Bitcoin Tweets Sentiment Analysis using Bidirectional Encoding Representational Transformers (BERT)

MSc Research Project
Data Analytics

Samir Huseynov
Student ID: x21245070

National College of Ireland

School of Computing

MSc in Data Analytics

Supervisior:

Bharat Agarwal

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Samir Huseynov
Student ID:	x21245070
Programme:	MSc Data Analytics
Year:	2023-2024
Module:	MSc Research Project
Supervisor:	Bharat Agarwal
Submission Due Date:	31/01/2024
Project Title:	Bitcoin Tweets Sentiment Analysis using Bidirectional Encoding Representational Transformers (BERT)
Word Count:	8763
Page Count:	29

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Samir Huseynov
Date:	30 January 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Bitcoin Tweet Sentiment Analysis Using BERT

Samir Huseynov
X21245070

Abstract

In recent years, the cryptocurrency industry has experienced unprecedented growth and has gained attention of people across globe. Similar to traditional currencies, virtual transactions for commodities and facilities are conducted without the involvement of a central authority, utilizing cryptocurrencies. Despite the assurance of authentic and unique transactions through cryptographic techniques, the cryptocurrency business is still in its early stages, and significant concerns have been raised regarding the usage and sustainability. In order to attain a brief and descriptive understanding of public sentiments towards cryptocurrencies, there is a particular interest in employing sentiment analysis. In light of this, the proposed research aims to leverage BERT, an advanced transformer-based model, for sentiment analysis on a dataset of tweets related to Bitcoin. The proposed methodology encompasses several crucial phases, including data preprocessing, BERT model optimization, and the evaluation of model effectiveness. By implementing this suggested approach, we aim to develop sentiment analysis model capable of accurately categorizing the sentiment expressed in Bitcoin-related tweets. The study's findings are anticipated to participate for a wider understanding of sentiment analysis methodologies applied to social media data in the context of cryptocurrencies.

1. Introduction

In ever-evolving landscape of the modern world, few themes have been as captivating and transformative as the rise of cryptocurrency. Bitcoin, the pioneering cryptocurrency, emerged in 2009, revolutionizing our perceptions of money. This virtual currency operates through a blockchain-based cryptographic mechanism, as noted by (3024330 & 2022, n.d.), and its pervasive usage has significantly expanded its real-world applications. Bitcoin's decentralized nature has made it a prominent choice for digital transactions, but it is also renowned for its extreme volatility, characteristic often influenced by social perceptions. Bitcoin, the pioneer of cryptocurrencies, has seen its market capitalization soar to over $150 billion. Individuals from various backgrounds, including programmers, businesspeople, enthusiasts, and even malicious actors, engage in discussions about Bitcoin's merits and drawbacks on the internet. It is worth mentioning that Bitcoin's price fluctuations have been linked to significant events in China, as revealed by (one & 2015, 2015). In today's hyper-connected world, the internet and social media have facilitated an unprecedented exchange of ideas, feelings, and opinions. Platforms like social media and e-commerce have become the stage for individuals and businesses to share their thoughts through textual and multimedia data, a phenomenon documented by researchers such as Behera, Gupta & Gupta, Wang, and Z. Zhang (Behera et al., 2020), ((Gupta et al., 2018), (Wang et al., 2020), and (Z. Zhang et al., 2017). This vast wealth of data has given rise to sentiment analysis, a powerful tool for discerning public sentiment, a field that was discovered by (Liu,2020) and has become increasingly relevant.

Understanding the sentiment expressed in tweets about Bitcoin holds substantial significance for various reasons. Firstly, sentiment analysis provides invaluable insights into public opinion, which can guide investors in assessing general sentiment toward Bitcoin. Moreover, traders can adapt their strategies based on emerging trends and market shifts identified through sentiment analysis. Financial experts can employ sentiment analysis to evaluate the impact of public sentiment on the Bitcoin market, enabling more precise market predictions and risk assessments. An extensive review of the existing literature on sentiment analysis and its application to cryptocurrencies reveals widespread utilization across various domains, encompassing news articles, social media posts, and consumer evaluations. Despite its extensive application, there is a noticeable dearth of research specifically addressing sentiment analysis on tweets related to Bitcoin.

This gap underscores the need for a dedicated exploration of sentiment analysis techniques tailored for the unique characteristics of Bitcoin-related tweets. While sentiment analysis has been approached through various methods, such as hybrid approaches, machine learning, and lexicon-based, it is noteworthy that supervised machine learning methods have exhibited dominant accuracy in contrast to alternative processes. However, it is important to acknowledge the inherent challenges associated with this approach, as it necessitates labeled data(CSUR) & 2002, n.d.), which can be resource-intensive, costly, and prone to errors (Zhu, 2005). Particularly within the context of platforms like Twitter, characterized by microblogging and vast amounts of unlabeled data, the acquisition of labeled datasets for training classifier models becomes a formidable task. Microblogging messages on platforms like Twitter introduce an additional layer of complexity, as they tend to be more chaotic, random, and ambiguous compared to traditional media texts (Feng et al., 123 C.E.) (Saif et al., 2012)

This inherent nature poses challenges for supervised classification using machine learning, making it difficult to derive meaningful conclusions. Furthermore, when using text representation methods like term frequency-inverse document frequency ("TF-IDF") or "n-gram," the scale and lexical diversity of microblogging datasets frequently result in a high-dimensional feature space. The development of an interpretable model with high prediction accuracy is complicated by the sparse representation, which is typified by short and noisy words. Despite these challenges, the vast scale of microblogging data provides an abundance of unprocessed information that can be leveraged for feature extraction, thereby enhancing the complexity of machine learning models (Silva et al., 2016). Regrettably, only a limited number of studies have delved into this particular aspect of sentiment analysis, overlooking the potential impact of dataset size on model performance during training and its influence on determining optimal model parameters. Given the time-intensive nature of labeling and processing extensive raw data, determining an appropriate dataset size becomes a pivotal consideration in capturing all essential features for constructing an effective sentiment classification model.

The primary focus of our research is to develop a personalized sentiment analysis model for categorizing tweets related to Bitcoin into positive, negative, and neutral sentiment classes. We aim to achieve precise classification by leveraging modified BERT or Transformer models with transfer learning strategies, considering the exemplary performance of these models in sentiment analysis and other natural language processing applications. Our study aims to make significant contributions to the sentiment analysis field in several ways. Firstly, we narrow our focus to tweets specifically discussing Bitcoin, recognizing their significance and influence on the cryptocurrency market. This targeted approach provides valuable insights into sentiment dynamics within this context. Secondly, we employ state-of-the-art preprocessing techniques to enhance the accuracy and reliability of our sentiment analysis model.

This involves the removal of abnormalities, irrelevant data, and noisy language to refine sentiment categorization. To further boost the precision of sentiment classification, we incorporate transfer learning techniques. By utilizing a tailored BERT or Transformer model, the proposed methodology allows the model to learn from an extensive corpus of text data and adapt specifically to sentiment analysis related to Bitcoin. This adaptation is expected to result in improved accuracy and overall performance. Additionally, our study explores the influence of data quantity on the performance of the classification model. Understanding how the model responds to varying amounts of data is crucial for optimizing its effectiveness. This holistic approach, combining targeted analysis, advanced preprocessing, and transfer learning, aims to create a robust sentiment analysis model tailored for the unique characteristics and emotions conveyed in Bitcoin-related tweets.

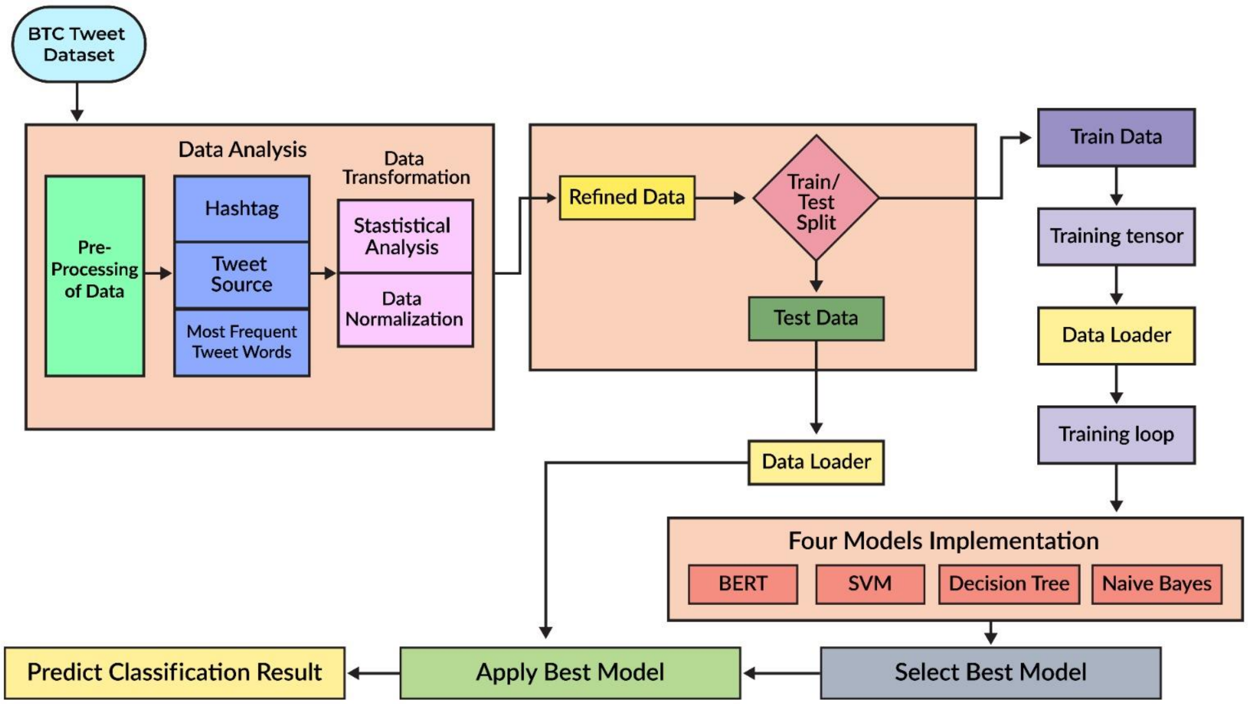

Figure 1:Processflow of Bitcoin Tweet Sentiment Analysis

2. Related Work

The proliferation of information regarding cryptocurrencies via social media and online platforms has significantly broadened alongside their increasing popularity (Kim et al., 2016). Social media platforms serve as potential indicators for predicting forthcoming events and advancements by gauging public sentiments on socioeconomic matters (Schoen et al., n.d.). Prior studies (Garcia et al., 2015) (Phillips et al., n.d.)have affirmed the correlation between Twitter discussions and the predictive nature of bitcoin prices. A hybrid sentiment analysis approach, amalgamating supervised machine learning with

semantic lexicons, has garnered heightened attention in recent years (V. D, 2019); (Fan et al., n.d.); (Imanuel Tanesab et al., 2017).VADER, introduced in 2014, stands out as one of the most frequently utilized lexical semantic methods for assessing sentiment polarity ratings. This rule-based sentiment analysis model, reliant on the VADER lexicon, delineates emotional polarities (positive/negative) and their intensities (strong) to derive a sentiment score. VADER boasts several advantages, including its open-source nature, human-centric design, and tailored suitability for social media content analysis (Valencia et al., 2019).Supervised machine learning techniques like Support Vector Machines (SVM) and Naïve Bayes (NB) remain the most commonly employed methodologies for sentiment analysis, either independently or in conjunction with VADER. In contrast, sentiment lexicons and unsupervised learning approaches have demonstrated comparatively lesser accuracy in conducting sentiment analysis (Bermingham et al., 2010).
(Saif et al., 2012) demonstrated that Twitter text is notably sparser compared to other data forms, such as movie reviews. This sparsity can be attributed to the prevalence of uncommon terms, grammar errors, and the use of slang. Twitter also harbors a substantial amount of noise, including URLs, emojis, punctuation, stopwords, and special characters. These elements, functioning as outliers, can potentially alter the sentiment or information entropy within tweets by introducing unrelated or randomly occurring data. Given the sheer volume of raw data, it becomes crucial to automatically discern significant information from this vast pool. Consequently, various researchers (C. Huang et al., 2019); (Kübler et al., 2018); (Kumar et al., n.d.); (Tommasel & Godoy, 2018) have delved into different feature selection techniques and classifier models.The bag-of-words model, due to its simplicity and processing efficiency, represents text by compiling a list of words alongside their frequencies using a document-term matrix (C. C. Aggarwal & Zhai, 2012). This matrix groups words based on their relative distances, proving advantageous in tasks such as information retrieval, text grouping, and text categorization. However, most document-term matrices are high-dimensional and sparse since each document typically contains only a small subset of the corpus's unique phrases. Thus, these matrices comprise many zeros for phrases absent in a given document. Consequently, a strategy is needed to mitigate this high dimensionality.A commonly employed technique for evaluating word significance in a text collection is TF-IDF (Havrlant et al., 2017), (documentation & 2004, 2004). TF-IDF illustrates the distribution of each word across the entire document or corpus.

