature of sentiment in cryptocurrency markets, accurately capturing shifts in sentiment over time.

Discussion:

The success of our cryptocurrency sentiment analysis using machine learning underscores the importance of sentiment analysis in understanding market dynamics and informing decision-making processes within the cryptocurrency domain. By analyzing sentiment expressed in textual data from various sources, our approach enables stakeholders to gain valuable insights into market trends, investor sentiment, and potential price movements.

One key finding of our study is the correlation between sentiment and cryptocurrency prices. We observed that periods of positive sentiment often coincide with bullish market trends, while negative sentiment tends to precede market downturns. This highlights the predictive power of sentiment analysis in forecasting cryptocurrency price movements and identifying potential trading opportunities.

Furthermore, our analysis revealed the impact of external factors, such as regulatory announcements, technological developments, and market sentiment in traditional financial markets, on cryptocurrency sentiment. By incorporating these factors into our analysis, we were able to provide a more holistic understanding of sentiment dynamics within the cryptocurrency market.

Overall, our study demonstrates the value of machine learning techniques in analyzing cryptocurrency sentiment and its implications for market participants. Moving forward, further research in this area could focus on enhancing the accuracy and robustness of sentiment analysis models, as well as exploring new sources of data and methodologies to gain deeper insights into cryptocurrency markets..

