# Ensemble and Multimodal Approach for Forecasting Cryptocurrency Price

#### **ABSTRACT**

Since the birth of Bitcoin in 2009, cryptocurrencies have emerged to become a global phenomenon and an important decentralized financial asset. Due to this decentralization, the value of these digital currencies against fiat currencies is highly volatile over time. Therefore, forecasting the crypto-fiat currency exchange rate is an extremely challenging task. For reliable forecasting, this paper proposes a multimodal AdaBoost-LSTM ensemble approach that employs all modalities which derive price fluctuation such as social media sentiments, search volumes, blockchain information, and trading data. To better support investment decision making, the approach forecasts also the fluctuation distribution. The conducted extensive experiments demonstrated the effectiveness of relying on multimodalities instead of only trading data. Further experiments demonstrate the outperformance of the proposed approach compared to existing tools and methods with a 19.29% improvement.

#### **CCS CONCEPTS**

- Computing methodologies  $\rightarrow$  Boosting; Supervised learning.

#### **KEYWORDS**

cryptocurrency price prediction, ensemble learning, multimodal ML

#### 1 INTRODUCTION

Since the first Bitcoin was mined in 2009, cryptocurrencies have changed the concept of financial assets, where most of the prominent ones are measured to have a combined value of almost \$1 trillion in terms of market capitalization by January 2022 <sup>1</sup>. The global cryptocurrency market grows at a compound annual growth rate (CAGR) of 30% from 2019 to 2026<sup>2</sup>. This growth represents a profitable investment and trading opportunity which in return affects the growth of the cryptocurrency market yielding a virtuous circle. Unlike FIAT currencies which are controlled by central authorities and banks, cryptocurrencies are completely decentralized and their transactions are validated and processed through a cryptographical network of nodes and recorded in a digital ledger of transactions (i.e. Blockchain) [7]. All these characteristics and others make the forecasting of the fiat-crypto currency exchange rate

(or simply cryptocurrency price forecasting) extremely challenging and often inaccurate [2]. Therefore, financial and AI experts are constantly analysing the market to better understand the trends of price fluctuations [1, 2, 5, 6, 13, 16, 17].

Most of the recent studies on forecasting the price of cryptocurrencies [17, 19, 23, 25, 32] employed neural networks due to their remarkable performance on similar tasks. Consequently, these approaches could achieve better results than traditional machine learning and statistical approaches [17, 31]. However, most of the existing approaches consider only a few factors that influence the cryptocurrency market. The problem is usually formalized as inferring a forecast function based on the available training sets and then evaluating the obtained functions by how well it generalizes assuming that price series often exhibit a homogeneous nonstationary [25]. In reality, unlike foreign exchange rates, many factors contribute to the volatility of cryptocurrency prices, such as the hash rate of the cryptocurrency mining process and public awareness[16]. For instance, it has been found that the price of cryptocurrencies is sensitive to public opinion sentiments [1, 18]. Yang, et al[31] assume that social media sentiment is an important indicator of future bitcoin price volatility.

To improve the forecasting results, this paper proposes to exploit all accessible factors influencing cryptocurrency prices. Specifically, we propose, for better investment decision-making, an ensemble of multimodal cryptocurrency price prediction LSTM models that employ the trading data, sentiments from social media, the blockchain data (i.e. hash rate, network difficulty, etc.) and search volumes from search engines. The cryptocurrency use case is Bitcoin (B) because of its significant market domination and popularity among the other 7812 existing cryptocurrencies and because of the extensive data availability needed in this research. In summary, we aim to support the investor decision making process with reliable forecasting of price and fluctuation distribution by answering the following research questions:

- RQ1: Which of the data modality combinations has a direct impact on the cryptocurrency price?
- RQ2: Is ensemble learning on multimodal data effective in cryptocurrency price forecasting?
- RQ3: How to interpret price forecasting with fluctuation distribution?

To this end, the main contributions of this paper can be summarized as follows:

 We exploit all accessible factors that drive the cryptocurrency market, such as trading data, social media sentiments, blockchain data, and search volumes.

 $<sup>^1</sup>https://gadgets.ndtv.com/cryptocurrency/news/bitcoin-price-btc-cryptocurrency-market-crash-usd-1-trillion-coinmarketcap-2726233$ 

<sup>&</sup>lt;sup>2</sup>https://www.globenewswire.com/news-release/2021/04/12/2208331/0/en/At-30-CAGR-CryptoCurrency-Market-Cap-Size-Value-Surges-to-Record-5-190-62-Million-by-2026-Says-Facts-Factors.html

- The proposed approach outperforms existing models and provide reliable help for decision-making to investors by proposing an AdaBoost-LSTM ensemble-learning architecture.
- In addition to the predicted price, our model outputs a fluctuation distribution to help the investor understand the certainty of the model's forecasting.
- We conducted extensive experiments and analyses to validate the effectiveness of the approach.
- For further improvements, we provide an open-source implementation

Following this section, Section 2 provides an overview of the existing related works. Section 3 presents the proposed approach and the model's proposed architecture. Next, all experiments details and results are provided in Section 4. Finally, Section 5 concludes this paper and discusses possible future work improvements.

#### 2 RELATED WORK

In this section, we review the related works divided into three categories:

#### 2.1 Traditional Market Price Forecasting

The task of forecasting stock price has been addressed for decades by scholars in finance, statistics [9] and data science in order to support investors in maximizing their profits and minimizing their losses. One of the earliest solutions is adopting a statistical model. Haviluddin et al. [4] conducted an in-depth comparison between statistical and machine learning techniques in near future forecasting from time-series data. The comparison mainly covers the statistical method ARIMA, Neural Networks, and genetic algorithms. The obtained results show that Neural Networks are more efficient and reliable in forecasting short-term time series.

In the last decade, with all the improvements that deep neural networks proved on different applications, they became a primary focus for financial and economic forecasting and time-series predictions [10, 14, 24, 27]. Deep learning approaches demonstrated their capability to significantly outperform other methods due to their ability to learn hidden features of time-series and historical market trends. Traditional markets have relatively well-known factors and rules that directly influence the price trends, including the number of asks, bids, the number of transactions, and more. This makes it easier for a deep learning model to learn the patterns from historical data and have a much more reliable forecasting, which is reflected in the high accuracy and low error of the forecasting by just applying prices, asks, and bids time series [24, 27].

