Convergence of Twitter Sentiment Analysis and Optimized Learning Models for Predicting Bitcoin Price Volatility

Hasindu Rathnayake*, Muditha Tissera
University of Kelaniya, Sri Lanka

Abstract—**Bitcoin has attained increasing recognition and interest from individuals and corporations, with more than $1 billion market capitalization. Twitter users' sentiment on the topic is a major factor that influences volatility of Bitcoin's price. Compared to other financial markets, there are a limited number of studies that discuss the price fluctuation prediction of Bitcoin using Twitter sentiment. A dataset with 16 million tweets from August 2018 to October 2019 was utilized for finding the correlation between the daily close price of Bitcoin and Twitter sentiment. This dataset was pre-processed by following steps such as removing null, duplicate and non-English tweets. The sentiment analysis was carried out using VADER sentiment analyzer. This research utilized hyperparameter optimization and improved two deep learning models (with Long Short-Term Memory and Convolutional Neural Network architectures), for the tasks of direction and magnitude prediction with accuracies of 82.35% and 72.06%, respectively on test datasets. With hyperparameter optimization this research addresses a gap in the existing research of this research area, which was not utilizing hyperparameter optimization to improve deep learning models.**

Index Terms—**bitcoin fluctuations, deep learning, bitcoin predictions, twitter sentiment, hyperparameter optimization**

I. Introduction

Blockchain technologies have emerged as substantial technological trends within the past years and continue to stay prominent in the year 2023 [1]. A blockchain is a distributed and decentralized ledger that keeps the records of immutable transactions. All the users of the blockchain can access the ledger, thereby providing transparency to every transaction that occurred within the blockchain. Cryptocurrency is the most popular application of blockchains [2]. Bitcoin is the most popular as well as the first cryptocurrency, which was introduced in 2009 by Satoshi Nakamoto [3], to solve the issues in trust-based transaction system which involved third-party entities such as banks, card providers and other payment processing organizations. According to Nakamoto, some of the inherent issues of this model are mediation costs, need for enforcement of minimum transaction size and reversibility of transactions. Since the financial institutes acting as the trusted third parties must generate profit, the mediation costs were added to the overall transaction costs. Minimum transaction sizes were enforced by the transaction processing third party institutes, which eliminated the possibility of small casual transactions. The third issue with this model is that it was not able to provide irreversible payments for irreversible services. Bitcoin utilizes the immutability of records and decentralized nature of the blockchain to provide a cryptographic proof-based transaction system that mitigates the limitations of the trust-based system. Bitcoin provided transactions that are computationally impractical to reverse, which provided the protection from frauds to the sellers and easily implemented routine escrow mechanisms, which provided the protection for the buyers.

With the increase of interest and investments in Bitcoin, several governing bodies including the US government and the European Union have recognized Bitcoin as a legal currency and introduced regulations and tax principles [4]. International corporations such as Overstock, PayPal and Microsoft have embraced Bitcoin as a viable payment method [5].

However, with this increase of interest, investments and legal recognition, the losses of investments have also increased [6]. A 2022 study explained that there was a sudden increase of the number of downloads of crypto apps in October 2022. The study further revealed if the users who downloaded crypto apps and invested in Bitcoin during that period, 81% of them would incur losses. Furthermore, the study explained that 1/3 of the investors have experienced losses of their investments during the period between 2015 and 2022.

Most of these losses of investments are caused by the volatile nature of Bitcoin i.e., the high frequency and magnitude of price movements, up or down [7].

A major factor of Bitcoin price volatility is the phenomenon known as the media effect, which is the ability of news and media outlets to influence price changes of a particular asset class, sector, or overall market [8]. A report in 2017 illustrates that 67% of Americans get at least a portion of their news from social media platforms [9]. Therefore, social media platforms have become major news outlets. Twitter is one of the key outlets of news and information regarding cryptocurrencies including Bitcoin as well as other financial markets [10]. Therefore, Twitter sentiment i.e., the feeling or attitude users express on Twitter regarding Bitcoin has shown to be a major factor in Bitcoin price fluctuations [11].

Elon Musk's tweets from 2021 can be considered as exam-

Correspondence: Hasindu Rathnayake (E-mail: hasindushanuka@gmail.com)
Received: 16-06-2024 **Revised:** 12-08-2024 **Accepted:** 09-09-2024
Hasindu Rathnayake and Muditha Tissera are from University of Kelaniya (hasindushanuka@gmail.com, mudithat@kln.ac.lk)
DOI: https://doi.org/10.4038/icter.v18i2.7298
The 2025 Special Issue contains the full papers of the abstracts published at the 24th ICTer International Conference.

ples for price fluctuations caused by social media sentiment. In his tweet from March 24, 2021, Musk stated that Tesla was accepting Bitcoin as payment for Tesla cars. The positive sentiment caused by this tweet influenced Bitcoin price to increase, as shown in Figure 1. After two months of his initial tweet on Bitcoin and Tesla, on May 13, 2021, Musk tweeted that Tesla is no longer accepting payments via Bitcoin, because Bitcoin mining consume large amounts of energy and their company was concerned about environment pollution and did not want to encourage pollution caused by the mining. This tweet caused Bitcoin price to drop around 15% [12]. The subsequent price fluctuation is illustrated in Figure 2.