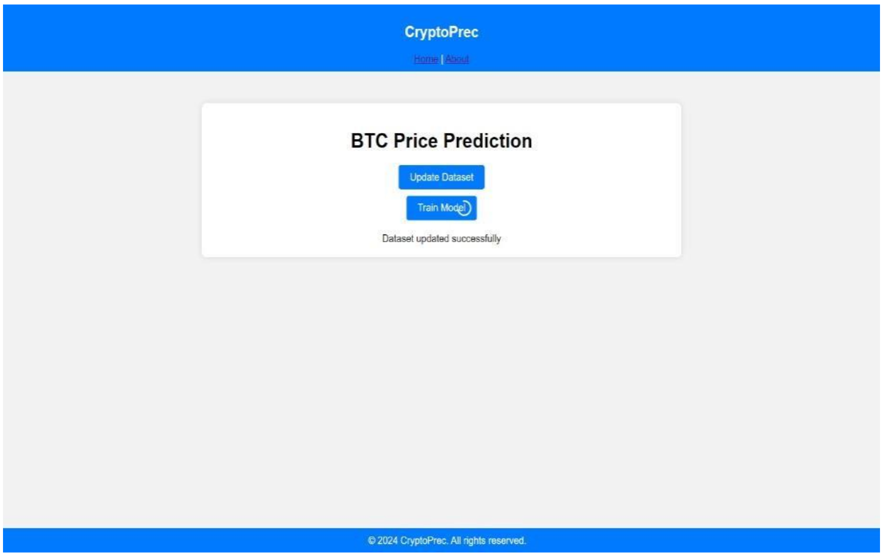

Fig. Update and training the dataset

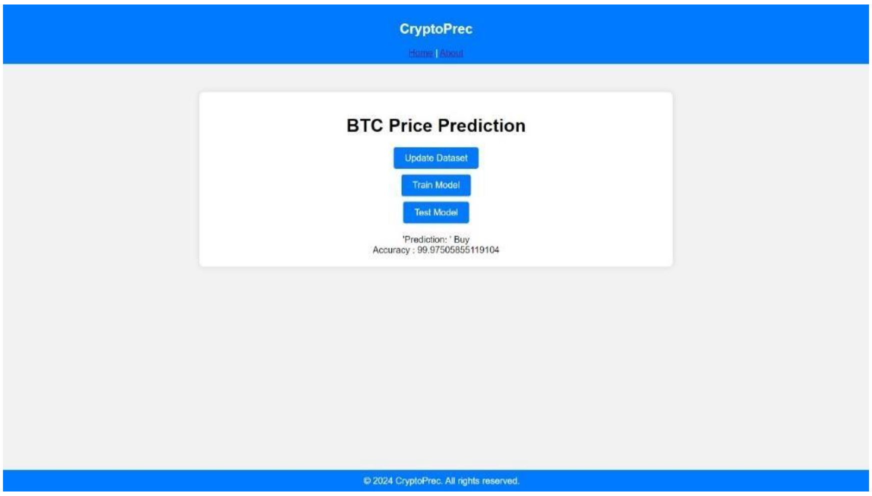

Fig. Final result of project

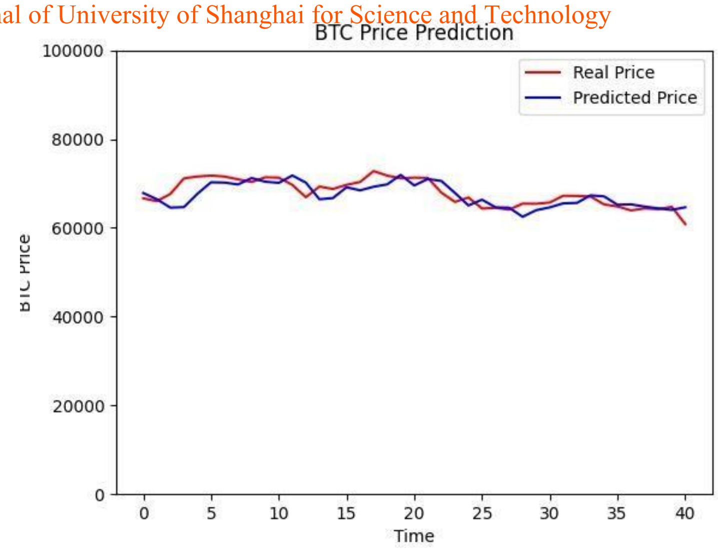

Fig. Real Price and Predicted Price of Bitcoin on 1 May 2024

VI. Conclusion and Future Scope

This study performs sentiment analysis and emotion detection on tweets related to cryptocurrency. Sentiment analysis of cryptocurrency holds potential significance as it is widely used for predicting the market price of the cryptocurrency which necessitates sentiments classification with high accuracy. For experiments, tweets are extracted from Twitter TM, and the dataset is annotated using Text Blob and Text2Emotion for sentiments and emotions, respectively. Besides the use of several machine learning and deep learning models for classification, this study leverages recurrent neural networks LSTM and GRU to form an ensemble model to enhance classification performance. In addition, BoW, TFIDF, and Word2Vec features are used as feature extraction techniques for the machine learning models. Results indicate that machine learning models perform well with BoW features compared with TF-IDF and Word2Vec. The proposed model achieves the highest performance for sentiment analysis with a 0.99 accuracy score and the highest precision and recall of 0.99 and 0.98, respectively. Similarly, LSTMGRU outperforms all other models in terms of correct and wrong predictions for both sentiment analysis and emotion detection. Dataset balancing using the random under sampling suggests that LSTM-GRU performance is decreased due to fewer training data. This study considers the sentiment analysis for cryptocurrency related tweets, we intendto perform cryptocurrency market price prediction based on the analyzed sentiments in the future.

VII. REFERENCES

[1] Naila Aslam, Furquan Rustam , Ernesto Lee "Sentiment Analysis and Emotion Detection on Cryptocurrency Related Tweets Using Ensemble LSTM-GRU Model" [March 19, 2022].

[2] J. Abraham, D. Higdon, J. Nelson, and J. Ibarra, ''Cryptocurrency price prediction using tweet volumes and sentiment analysis,'' SMU Data Sci. Rev., vol. 1, no. 3, p. 1, 2018.

[3] S. Colianni, S. Rosales, and M. Signorotti, ''Algorithmic trading of cryptocurrency based on Twitter sentiment analysis,'' CS229 Project, Stanford Univ., Stanford, CA, USA, Tech. Rep., 2015, pp. 1–5.

[4] A. Inamdar, A. Bhagtani, S. Bhatt, and P. M. Shetty, ''Predicting cryptocurrency value using sentiment analysis,'' in Proc. Int. Conf. Intell. Comput. Control Syst. (ICCS), May 2019, pp. 932–934.

[5] D. L. K. Chuen, L. Guo, and Y. Wang, ''Cryptocurrency: A new investment opportunity?'' J. Alternative Inve.stments, vol. 20, no. 3, pp. 16–40, 2017.

[6] K. Wołk, ''Advanced social media sentiment analysis for short-term cryptocurrency price prediction,'' Expert Syst., vol. 37, no. 2, p. e12493, Apr. 2020.

[7] C. Lamon, E. Nielsen, and E. Redondo, ''Cryptocurrency price prediction using news and social media sentiment,'' SMU Data Sci. Rev, vol. 1, no. 3, pp. 1–22, 2017.

[8] S. Dhar and I. Bose, ''Emotions in Twitter communication and stock prices of firms: The impact of COVID- 19 pandemic,'' Decision, vol. 47, no. 4, pp. 385– 399, Dec. 2020.

[9] J. Zhang, Y. Li, W. Xiao, and Z. Zhang, ''Non- iterative and fast deep learning: Multilayer extremelearning machines,'' J. Franklin Inst., vol. 357, no. 13, pp. 8925– 8955, 2020.

[10] J. Zhang, W. Xiao, Y. Li, S. Zhang, and Z. Zhang, ''Multilayer probability extreme learning machine for device-free localization,'' Neurocomputing, vol. 396, pp. 383–393, Jul. 2020.

[11] VOLUME 10, 2022 39323N. Aslam et al.: Sentiment Analysis and Emotion Detection on Cryptocurrency Related Tweets.