In addition to analyzing sentiments, certain studies delve into emotional analysis using Bitcoin-related tweets. Typically, positive emotions align with positive sentiments, and vice versa. Twitter sentiment analysis has gained prominence as a significant subject of study for experts in Natural Language Processing (NLP) and sentiment analysis. However, due to the vast and diverse nature of social media data, conducting sentiment analysis manually becomes challenging. This challenge prompted the development of automated systems for analyzing sentiment in Bitcoin-related tweets. Consequently, there exists a substantial body of literature dedicated to sentiment and emotion analysis. For example, a work by (Hasan et al., 2019) presents a machine-learning strategy for automatically recognising emotions from social media posts. Sentiments are discerned through text categorization. The study covers a number of issues, including message semantic complexity, the informal character of microblogs, the existence of many emotions in text, and the wide range of emotional states. Tweets with and without emotions are distinguished using binary classifiers. Online categorization and offline training are the two main goals of the approach. Emotex, an advanced emotion categorization system, demonstrates an impressive accuracy of 90% in classifying texts.

(Sharifirad et al., 2019) expects and analyses the amount of emotional intensity and emotion detection, similar to (Hasan et al., 2019). Natural language processing (NLP) technologies are used in their study to analyse sexist tweets, which are then classified into three categories: physical harassment, sexual harassment, and indirect harassment. The investigation also delves into the examination of low, medium, and high-intensity emotions encompassing joy, sadness, and fear. The study combines several NLP tools such as Word2Vec, Glove, and FastText in conjunction with SVM, Naive Bayes (NB), KNN, Multi-layer Perceptron (MLP), LSTM, and convolutional neural network (CNN) to achieve accurate multilabel categorization. Furthermore, three types of speeches containing sexist remarks are scrutinized to ascertain the intensity of each emotion.The findings highlight a direct correlation between happiness and indirect harassment. Additionally, when considering emotional intensity, anger demonstrates a notable connection to instances of sexual harassment. The parallels can be drawn to how emotions like anger, happiness, and grief are linked to cases of physical abuse.

(Shah et al., 2019) run experiments with the AIT-2018 dataset to determine the emotions expressed in tweets. The experiment presents a unique approach to emotion detection that utilizes the WordNetAffect and EmoSenticNet. However, the results reveal that the limited nature of the dataset and challenges associated with linguistic ambiguity adversely affect performance. The abundance of varied emotions within the material contributes to reduced accuracy. X. Zhang et al. (2020) conducted a research whereby they employed both lexicon-based and machine-learning approaches for sentiment analysis on online social networks. A method for multilabel learning is presented in the paper. The suggested approach builds a multilabel emotion detection system using machine learning techniques with the goal of comprehensively identifying emotions from the viewpoint of the user. Aside from emotion label relationships, the authors investigate social and temporal correlations. However, the method's potential is limited due to the availability of a short dataset, which results in diminished accuracy. Furthermore, in the work by (Haryadi et al., n.d.), a stacked LSTM model for emotion detection, based on deep learning, is introduced. The objective is to enhance the accurate classification of Ekman's emotions. This study utilizes a substantial dataset comprising 144,160 testing samples and 980,549 training samples. The nested LSTM attains the highest accuracy rating of 99.16%. However, the findings do not significantly differ from other models, indicating a noteworthy flaw that necessitates resolution.

Certain studies incorporate sentiment analysis of cryptocurrency-related tweets alongside emotional research to predict cryptocurrency market values. Positive remarks often correlate with an increased interest in the Bitcoin market, and conversely. For instance, a study by (X. Huang et al., n.d.)proposes the utilization of machine learning and sentiment analysis to forecast cryptocurrency values. Mining cryptocurrency-related content from the Chinese social networking site Sina-Weibo is part of the research. Both a proposed sentiment dictionary specific to crypto and an LSTM model are employed for prediction. This technique demonstrates a superior performance, surpassing previous methods by 18.5% in accuracy and 15.4% in recall. Similarly, in another study by (Şaşmaz et al., n.d.), sentiment analysis is conducted specifically on bitcoin currencies using tweets and machine learning methodology. The experiments, conducted on the NEO dataset and a manually labeled dataset, yield a classification accuracy of 77%. The literature review detail is given in the table below:

Table 1: Literature Review of the related work for Bitcoin Tweet Sentiment Analysis

Publication	Dataset	Objective	Technique	Results
HyVADRF: Hybrid VADER–Random Forest and GWO for Bitcoin Tweet Sentiment Analysis by Anny Mardjo and Chidchanok Choksuchat (Mardjo & Choksuchat, 2022)	Collected 3,625,091 265 tweets.	Analyze relation of Twitter sentiments and Bitcoin behavior	HyVADRF (hybrid valence aware dictionary and sentiment reasoner (VADER)–random forest) and gray wolf optimizer (GWO) model.	Accuracy=75.29%, precision =70.22%, recall = 87.70%, F1-score = 78%
Crypto-currencies narrated on tweets: a sentiment analysis approach by Saeed Rouhani, Ehsan Abedin (Rouhani & Abedin, 2020)	Collected almost 10 Million Tweets for various crypto currencies	Investigate the social opinions about different kinds of crypto-currencies and tune the best-customized classification technique	Lexicon-based approach for analyzing the tweets, by utilizing supervised machine learning techniques with SVM Classifier	Accuracies for different cryptocurrencies: Bitcoin: 81.35% Cardano: 83.48% Ethereum:80.97% Litcoin:83.65% Ripple: 85.35%
Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning (Valencia et al., 2019)	Collected almost 15 million Tweets for various crypto currencies	Investigate the sentiments to predict the market movement of cryptocurrencies	Applied machine learning algorithms RF, MLP and SVM	Bitcoin accuracies by: MLP:72% SVM:55% RF:44%
Predicting bitcoin price movements using sentiment analysis: a machine learning approach (Gurrib et al., n.d.)	Collected almost 9,256 observations of Tweets	Latest method for predicting the direction of BTC price using linear discriminant analysis (LDA) together with sentiment analysis.	LDA-based classifier integrated with SVM classifier	SVM Classifier Precision:58.5% Recall:58.5% F1-Score:58.5% LDA Classifier Precision: 54% Recall:54% F1-Score: 54%

Forecasting Bitcoin Price Fluctuation by Twitter Sentiment Analysis (Sattarov et al., 2020)	Customized dataset collected from several resources consisted of 92550 tweets	Analyzing the Twitter's impact on financial market focused on cryptocurrency targeting Bitcoin	Executed sentiment analysis with VADER and apply Random Forest Regression Classifier	Accuracy: 62.48%
Sentiment Analysis Based Direction Prediction in Bitcoin using Deep Learning Algorithms and Word Embedding Models (and & 2020, n.d.)	Collected almost 17629 tweets from several resources	To forecast the direction of Bitcoin price by analyzing user opinions in social media such as Twitter.	Applied RNN ,LSTM, and CNN are used as deep learning architectures and Word2Vec, GloVe, and FastText are employed as word embedding models	FastText and word embedding outperformed Accuracy: 89.13%
Forecasting the Early Market Movement in Bitcoin Using Twitter's Sentiment Analysis: An Ensemble-based Prediction Model (IOT et al., n.d.)	Collected almost 4 Million Tweets to create custom dataset	Gathered Tweets' to analyze manipulation and interpretation of cryptocurrency for predicting early market movements	Developed CEPM which is XGBoost-Composite ensemble model that achieve higher performance than the state-of-the-art prediction models.	Accuracy: 88% Precision: 89% Recall:88% F1-Score:88%
Evaluating Sentiment Classifiers for Bitcoin Tweets In Price Prediction Task (Balfagih et al., n.d.)	Customized dataset consist of 282K tweets	Explores the relationship between Twitter feed on Bitcoin and sentiment analysis	Employed language models tweet embedding and N-Gram, and then apply classification models MLP, WiSARD and Decision Tree	WiSARD outperformed shows accuracy of 89.61%
Sentiment Analysis and Emotion Detection on Cryptocurrency Related Tweets Using Ensemble LSTM-GRU Model (Aslam et al., n.d.)	Custom dataset collected from hashtags consists of 34K training and 6K test dataset	Predicted price of cryptocurrency by executing sentiment analysis and emotion detection	Custom LSTM-GRU model being utilized for classification alongside employed TextBlob and Text2Emotion for emotion detection	Accuracy Emotion Detection: 91% Sentiment Analysis: 99%

Recurrent Neural Network Based Bitcoin Price Prediction by Twitter Sentiment Analysis (Raj Pant et al., n.d.)	Custom dataset consist of 2585 positive, 1669 negative and 3200 irreverent tweets	Predicting volatile price of Bitcoin by analyzing sentiment in Twitter	Employed RNN model	Accuracy: 81.39% Precision:82.90% Recall:84.86% F1-Score: 83.86%
Tweet Sentiment Analysis for Cryptocurrencies (Şaşmaz et al., n.d.)	Collected 3 Million Tweets and refined to get 846,790 samples	Investigated the feasibility of automated sentiment analysis for cryptocurrencies	Trained and tested Random Forest Classifier	Accuracy: 77%
Automatic emotion detection in text streams by analyzing twitter data (Hasan et al., 2019)	Collected and refined almost 135,000 tweets data	To analyze emotion detection is that emotions are subjective concepts with fuzzy boundaries and with variations in expression and perception.	Trained and test state-of-art Machine learning Model SVM	Accuracy: 90%
Emotion Detection from Tweets using AIT-2018 Dataset (Shah et al., n.d.)	AIT-2018, Collected almost 71,816 Tweets for 4 emotions	Emotion detection through text to analyze the writer's mood and real-time sitauion of the sentiment	Trained Naives Bayes, Decision Tree and SVM Classifier	Accuracy Naïve Bayes: 86.42% Decision Tree: 80.09% SVM: 88.23%
Emotion Detection in Online Social Networks: A Multilabel Learning Approach (X. Zhang et al., n.d.)	Collected almost 16424 Tweets from 100 users	Analyzing emotions from user perspective through sentiment analysis	Trained KNN and Proposed CNN for Multi-Label	Accuracy: 90%
Emotion Detection in Text using Nested Long Short-Term Memory (Haryadi et al., n.d.)	Collected almost 1 Million Tweets Data utilize 980,549 for training and	Classifying 7 emotions using online tweets data	Utilize nested LSTM and LSTM model for classification on 7 emotions	Accuracy 99%