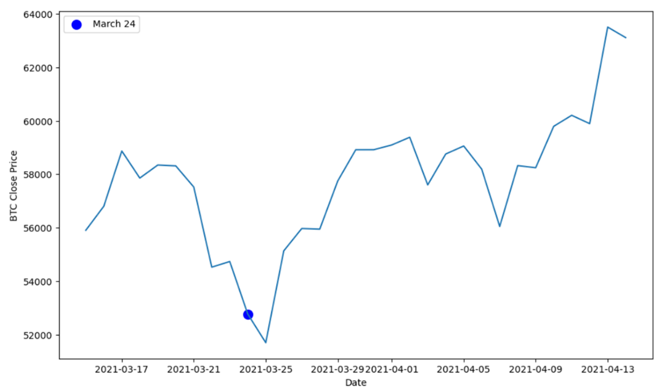

Fig. 1: BTC Price Fluctuation After Musk's Positive Tweet

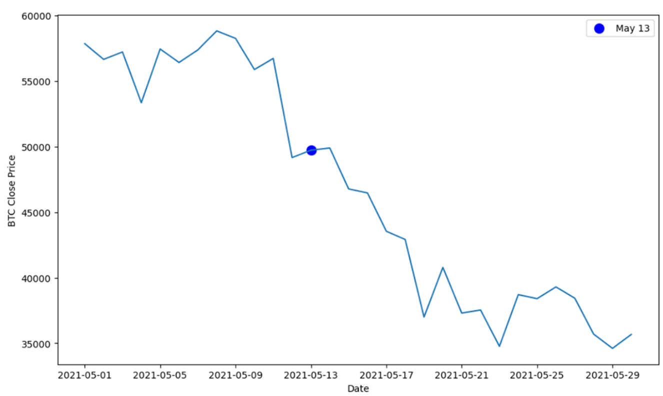

Fig. 2: BTC Price Fluctuation After Musk's Negative Tweet

In recent years, deep learning has become a powerful tool for tackling complex prediction tasks in various fields, including finance [13]. However, the success of these models depends not only on the data for the neural network but also on the careful selection of hyperparameters [14].

Hyperparameters are the configurations that are not learned from the data but are set prior to training a neural network. These hyperparameters govern various aspects of the training process and the architecture of the neural network, such as learning rate, batch size and activation functions.

Hyperparameter optimization refers to the process of automatically searching for the best combination of hyperparameters to improve model performance. In the context of deep learning, hyperparameter tuning can significantly influence the model's ability to generalize and make accurate predictions, especially in complex and volatile environments like cryptocurrency markets.

A. *Research Question*

Can hyperparameter optimization enhance predictive models for forecasting the direction and magnitude of Bitcoin price changes?

B. *Objective*

This paper aims to address the identified gap in existing studies by utilizing hyperparameter optimization for deep learning models that predict the direction and magnitude of price changes, and compare the models with the same structure and trained on the same data but have not used hyperparameter optimization. The comparison is aimed to identify the impact of hyperparameter optimization.

II. Related work

Several studies have identified this relationship and introduced predictive models for predicting Bitcoin price fluctuations primarily utilizing Twitter sentiment.

Authors of the research [15] have tested the hypothesis that Bitcoin price forecasting must consider crowd sentiments, because exchange rates depend on behavioral signal rather than any fundamental conditions. This study was carried out by retrieving English tweets with keyword 'Bitcoin' and 'exchange rate' via Twitter API (Application Programming Interface) and Bitcoin exchange rates for the period from January 2014 to September 2017. The sentiment analysis of tweets with Pattern package (a package in Python language) revealed a distribution of 39% of positive, 25% of negative and 36% of neutral tweets. A CNN (Convolutional Neural Network) model, which is a type of artificial neural network, that was implemented with TensorFlow was used for predicting the daily directional change of Bitcoin price. Max pooling and dropout computing techniques were used for reducing computational complexity and improving generalization. Upon training and testing, the best performing model achieved 68.6% of accuracy on the test data. Furthermore, the researchers used exchange rates data (without sentiment scores) to forecast price changes. However, this attempt only provided an accuracy of 52.6%. Therefore, the researchers' hypothesis about crowd sentiment influencing the exchange rates was validated.

In 2019, a pioneering study was conducted to predict Bitcoin price fluctuations using deep learning as well as word-embedding models [16]. English tweets related to Bitcoin were gathered from May 1 to August 1, 2019, totalling 17,629 tweets from both individual and organizational accounts. The tweets were labelled as positive or negative sentiment using TextBlob. Various preprocessing steps were applied to clean the data, and word embedding models such as Word2Vector, GloVe (Global Vector), FastText and deep learning models i.e., models that utilize neural networks such as CNN, RNN (Recurrent Neural Network), LSTM (Long Short-Term Memory) were implemented using Python's Keras library. FastText outperformed other models, achieving 89.13% of accuracy, while LSTM followed with 87.45%. The remaining models ranked as CNN, RNN, GloVe, and Word2Vector in descending order of accuracy.

In a 2018 study, researchers went further than predicting just Bitcoin's price direction and forecasted the price amount [17]. Tweets from well-known crypto news accounts between January 2015 and December 2017 were collected and manually labelled as positive, negative, or neutral. The dataset consisted of 2,585 positive, 1,669 negative, and 3,200 neutral tweets. FuzzyWuzzy, regular expressions and Stanford Named Entity Recognizer were used for tweet filtering, hyperlink and emoji removal, and entity extraction, respectively. Feature extraction was done using Word2Vector and Bag-of-Words. Five classification models were trained, including Naïve Bayes, Bernoulli Naïve Bayes, Multinomial Naïve Bayes, Linear Support Vector Classifier, and Random Forest. A voting classifier on these five models achieved 81.39% of accuracy. The voting classifier was compared with an LSTM model which had an accuracy of 77.62%.

A comparison of ARIMAX (Autoregressive Moving Average with Exogeneous Variables) and LSTM models were conducted with the aim of showing that crowd sentiment is one of the most important factors in predicting the price fluctuation of Bitcoin [18]. Tweets spanned the period from April 2017 to October 2019. Sentiment scoring was done with VADER (Valence Aware Dictionary and sEntiment Reasoner) library. Granger's causality test confirmed the relationships between Bitcoin price and input features. Among various feature combinations, sentiment-based inputs produced the lowest MSE (Mean Squared Error). The ARIMAX model outperformed the LSTM, despite LSTM's popularity in stock market prediction. The ARIMAX model achieved a 0.03% MSE, while the LSTM model had a 0.14% MSE.