[12] W. Xiao, J. Zhang, Y. Li, S. Zhang, and W. Yang, ''Class-specific cost regulation extreme learningmachine for imbalanced classification,'' Neurocomputing, vol. 261, pp. 70–82, Oct. 2017.

[13] R. Jamil, I. Ashraf, F. Rustam, E. Saad, A. Mehmood, and G. S. Choi, ''Detecting sarcasm in multi-domain datasets using convolutional neural networks and long short-term memory network model,'' PeerJComput. Sci., vol. 7, p. e645, Aug. 2021.

[14] M. Mujahid, E. Lee, F. Rustam, P. B. Washington, S. Ullah, A. A. Reshi, and I. Ashraf, ''Sentiment analysis and topic modeling on tweets about online education during COVID-19,'' Appl. Sci., vol. 11, no.18, p. 8438, Sep. 2021.

[15] R. P. Espíndola and N. F. Ebecken, ''On extending Fmeasure and Gmean metrics to multi-class problems,''

[16] M. Wimalagunaratne and G. Poravi, ''Deep Learning and Sentiment Analysis-Based Cryptocurrency Price Prediction'' in Proc. 8th Int. Conf. Intell. Syst., Modelling Simulation (ISMS) [May 2020].

[17] F. M. Shah, A. S. Reyadh, A. I. Shaafi, S. Ahmed, and F. T. Sithil, ''Emotion detection from tweets usingAIT-2018 dataset,'' in Proc. 5th Int. Conf. Adv. Electr. Eng. (ICAEE), Sep. 2019, pp. 575–580.

[18] D. Haryadi and G. Putra, ''Emotion detection in text using nested long short-term memory,'' Int. J. Adv.Comput. Sci. Appl., vol. 10, no. 6, pp. 1–7, 2019.

[19] X. Zhang, W. Li, H. Ying, F. Li, S. Tang, and S. Lu, ''Emotion detection in online social networks: A multilabel learning approach,'' IEEE Internet Things J., vol. 7, no. 9, pp. 8133–8143, Sep. 2020.

[20] E. Şaşmaz and F. B. Tek, ''Tweet sentiment analysis for cryptocurrencies,'' in Proc. 6th Int. Conf. Comput.Sci. Eng. (UBMK), Sep. 2021, pp. 613–618.

VIII. REVIEW OF CONFERENCE/JOURNAL PAPERS SUPPORTING PROJECT IDEA

Ref. No	Publication Details	Method Used	Dataset Used	Accuracy (%)	Research Gap Identified/future Scope
1	Sentiment Analysis and Emotion Detection on Cryptocurrency Related Tweets Using Ensemble LSTM-GRU Model	The proposed ensemble model achieves the highest performance for sentiment analysis, with LSTM-GRU outperforming all other models for both sentiment analysis and emotion detection.	Social Media	78%	intend to perform cryptocurrency market price prediction based on the analyzed sentiments in the future.
2	Tweet Sentiment Analysis for Cryptocurrencies	In this paper, Sentiment Analysis is done to determine whether tweets posted by people on Twitter influence the price of the altcoin NEO. Depending on the sentiment of the tweets,	Tweet Data	72%	Just a Single cryptocurrency's analysis is done. We intend to implement our model for multiple cryptocurrencies

		i.e. positive, negative and neutral. correlation is done with the prices of the cryptocurrency. Correlation is also done with respect to Bitcoin and Ethereum prices.			The accuracy of the model used here was 77%. We will improve the accuracy by using better models.
3	Deep Learning and Sentiment Analysis-Based Cryptocurrency Price Prediction	This paper proposes a fusion-based model for cryptocurrencies price prediction, i.e., DLGuesS. It ims to predict the price of a specific coin considering their price history and tweet sentiments of the other dependent or alternate coins.	Dataset (‘Intern ational Survey on Emotio n Detectio n Anteced ents & Reactio ns’ (ISEAR))	64%	VADER model is used over here
4	Evaluating Sentiment Classifiers for Bitcoin Tweets in Price Prediction Task	Partial Correlation between Bitcoin price fluctuation and fluctuation of sentiment classes using different ML algorithms MLP,WiSARO and decision tree methods have better correlation also use NGram data modelling and tweet Embedding	Twitter & Reddit	71%	Future study given as to incorporate other features from text abstraction as hashtags, twitter user. number Of tweets and emoticons
5	A Methodology for Securities and Cryptocurrency Trading using Exploratory Data Analysis and Artificial Intelligence	Larger focus On EDA and more practical use regression ANN model with reward loss introduced in this paper and multistep ahead prediction will result in better performance for profit generation	Tweet Data, 980 549, training data, 144160 testing data	68%	The data used to train and evaluate sentiment analysis models is often not very accurate. This can be due to a number of factors, such as the difficulty of