Several studies tried to apply the same state-of-the-art deep learning approaches used in traditional markets to forecast cryptocurrency prices[20]. However, their results were not satisfactory due to the different characteristics of the cryptocurrency market.

## 2.2 Machine Learning for Cryptocurrency Price Forecasting

Due to the specific characteristics and high volatility of the cryptocurrency market, specific approaches have been proposed to forecast their prices. Yiying and Yaze [32] proposed various *LSTM* and classical *ANN* architectures that take the price, ask, and bids time series as input to forecast long and short-term price, focusing

on the price non-stationary dynamics of three cryptocurrencies (Bitcoin, Ethereum, Ripple). The obtained results demonstrate that applying the correct *ANN* architecture can learn long-term patterns, while *LSTM* tends to rely more on the short-term dynamics of time series.

Similarly, McNally, et al. [21] compared two deep learning models, namely Bayesian-optimized Recurrent Neural Network and an *LSTM* on long and short-term Bitcoin price forecasting. Both models are fed with trading data (prices, asks, bids, and the number of transactions) to outputs a binary price trend classification (price goes up or down). In their results, *LSTMs* achieved a higher accuracy of 52% that marginally outperforms the Bayesian-optimized Recurrent Neural Network and the classical statistical method ARIMA. However, the results are still cannot still be helpful for the investors due to the high volatility and variance of cryptocurrency time series over time, which make the line between overfitting and underfitting very fine [21].

Kumar and Rath [17] also proposed to use only historical trading data to forecast the trends of Ethereum prices using machine learning and deep learning models. The conducted experiments showed that *LSTM* marginally outperforms the *MLP* in short-term predictions. Pintelas et al. [25] conducted extensive experiments using *LSTM*-based and *CNN*-based models and their analysis conducted that both models follow an almost random walk process. To overcome this, they suggest that new approaches and new validation metrics should be explored to capture the hidden patterns.

Recently, Chevallier et al. [6] proposed a novel approach for cryptocurrency forecasting that improved the performance significantly while keeping the approach simple, which is an *AdaBoost* that uses multiple decision trees weak learners. Surprisingly, the results have demonstrated that *AdaBoost* outperforms all *ANNs*, *LSTMs*, *KNN*, and *SVMs* by an average *RMSE* of \$23.42 per 1 Bitcoin. In addition to the promising results, the simplicity of the approach allows for more generalizability and good interpretability of the outcome.

### 2.3 Sentiment Analysis for Cryptocurrency Price Forecasting

Given the fact that the cryptocurrency market is influenced by more factors than traditional markets, Krisoufek [16] studied the price fluctuations to determine the driving factors of the prices for multiple cryptocurrencies. Consequently, it has been found that the cryptocurrencies market is affected by multiple factors, mainly: the correlation between the number of asks, bids to the rate exchange, The hash rate and other blockchain information and public awareness. According to Krisoufek [16], all these factors contribute equally to the price fluctuations, but public awareness has proven a much stronger correlation. Due to the complexity of capturing public awareness, people sentiments can be a good alternative. These sentiments are supposed to partially influence the cryptocurrency price trend, especially sudden short-term changes. For instance, In 2021, when news, social media influencers, and governments were promoting against cryptocurrencies due to the hashing consumption of electricity and the harm that is caused to the environment [26], Bitcoin prices experienced a 41% drop in just 15 days. The negative sentiment generated for such announcements

resulted in much fear within the market and, therefore, a further collapse of the prices.

Young, et al. [15] has first introduced the hypothesis that cryptocurrency forums' sentiments influence Bitcoin prices. To validate this hypothesis, they proposed a model that is fed with the sentiment data only. Using the VADER sentiment analysis tool, the study analyzes user comments from the three most popular cryptocurrency forums that are first tagged with a sentiment score from 0 being very negative and 1 being very positive. The experimental results validated the existence of a correlation between the price trends and sentiments on the most popular cryptocurrency forums, suggesting that fluctuations within the forum's sentiments are an early indication of near-future price fluctuations.

Inspired by these findings, several approaches [2, 11, 23] proposed to employ sentiment analysis along with LSTM models to predict the prices on the next trading day. Commonly, social media and forum posts that contain keywords related to cryptocurrencies (Bitcoin, Ethereum, and Ripple) are first crawled. Subsequently, each post is assigned a score from 0 to 1 depending on its sentiment using the VADER tool. Next, all the sentiment scores are combined with the prices, asks, and bids time series into a single vector which serves as input to the LSTM architecture. The experiments of these approaches demonstrated that applying sentiment analysis marginally improves the forecasting results over other previous deep learning models that rely purely on trading data. For a better understanding of these improvements, Huang, et al. [11] compared their approach and an autoregressive that use only trading data on binary forecasting of cryptocurrency prices. The experiments demonstrated that embedding sentiments into the input increase the accuracy by 18.5% and the recall by 18.5% compared to the autoregressive model. However, despite these improvements being non-negligible, they are not enough to achieve reliable results that can be used and helpful for financial decision-making, as concluded by the study.

The main limitation of these previously mentioned studies [2, 6, 11, 17, 23, 25] is that all tweets and posts are equally treated without considering the level of interactions (likes, comments, and shares) that can significantly influence the impact of a particular post on the overall sentiments. Another limitation is the focus on only online communities and social media posts as a fluctuation factor despite the fact that other factors are proven to be as crucial as sentiments to the fluctuation changes [16].

#### 3 **APPROACH**

Inspired by the work of Chevallier et al. [6] and Sun et al. [29], this paper proposes to forecast cryptocurrency price in the next day using taking advantage of all indicating factors (RQ1). To tackle RQ2, we propose an ensemble learning approach that adopts adaptive boosting and LSTM architectures. This is because LSTMs are known for their capability to learn hidden features and trends within sequential data, especially for short-term goals [25]. Figure 1 illustrates an overview of our system, which consists of multiple LSTM weak learners (i.e. forecasters) that together combine a cryptocurrency price forecasting model. Each of these LSTM models in trained on a sampled dataset from the original one, and then gets

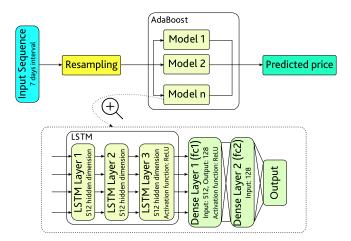


Figure 1: The AdaBoost-LSTM ensemble Learning Architecture for Cryptocurrency Price Prediction

assigned a weighted score according to its performance to be used in the inference phase. Here, all LSTM outputs are multiplied with their weights and divided by the total number of LSTM models to calculate the predicted final price. Specifically, our system goes through six main steps:

- (1) Initialize the sampling weights  $s_{i=1}^N$  for the N training samples, such that  $s_i = \frac{1}{N}$ .
- (2) The LSTM forecaster  $M_j$  is trained on the training samples sampled using the sampling weights  $s_{i=1}^{N}$ .
- (3) Calculate the errors  $e_{i=1:N}^{(j)}$  of the *LSTM* forecaster  $M_j$ . (4) Calculate the forecaster's weight  $w_j$  as follows:

$$w_j = \frac{1}{2} ln(\frac{1 - \sum_{i=0}^{N} e_i^{(j)}}{\sum_{i=0}^{N} e_i^{(j)}}).$$
 (1)

(5) Update the sampling weights  $s_{i=1}^N$  of all samples as follows:

$$s_i^* = \frac{s_i \times exp(e_i)}{\sum_{u=0}^{N} s_u \times exp(e_u)}.$$
 (2)

(6) Go to (1) and repeat the steps until all the LSTM forecasters are trained.

Once, all the LSTM forecasters are trained, we can compute the final price prediction  $\hat{y}_t^{(f)}$  at time t by combining the *LSTMs*' outputs with their weights to obtain a final prediction as follows:

$$\hat{y}_t^{(f)} = \frac{\sum_{j=0}^{J} \hat{y}_t^{(j)} w_j}{J},\tag{3}$$

where *J* denotes the total number of *LSTM* forecasters,  $\hat{y}_t^{(j)}$  denotes the price prediction of the LSTM model j, and  $w_j$  denotes the weight of the *j*th *LSTM* forecaster.

Each of the LSTM forecasters outputs a price forecast of 1 Bitcoin  $(\beta)$  for the next day (t + 1) in US dollars (\$). For this, the forecaster is fed with features of the last seven days  $\{\mathbf{x}\}_{t-7}^t$ , where  $\mathbf{x}_t$  is the normalized feature vector of length 18 for the date t, and is the concatenation of four feature representations denoting four modalities, namely trading data, sentiments, blockchain data and search volumes. In the following, we discuss the modalities:

#### 3.1 Input modalities

Although there exists a lot of cryptocurrencies, B is the most dominant one and its related data is easily accessible. Therefore, Bitcoin is the cryptocurrency use case considered in this paper, where we collected data from 2012 until the end of 2020, consisting of:

- 3.1.1 Trading data: Trading data is a sequential set of attributes that characterize the Bitcoin market. A time stamp t represents a single day where the open price is at 00:00 and the close price is at 23:59. At each t, the market is characterized with:
  - *Open (\$)*: the price of one  $\Bar{B}$  at the start of t.
  - *High* (\$): the highest price of one B within *t*.
  - Low (\$): the lowest price of one  $\beta$  within t.
  - Close (\$): the price of one  $\beta$  at the end of t.
  - Volume (B): the volume of B transacted during the interval of t.
  - *Volume* (\$): the volume of \$\beta\$ transacted during the time window of t
  - Weighted (\$): the ratio of the value of ₿ traded during the time window of t.
  - Average (\$): the average fees charged for transactions in the top 20 trading and exchange platforms and services.
  - *Transactions*: the total number of transactions performed during the time window of *t*.

3.1.2 Sentiment: According to the official statistics collected from all social media platforms, 4.48 billion people are actively using social media daily, roughly 57.7% of the world's population<sup>3</sup>. Therefore, we consider posts/tweets as an essential indicator of the price trend and associate it with a sentiment score ([-1, +1]) and class (from extremely negative to extremely positive).

Clearly, the cryptocurrency-related tweets do not reflect the market equally, where the main indicator is the engagement (i.e. the tweet's reach and the number of interactions) with these tweets. According to a recent statistic about Twitter Engagements <sup>4</sup>, the median number of likes and comments is 0, which means that a large number of tweets receive little to no engagement at all. Therefore, an equal sentiment weighting for all the tweets will not reflect the actual scenario. Consequently, we propose the assign a weight to each tweet, such that important tweets are highly weighted compared to those with little engagement. Since it is complex to assess which of the engagement factors is more important, the weight is computed simply as the harmonic mean of the number of likes, comments, retweets and quotes. Afterwards, the weights of all tweets are normalized using min-max normalization. To form a daily-interval sequence of sentiment scores, all weighted sentiments of the published tweets for the date t are averaged to form a single value that represents the overall Twitter's public sentiments on the

3.1.3 Blockchain details: According to the financial study conducted by Krisoufek [16], Blockchain plays an important role in

affecting the price fluctuations of cryptocurrencies, especially  $\beta$ . Therefore, we aim to embed multiple blockchain data that financial studies have proved to affect the market [16]. Using several blockchain sources, we consider the following details for a given day:

- Hash rate: the estimated hash rate that the B network is performing.
- *Block size:* the size of a complete B block.
- *Block time*: the time required to mine and produce a new ₿ block.
- Network difficulty: the difficulty level of mining B blocks through the network.
- Active Addresses: the total number of active addresses.

All the beforementioned blockchain attributes are sequential with a one-day interval between each data point.

3.1.4 Search volumes. People's curiosity is generally reflected through various actions and nowadays, it is easier than ever to search online, which makes it possible to understand public opinion through search statistics. To this end, we consider the search volumes such that any search containing either the word "Bitcoin" or the word "BTC" is collected (see Appendix 6.3).

#### 3.2 Fluctuation Analysis

As the main goal of this paper is to support investors' decision-making, only price forecasting is not sufficient because the cryptocurrency market's volatility is not uniform over time. Precisely, if the fluctuation level is low, the model's prediction is supposed to be close to the actual one. In contrast, if the fluctuation level is very high, the predicted price can be off from the actual value. However, this fluctuation level is latent and up to our knowledge, no existing approach proposes to capture it. Since the strength of our approach is the combination of different modalities, we assume that each of these modalities correlates independently to the price. Intuitively, the fluctuation is low when all modalities conform and high otherwise. Therefore, we aim to tackle **RQ3** by analysing the fluctuation with multiple varieties of the model to make multiple predictions, such that each model is trained with different modalities and dropout rate.