A 2022 study investigated the optimal time interval for Twitter sentiment to reliably predict Bitcoin price fluctuations [19]. They obtained tweet and Bitcoin price datasets from Kaggle. The pre-processed dataset spanned from August 30, 2018 to November 23, 2019. VADER model was used for sentiment analysis. Sentiment and price datasets were merged with different lag times, creating three datasets. These were used to predict price change direction and magnitude. BiLSTM (Bidirectional LSTM) outperformed CNN and LSTM in direction prediction (60.9% accuracy), while CNN achieved the best F1-score for magnitude prediction (14.21%). A voting classifier combined direction and magnitude predictions, achieving a mean accuracy of 68.4%, demonstrating the importance of sentiment in predicting Bitcoin price in a highly volatile environment.

An overview of the existing studies is given in Table I.

TABLE I: Overview of Related Studies.

Study	**Date Range**	**Performance Metrics**	**Direction / Magnitude**
Galeshchuk et al., 2018	January 2014 to September 2017 (912 days)	Accuracy – 68.6%	Direction
Kilimci, 2020	May 1, 2019 to August 1, 2019 (93 days)	Accuracy – 89.13 F-score – 84.06% Precision – 86.97% Recall – 90.28%	Direction
Pant et al., 2018	January 1, 2015 to December 31, 2017 (1096 days)	Accuracy – 77.62%	Both
Serafini et al., 2020	April 2017 to October 2019 (944 days)	MSE – 0.03%	Both
Critien et al., 2022	August 30, 2018 to November 23, 2019 (450 days)	Accuracy – 68.4%	Both

The review of existing studies revealed gaps for improvement, underutilization of optimization techniques such as hyperparameter optimization to find the best predictive models.

The following section reviews studies focused on the impacts of hyperparameter optimization on the performance of predictive models in various domains.

In 2021, a study used conventional machine learning models (non-neural network models) to find the impact of hyperparameter optimization and the predicting capabilities of Saudi stock market prices [20]. They utilized the daily closing stock prices of 11 companies from the listing date to 31st December 2020. The models were implemented with SciKit-Learn library. The models that were utilized in this study included models such as Support Vector Regression (SVR) and Kernel Ridge Regression (KRR). The performances were measured by Root Mean Square Error (RMSE) and Mean Magnitude of Relative Error (MMRE). They have trained each model 30 times with the default hyperparameter values to get more reliable and robust performance values. Among the un-tuned models, KRR achieved the highest performance with an average RMSE of 0.71 and and average MMRE of 1.17.In comparison, SVR recorded an average RMSE of 4.89 and average MMRE of 6.12. The grid search was utilized to find the best hyperparameter combinations. For SVR, kernel, C and epsilon hyperparameters, and for KRR, alpha and kernel hyperparameters were optimized during the study. When comparing the optimized models, SVR achieved the highest performance of an average RMSE of 0.70 and 1.14 average MMRE. It has outperformed the optimized KRR, which achieved average RMSE and MMRE of 0.71 and 1.16, respectively.

An empirical study was conducted in 2022 to observe the impact of hyperparameter optimization on performance properties such as inference accuracy, inference latency, model size and battery consumption in mobile platforms and cloud servers [21]. The authors used four types of models (subject models), namely CNN image classification, Resnet-50, CNN text classification and LSTM sentiment classification. Hyperparameter optimization was carried out with Keras Tuner. For each subject model, they have created 100 models (400 total models), each with different hyperparameter combination. Out of these 100 models, 10 models with the highest accuracies were selected to be compared with baseline (unoptimized) models (40 total models). The authors observed that these top 10 models achieved higher inference accuracies with respect to their baseline counterparts. For each subject model, their top 10 model exhibited inference accuracies with minimal differences. But other performance properties were significantly different between the top 10 models of the same subject model. For example, the inference accuracy (normalized by the minimum value) of CNN text classification model had a first quartile (Q1) of 1.01 and a third quartile (Q3) of 1.02, while achieving a Q1 closer to 1 and Q3 more than 10 for model size (size in megabytes). With these observations, authors recommend to consider not only the inference accuracy, but also other performance aspects discussed in the study when selecting the optimized model.

The potential of using optimized hyperparameters on a variational autoencoder including LSTM layers (LSTM-VAE) for time series in a physical system was evaluated in the study [22]. A genetic algorithm was used as the hyperparameter optimization technique. The goal was to determine whether this method was suitable to identify anomalies in real-time. This study used time-series sensor data recorded during geo-drills at different construction sites. Optimized hyperparameters include number of train epochs, number of neurons in the encoder and decoder, learning rate and mini batch size. The results of this study suggested that the clusters of anomalous and non-anomalous data can be distinguished with much clarity when using models with optimized hyperparameters compared to models trained on random hyperparameters.

III. Methods

A. Data Gathering

'Bitcoin Tweets – 16M Tweets' [23] dataset was used for this study, which is the same dataset used in [19] study. The Python package, Yfinance was used for retrieving Bitcoin price data from Yahoo Finance platform, which provides data on various financial markets such as stock prices, exchange rates and cryptocurrency prices.

B. Data Preprocessing

Preprocessing tweets dataset followed steps such as removing null, duplicates, non-English tweets (with FastText classifier [24]), replacing user mentions with generic text '@user', removing new line characters, URLs, records that do not contain '#bitcoin' and '#btc' hashtags, replacing HTML characters with their corresponding ASCII characters and selecting a date range without missing tweets. This resulted in tweets spanning from August 28, 2018 to November 23, 2019. Using state-of-the-art transformer-based sentiment analyzers such as Twitter-RoBERTa-base was not practical for this dataset with 9.8 million tweets because the model did not show adequate speed. The estimated time for sentiment analysis was 433 hours (18 days). Therefore, VADER model was utilized which calculated sentiment scores for the dataset within 1.5 hours. After sentiment analysis, datasets were processed to include daily negative, neutral, positive sentiment scores and daily tweets counts.

When considering preprocessing of Bitcoin price dataset, the index (which was originally the date) was reset to be numerical, and the date was added as a column to the data frame for the ease of future operations. The daily Bitcoin close price column was kept, and other price columns were dropped.