	144,160 for test data			
LSTM Based Sentiment Analysis for Cryptocurrency Prediction (X. Huang et al., n.d.)		Emotion Detection for the Chinese Language	Chinese Text Data (Sina-Weibo Dataset) using Deep Learning model LSTM	Precision: 18.5% Recall: 15.4%
Predicting Cryptocurrency Price Bubbles Using Social Media Data and Epidemic Modelling (Phillips et al., n.d.)	Selected time period from April 2015 to September 2016 to analyze data points through sliding window technique	Predict bubbles for number of cryptocurrencies using a hidden Markov model	Hidden Markov Model Technique	Predicted price achieved $7K profit
Forecasting Price of Cryptocurrencies using Tweets Sentiment Analysis (Jain et al., n.d.)	Collected almost 1.4 Million Bitcoin and Litecoin Tweets	Predict the two-hour price of cryptocurrencies on the basis of the Social Factors, which are increasingly used for online transactions worldwide.	Trained and test Multiple Linear Regression Model	Accuracy Bitcoin:44% Litecoin: 59%
Analysis on relationship between bitcoin price trend and sentiment of bitcoin related tweets by ML and NLP (Jiang et al., n.d.)	Collected and cleaned almost 1.3 Million Tweets	Research focuses on bitcoin-related tweets and bitcoin price	Trained and Tested Naïve Bayes with TextBob	Accuracy: 62.3% Precision: 64.16% Recall: 86.3% F1-Score: 72.9%
Deep Learning Approach to Determine the Impact of Socio Economic Factors on Bitcoin Price Prediction (A. Aggarwal et al., n.d.)	Dataset collected from Poloniex and gold price predicted from datahub.io	Comparative study of the various parameters affecting bitcoin price prediction	Trained and Tested CNN, LSTM and GRU	For Bitcoin RMSE value: CNN: 61.23 LSTM: 47.91 GRU: 55.98

3. Research Methodology

Natural language processing spans a wide range of applications, from conversational bots and machine translation to voice assistants and online speech translation. This business has seen tremendous expansion in recent years, both in terms of quantity (with a rise in market applications and goods) and quality (with breakthroughs in efficacy of latest models nearing a level of language understanding close to that of humans).

The important job of text representation is at the centre of natural language processing. Text representation functions as a collection of instructions for converting input data in natural language to machine-readable format. Although a representation may be thought of simply a computer encoding of text, representations that capture the text's inherent content and conceptual structure are more useful in field of applied machine learning and algorithms.

Several machine learning models have been used in the past to accomplish the job, as mentioned in (Colianni et al., n.d.) that the Nave Bayes Model, which is a generative algorithm. The Naive Bayes method is a generative learning technique that is extensively used in machine learning to handle text classification and sentiment analysis problems. This approach employs the initial feature vector format, in which the representation of a word is recorded using either a Bernoulli or multinomial distribution. The assumption in both rounds of this method is that the variable xi, provided Y in the Naive Bayes mathematical programme represented below, is conditionally independent of each other.

$$argmax_{y_j}\, P\left(Y = \ y_j\right)\textstyle\sum_{i=1}^{m} P(x_i|Y = y_j)$$

Here, yj signifies the classification indicating whether the Bitcoin price is experiencing an increase or decrease over a specified time interval. The variable xi represents the feature vector for tweet m, with a total of m tweets being collected. As a generative learning algorithm, it allows the creation of models for both positive and negative market variations. For each observation in the training set, the product of probabilities, as indicated above, is computed for each market trend. The results are of the models provide a subsequent accuracy of using Tweet Sentiment as feature vector on day-to-day basis as 55% and on hourly-basis as 52.42%.

In order to further strengthen, the research another state-of-art model has been trained on Tweet dataset the SVM model. Support Vector Machines (SVMs) represent supervised learning algorithms capable of nonlinear classifications by employing the kernel trick, which involves mapping data to higher dimensions. In sentiment analysis, SVMs have demonstrated effectiveness, as evidenced by (Go et al., 2009).

$$\min \frac{1}{2}\left||\omega|\right|\frac{2}{2} + C\textstyle\sum_{i=1}^{m} \xi_i$$

$$s.t\ y_{\xi_i}^{(i)}\left(\omega^T x^{(i)} + b\right) \geq 1 - \xi_i \qquad i = 1 \ldots., m$$

$$\xi_i \geq 0$$

The L1 Norm Soft Margin model, presented below, can be trained using either the feature vector for text classification or the feature vector for sentiment analysis.

This specific support vector machine employs the L1 norm soft margin formulation, including a penalty term for points that are not linearly separable. The penalty term is expressed as follows:

The term, ξi is a slack variable. The penalty equation serves as a trade-off, balancing a considerable separation between terms and the correct classification of observations. The feature vector x(i), is utilized in both the text classification and sentiment analysis forms. The variable y(i), denotes the observed class for a specific observation. The SVM model provide an accuracy for the Tweet sentiment on day-to-day basis as 53.5% and on hourly basis as 47.03%.

$$C\sum_{i=1}^{m}\xi_i$$

The research gap lies at many points ranging from uneven data distribution between classes to the models inability to capture the intriguing and minor details. The dataset should be systemize and designed accurately thus improving the accuracy of models. Such as the Naïve Bayes accuracy improves to upto 76.36% (Colianni, S. 2015). There is still room for a lot of improvement in order to bridge the gap we utilize a transformer based model called BERT.

A recent study leveraging the attention mechanism, specifically the Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., n.d.), has achieved cutting-edge performance across eleven natural language processing tasks. This success is attributed to fine-tuning the pre-trained BERT model with an additional output layer. In the research methodology, proposed assessing BERT's effectiveness in addressing the financial news sentiment analysis challenge with the ultimate goal of enhancing stock market prediction. This brief communication represents an early stage in ongoing research, presenting preliminary findings. The primary objective is to experimentally evaluate BERT in the context of Tweet sentiment analysis, with subsequent research phases dedicated to refining Bitcoin Price predict for market and stock enhancement methodologies.

BERT is a state-of-art model when applied on the dataset deliver the highest accuracy. The reason is BERT has been designed with Transformer based technology that encapsulates the attention mechanism. The attention mechanism assist in capturing the details and translating it with the highest probability thus the sentiment score matches with the word that has the highest probability of aligning with the specific sentiment get matched and placed in the specific place.

3.1 Exploratory Data Analysis

In the following research project, the plan is to utilize the BTC tweets sentiments dataset available on the Data-world platform. This dataset has been curated by aggregating tweets from various users, each accompanied by its corresponding sentiment analysis. The BTC tweets sentiments dataset is specifically focused on Twitter posts related to Bitcoin. The objective is to employ this dataset to develop a deep learning model for predicting the sentiment of tweets.

The original dataset comprises a specific number of rows and columns, which provide the foundational structure for the analysis.

Table 2: Dataset Count of Columns and Rows

Tweets Sentiments of Bitcoin	**Dataset Rows Count:**	50852
	Dataset Variables Count:	10

3.1.1 Dataset – Summary of Attributes

The BTC dataset is a tabular dataset that is store in csv file format. The chosen dataset contains the total ten features/columns. The first eight features of the BTC dataset were collected manually while the next two features (New-Sentiment-Score, New-Sentiment-State) were generated with the help of NLP model. The table described the features name, description of feature, and type of data in that feature.

Table 3: Dataset features and their description.

	Variable Name	Description	Type
1	ID	Serial Number / identification Number	Numeric
2	Date	Date of the tweet when tweet posted on tweeter.	Date
3	Tweet	Content of the tweet	Text
4	Screen-name	The name of the user who posted the tweet.	Text
5	Source	List of prominent words in tweet.	Text
6	Link	Http link of extracted tweet.	URL
7	Sentiment	Sentiment of the tweet in string categorical format. i.e., positive, negative	Categorical
8	Sent-score	Label encoded form of Sentiment feature.	Numeric
9	New-Sentiment-Score	Sentiment of the tweet in string categorical format. i.e., positive, negative by NLP model	Categorical
10	New-Sentiment-state	Label encoded form of New-Sentiment-Score feature.	Numeric

3.1.2 Target Variable Description

The BTC dataset contain the sentiment and Sent-Score that have the most suitable target variables for the proposed solution. Both features contain the sentiment of tweet in string or numeric format respectively. After the initial understanding of the dataset and by considering the problem statement, finalize the sentiment feature as the target variable. The target variable contains the three unique values that consider as classes for the prediction of sentiment.

Table 4: General Statistics about dataset.

Features	ID, Date, Tweet, Screen-name, Source, Link,
Target	Sentiment
Classes	Positive, Negative, Neutral

3.1.3 Most Frequent Words

The data analysis has been conducted on a series of preprocessing steps on the tweets within the dataset to identify the most frequently occurring words. To extract these high-frequency words, employed procedures such as eliminating stop words, punctuation, hashtag signs, special characters, and converting all text to lowercase. The resulting list of the top 10 most common words is visualized in Figure 2.

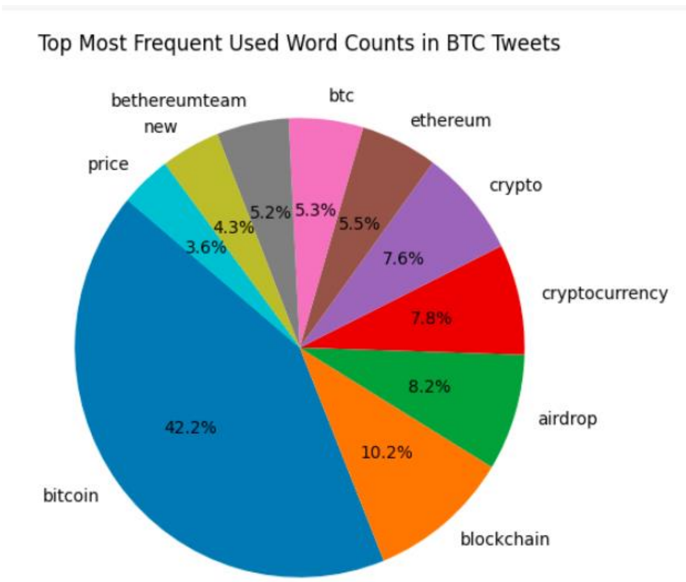

Figure 2Top Most used Words in Bitcoin Tweets

3.2 Most Frequent Hashtags in the list of top 10

In order to unveil the most frequently used hashtags in Bitcoin-related tweets, extracted all the words following the '#' symbol. The outcome, featuring the top hashtags, is displayed in Figure 2.

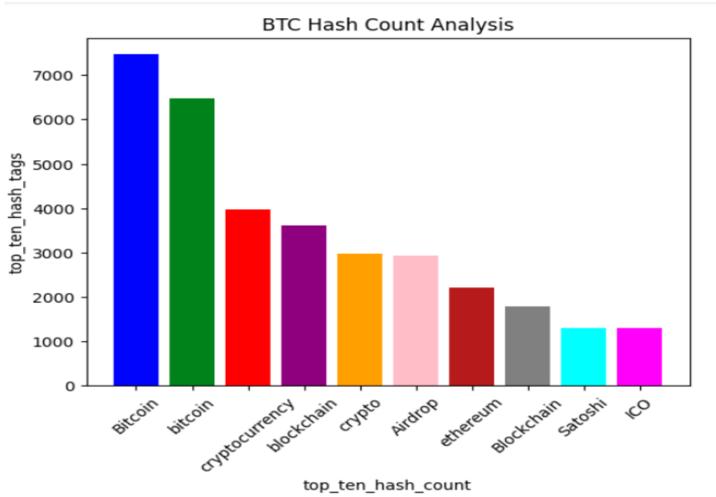

Figure 3: Top 10 Hash Tags in Tweets Dataset.