- 3.2.1 Input varieties: Given the AdaBoosting model described above and fed with the Trading data, Twitter sentiments, Blockchain data, and Online search volumes, several of its varieties are trained with different modality combinations, where each of the model varieties forecasts a slightly different price. Note that only the length of the input layer is different between the architectures of these varieties:
- Trading data: The main component of any market, including cryptocurrencies, is the historical trading data. Therefore, it is the minimal data that can be used in cryptocurrency price forecasting. The trading data consists of the 8 features mentioned earlier: ["Open", "High", "Low", "Close", "Volume BTC", "Volume Currency", "Weighted Price", "Average Fees"].
- Twitter sentiments: As discussed earlier, multiple research works proposed that social media sentiments and public awareness play a huge role in affecting the price fluctuation levels. Abraham, et al. [2] proposed a B price prediction by applying

 $<sup>^3</sup> https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/$ 

<sup>&</sup>lt;sup>4</sup>https://mention.com/en/reports/twitter/engagement/2

only Twitter sentiments without any external data. The results demonstrated that there is a strong correlation between social media sentiments and price. Therefore, we consider a model variety that is fed with only twitter sentiment, which is represented by ["The weighted twitter sentiments", "Tweet volumes"].

- Trading data with blockchain data: Blockchain data greatly influences cryptocurrency prices, especially the hash rate, because it is their backbone, after all. This model's variant takes as input the concatenation of trading data feature and blockchain details, namely ["Hash Rate", "Block Size", "Block Time", "Network Difficulty", "Number of Active Addresses", "Mining Profitability"]
- Trading data with search volumes: Because the online search volumes consist of only a single feature, it is nearly impossible to make an accurate price prediction for a highly volatile market. Therefore, it is advised to apply online search volumes along with the trading data. Therefore, this model variant is fed with the trading data concatenated with ["Online Search Volumes (Google searches)"]
- 3.2.2 Model dropouts: We applied different dropout rates to different layers of the final trained model. In addition to helping overcome overfitting, we regularize the model with dropout rates to obtain multiple variant outputs. With this, we aim at estimating an output distribution given the outputs of six models with different dropout rates:
- *Model V\_1*, *V\_2* and *V\_3*: A dropout with rates = 0.1, 0.2 and 0.35 at the *LSTM*'s last hidden layer.
- *Model V\_4, V\_5 and V\_6*: A dropout with rates = 0.1, 0.2 and 0.35 at at the fc1's output.

3.2.3 Predicted Price Distribution: We presume that fluctuation distributions are great clues that a decision-maker can use to accumulate information about the best possible actions to take. Together with the price forecast, they provide the uncertainty levels of the forecasting. To this end, we used Maximum Likelihood Estimation (MLA) to estimate the parameters (i.e. Mean and Variance) of the distribution given the sample of outputs. Given the the outputs of the ten model varieties  $O = \{o_1, o_2, \ldots, o_{10}\}$ , the goal is to estimate the set of parameters  $\hat{\theta} = \{\mu, \sigma^2\}$  that maximizes the function  $L(\theta; O) = \prod_{i=1}^{10} \mathcal{N}(o_i; \theta)$  such that  $\hat{\theta} = \arg\max L(\theta; O)$ .

#### 4 EXPERIMENTS

In this section, we evaluate the effectiveness of the proposed approach from different aspects and against several baseline approaches. All the implementations are publicly available and accessible on GitHub<sup>5</sup>. The experimental setup are discussed in Appendix 6.2.

#### 4.1 Datasets

Although there exists a lot of cryptocurrencies,  $\Bar{B}$  is the most dominant one and its related data is easily accessible. Therefore,  $\Bar{B}$  is the cryptocurrency use case considered in this paper, where we collected data from the beginning of 2016 until the end of 2020. , resulting in a total of 1825 data points (days) split into 70% training, 15% validation, 15% testing, respectively as demonstrated in Figure 2

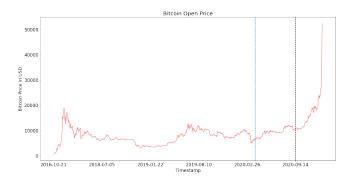


Figure 2: The training, testing, validation data split (75% training, 15% validation, 15% testing)

More details about the data can be found in Appendix 6.1.1

#### 4.2 Baselines

To compare the performance of the proposed approach against existing ones, we selected six methods, which are widely popular and recent. Each of them has a different architecture and/or deals with different data types to forecast the price of Bitcoin, namely:

- ARIMA19 [3]: This method is an ARIMA-based approach that
  uses only the trading data of B to forecast the prices on the next
  coming day.
- BNN17 [13]: A Bayesian Neural Networks (BNN) implementation to forecast β price on the near future using trading data collected from 2011 until 2017.
- LSTM20A [30]: This method employs a multivariate LSTM for cryptocurrency price forecasting from 2007 until 2017.
- LSTM20B [22]: Similarly, this method uses an LSTM architecture to make forecasting on the data from 2016 until 2019.
- GRU20 [8]: This method uses Gated Recurrent Unit on a data interval from 2010 until 2019.
- GRU21 [5]: This method also uses Gated Recurrent Unit but on a data interval from 2017 until 2020.
- AdaBoost21 [6]: This method uses traditional AdaBoost algorithm that takes advantage of multiple decision trees weak learners for price forecasting. Despite the simplicity of approach, it could prove significant improvements over other sophisticated approaches. This method covers covers the period from 2018 until 2020.

A primary limitation of comparing our approach against the above-mentioned baseline approaches is that their implementations are not publicly available. Consequently, we re-implemented these approaches from scratch and reproduce their results. In addition, because the baselines were developed and tested on a certain time interval using specific datasets, all the approaches and the proposed approach were applied on the same time interval and using the exact dataset to ensure fairness.

#### 4.3 Results and Discussion

In order to answer **RQ1**, we compared *AdaBoost-LSTM* ensemble learning against *LSTM* on different modality combinations, where both models are trained for 200 epochs and using the experimental

 $<sup>^5</sup> https://github.com/azeddinebouabdallah/DMCrypt \\$ 

setup discussed in Appendix 6.2. Note that *LSTM* has the same architecture as a single weak learner of *AdaBoost-LSTM* but trained on the entire dataset. Table 1 illustrates the obtained results of both models on (a) trading data only, (b) sentiments only and (c) the combination of trading and sentiments. As clearly shown, *AdaBoost-LSTM* ensemble learning significantly outperforms *LSTM* in the three experiments. In (a) and (c), it can be observed that the overall validation error of *LSTM* is lower, which can be explained by the non-uniformity of the price trend and volatility over time. Since *LSTM* tends to forecast smoothly compared to the weak learners of *AdaBoost-LSTM*, it achieves a lower error when the volatility is high. As shown in Figure 2, the period from 2012 until 2017 and from the end of 2018 until 2019 experienced minor fluctuations compared to other periods.