Several datasets were created by appending pre-processed tweets datasets with Bitcoin price columns. Each of the datasets was created such that there was a different number of days of lag between the sentiment score date and the corresponding Bitcoin price date. For an example, if the lag was 1 day, the row with sentiment scores of December 01, 2018 contained the Bitcoin price data of December 02, 2018. Following this method, datasets were created with 1, 2, 3, 5, 7, 14 and 30 days lags.

C. Exploratory Data Analysis

Exploratory data analysis was carried out to identify patterns in the dataset.

The dataset showed considerable correlation between close price and sentiment scores as well as daily tweet count. This can be viewed in Figures 3 and 4 which show correlation matrix for 7 days lag and time series plot between positive sentiment score and close price.

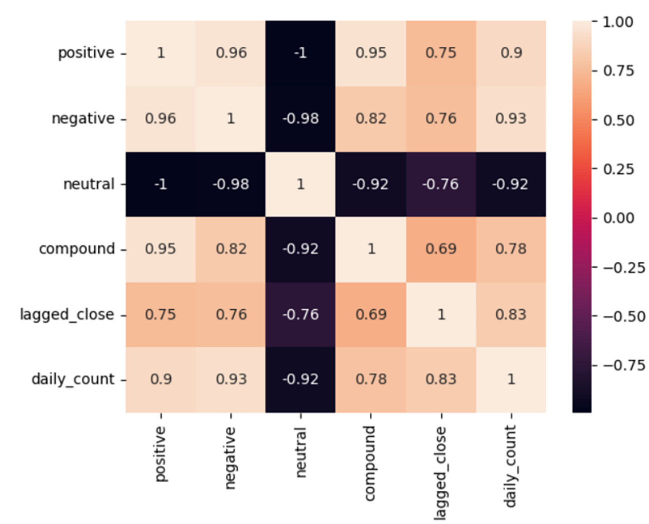

Fig. 3: 7 Days Lag Correlation Matrix of 16M Tweets Dataset

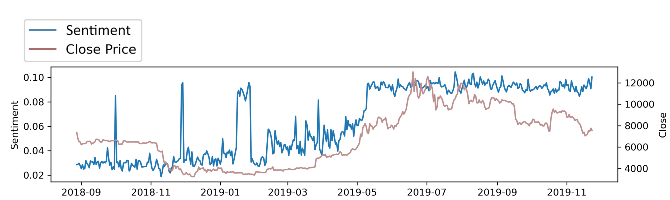

Fig. 4: Time Series of Sentiment Score and Close Price with 7 Days Lag

D. Model Development

Three types of deep learning models were experimented with for developing models. LSTM and BiLSTM models were used since they are two types of recurrent neural networks (RNN), and have the capability of remembering previous data and utilizing these data to predict future values with greater accuracy than time-series algorithms such as ARIMA or SARIMA [25], [26]. BiLSTM is an extension of the LSTM model. In contrast to the standard LSTM, BiLSTM can be trained to predict both positive and negative time directions at the same time. This characteristic can be useful for training models on Bitcoin price fluctuations, which allows the model to read input data from both the past and future. One-dimensional CNN models were also experimented with as they can be utilized for predictions on sequential data such as time series data. A key benefit of using these deep learning models is that they can extract features from inputs without human intervention [27], which allows researchers to significantly reduce the time spent on feature engineering. TensorFlow library was utilized for implementing models.

Since the primary objective of this study is properly tune hyperparameter optimization, Optuna framework [28] was utilized for hyperparameter optimization. Optuna utilizes the define-by-run principle to dynamically construct the search space. This framework combines efficient searching and pruning (automated early stopping) to improve the optimization cost. Optuna allows the users to define an objective value and whether the value should be maximized or minimized. In the context of this research, the training accuracy was the objective value that should be maximized.

In this study, one type of hyperparameter that was optimized is the activation function i.e., the mathematical function that determines whether the neuron provides an output. Hyperbolic tangent (Tanh), Rectified Linear Unit (ReLU), Scaled Exponential Linear Units (SELU) and Swish activation functions were the options given to Optuna, from which the most suitable activation functions were selected for input layer and hidden layers in each model. Softmax activation function was used for the output layers of the model to get the probability distribution, because the price fluctuation prediction will be approached as a classification problem. Furthermore, Adam optimizer and Categorical cross-entropy loss function were utilized. Early stopping mechanism was utilized to prevent overfitting.

This research has developed models for two tasks, predicting the direction and the magnitude of Bitcoin price change, and has evaluated to identify the best performing model in each task. By combining the best performing models (base models), a voting classifier was implemented. This approach is also known as ensemble learning, and it usually provides a robust model. However, there are several disadvantages such as increased expenses in terms of computability, lack of explainability and decreased performance [29]. Considering these disadvantages, the performance of the voting classifier was also evaluated.

1) Developing Direction Prediction Models: LSTM and BiLSTM models were experimented for developing the direction prediction model. Since there are only two directions (positive – price increase, negative – price decrease), the direction prediction problem can be approached as a binary classification problem.

BiLSTM models consisted of two bidirectional LSTM layers, each followed by a dropout layer, and a dense layer as output layer. The dropout layers were used for preventing overfitting and improving generalization. Hyperparameters that were optimized are the number of neurons of a hidden layer, activation functions for BiLSTM layers, dropout ratios for dropout layers and batch size. Seven BiLSTM models were trained on the 7 datasets with different lags. LSTM models consisted of an LSTM layer and a dropout layer. A dense layer with two neurons and the Softmax activation function was utilized as the output layer. Similar to BiLSTM, 7 models were trained on the 7 prepared datasets. Four hyperparameters in the model, such as the activation function of the LSTM layer, number of neurons in the LSTM layer, batch size and the dropout ratio of the dropout layer were optimized.

Among the 14 models that were developed, BiLSTM model which was trained on 7 days lag dataset achieved a high level of generalization as well as a higher performance, which will be discussed deeply in Results section.

2) Developing Magnitude Prediction Models: Since 7 days lag BiLSTM direction predicting model showed the highest performance, magnitude prediction models were trained on the same dataset. Price change magnitudes were binned into four bins, as shown in Table II. Inequal ranges for bins were selected due to the reason that there was a higher sample density in the range from -\$330 to \$330. These bin labels were used as a feature instead of changing direction.