3.3 Most Common Sources for Tweets top 10

This visual representation in the figure below illustrates the primary sources of tweets, with a notable predominance of entries originating from the Twitter Web application. It is noteworthy that tweets from Android systems and iPhones exhibit a similar pattern, as they closely align in the rankings. In order to analyze details list of the top 10 tweet sources is shown in figure 3:

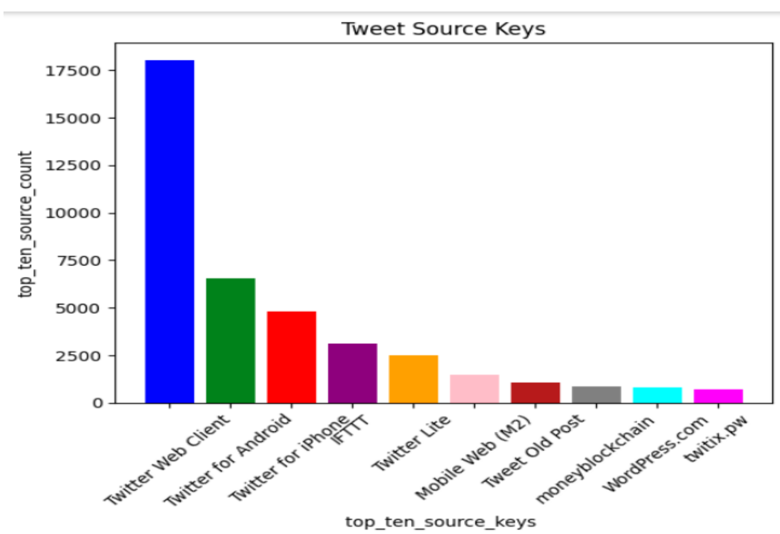

Figure 4:The top 10 Resources of Tweet

3.4 The Analysis of BTC Tweet Sentiments

In order to gain a better understanding of the dataset, executed an analysis to determine the distribution of tweets across various sentiment categories. The findings revealed the presence of three distinct sentiment types in the dataset. The below figure displays the respective counts of tweets for each of these sentiment categories, shown in figure 4:

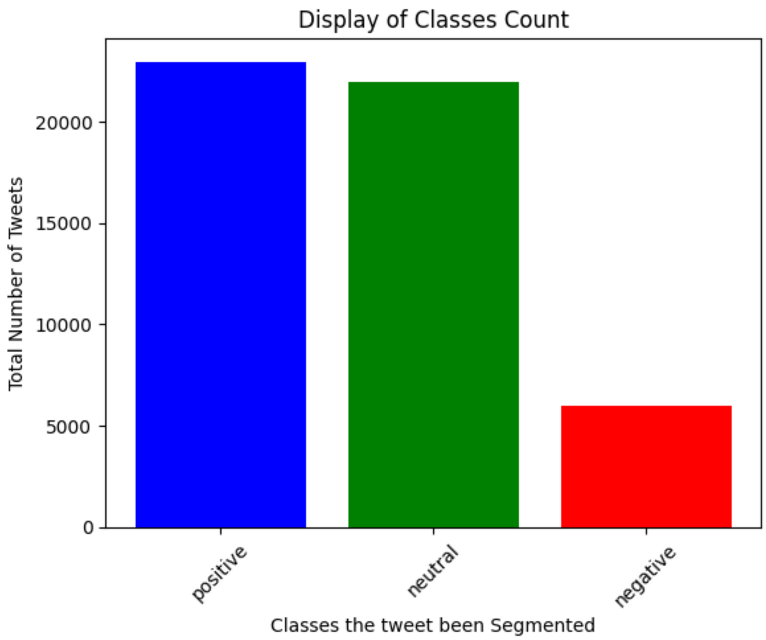

Figure 5:The Sentiment Analysis of BTC Tweets

4. Design Specifications

4.1 Architecture of BERT

BERT (Bidirectional Encoder Representations from Transformers) is a transformer-based approach announced by Google in 2018. Because of its outstanding performance on a variety of benchmarks, it has become a commonly utilised architecture for different natural language processing (NLP) tasks and experiments.

BERT is built on the transformer design, which was first described in the paper "Attention is All You Need" by Vaswani et al. The transformer design uses self-attention mechanisms to handle incoming data in parallel, which allows for more efficient sequence training. BERT's bidirectional processing of input sequences is a fundamental advance. Unlike standard models that read text from left to right or right to left, BERT processes the complete input sequence in both directions at the same time. This bidirectional context enables the model to acquire a more comprehensive grasp of word connections. BERT is pre-trained on huge corpora with two primary unsupervised pre-training goals.:

Masked Language Model (MLM): Randomly masks off part of the words in the input sequence and trains the model to predict the masked words based on the context supplied by the surrounding words.

Next Sentence Prediction (NSP): This function trains the model to predict whether a randomly chosen sentence will follow another sentence in the corpus.

A stack of transformer encoder layers makes up BERT. The "BERT base" model has 12 layers, but the bigger "BERT large" model has 24 layers in the basic architecture as it is usually employed. A feedforward neural network and a multi-head self-attention mechanism are present in every layer. BERT uses multi-head self-attention in each transformer layer so that the model may make predictions by varying the weights given to different sections of the input sequence. This aids the model in capturing relationships between words that are further apart. BERT uses embeddings to represent words and positional encodings to convey the order of words in a sequence. These embeddings are learned during pre-training and fine-tuned for specific downstream tasks. BERT is often used for downstream tasks, such as sequence classification. For these tasks, additional output layers (e.g., a classification layer) are added on top of the pre-trained BERT architecture. During fine-tuning, these layers are trained to adapt the model to the specific task. The core BERT architecture consists of an encoder with 12 transformer blocks, 12 attention heads, and a 768 textual representation size. The model creates its vector representation by taking an input text sequence with a maximum of 512 tokens. With the use of a unique token [SEP], the sequence can be divided into one or two segments. The token [CLS] at the start of each segment denotes a distinct classification representation. The architectural diagram of BERT model has been represented below:

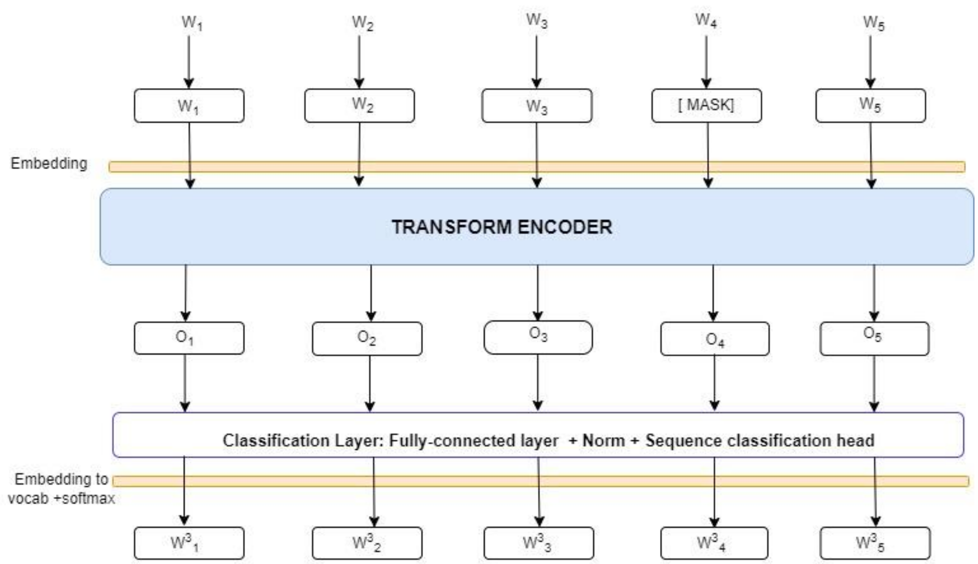

Figure 6:Modified BERT Model for Tweet Classification

4.2 Preprocessing of Tweet Data

Optimizing the performance of machine learning models requires a critical step called data preparation. Cleaning and tokenization are only two of the many approaches involved. Stopwords, punctuation, special characters, digits, and URLs are all removed from tweets during the data cleaning process. The tweet's remaining text is changed to lowercase and any emojis and hashtags are eliminated. Stripping unifies several word forms and is the last preprocessing step. The WordNetLemmatizer from the NLTK tool is used to do this. Lemmatization is done using the NLTK WordNetLemmatizer. Reducing words to their dictionary or basic form, or lemmas, while taking their context and meaning into account is a process called lemmatization. Unlike stemming, which often just removes suffixes or prefixes to obtain a root form, lemmatization aims to find the actual dictionary form of a word. The sample tweet after preprocessing is illustrated in Table 5.

Table 5: Sample of Tweet before and after preprocessing

Prior to pre-processing of Tweet	Latter to Preprocessing of Tweet
I added a video to a @YouTube playlist https://t.co/ntFJrNvSvZ How To Bitcoin Cloud Mining Free For Lifetime Urdu / Hindi	add video playlist bitcoin cloud mining free lifetime urdu hindi

To convert categorical labels, we utilize the Label Encoder function from the Scikit-Learn library. This function transforms 'Negative,' 'Neutral,' and 'Positive' labels into 0, 1, and 2, respectively. For dataset splitting into training and testing sets, we apply the train-test-split function from Scikit-Learn, resulting in a 70% training set and a 30% testing set.it function, we split the data with 70% and 30% in training and testing set respectively.

4.3. Extraction of features in Tweet

After leminization of words with WordNetLemmatizer the tweets, it's essential to extract meaningful features for classification while eliminating irrelevant words that don't significantly impact sentiment analysis.

According to reference [98], several key features are necessary to represent input tweet text effectively. The most fundamental and traditional technique for feature extraction is the "bag-of-words" (BoW) model. It includes both uni-grams (single-word terms) and bi-grams (two-word terms). BoW is essentially an unordered collection of words where word positions are disregarded, but their occurrences in the text are counted.

Another approach involves lexicon-based features, where each tweet's positive and negative terms are tallied using a predefined lexicon. During a preprocessing phase, the part of speech of extracted words can be preserved. The parts of speech considered for counting include nouns, verbs, adjectives, and adverbs.

Additionally, there are two common weighted approaches for feature extraction: TF-IDF (Term Frequency-Inverse Document Frequency) and Word2Vec. TF-IDF boosts the importance of rarely occurring words and diminishes the significance of frequently appearing terms in a text. It's the product of Term Frequency (TF), which measures how often a word appears in a text, and Inverse Document Frequency (IDF). TF-IDF aids in identifying meaningful words that contribute value to the text, as referenced in [99].

These various features are used to represent each tweet as a numeric vector. The dimension of this vector increases with the growth of distinct phrases in the input tweet collection. In this study, the scikit-learn Count Vectorizer class was employed to convert text into a bag of words representation, allowing word counting and frequency measurement.

4.4 Optimization

The optimizer employed in the experiment is Adam. Adam, short for Adaptive Moment Estimation, is a widely used optimization algorithm in deep learning. It combines features from RMSprop and Momentum by maintaining adaptive learning rates for each parameter. The algorithm computes the gradients of model parameters during training, updates first-order and second-order moving averages, and incorporates bias correction to counteract initial biases. The parameters are then updated using a combination of the first and second moments, along with a learning rate. The systemized model is been executed on T4 GPU model using the platform of Google collab.

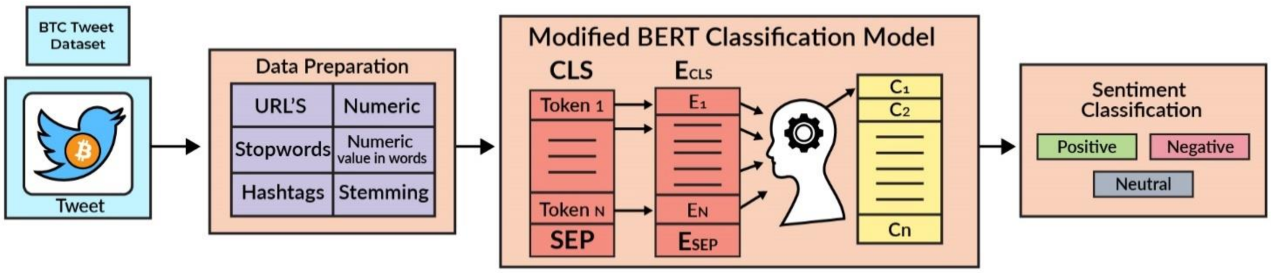

Figure 7:The Design Specification Diagram

5. Implementation Details

5.1 Libraries Applied to Execute the Task

The following libraries are applied to execute the task, their purpose and utilization has been explained as below:

NumPy:

For convenience, the numpy package is frequently imported as np in scientific computing and data processing Python. The precise attributes are listed as follows:

Numerical Computing: A key library for numerical computations in Python is called NumPy. Random number creation, multidimensional arrays, and sophisticated mathematical functions are all supported.