Figures 3 illustrate qualitative results of *AdaBoost-LSTM* and *LSTM* on randomly sampled testing interval of 100 days. Here, the blue curve represents the real price, the red curve represents the *LSTM* price forecast and the green curve represents the forecasting of *AdaBoost-LSTM* ensemble learning model. The Figures clearly show that *AdaBoost-LSTM* follows better the price trend compared to *LSTM* in all experiments.

Although the results of (a) validate the effectiveness of *AdaBoost-LSTM*, they are still far from being reliable and support the state-of-the-art finding that indicates the trading data cannot be enough alone to reliably forecast cryptocurrency prices.

The results of (b) show an impressive discovery supporting the hypothesis that social media sentiments correlate well with cryptocurrency prices. However, the forecasting error is very high compared to the results of trading data only, but it is important to note that despite both models are fed with only a single input per day (i.e. Twitter sentiments), they were able to capture the relation between the prices and sentiments and make a meaningful price forecast. The price forecast example shown in Figure 3b demonstrates that AdaBoost-LSTM can follow the price trend using only sentiments. In contrast, LSTM model struggles at learning any patterns within the given input and appears to give a relatively constant forecasts over time.

The results of (c) prove that the combination of trading data with Twitter sentiments achieve the best results. This suggests that social media sentiments can capture the cryptocurrency price fluctuation and help the model make accurate price forecasts. Specifically, incorporating sentiment analysis to both models improved the results by 15% compared to using only trading data.

Furthermore, we conducted additional experiments using *LSTM*, *AdaBoost-LSTM* ensemble learning on trading data with the inclusion of different other modalities; (d) the hash rate, (e) search volume and (f) blockchain data (i.e. all six attributes). Table 2 represents the obtained evaluation results of both models. Although the error of (d) is slightly higher for *AdaBoost-LSTM*, which might be caused by the boosting process, the low error obtained by *LSTM* when including the hash rate indicates that it is a good factor to accurately forecast the price.

The results of (e) show that search volumes did not rally affect the performance of both *LSTM* and *AdaBoost-LSTM* when merged with trading data. However, search volumes can be hypothesized to perform better when merged with other attributes such as social media sentiments because both represent "public awareness".

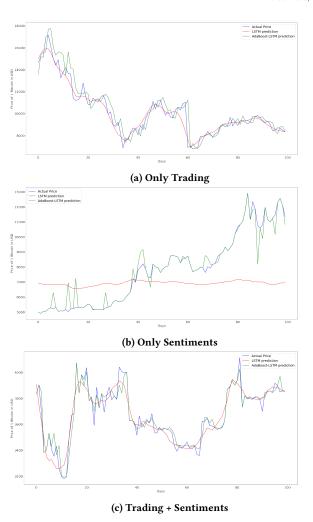


Figure 3: Qualitative forecasting results of *LSTM* vs. *AdaBoost-LSTM* in different 100-days time window on (a) trading data only, (b) sentiments only and (c) the combination of sentiments and trading data.

The results of (f) for both models demonstrate that applying blockchain information and trading data significantly improves the forecasting performance compared to applying only the hash rate. However, *AdaBoost-LSTM* ensemble learning testing errors are considerably higher than those of *LSTM*. When using the blockchain data as a single input to both models as illustrated in Figure 4, *AdaBoost-LSTM* did not forecast the price accurately but could follow the price trend. These results demonstrate the existence of a correlation between blockchain data and cryptocurrency prices to the extent that they can help models forecast the prices. Other qualitative results are shown in Appendix 6.4.

To answer **RQ2**, we applied both *LSTM* and the *AdaBoost-LSTM* ensemble learning model using the concatenation of all attributes. Table 3 presents the obtained results, demonstrating that embedding all four attribute categories (trading data, social media sentiments,

	(a) Only Trading		(b) Only Sentiments		(c) Trading + Sentiments	
	LSTM	AdaBoost-LSTM	LSTM	AdaBoost-LSTM	LSTM	AdaBoost-LSTM
Training RMSE (\$)	346.507	100.593*	2,617.935	221.707	344.082	104.497
Training MAE (\$)	204.773	64.651	2,120.712	59.85*	209.094	61.134
Validation RMSE (\$)	502.473	1,097.734	8,233.91	8,650.681	436.684*	1,098.085
Validation MAE (\$)	321.106	709.154	7,210.621	6,068.537	309.384*	708.485
Testing RMSE (\$)	502.473	272.027	8,233.91	4,006.787	354.071	243.47*
Testing MSE (\$ <sup>2</sup> )	252,479.5	73,999.136	67,797,288.0	16,054,343.133	125,366.981	59,258.164*
Testing MAE (\$)	321.106	207.332	7,210.621	2,969.586	312.009	201.568*

Table 1: Comparison results between LSTM and AdaBoost-LSTM ensemble learning applied on (a) trading data only, (b) sentiments only and (c) the combination of sentiments and trading data.

	(d) Trading + Hash rate		(e) Trading + Search volume		(f) Trading + Blockchain data	
	LSTM	AdaBoost-LSTM	LSTM	AdaBoost-LSTM	LSTM	AdaBoost-LSTM
Training RMSE (\$)	299.818	41.201*	352.823	89.093	280.944	96.525
Training MAE (\$)	178.795	14.433*	206.466	30.981	171.552	62.029
Validation RMSE (\$)	433.68*	1,156.835	522.815	1,438.16	1,052.821	1,151.279
Validation MSE (\$ <sup>2</sup> )	188,078.828*	1,338,268.132	273,336.187	2,068,305.97	1,108,433.75	1,325,444.897
Validation MAE (\$)	299.766*	782.773	369.701	940.658	707.377	765.778
Testing RMSE (\$)	433.680	356.554	519.175	277.861	200.263*	281.607
Testing MSE (\$ <sup>2</sup> )	188,078.828	127,131.12	274,476.23	77,207.124	40,105.472*	79,302.879
Testing MAE (\$)	299.766	291.09	365.0	201.177	156.694*	213.981

Table 2: Comparison results between *LSTM* and *AdaBoost-LSTM* ensemble learning applied on trading data combined with (d) the hash rate, (e) search volume and (f) blockchain data

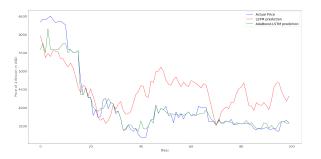


Figure 4: Qualitative forecasting results of *LSTM* vs. *AdaBoost-LSTM* in different 100-days time window on blockchain data.