TABLE II: Price Change Bins.

Price change range	Bin	Number of elements in the bin
$change <= -\$330$	0	124
$-\$330 < change <= \$0$	1	123
$\$0 < change <= \$330$	2	111
$\$330 < change$	3	95

The structure of the LSTM model consisted of three LSTM layers each followed by a dropout layer and dense layer. Eight hyperparameters were optimized such as number of neurons in a hidden layer, activation functions for each of the three LSTM layer, dropout ratios for dropout layers and the batch size.

Two BiLSTM layers, each followed by a dropout layer and a dense output layer were used for developing magnitude predicting BiLSTM model. Number of neurons in a hidden layer, activation functions for BiLSTM layers, dropout ratios and the batch size were the hyperparameters that were optimized.

CNN model was structured as, two one-dimensional convolutional layers, each followed by a one-dimensional max pooling layer and a dropout layer. After these layers, a flatten layer was placed for the purpose of converting multi-dimensional feature maps into one dimension before passing to the output layer. A dense layer was used as the output layer. Number of neurons in a hidden layer, activation functions for convolutional layers, dropout ratios and the batch size were tuned using Optuna.

Among the considered models for magnitude prediction, CNN model outperformed other models by obtaining higher average metrics on 20 random datasets that were created from test dataset.

3) Developing Voting Classifier: Selected direction prediction model (BiLSTM) and magnitude prediction model (CNN) were combined to develop a voting classifier. When an input is given, BiLSTM and CNN models independently predict the direction and the magnitude of the price change. If the direction predicted by the direction prediction model matches the direction that corresponds to the bin predicted by the magnitude prediction model, the output is valid, and it is accepted. Otherwise, the output is discarded. The average metrics of the voting classifier such as precision, recall and F1-score were also calculated using 20 random subsets created from the dataset. The flow chart of the voting classifier is depicted in Figure 5.

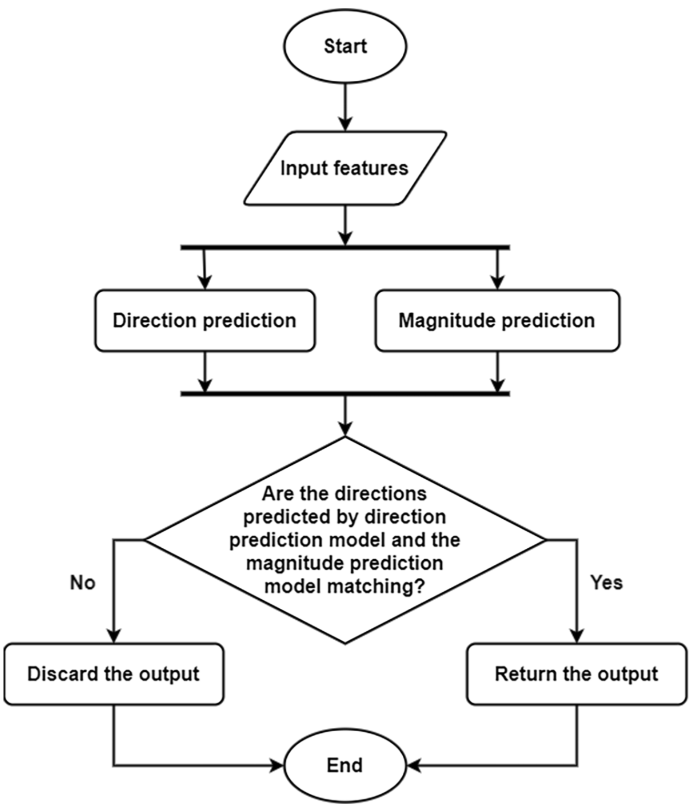

Fig. 5: Flow Chart of the Voting Classifier

IV. Results and Discussion

When considering the direction prediction, BiLSTM model trained on 7 days lag dataset achieved accuracy of 82.35%, with stable decrease and increase of loss and accuracy against the number of epochs, as shown in Figures 6 and 7.

This model achieved 81.47%, 80.99%, 81.8% and 80.87% average accuracy, macro precision, recall and F1-score, respectively on 20 random subsets created from test dataset.

Tanh and SELU as activation functions of BiLSTM layers, 64 as batch size, 64 and 0.5 as the drop ratios for dropout layers were the hyperparameters that achieved this result.