Effective Data Structures: When doing numerical operations, NumPy arrays are quicker and more memory-efficient than Python lists. For data processing and scientific computing, they are an essential data structure.

Mathematical Operations: NumPy is a crucial tool for a variety of mathematical and statistical activities, such as linear algebra and Fourier transformations. It encapsulated wide range of functions and operations to be easily executed in code without calling the individual functions.

Pandas:

One popular Python module that helps with maintaining structures like DataFrames and reading from CSV and XML files is the pandas package. The following is a list of traits associated with pandas:

Data Analysis and Manipulation: Pandas is an effective library for handling and analysing data. It offers functions and data structures (like DataFrame) for effectively working with structured data in CSV files, Excel spreadsheets, SQL databases, and other formats.

Data Cleaning: Pandas makes jobs like managing duplicates, cleaning up missing data, and converting data for analysis easier.

Data Exploration: Grouping and aggregating data, creating pivot tables, and exploring and summarising datasets are all made simple.

Time Series Analysis: Pandas provides strong support for operations and data related to time series.

Matplotlib:

Python users may prepare to build plots and visualizations by importing matplotlib. The matplotlib package is used for this, and matplotlib.pyplot in particular provides an easy-to-use interface for creating charts and plots. Among its primary objectives are:

Data Visualization: A variety of data visualizations, including line charts, bar charts, scatter plots, histograms, pie charts, and more, may be created by users with the help of Matplotlib.pyplot. The comprehension and analysis of complicated datasets are facilitated by these data visualizations, which increase data accessibility.

Customization: Plot look may be greatly customized using this module. Plots may be made more aesthetically pleasing and useful by users by customizing the colors, labels, titles, axis scales, markers, and other visual components.

Interactive Plotting: Interactive plots may be made with matplotlib. This implies that viewers may interact with the plots, allowing for operations like panning, zooming, and inspecting individual data points to provide more information.

Publication-Quality Graphics: Matplotlib is a master at creating graphics of a calibre appropriate for use in reports, presentations, and research papers. Because of its accuracy and adaptability, it's the best option for producing high-quality graphics.

Seaborn:

While preparing and producing statistical data visualisations in Python, the Seaborn module is frequently imported as sns. Dataframes from the pandas library may be easily utilised with Seaborn, a data visualisation framework built on top of matplotlib. Seaborn's methods and functionalities may be more conveniently integrated by importing it as an SNS. Seaborn is being imported as SNS for the following reason:

Data Visualisation using Statistics: Seaborn simplifies the process of making complex statistical charts with its high-level matplotlib interface. Its features allow for the creation of a large variety of statistical graphics, such as heatmaps, pair plots, scatter plots, bar plots, and violin plots. It could be helpful for examining distributions, seeing patterns and trends, and visualising connections within data.

Enhanced Aesthetics: Plots are aesthetically pleasing without requiring a lot of customising thanks to Seaborn's lovely preset colour schemes and themes. It makes it simple to alter plot aesthetics to suit personal tastes or blend in with the presentation and report styles.

Seamless Integration with Pandas: Seaborn works well with pandas dataframes, making it easy to create visualizations directly from data. It can automatically handle the inclusion of column names as axis labels and can use pandas data structures for plotting.

NLTK:

The Natural Language Toolkit (NLTK) is a large Python toolkit built for natural language processing (NLP) and text analysis activities. Its main goal is to let academics, developers, and linguists work with human language data. The following are some of the main goals and functions of the NLTK library:

Text Preparation: Tokenization (splitting text into words or phrases), stemming (reducing words to their base or root form), and lemmatization (reducing words to their dictionary form) are all available in NLTK. These characteristics are critical for preparing text data for analysis.

Corpus and Lexical Resources: NLTK provides access to a variety of corpora, which are enormous collections of text documents, as well as lexical resources such as WordNet, which is a lexical database that contains semantic associations between words. These materials are useful for linguistic analysis and study.

Part-of-speech Tagging: NLTK provides tools for part-of-speech tagging, which entails labelling words in a text with their grammatical category, such as nouns, verbs, adjectives, and so on. This is critical for analyzing syntactic and grammatical structures.

NER: NLTK supports NER, a method that recognizes and categorizes named entities inside text, such as names of individuals, organizations, locations, and dates.

Text Classification: NLTK includes tools for developing and training text classification models, which are used in tasks such as sentiment analysis, spam detection, and document classification.

Parsing and Grammar Analysis: The library contains parsers and syntactic analysis tools that allow you to design and analyze sentence structures and grammars.

Language Modelling: NLTK makes it easier to create language models like as n-gram models and Hidden Markov Models, which are utilized in tasks such as speech recognition and machine translation.

Sentiment Analysis: NLTK may be used for sentiment analysis, which is useful for social media monitoring and consumer feedback analysis.

Machine Learning Integration: NLTK interfaces with machine learning libraries such as scikit-learn, allowing you to develop NLP models and perform text categorization and other tasks.

Education and Research: NLTK is widely used in the academic and research communities to study and develop NLP algorithms and techniques. It serves as a valuable educational resource for those learning NLP concepts.

5.2 Framework:

The Pytorch framework was used for the challenge, and the specifics and attributes of the framework are as follows:

Pytorch: PyTorch is a deep learning framework that was created by Facebook's AI Research lab (FAIR). Its versatility, dynamic processing structure, and user-friendly interface have made it popular in the machine learning and deep learning sectors. Because of the following properties and attributes, PyTorch is especially well-suited for research projects:

Graph of Dynamic Computation: PyTorch employs a dynamic computation graph, which implies that graph is constructed as operations are executed. In contrast, several other frameworks, like as TensorFlow, employ static computation graphs. The dynamic nature of PyTorch allows for more flexibility, easy debugging, and a more intuitive way to work with complex models and prototypes.

User-Friendly and Pythonic Interface: Researchers may experiment with various models and ideas more easily with PyTorch's API as it is said to be more intuitive and Pythonic. It offers a simple and natural method for creating neural networks and working with tensors. This is particularly helpful in a research setting where quick prototyping is crucial.

Rich ecosystem: PyTorch boasts a sizable user and contributor community, as well as a thriving ecology. It gains from the wide range of pre-trained models and libraries that are readily available, such as Hugging Face Transformers, offering quick access to cutting-edge models for a variety of applications, such as computer vision and natural language processing.

Ease-of-Debugging: Debugging is made easier using PyTorch because of its dynamic nature, which makes it easier to insert print statements, examine intermediate values, and debug models. This is important in research since sophisticated and experimental models are common.

Support for Custom Models: It's common for researchers to have to try out new models and architectures. Because of PyTorch's versatility, researchers can more easily develop and train bespoke models, enabling them to push the limits of deep learning.

Active Research Community: A large portion of the research community uses PyTorch. PyTorch is therefore a perfect choice for reproducing and expanding upon the most recent research, as seen by the large number of cutting-edge research papers, models, and methodologies that are published alongside PyTorch implementations.

Detailed lessons and Documentation: PyTorch is user-friendly for both novice and seasoned deep learning practitioners, thanks to its extensive lessons and documentation.

Data Handling: The `torch.utils.data` and `torch-vision` packages in PyTorch offer tools for effective data handling, data augmentation, and data loading. Large datasets are often handled and preprocessed in research projects, therefore this is crucial.

GPU Acceleration: PyTorch facilitates GPU acceleration, enabling researchers to take advantage of GPU capability to expedite the training and testing of deep learning models.

Model Deployment: PyTorch is mostly used in research, but it also offers tools and conversion techniques for deploying models to production, testing, and deployment, which makes it a good tool for moving from theory to real-world use.

6. Evaluation

The data is now prepared after stemming and tokenization, the next step now is the implementation of Machine learning models for classification of data into the respective classes and analyzing the model performances, we have employed variety of pre-trained Machine learning models including SVM, Decision Tree, and Naïve Bayes models. Their accuracies are compared to BERT(Bidirectional

Encoder Representations from Transformers) using the same dataset. The dataset employed for this purpose consist of 50866 Tweet samples. The models are been trained on 38139 and in order to analyze their results, the models been test on 12713 samples.

6.1 Support Vector Machine Classifier

In the realm of applying machine learning classifier, we first applied the Support Vector Machine Classifier on the Bitcoin Tweet Dataset, the model provides us with an accuracy of . The complete classification report of SVM model is show below:

Table 6:Sentiment Classification - SVM Report

Classification Report for SVM Model				
	precision	recall	f1-score	support
Neutral	0.62	0.70	0.66	6623
Positive	0.77	0.60	0.68	6857
Negative	0.39	0.54	0.45	1776
accuracy			0.64	15256
macro avg	0.60	0.62	0.60	15256
weighted avg	0.66	0.64	0.64	15256
Validation accuracy Score: 0.6380				

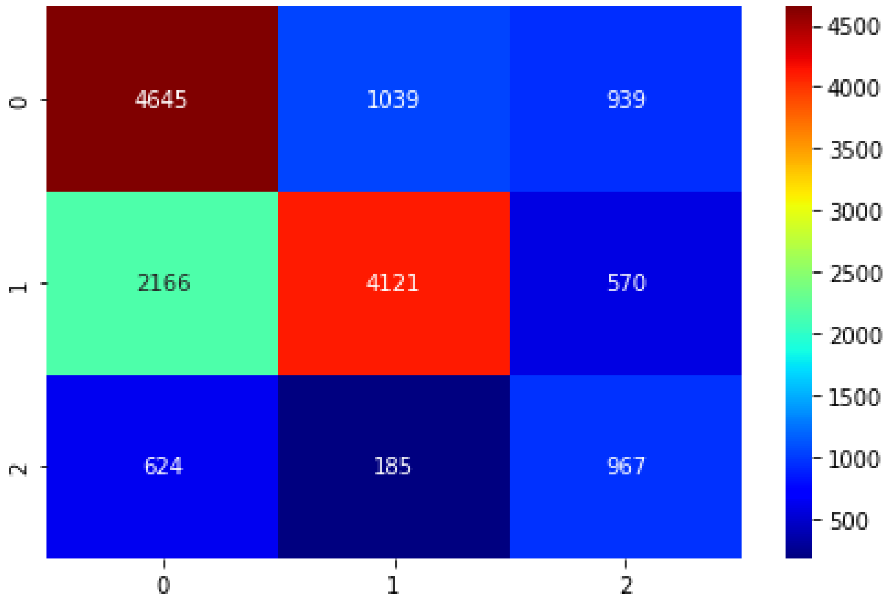

Figure 8:Sentiment Classification - SVM Confusion Matrix

6.2 Decision Tree

Decision tree is a supervised machine learning algorithm suitable for both classification and regression tasks. It offers a straightforward and relatively straightforward implementation for small to moderately sized datasets. Decision tree was the second machine learning algorithm employed in our sentiment classification of Bitcoin-related tweets. It demonstrated a test accuracy of 0.64%. Below, you can find the comprehensive classification report for the Decision Tree model.