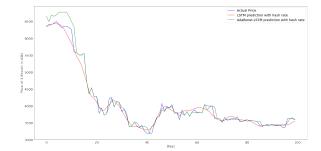


Figure 5: Qualitative forecasting results of *LSTM* vs. *AdaBoost-LSTM* in different 100-days time window on all modalities.

search volumes, and blockchain data) contribute to obtaining significantly better results as there is an improvement of \$75.312 in MAE, resulting in a 36.32% improvement in the price forecasting performance. The qualitative results on a 100-days interval for both models are shown in Figure 5, where it is evident that obtained forecasts are close to the actual price.

To further understand the performance of the model, an *MAE* distribution for the whole testing set was calculated and visualized in Figure 6. As can be seen that 68% of the times the price forecast varies by +/-500\$ of the price of 1  $\Breve{B}$ , making the whole model's forecast extremely close to what the actual price is considering the average price of  $\Breve{B}$  in the testing set is 16553.59.



Figure 6: MAE distribution of the testing set

	LSTM using All	AdaBoost-
		LSTM using
		All
Training RMSE (\$)	280.013	83.564
Training MSE (\$ <sup>2</sup> )	78,407.63	6,982.959
Training MAE (\$)	173.698	25.780
Validation RMSE (\$)	389.245\$	234.718
Validation MAE (\$)	281.399	234.666
Testing RMSE (\$)	389.245	158.929
Testing MSE (\$ <sup>2</sup> )	151,512.093	25,258.60
Testing MAE (\$)	281.399	132.027

Table 3: Comparison results between LSTM and AdaBoost-LSTM ensemble learning applied on all modalities

Furthermore, we compare AdaBoost-LSTM trained on all modalities against baseline approaches on the same dataset over the period ranging from (the 1st of January 2020 until the 1st of July 2020). It is important to note that all the approaches in this comparison price forecasts of  $\beta$  for the next  $24^{\rm th}$  hours. The results shown in Table 4 clearly outline that the proposed approach outperforms the other models.

Method	MSE (\$ <sup>2</sup> )	RMSE (\$)	MAE (\$)
AdaBoost-LSTM	25,258.60	158.929	132.027
BNN17	49,125.346	221.642	184.124
LSTM20A	27,847.415	166.875	148.628
LSTM20B	41,122.567	202.787	177.02
AdaBoost21	39,023.631	197.544	156.346
GRU21	43,266.496	208.006	160.44
ARIMA19	341,504.659	584.384	542.73
GRU20	34,282.003	185.154	157.632

Table 4: Evaluation Comparison Between the results of *AdaBoost-LSTM* and baseline approaches on test data.

To tackle **RQ3**, we propose to provide the investor with the price forecast together with the fluctuation distribution. As it is very challenging to evaluate this quantitatively, we show in Figure 7 a few results obtained by this technique for 6 random days. It is noticeable that the price forecast certainty can differ from one day to another, where if there is a massive volatility level on that particular day, the price distribution is closer to be flat and has a high standard deviation, which indicates that the price is highly likely to fluctuate from the single model's prediction. In contrast, when the model gives a prediction with a high certainty level, it is apparent that the price distribution has a high peak and low standard deviation. All these prediction distributions are an excellent source for an investor for decision-making as it provides a clear interpretation of how confident the model is for a given prediction on a particular day.

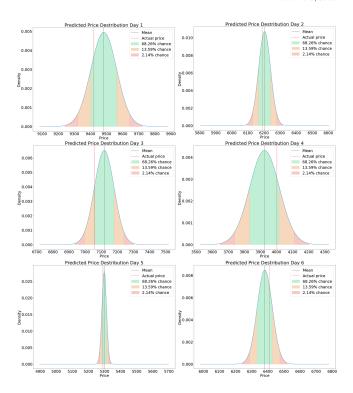


Figure 7: Predicted price distribution of six randomly selected days

#### 4.4 Long-term forecasting

Due to the good results of the proposed approach, we found it interesting to push this approach to its limits and test to what extent it can make predictions and what is its performance in making longterm predictions. To this end, the model predicts the price of the next 24th hour, and then the prediction is appended to the original input to make another prediction for the second day, and repeat this process for the next days until the 30th day. Please note that for this experiments only price history is used as the other data are assumed to by in the future. This automatically means that the performance will drop given the fact that less data is used to infer the next day's price. Figure 8 illustrates the obtained results, where the red curve represents the average MAE of each time-window prediction, and the maximum and minimum MAE are shown by the higher and lower boundaries of the blue area respectively. As expected, the prediction is at its best in the first day and drops for long-term predictions.

#### 5 CONCLUSION

This paper presents a novel cryptocurrency price forecasting approach to support the investor decision-making process by providing reliable short-term price forecasts derived from all sources proved to correlate with the price (**RQ1**). The proposed approach is based on ensemble learning of multiple weak LSTM learners (**RQ2**), which proved its outperformance over existing approaches. In addition to forecasting the price, the approach supports investors with

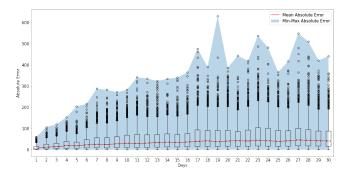


Figure 8: Bitcoin price prediction absolute testing errors over different time windows

fluctuation distributions that give a wider overview of the market and how reliable is the price forecast (RQ3).

For future work, we aim to generate trading data with extremely high volatility to be used in the training process. With this, we assume that the model can be well prepared for unexpected market volatility in the future that were never seen in the training process. Our motivation is drawn from the fact that cryptocurrency markets have been surprising us since their existence. In addition, we aim to analyse the sentiments deeply because they have shown a great support to the forecasting model.