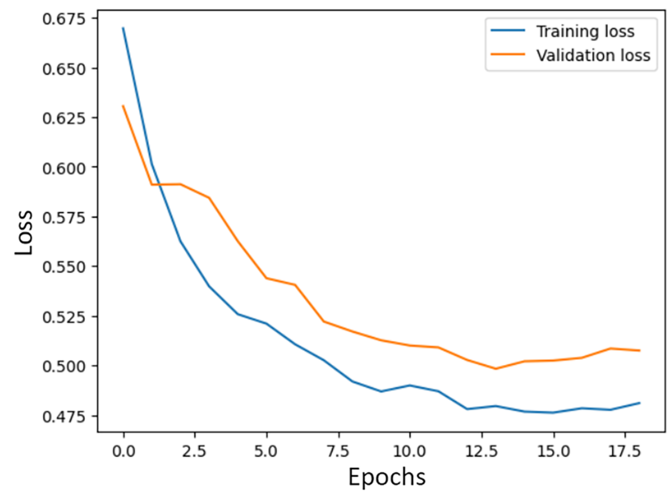

Fig. 6: Loss Against Epochs of 7 Days Lag BiLSTM Model

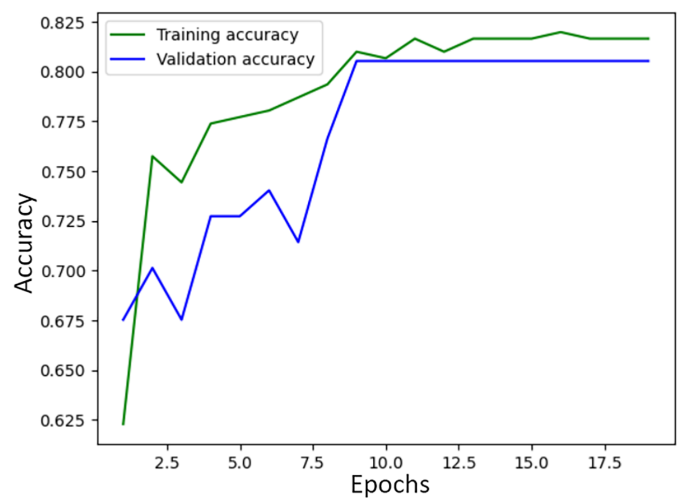

Fig. 7: Accuracy Against Epochs of 7 Days Lag BiLSTM Model

Figure 8 shows the importance of hyperparameters for the selected BiLSTM, which is provided by Optuna. The hyperparameter importance is calculated by fANOVA hyperparameter importance evaluation algorithm proposed by [30].

In magnitude prediction models, LSTM, BiLSTM and CNN achieved 64.71%, 72.06% and 72.06% accuracies, respectively. Between BiLSTM and CNN, the latter was selected since it showed a more stable decrease of the loss and increase of accuracy against number of epochs, which are shown in Figures 9 and 10.

CNN model also achieved highest average macro precision, recall and F1-score of 68.24%, 68.39% and 65.09%, respectively.

These results were given by the hyperparameters, Swish as the activation functions of convolutional layers, 32 as batch size, 0.5 as dropout ratios for dropout layers and 64 as the number of neurons in each hidden layer.

The importance of each optimized hyperparameter of CNN model is shown in Figure 11.

The voting classifier achieved an average accuracy of

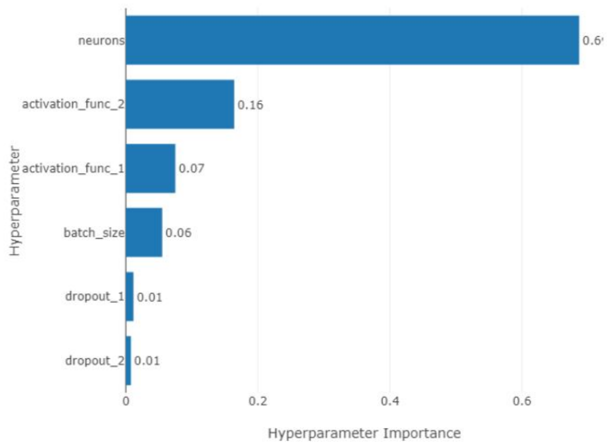

Fig. 8: Hyperparameter Importance of Selected Direction Prediction Model

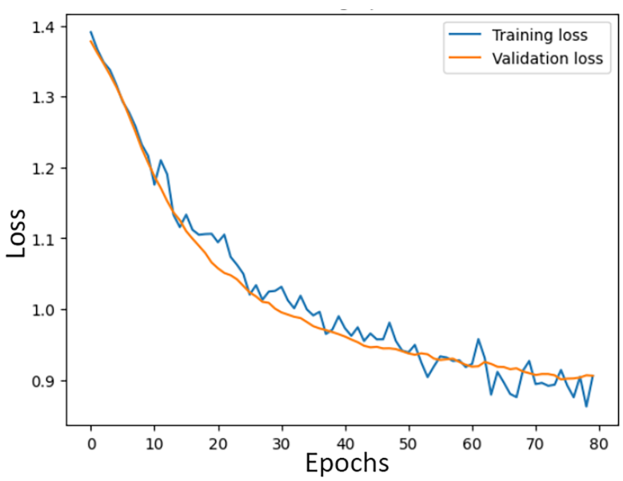

Fig. 9: Loss Against Epochs of CNN Model

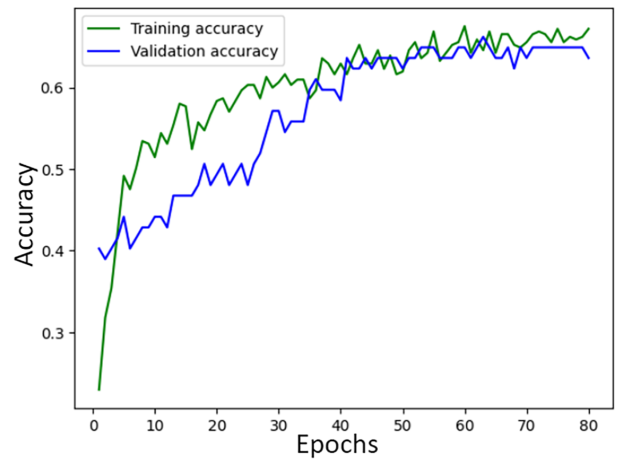

Fig. 10: Accuracy Against Epochs of CNN Model

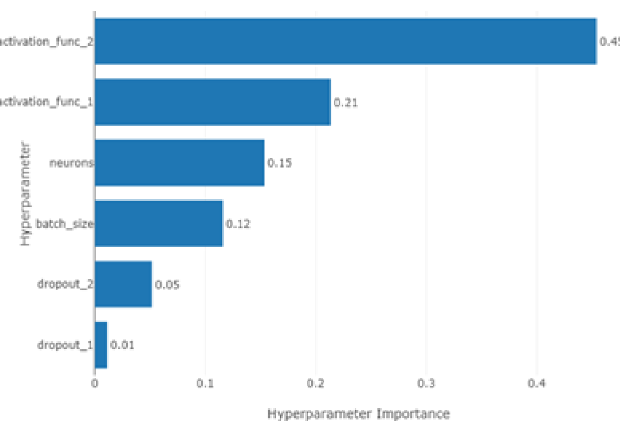

Fig. 11: Importance of Hyperparameters in Selected Magnitude Prediction Model

67.11% and it discarded around 4% samples. Actual vs. predicted bins are shown in Figure 12, with red markers denoting the outputs that were discarded. The pie chart in Figure 13 represents the outputs of the voting classifier as percentages in each bin.