Table 7:: Sentiment Classification – Decision Tree Report

Classification Report for Decision Tree Model				
	precision	recall	f1-score	support
Neutral	0.62	0.70	0.66	6623
Positive	0.77	0.60	0.68	6857
Negative	0.39	0.54	0.45	1776
accuracy			0.64	15256
macro avg	0.60	0.62	0.60	15256
weighted avg	0.66	0.64	0.64	15256
Validation accuracy Score: 0.6380				

Here the confusion matrix of Decision Tree:

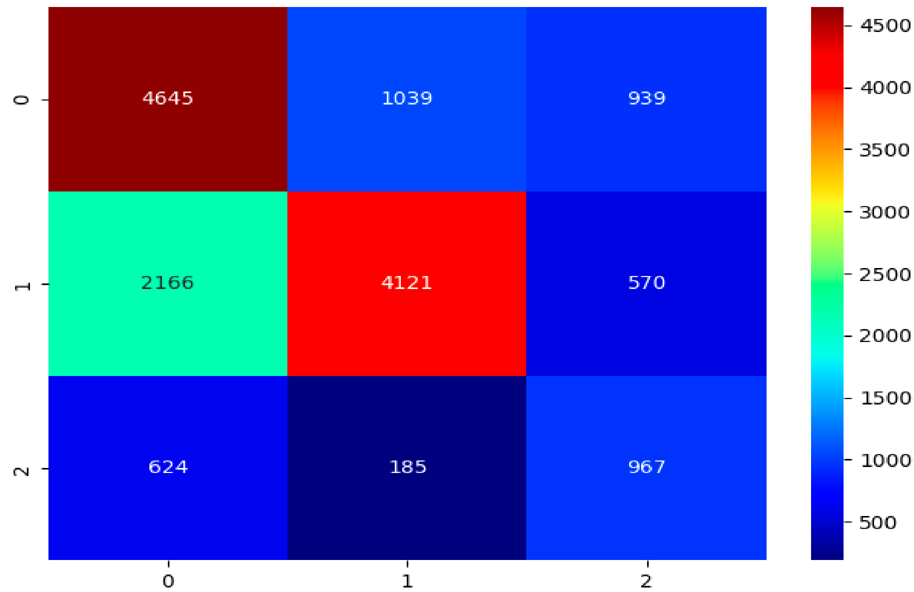

Figure 9:Sentiment Classification – Decision Tree Confusion Matrix

6.3 Naïve Bayes Model

Naive Bayes is a supervised machine learning algorithm well-suited for both classification and regression tasks. Naive Bayes was the second machine learning algorithm employed in our sentiment classification of Bitcoin-related tweets. It demonstrated a test accuracy of 0.91%. Below, you can find the comprehensive classification report for the Naive Bayes model.

Table 8:Sentiment Classification – Multinominal Naïve Bayes Report

Classification Report for Multinominal Naïve Bayes Model				
	precision	recall	f1-score	support
Neutral	0.94	0.90	0.92	6623
Positive	0.95	0.91	0.93	6857
Negative	0.68	0.91	0.78	1776
accuracy			0.91	15256
macro avg	0.86	0.91	0.88	15256
weighted avg	0.92	0.91	0.91	15256
Validation accuracy Score: 0.9067				

Here the confusion matrix of MultinominalNB

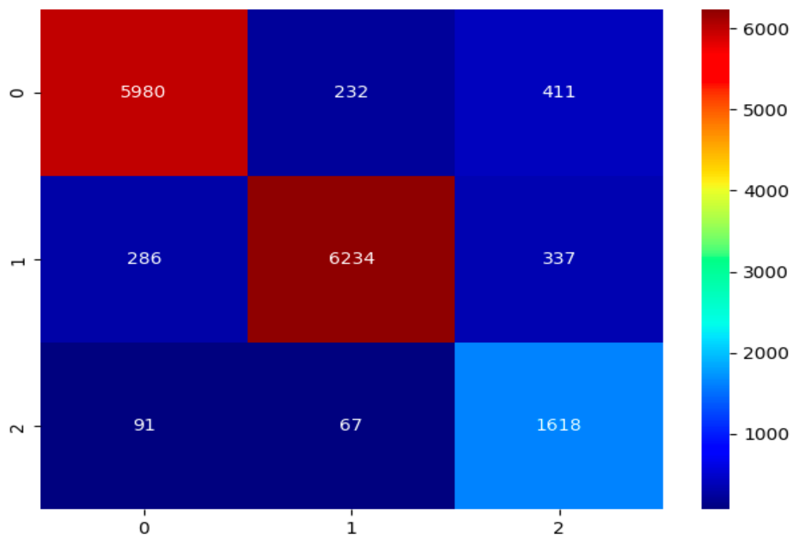

Figure 10:Sentiment Classification – Multinominal Naïve Bayes Confusion Matrix

6.4 Deep Neural Network Model

The Deep Neural Network is a supervised Deep learning architecture well-suited for analyzing the classification tasks. Deep Neural Network was the fourth model applied to the dataset to execute the sentiment analysis task for Bitcoin-related tweets. It demonstrated a test accuracy of %. Below, you can find the comprehensive classification report for the model.

Table 9: Sentiment Classification – Deep Neural Network Report

Classification Report for Deep Neural Network Model				
	precision	recall	f1-score	support
Neutral	0.96	0.88	0.92	1186
Positive	0.95	0.98	0.96	4331
Negative	0.97	0.96	0.97	4654
accuracy			0.96	10171
macro avg	0.96	0.94	0.95	10171
weighted avg	0.96	0.96	0.96	10171
Validation accuracy Score: 0.9588				

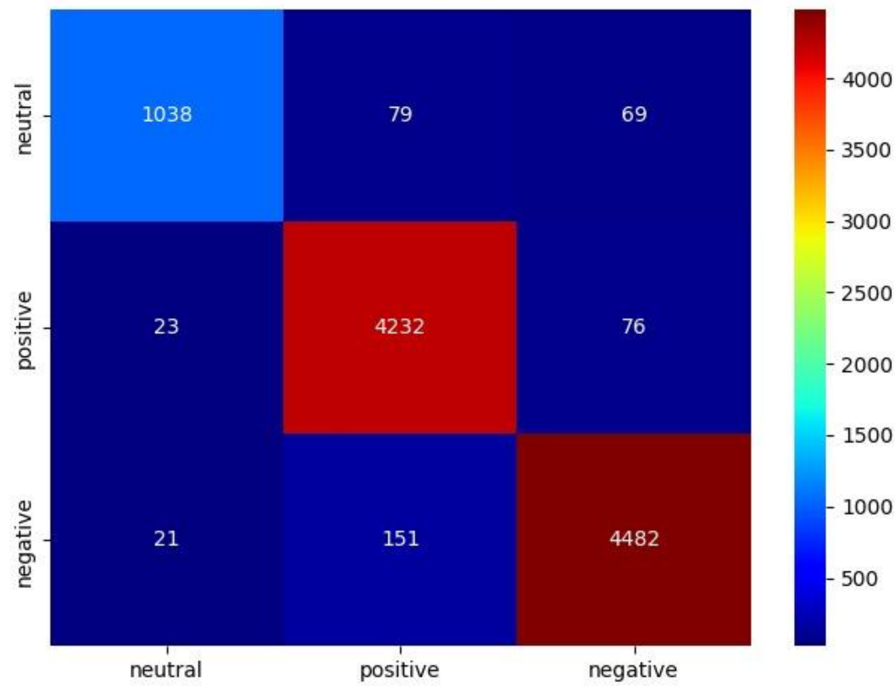

Figure 11: Sentiment Classification – Deep Neural Network Confusion Matrix

6.5 BERT Model

Now applied a pre-trained BERT model developed by Hugging Face. Additionally, Thilina Rajapakse created a transformer library that harnessed the power of the pre-trained model from Hugging Face. Our team utilized this transformer library to train the BERT model for the sentiment classification of tweets. The trained BERT model achieved an impressive training accuracy of 0.98%. The comprehensive classification report for the BERT model below:

Table 10:Sentiment Classification - BERT Report:

Classification Report for BERT Model				
	precision	recall	f1-score	support
Neutral	0.98	0.98	0.98	5483
Negative	0.98	0.99	0.99	5734
Positive	0.97	0.95	0.96	1496
accuracy			0.98	12713
macro avg	0.98	0.97	0.98	12713
weighted avg	0.98	0.98	0.98	12713

Validation accuracy Score: 0.9817

Here the confusion matrix of BERT model:

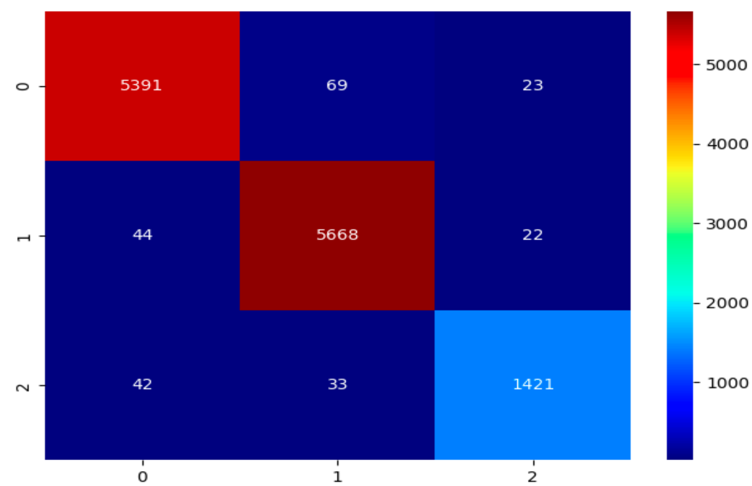

Figure 12: Sentiment Classification - BERT Confusion Matrix

6.6 Compare result of trained models

In the research experiment, crafted the best results by BERT model followed by the Decision Tree, Navie Bayes and SVM BERT model showed the 0.98% testing accuracy and the highest value for other evaluation measures. The comparison of all evaluation measures is also presented in the below table.

Table 12:Results of all trained models

Metrics	BERT	SVC	Decision Tree	MultinominalNB
accuracy	0.981672	0.637979	0.637979	0.90666
precision	0.978642	0.595433	0.595433	0.859594
Recall	0.973859	0.615606	0.615606	0.907698
f1-score	0.976214	0.597047	0.597047	0.877942

The comparison of all evaluation measure is also presented by bar graph. The bar shows the four groups of bars relevant to accuracy, precision, recall and f1-score for each model respectively. The bar is presented in below figure:

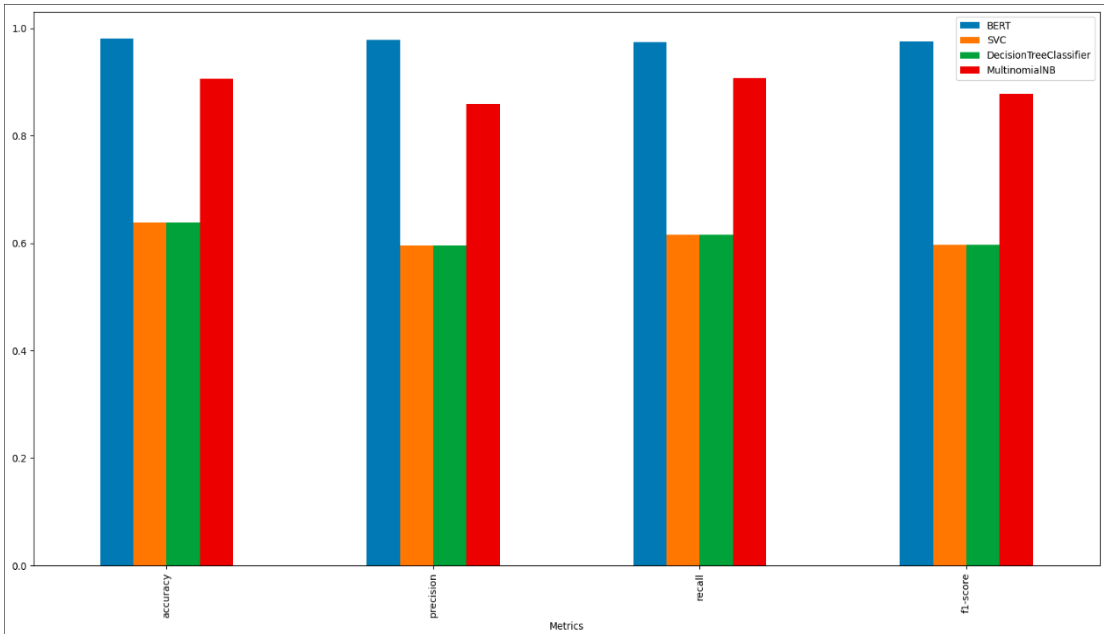

Figure 13: Evaluation Scores of trained models

6.6 Discussion

To widen our discussion, we will now compare the BERT Architectural Model with the machine learning models. The machine learning models deliver a high level of accuracy and produce satisfying outcomes. We generated a comparative bar graph exhibiting the accuracies of each trained model to offer a more thorough comparison of the accuracy across the models and to determine the best-performing model. The following is a glimpse of a bar graph that highlights the performance differences between the models:

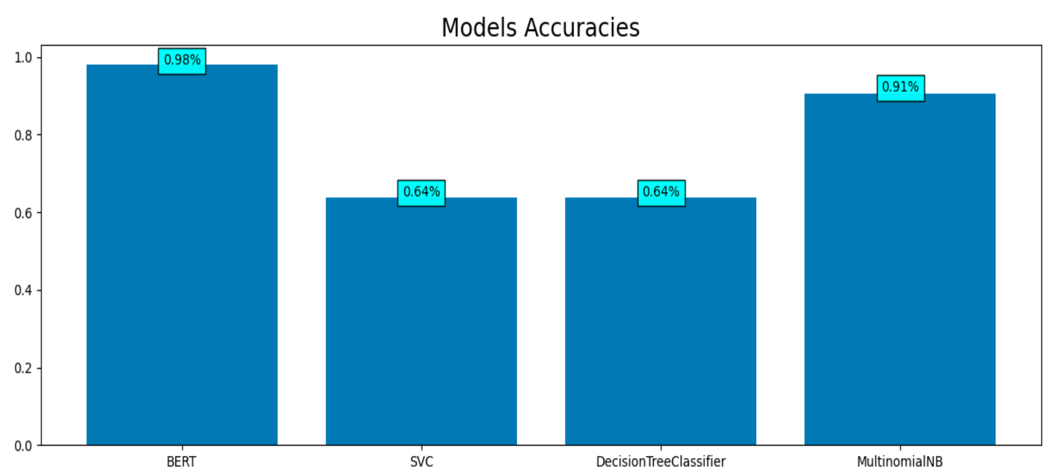

Figure14 :Accuracy Comparison of all models.

The graph clearly shows that the BERT model outperformed all other models with **98%** classification accuracy. The other state-of-art models however provide the optimum results but still less than our purposed architecture. The reason BERT Model was able to deliver the highest accuracy results was the integration of attention mechanism that assist in capturing the minor details of the text by analyzing the most probable matching score with the word, and the other is the Sequence Classification Head. The pooled representation from the pooler layer is passed through a linear layer with softmax activation for sequence classification. The number of output units in this layer corresponds to the number of labels in the classification task.

7. Conclusion and Future Work

In conclusion, the BERT model was not only able to capture the minor details, the architectural model assist in performing the sentiment classification and providing the highest classification results. The BERT model architecture provides us an accuracy of 98%, while the machine learning models, e.g, Naïve Bayes provides an accuracy of 91%, while Support Vector Machine and Decision Tree both provides an accuracy of 64%. The Transformer based architecture provide a stepping stone for the Natural Language Processing Task since the BERT was primarily designed o execute the huge dataset and been trained on Large Language Dataset and provide the best possible outcomes. The model has its huge participation in domain of NLP processing task since it's the main concern for the scientists since 2019. The NLP task found in application in the diversifying domains ranging from chat bot assistance to virtual assistance online help. The natural language processing is the main point of concern for the programmers and developers. The BERT model will assist in the real-time-management and classification of tasks where it is mandatory to monitor the ongoing scenarios and provide the on time solution ranging from developers to the people belonging to all domains. The BERT model can assist in providing the real time solution much faster and greater than any other model since it has a capability to manage huge volume of data that can provide beneficial solution for the many applications that require the real time monitoring of data.

8. Referneces

3024330, U. C.-A. at S., & 2022, undefined. (n.d.). Cryptocurrencies: A brief thematic review. *Papers.Ssrn.ComUW ChohanAvailable at SSRN 3024330, 2022•papers.Ssrn.Com*. Retrieved 23 November 2023, from https://papers.ssrn.com/SOL3/PAPERS.CFM?ABSTRACT_ID=3024330

Aggarwal, A., Gupta, I., … N. G.-2019 twelfth international, & 2019, undefined. (n.d.). Deep learning approach to determine the impact of socio economic factors on bitcoin price prediction. *Ieeexplore.Ieee.OrgA Aggarwal, I Gupta, N Garg, A Goel2019 Twelfth International Conference on Contemporary Computing (IC3), 2019•ieeexplore.Ieee.Org*. Retrieved 23 November 2023, from https://ieeexplore.ieee.org/abstract/document/8844928/

Aggarwal, C. C., & Zhai, C. X. (2012). A survey of text classification algorithms. *Mining Text Data*, *9781461432234*, 163–222. https://doi.org/10.1007/978-1-4614-3223-4_6

and, Z. K.-I. J. of I. S., & 2020, undefined. (n.d.). Sentiment analysis based direction prediction in bitcoin using deep learning algorithms and word embedding models. *Ijisae.OrgZH KilimciInternational Journal of Intelligent Systems and Applications in Engineering, 2020•ijisae.Org*, *2020*(2). Retrieved 23 November 2023, from https://ijisae.org/IJISAE/article/view/1062

Aslam, N., Rustam, F., Lee, E., … P. W.-I., & 2022, undefined. (n.d.). Sentiment analysis and emotion detection on cryptocurrency related tweets using ensemble LSTM-GRU model. *Ieeexplore.Ieee.OrgN Aslam, F Rustam, E Lee, PB Washington, I AshrafIeee Access, 2022•ieeexplore.Ieee.Org*. Retrieved 23 November 2023, from https://ieeexplore.ieee.org/abstract/document/9751065/

Balfagih, A., Data), V. K. on big data (Big, & 2019, undefined. (n.d.). Evaluating sentiment c1assifiers for bitcoin tweets in price prediction task. *Ieeexplore.Ieee.OrgAM Balfagih, V Keselj2019 IEEE International Conference on Big Data (Big Data), 2019•ieeexplore.Ieee.Org*. Retrieved 23 November 2023, from https://ieeexplore.ieee.org/abstract/document/9006140/

Behera, R., Naik, D., Rath, S., and, R. D.-N. C., & 2020, undefined. (2020). Genetic algorithm-based community detection in large-scale social networks. *SpringerRK Behera, D Naik, SK Rath, R DharavathNeural Computing and Applications, 2020•Springer*, *32*(13), 9649–9665. https://doi.org/10.1007/s00521-019-04487-0

Bermingham, A., international, A. S.-P. of the 19th A., & 2010, undefined. (2010). Classifying sentiment in microblogs: is brevity an advantage? *Dl.Acm.OrgA Bermingham, AF SmeatonProceedings of the 19th ACM International Conference on Information and, 2010•dl.Acm.Org*. https://dl.acm.org/doi/abs/10.1145/1871437.1871741

Colianni, S., Rosales, S., Project, M. S.-C., & 2015, undefined. (n.d.). Algorithmic trading of cryptocurrency based on Twitter sentiment analysis. *Cs229.Stanford.EduS Colianni, S Rosales, M SignorottiCS229 Project, 2015•cs229.Stanford.Edu*. Retrieved 23 November 2023, from http://cs229.stanford.edu/proj2015/029_report.pdf

(CSUR), F. S.-A. computing surveys, & 2002, undefined. (n.d.). Machine learning in automated text categorization. *Dl.Acm.OrgF SebastianiACM Computing Surveys (CSUR), 2002•dl.Acm.Org*. Retrieved 23 November 2023, from https://dl.acm.org/doi/abs/10.1145/505282.505283

Devlin, J., Chang, M.-W., Lee, K., Google, K. T., & Language, A. I. (n.d.). Bert: Pre-training of deep bidirectional transformers for language understanding. *Arxiv.OrgJ Devlin, MW Chang, K Lee, K ToutanovaarXiv Preprint ArXiv:1810.04805, 2018•arxiv.Org*. Retrieved 23 November 2023, from https://arxiv.org/abs/1810.04805 documentation, K. S. J.-J. of, & 2004, undefined. (2004). A statistical interpretation of term specificity and its application in retrieval. *Emerald.Com*, *60*, 11–21. https://www.emerald.com/insight/content/doi/10.1108/00220410410560573/full/html

Fan, X., Li, X., Du, F., Li, X., international, M. W.-2016 3rd, & 2016, undefined. (n.d.). Apply word vectors for sentiment analysis of APP reviews. *Ieeexplore.Ieee.OrgX Fan, X Li, F Du, X Li, M Wei2016 3rd International Conference on Systems and Informatics (ICSAI), 2016•ieeexplore.Ieee.Org*. Retrieved 23 November 2023, from https://ieeexplore.ieee.org/abstract/document/7811108/

Feng, S., Reports, A. K.-S., & 2021, undefined. (123 C.E.). Integrating online and offline data for crisis management: Online geolocalized emotion, policy response, and local mobility during the COVID crisis. *Nature.ComS Feng, A KirkleyScientific Reports, 2021•nature.Com*, *11*(1), 8514. https://doi.org/10.1038/s41598-021-88010-3

Garcia, D., science, F. S.-R. S. open, & 2015, undefined. (2015). Social signals and algorithmic trading of Bitcoin. *Royalsocietypublishing.OrgD Garcia, F SchweitzerRoyal Society Open Science, 2015•royalsocietypublishing.Org*, *2*(9). https://doi.org/10.1098/rsos.150288

Go, A., Huang, L., Entropy, R. B.-, & 2009, undefined. (2009). Twitter sentiment analysis. *Www-Nlp.Stanford.EduA Go, L Huang, R BhayaniEntropy, 2009•www-Nlp.Stanford.Edu*. https://www-nlp.stanford.edu/courses/cs224n/2009/fp/3.pdf

Gupta, S., Applications, B. G.-M. T. and, & 2018, undefined. (2018). XSS-secure as a service for the platforms of online social network-based multimedia web applications in cloud. *SpringerS Gupta, BB GuptaMultimedia Tools and Applications, 2018•Springer*, *77*(4), 4829–4861. https://doi.org/10.1007/s11042-016-3735-1

Gupta, S., & Gupta, B. B. (2018). XSS-secure as a service for the platforms of online social network-based multimedia web applications in cloud. *Multimedia Tools and Applications*, *77*(4), 4829–4861. https://doi.org/10.1007/S11042-016-3735-1/TABLES/21

Gurrib, I., Finance, F. K.-S. in E. and, & 2022, undefined. (n.d.). Predicting bitcoin price movements using sentiment analysis: a machine learning approach. *Emerald.ComI Gurrib, F KamalovStudies in Economics and Finance, 2022•emerald.Com*. https://doi.org/10.1108/SEF-07-2021-0293

Haryadi, D., … G. K.-J. of A. C. S. and, & 2019, undefined. (n.d.). Emotion detection in text using nested long short-term memory. *Academia.EduD Haryadi, GP KusumaInternational Journal of Advanced Computer Science and Applications, 2019•academia.Edu*. Retrieved 23 November 2023, from https://www.academia.edu/download/73117884/Paper_45-Emotion_Detection_in_Text.pdf