#### **REFERENCES**

- David Aboody, Omri Even-Tov, Reuven Lehavy, and Brett Trueman. 2018.
   Overnight returns and firm-specific investor sentiment. *Journal of Financial and Quantitative Analysis* 53, 2 (2018), 485–505.
- [2] Jethin Abraham, Daniel Higdon, John Nelson, and Juan Ibarra. 2018. Cryptocurrency price prediction using tweet volumes and sentiment analysis. *SMU Data Science Review* 1. 3 (2018). 1.
- [3] Saad Ali Alahmari. 2019. Using machine learning ARIMA to predict the price of cryptocurrencies. The ISC International Journal of Information Security 11, 3 (2019), 139–144.
- [4] Rayner Alfred, Joe Henry Obit, Mohd Hanafi Ahmad Hijazi, Ag Asri Ag Ibrahim, et al. 2015. A performance comparison of statistical and machine learning techniques in learning time series data. Advanced Science Letters 21, 10 (2015), 3037–3041.
- [5] Reem K Alkhodhairi, Shahad R Aljalhami, Norah K Rusayni, Jowharah F Alshobalil, Amal A Al-Shargabi, and Abdulatif Alabdulatif. 2021. Bitcoin Candlestick Prediction with Deep Neural Networks Based on Real Time Data. CMC-COMPUTERS MATERIALS & CONTINUA 68, 3 (2021), 3215–3233.
- [6] Julien Chevallier, Dominique Guégan, and Stéphane Goutte. 2021. Is It Possible to Forecast the Price of Bitcoin? Forecasting 3, 2 (2021), 377–420.
- [7] David LEE Kuo Chuen, Li Guo, and Yu Wang. 2017. Cryptocurrency: A new investment opportunity? The Journal of Alternative Investments 20, 3 (2017), 16-40.
- [8] Aniruddha Dutta, Saket Kumar, and Meheli Basu. 2020. A gated recurrent unit approach to bitcoin price prediction. *Journal of Risk and Financial Management* 13, 2 (2020), 23.
- [9] James V Hansen, James B McDonald, and Ray D Nelson. 1999. Time series prediction with Genetic-Algorithm designed neural networks: An empirical comparison with modern statistical models. *Computational Intelligence* 15, 3 (1999), 171–184.
- [10] Mohammad Asiful Hossain, Rezaul Karim, Ruppa Thulasiram, Neil DB Bruce, and Yang Wang. 2018. Hybrid deep learning model for stock price prediction. In 2018 ieee symposium series on computational intelligence (ssci). IEEE, 1837–1844.
- [11] Xin Huang, Wenbin Zhang, Yiyi Huang, Xuejiao Tang, Mingli Zhang, Jayachander Surbiryala, Vasileios Iosifidis, Zhen Liu, and Ji Zhang. 2021. LSTM Based Sentiment Analysis for Cryptocurrency Prediction. arXiv preprint arXiv:2103.14804 (2021).
- [12] Clayton Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the International

- AAAI Conference on Web and Social Media, Vol. 8.
- [13] Huisu Jang and Jaewook Lee. 2017. An empirical study on modeling and prediction of bitcoin prices with bayesian neural networks based on blockchain information. *Ieee Access* 6 (2017), 5427–5437.
- [14] Kaustubh Khare, Omkar Darekar, Prafull Gupta, and VZ Attar. 2017. Short term stock price prediction using deep learning. In 2017 2nd IEEE international conference on recent trends in electronics, information & communication technology (RTEICT). IEEE, 482–486.
- [15] Young Bin Kim, Jun Gi Kim, Wook Kim, Jae Ho Im, Tae Hyeong Kim, Shin Jin Kang, and Chang Hun Kim. 2016. Predicting fluctuations in cryptocurrency transactions based on user comments and replies. PloS one 11, 8 (2016), e0161197.
- [16] Ladislav Kristoufek. 2015. What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis. PloS one 10, 4 (2015), e0123923.
- [17] Deepak Kumar and SK Rath. 2020. Predicting the trends of price for ethereum using deep learning techniques. In Artificial Intelligence and Evolutionary Computations in Engineering Systems. Springer, 103–114.
- [18] Connor Lamon, Eric Nielsen, and Eric Redondo. 2017. Cryptocurrency price prediction using news and social media sentiment. SMU Data Sci. Rev 1, 3 (2017), 1–22
- [19] Ioannis E Livieris, Emmanuel Pintelas, Stavros Stavroyiannis, and Panagiotis Pintelas. 2020. Ensemble deep learning models for forecasting cryptocurrency time-series. Algorithms 13, 5 (2020), 121.
- [20] Isaac Madan, Shaurya Saluja, and Aojia Zhao. 2015. Automated bitcoin trading via machine learning algorithms. URL: http://cs229. stanford. edu/proj2014/Isaac% 20Madan 20 (2015).
- [21] Sean McNally, Jason Roche, and Simon Caton. 2018. Predicting the price of bitcoin using machine learning. In 2018 26th euromicro international conference on parallel, distributed and network-based processing (PDP). IEEE, 339–343.
- [22] Mohammed Mudassir, Shada Bennbaia, Devrim Unal, and Mohammad Hammoudeh. 2020. Time-series forecasting of Bitcoin prices using high-dimensional features: a machine learning approach. Neural Computing and Applications (2020), 1–15.
- [23] Sergey Nasekin and Cathy Yi-Hsuan Chen. 2019. Deep learning-based cryptocurrency sentiment construction. Available at SSRN 3310784 (2019).
- [24] Mahla Nikou, Gholamreza Mansourfar, and Jamshid Bagherzadeh. 2019. Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms. *Intelligent Systems in Accounting, Finance and Management* 26, 4 (2019), 164–174.
- [25] Emmanuel Pintelas, Ioannis E Livieris, Stavros Stavroyiannis, Theodore Kotsilieris, and Panagiotis Pintelas. 2020. Investigating the problem of cryptocurrency price prediction: a deep learning approach. In IFIP International Conference on Artificial Intelligence Applications and Innovations. Springer, 99–110.
- [26] Christophe Schinckus. 2020. The good, the bad and the ugly: An overview of the sustainability of blockchain technology. Energy Research & Social Science 69 (2020), 101614.
- [27] Sreelekshmy Selvin, R Vinayakumar, EA Gopalakrishnan, Vijay Krishna Menon, and KP Soman. 2017. Stock price prediction using LSTM, RNN and CNN-sliding window model. In 2017 international conference on advances in computing, communications and informatics (icacci). IEEE, 1643–1647.
- [28] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In Proceedings of the 2013 conference on empirical methods in natural language processing. 1631–1642.
- [29] Shaolong Sun, Yunjie Wei, and Shouyang Wang. 2018. AdaBoost-LSTM ensemble learning for financial time series forecasting. In *International Conference on Computational Science*. Springer, 590–597.
- [30] Nicola Uras, Lodovica Marchesi, Michele Marchesi, and Roberto Tonelli. 2020. Forecasting Bitcoin closing price series using linear regression and neural networks models. PeerJ Computer Science 6 (2020), e279.
- [31] Liuqing Yang, Xiao-Yang Liu, Xinyi Li, and Yinchuan Li. 2019. Price Prediction of Cryptocurrency: An Empirical Study. In International Conference on Smart Blockchain. Springer, 130–139.
- [32] Wang Yiying and Zang Yeze. 2019. Cryptocurrency price analysis with artificial intelligence. In 2019 5th International Conference on Information Management (ICIM). IEEE, 97-101.