By comparing the results of this study with existing studies discussed in the literature review, the selected direction prediction model has outperformed all the direction prediction models in discussed existing studies, except one model. The model developed by Zeynep Kilimci achieved higher accuracy than the direction prediction model in this research [16]. However, their model was built on data collected over a duration of 93 days, which exhibited a general upward trend. In contrast, this research utilized data from 450 days with higher price fluctuations. The data structuring for the models was changed and improved based on the method in Critien et al., and this research obtained higher average direction prediction accuracy (81.47%) than the direction prediction model they developed, which achieved average accuracy of 60.9% [19].

The magnitude predicting CNN model in this study outperformed most models but fell short of the model introduced in Pant et al., which achieved a 77.62% accuracy for magnitude prediction [17]. However, CNN's lower accuracy could be due to the higher volatility in the data collected from August 2018 to October 2019, compared to the less volatile period they used from 2015 to 2017. The F1-score for magnitude prediction improved significantly to 65.09%, compared to 14.21% in the study by Critien et al. [19].

The voting classifier achieved a lower accuracy than both base models. Therefore, the selected BiLSTM model can be recommended for use cases where only the direction change of the next-day close price of Bitcoin required. For use cases where the magnitude change of the next-day close price is required, the selected CNN model is recommended for predicting with a higher granularity. However, the magnitude prediction model provides higher granularity, it does so at the expense of accuracy. Conversely, the direction prediction model, while offering higher accuracy, tends to sacrifice granularity in comparison.

A. Limitations

This study faced limitations in obtaining up-to-date tweets and tweets for a longer period due to Twitter API restrictions.

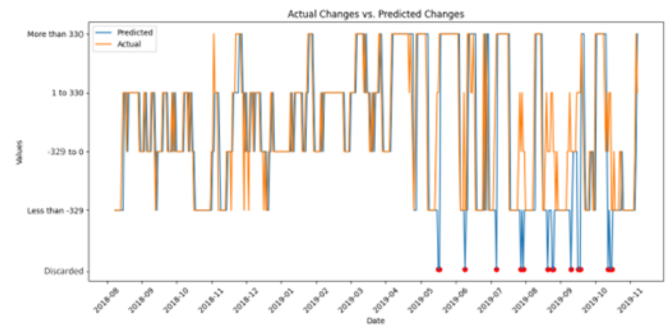

Fig. 12: Actual vs. Predicted Bins and Discarded Values by Voting Classifier

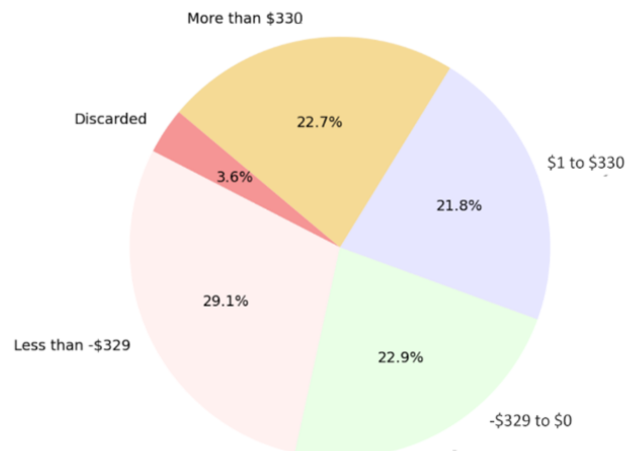

Fig. 13: Percentage of Predicted Bins and Discarded Values by Voting Classifier

Additionally, state-of-the-art transformer-based sentiment analyzers, such as Twitter-RoBERTa-base were not able to provide necessary speed to analyze the sentiments of large number of tweets that were used in this study. Furthermore, only English tweets were utilized. This may introduce bias and loss of information in non-English tweets that can further improve the models.

V. Conclusion

This study explored the link between Twitter sentiment and Bitcoin price volatility and aimed to utilize hyperparameter optimization to obtain models with higher accuracy. Sentiment analysis was performed using VADER. The study was able to identify a reasonably strong relationship between Twitter sentiment and Bitcoin price fluctuations.

The study developed deep learning models, experimenting with CNN, LSTM, and BiLSTM architectures. The BiLSTM model with a 7 day lag achieved an 82.35% accuracy for predicting whether next day price will increase or decrease, while the CNN model reached a 72.06% accuracy for predicting if the next day price change falls between in a range of $(-\infty, -\$330]$ or (-$330, $0] or ($0, $330] or $(\$330, \infty)$.

A voting classifier combining both models achieved 67.11% accuracy, and it was less effective than the direction and magnitude prediction models alone, suggesting the models should be used independently based on the task.

Optuna was utilized for hyperparameter optimization and optimized hyperparameters achieved higher performance than most of the related studies. Therefore, the objective of the study, to use hyperparameter optimization and to compare with the not-optimized models to identify the effects of hyperparameter optimization, was successfully achieved.

A. Recommendations

Future research should consider the possibility of using methods like quantization to enhance transformer-based sentiment analyzers, since this study refrained from employing such models due to their sluggish computational performance. Additionally, Twitter's paid APIs can be utilized for obtaining up-to-date tweets for periodically retraining the models. Furthermore, hyperparameter optimization can be utilized to further improve the models described in related studies and compare with current performance to understand the impact of hyperparameter optimization.