Hasan, M., Rundensteiner, E., Analytics, E. A.-J. of D. S. and, & 2019, undefined. (2019). Automatic emotion detection in text streams by analyzing twitter data. *SpringerM Hasan, E Rundensteiner, E AguInternational Journal of Data Science and Analytics, 2019•Springer*, *7*(1), 35–51. https://doi.org/10.1007/s41060-018-0096-z

Havrlant, L., Systems, V. K.-I. J. of G., & 2017, undefined. (2017). A simple probabilistic explanation of term frequency-inverse document frequency (tf-idf) heuristic (and variations motivated by this explanation). *Taylor & FrancisL Havrlant, V KreinovichInternational Journal of General Systems, 2017•Taylor & Francis*, *46*(1), 27–36. https://doi.org/10.1080/03081079.2017.1291635

Huang, C., Zhu, · Jia, Liang, Y., Yang, M., Pui, G., Fung, C., & Luo, J. (2019). An efficient automatic multiple objectives optimization feature selection strategy for internet text classification. *SpringerC Huang, J Zhu, Y Liang, M Yang, GPC Fung, J LuoInternational Journal of Machine Learning and Cybernetics, 2019•Springer*, *10*(3), 1151–1163. https://doi.org/10.1007/s13042-018-0793-x

Huang, X., Zhang, W., Tang, X., Zhang, M., Surbiryala, J., Iosifidis, V., Liu, Z., & Zhang, J. (n.d.). Lstm based sentiment analysis for cryptocurrency prediction. *SpringerX Huang, W Zhang, X Tang, M Zhang, J Surbiryala, V Iosifidis, Z Liu, J ZhangDatabase Systems for Advanced Applications: 26th International Conference, 2021•Springer*. Retrieved 23 November 2023, from https://link.springer.com/chapter/10.1007/978-3-030-73200-4_47

Imanuel Tanesab, F., Sembiring, I., Dwi Purnomo, H., Teknologi Informasi, F., Kristen Satya Wacana, U., & - Jawa Tengah, S. (2017). *Sentiment analysis model based on Youtube comment using support vector machine*. https://repository.uksw.edu/handle/123456789/13371

IOT, A. I.-2021 I. I., and, E., & 2021, undefined. (n.d.). Forecasting the early market movement in bitcoin using twitter's sentiment analysis: An ensemble-based prediction model. *Ieeexplore.Ieee.OrgA Ibrahim2021 IEEE International IOT, Electronics and Mechatronics, 2021•ieeexplore.Ieee.Org*. Retrieved 23 November 2023, from https://ieeexplore.ieee.org/abstract/document/9422647/

Jain, A., Tripathi, S., … H. D.-2018 eleventh, & 2018, undefined. (n.d.). Forecasting price of cryptocurrencies using tweets sentiment analysis. *Ieeexplore.Ieee.OrgA Jain, S Tripathi, HD Dwivedi, P Saxena2018 Eleventh International Conference on Contemporary Computing (IC3), 2018•ieeexplore.Ieee.Org*. Retrieved 23 November 2023, from https://ieeexplore.ieee.org/abstract/document/8530659/

Jiang, N., Ti, C., Mao, Y., International, T. W.-I. 2021; 2nd, & 2021, undefined. (n.d.). Analysis on relationship between bitcoin price trend and sentiment of bitcoin related tweets by ML and NLP. *Ieeexplore.Ieee.OrgN Jiang, C Ti, Y Mao, T WuICMLCA 2021; 2nd International Conference on Machine Learning and, 2021•ieeexplore.Ieee.Org*. Retrieved 23 November 2023, from https://ieeexplore.ieee.org/abstract/document/9736780/

Kim, Y. Bin, Kim, J. G., Kim, W., Im, J. H., Kim, T. H., Kang, S. J., & Kim, C. H. (2016). Predicting fluctuations in cryptocurrency transactions based on user comments and replies. *Journals.Plos.OrgYB Kim, JG Kim, W Kim, JH Im, TH Kim, SJ Kang, CH KimPloS One, 2016•journals.Plos.Org*, *11*(8). https://doi.org/10.1371/journal.pone.0161197

Kübler, S., Liu, C., & Sayyed, Z. A. (2018). To use or not to use: Feature selection for sentiment analysis of highly imbalanced data. *Natural Language Engineering*, *24*(1), 3–37. https://doi.org/10.1017/S1351324917000298

Kumar, A., … R. K.-I. in D. M. P. of the, & 2017, undefined. (n.d.). Firefly algorithm for feature selection in sentiment analysis. *SpringerA Kumar, R KhorwalComputational Intelligence in Data Mining: Proceedings of the International, 2017•Springer*. Retrieved 23 November 2023, from https://link.springer.com/chapter/10.1007/978-981-10-3874-7_66

Mardjo, A., & Choksuchat, C. (2022). HyVADRF: Hybrid VADER-Random Forest and GWO for Bitcoin Tweet Sentiment Analysis. *IEEE Access*, *10*, 101889–101897. https://doi.org/10.1109/ACCESS.2022.3209662

one, L. K.-P., & 2015, undefined. (2015). What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis. *Journals.Plos.Org*, *10*(4). https://doi.org/10.1371/journal.pone.0123923

Phillips, R., on, D. G.-2017 I. symposium series, & 2017, undefined. (n.d.). Predicting cryptocurrency price bubbles using social media data and epidemic modelling. *Ieeexplore.Ieee.OrgRC Phillips, D Gorse2017 IEEE Symposium Series on Computational Intelligence (SSCI), 2017•ieeexplore.Ieee.Org*. Retrieved 23 November 2023, from https://ieeexplore.ieee.org/abstract/document/8280809/

Raj Pant, D., Neupane, P., Poudel, A., Kumar Pokhrel, A., & Kumar Lama, B. (n.d.). Recurrent neural network based bitcoin price prediction by twitter sentiment analysis. *Ieeexplore.Ieee.OrgDR Pant, P Neupane, A*

Poudel, AK Pokhrel, BK Lama2018 IEEE 3rd International Conference on Computing, Communication, 2018•ieeexplore.Ieee.Org. https://doi.org/10.1109/CCCS.2018.8586824

Rouhani, S., & Abedin, E. (2020). Crypto-currencies narrated on tweets: a sentiment analysis approach. *International Journal of Ethics and Systems*, *36*(1), 58–72. https://doi.org/10.1108/IJOES-12-2018-0185/FULL/PDF

Saif, H., He, Y., & Alani, H. (2012). *Alleviating data sparsity for twitter sentiment analysis*. 2–9. https://oro.open.ac.uk/38501/

Şaşmaz, E., on, F. T.-2021 6th I. C., & 2021, undefined. (n.d.). Tweet sentiment analysis for cryptocurrencies. *Ieeexplore.Ieee.OrgE Şaşmaz, FB Tek2021 6th International Conference on Computer Science and, 2021•ieeexplore.Ieee.Org*. Retrieved 23 November 2023, from https://ieeexplore.ieee.org/abstract/document/9558914/

Sattarov, O., Jeon, H., … R. O.-… C. on I., & 2020, undefined. (2020). Forecasting bitcoin price fluctuation by twitter sentiment analysis. *Ieeexplore.Ieee.OrgO Sattarov, HS Jeon, R Oh, JD Lee2020 International Conference on Information Science and, 2020•ieeexplore.Ieee.Org*. https://doi.org/10.1109/ICISCT50599.2020.9351527

Schoen, H., Gayo-Avello, D., Metaxas, P. T., Mustafaraj, E., Strohmaier, M., & Gloor, P. (n.d.). The power of prediction with social media. *Emerald.Com*. https://doi.org/10.1108/IntR-06-2013-0115

Shah, F. M., Reyadh, A. S., Ahmed, S., Shaafi, A. I., & Sithil, F. T. (n.d.). Emotion detection from tweets using AIT-2018 dataset. *Ieeexplore.Ieee.OrgFM Shah, AS Reyadh, AI Shaafi, S Ahmed, FT Sithil2019 5th International Conference on Advances in Electrical, 2019•ieeexplore.Ieee.Org*. https://doi.org/10.1109/ICAEE48663.2019.8975433

Shah, F. M., Reyadh, A. S., Ahmed, S., Shaafi, A. I., & Sithil, F. T. (2019). Emotion detection from tweets using AIT-2018 dataset. *Ieeexplore.Ieee.OrgFM Shah, AS Reyadh, AI Shaafi, S Ahmed, FT Sithil2019 5th International Conference on Advances in Electrical, 2019•ieeexplore.Ieee.Org*, 575–580. https://doi.org/10.1109/ICAEE48663.2019.8975433

Sharifirad, S., Jafarpour, B., & Matwin, S. (2019). *How is Your Mood When Writing Sexist tweets? Detecting the Emotion Type and Intensity of Emotion Using Natural Language Processing Techniques*. http://arxiv.org/abs/1902.03089

Silva, N., Coletta, L., Surveys, E. H.-A. C., & 2016, undefined. (2016). A survey and comparative study of tweet sentiment analysis via semi-supervised learning. *Dl.Acm.OrgNFFD Silva, LFS Coletta, ER HruschkaACM Computing Surveys (CSUR), 2016•dl.Acm.Org*, *49*(1). https://doi.org/10.1145/2932708

Tommasel, A., & Godoy, D. (2018). A Social-aware online short-text feature selection technique for social media. *Information Fusion*, *40*, 1–17. https://doi.org/10.1016/J.INFFUS.2017.05.003

Valencia, F., Gómez-Espinosa, A., Entropy, B. V.-A.-, & 2019, undefined. (2019). Price movement prediction of cryptocurrencies using sentiment analysis and machine learning. *Mdpi.ComF Valencia, A Gómez-Espinosa, B Valdés-AguirreEntropy, 2019•mdpi.Com*, *21*(6). https://doi.org/10.3390/e21060589

V. D, C. (2019). Hybrid approach: naive bayes and sentiment VADER for analyzing sentiment of mobile unboxing video comments. *International Journal of Electrical and Computer Engineering (IJECE)*, *9*(5), 4452. https://doi.org/10.11591/IJECE.V9I5.PP4452-4459

Wang, H., Li, Z., Li, Y., Gupta, B. B., & Choi, C. (2020a). Visual saliency guided complex image retrieval. *Pattern Recognition Letters*, *130*, 64–72. https://doi.org/10.1016/J.PATREC.2018.08.010

Wang, H., Li, Z., Li, Y., Gupta, B. B., & Choi, C. (2020b). Visual saliency guided complex image retrieval. *Pattern Recognition Letters*, *130*, 64–72. https://doi.org/10.1016/J.PATREC.2018.08.010

Zhang, X., Li, W., Ying, H., Li, F., … S. T.-I. I. of T., & 2020, undefined. (n.d.). Emotion detection in online social networks: a multilabel learning approach. *Ieeexplore.Ieee.OrgX Zhang, W Li, H Ying, F Li, S Tang, S LuIEEE Internet of Things Journal, 2020•ieeexplore.Ieee.Org*. Retrieved 23 November 2023, from https://ieeexplore.ieee.org/abstract/document/9123365/

Zhang, X., Li, W., Ying, H., Li, F., Tang, S., & Lu, S. (2020). Emotion Detection in Online Social Networks: A Multilabel Learning Approach. *IEEE Internet of Things Journal*, *7*(9), 8133–8143. https://doi.org/10.1109/JIOT.2020.3004376

Zhang, Z., Sun, R., Zhao, C., Wang, J., Chang, C. K., & Gupta, B. B. (2017). CyVOD: a novel trinity multimedia social network scheme. *Multimedia Tools and Applications*, *76*(18), 18513–18529. https://doi.org/10.1007/S11042-016-4162-Z/TABLES/1

Zhu, X. (2005). *Semi-supervised learning literature survey*. https://minds.wisconsin.edu/handle/1793/60444