References

[1] B. Marr, “The top 10 tech trends in 2023 everyone must be ready for,” 2023. [Online]. Available: https://www.forbes.com/sites/bernardmarr/2022/11/21/the-top-10-tech-trends-in-2023-everyone-must-be-ready-for/?sh=53564a827df0

[2] A. Hayes, “Blockchain facts: What is it, how it works, and how it can be used,” 2023. [Online]. Available: https://www.investopedia.com/terms/b/blockchain.asp

[3] S. Nakamoto, “Bitcoin: A peer-to-peer electronic cash system,” 2008. [Online]. Available: https://doi.org/10.2139/ssrn.3977007

[4] P. Bajpai, “Countries where bitcoin is legal and illegal,” 2021. [Online]. Available: https://www.investopedia.com/articles/forex/041515/countries-where-bitcoin-legal-illegal.asp

[5] A. Lisa, “14 major companies that accept bitcoin,” 2022. [Online]. Available: https://www.gobankingrates.com/money/business/major-companies-that-accept-bitcoin/

[6] CNBCTV18, “Around 80

[7] D. Zinoviev, “Why is bitcoin volatile? an overview of bitcoin price fluctuations,” 2024. [Online]. Available: https://www.vaneck.com/us/en/blogs/digital-assets/bitcoin-volatility/

[8] J. Chen, “Media effect,” 2022. [Online]. Available: https://www.investopedia.com/terms/m/media_effect.asp

[9] E. Shearer and J. Gottfried, “News use across social media platforms 2018,” 2018. [Online]. Available: https://www.pewresearch.org/journalism/2018/09/10/news-use-across-social-media-platforms-2018/

[10] M. Dodaro, “The relationship between social media, cryptocurrency and blockchain,” 2024. [Online]. Available: https://topdogsocialmedia.com/social-media-cryptocurrency-and-blockchain/

[11] C. Tandon, S. Revankar, H. Palivela, and S. S. Parihar, “How can we predict the impact of the social media messages on the value of cryptocurrency? insights from big data analytics,” *Journal of Information Management Data Insights*, vol. 1, no. 2, 2021. [Online]. Available: https://doi.org/10.1016/j.jjimei.2021.100035

[12] R. Molla, “When elon musk tweets, crypto prices move,” 2021. [Online]. Available: https://www.vox.com/recode/2021/5/18/22441831/elon-musk-bitcoin-dogecoin-crypto-prices-tesla

[13] J. Huang, J. Chai, and S. Cho, “Deep learning in finance and banking: A literature review and classification,” *Frontiers of Business Research in China*, 2020.

[14] M. Srivastava, R. A., J. Singh, P. Chavriya, S. Chavriya, and S. Singh, “What do the ai methods tell us about predicting price volatility of key natural resources: Evidence from hyperparameter tuning,” *Resources Policy*, 2023.

[15] S. Galeshchuk, O. Vasylchyshyn, and A. Krysovatyy, “Bitcoin response to twitter sentiments,” ser. CEUR Workshop Proceedings. Sun SITE Central Europe, 2018. [Online]. Available: https://ceur-ws.org/Vol-2104/paper_199.pdf

[16] Z. H. Kilimci, "Sentiment analysis based direction prediction in bitcoin using deep learning algorithms and word embedding models," *International Journal of Intelligent Systems and Applications in Engineering*, 2020. [Online]. Available: https://doi.org/10.18201/ijisae.2020261585

[17] D. R. Pant, P. Neupane, A. Poudel, A. K. Pokhrel, and B. K. Lama, "Recurrent neural network based bitcoin price prediction by twitter sentiment analysis," ser. Proceedings on 2018 IEEE 3rd International Conference on Computing, Communication and Security. IEEE, 2018. [Online]. Available: https://doi.org/10.1109/CCCS.2018.8586824

[18] G. Serafini, P. Yi, Q. Zhang, M. Brambilla, J. Wang, Y. Hu, and B. Li, "Sentiment-driven price prediction of the bitcoin based on statistical and deep learning approaches," ser. Proceedings of the International Joint Conference on Neural Networks. IEEE, 2020. [Online]. Available: https://doi.org/10.1109/IJCNN48605.2020.9206704

[19] J. V. Critien, A. Gatt, and J. Ellul, "Bitcoin price change and trend prediction through twitter sentiment and data volume," *Financial Innovation*, vol. 8, no. 45, 2022. [Online]. Available: https://doi.org/10.1186/s40854-022-00352-7

[20] K. E. Hoque and H. Aljamaan, "Impact of hyperparameter tuning on machine learning models in stock price forecasting," *IEEE Access*, vol. 9, 2021.

[21] L. Liao, H. Li, W. Shang, and L. Ma, "An empirical study of the impact of hyperparameter tuning and model optimization on the performance properties of deep neural networks," *ACM Trans. Softw. Eng. Methodol.*, vol. 31, no. 3, 2022. [Online]. Available: https://doi.org/10.1145/3506695

[22] A. Terbuch, "Lstm hyperparameter optimization: Impact of the selection of hyperparameters on machine learning performance when applied to time series in physical systems," Master's thesis, Montanuniversitaet Leoben (000), 2021.

[23] A. Bouillet, "Bitcoin tweets - 16m tweets," 2019. [Online]. Available: https://www.kaggle.com/datasets/alaix14/bitcoin-tweets-20160101-to-20190329

[24] A. Joulin, E. Grave, P. Bojanowski, M. Douze, H. Jégou, and T. Mikolov, "Fasttext.zip: Compressing text classification models," ser. International Conference on Learning Representations, 2016.

[25] A. K. Dubey, A. Kumar, V. García-Díaz, A. Kumar Sharma, and K. Kanhaiya, "Study and analysis of sarima and lstm in forecasting time series data," *Sustainable Energy Technologies and Assessments*, 2021. [Online]. Available: https://doi.org/10.1016/j.seta.2021.101474

[26] Y. M. and W. J., "Adaptability of financial time series prediction based on bilstm," *Procedia Computer Science*, 2022. [Online]. Available: https://doi.org/10.1016/j.procs.2022.01.003

[27] I. H. Sarker, "Deep learning: A comprehensive overview on techniques, taxonomy, applications and research direction," *SN COMPUT. SCI.*, 2021. [Online]. Available: https://doi.org/10.1007/s42979-021-00815-1

[28] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," ser. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, 2019.

[29] P. Tannor and S. Tannor, "When you shouldn't use ensemble learning," 2021. [Online]. Available: https://deepchecks.com/when-you-shouldnt-use-ensemble-learning/

[30] F. Hutter, H. Hoos, , and K. Leyton-Brown, "An efficient approach for assessing hyperparameter importance," ser. International Conference on Machine Learning. JMLR.org, 2014.