#### 6 APPENDIX

#### 6.1 Data details

6.1.1 Trading Data: There are several sources that provide this sequential data. However, we noticed that it is likely that the data of one source contain missing values and some days are simply skipped. To overcome this, we collected the data from several sources and combined them all together such that each one complete the other, taking into consideration that the values might be slightly different due to the fee differences. Our data is collected from the following sources:

- Kaggle: an open-source dataset containing a 1-minute interval data of β prices, collected from January 1, 2012 March 1, 2021.
- **Binance:** a large cryptocurrency exchange, whose data is publicly available. Since no API is provided, the data was crawled from the website using a custom python script to translate the data from a visual format to a suitable csv format.
- Coinbase: a large cryptocurrency exchange, whose data is publicly available. In contrast to Binance, Coinbase provides APIs to collect the trading data, which is widely used (e.g. by Google)

6.1.2 Sentiments: Twitter is one of the top 5 social media platforms in 2021 according to the number of active users with almost 202 million daily users and over 500 million tweets a day<sup>6</sup>. Therefore, we consider Twitter as a source to analyze the public's sentiments, where the tweets are collected using *Twitter Academic Research API* and *TweetScraper*.

A tweet is collected only if it belongs to the results of the query: "q ='Bitcoin' or 'BTC' or \*Bitcoin or \*BTC", which results in more than 120 million tweets in the time window [Jan 2012, Dec 2021]. As per any social media data, the tweets contain a good portion of spam tweets (Tweets that use hashtags to gain a more extensive audience). Therefore, we removed every tweet that contains the word Bitcoin as a hashtag and not as text. Furthermore, all the emojis used are kept in their Unicode format to be used later with a fine-tuned model for sentiment analysis.

As the goal of collecting these tweets is to analyse their sentiments, which is a separate subdiscipline by itself, we use in this paper, two popular pretrained and publicly available sentiment analysis tools, namely *Vader* [12] and *Deeply Moving* [28].

 $Vader\ [12]$  is a rule-based model for social media sentiment analysis. According to several recent evaluations, Vader outperforms multiple state-of-the-art approaches that use rule-based or machine learning. Also, Vader is supposed to generalize better across contexts than other approaches. The model outputs a sentiment score ranging between [-1,1], which reflects the sentiment of a given post/tweet. The higher the sentiment score (i.e. up to +1), the more the sentiment is positive, while 0 stands for neutral. Using this score, the tweet is classified into one of the nine sentiment classes (i.e. from extremely negative to extremely positive).

Deeply Moving[28] is another tool that analyses the text as a whole entity to preserve the correlation between words. Mainly, this approach is proposed to analyze movies' reviews sentiments, meaning that the text is formally represented, which is different from the structure used in Twitter's tweets (i.e. use of emojis, hashtags, and abbreviations). Fine-tuning the model on tweets sentiments

is challenging due to the lack of labelled data. Therefore, we aim to use both models to get the best of both worlds (a model that is explicitly trained on tweets and a model that can understand the sentiments of a whole sentence). Similarly to Vader, the Deeply Moving model outputs a sentiment score in the range of [-1,1], where +1 indicates positive sentiment. Also, the sentiment score is used to classify the tweet into one of the nine classes mentioned previously.

Both sentiment analysis models (Vader, Deeply Moving) are applied to all the tweets that were previously collected, where the final sentiment score is computed as the average of both models' scores.

*6.1.3 Blockchain data:* We collected the blockchain data from the following source:

- Blockchain.com <sup>7</sup>.
- Bitinfocharts <sup>9</sup>.

Ycharts <sup>8</sup>.

Nasdaq Data Link <sup>10</sup>.

6.1.4 Search volumes: Google is the largest search engine with over 52% of the world's population using it every day<sup>11</sup>. Therefore, it is one of the most excellent sources for collecting search statistics because the results will cover a large portion of the total searches made online. Therefore, we used Google API to collect this data.

#### 6.2 Experimental Setup

All conducted experiments have multiple common setups outlined as follows:

- Using the *MSE* as the training loss function.
- Using the Adam optimizer for all the experiments training as an optimization algorithm for the convergence of the model.
- An initial learning rate of 0.0003.
- Training for 200 epochs.
- The dataset was normalized using the technique proposed on the dataset subsection.
- The training was done on a GPU server with the following specs: AMD Ryzen Threadripper 1950X 16-Core Processor, 128GiB System memory, and NVIDIA GV100.
- All experiments use the MAE as a validation loss for every ten epochs.

#### 6.3 Search volume

Figure 9 illustrates the search volumes of the query "Bitcoin" from the year 2013 until the year 2021.

### 6.4 Qualitative results on different modality combinations

Figure 10c illustrate qualitative results of *AdaBoost-LSTM* and *LSTM* applied on trading data combined with (d) the hash rate, (e) search volume and (f) blockchain data for randomly sampled testing interval of 100 days.

<sup>&</sup>lt;sup>6</sup>https://www.omnicoreagency.com/twitter-statistics/

<sup>&</sup>lt;sup>7</sup>https://www.blockchain.com/charts

<sup>&</sup>lt;sup>8</sup>https://ycharts.com/indicators/bitcoin\_network\_hash\_rate

<sup>9</sup>https://bitinfocharts.com/bitcoin/

<sup>10</sup> https://data.nasdaq.com/data/BCHAIN-blockchain

<sup>11</sup> https://review42.com/resources/google-statistics-and-facts/



Figure 11: Qualitative prediction results of *LSTM* vs. *AdaBoost-LSTM* in different 100-days time window on all modality combinations

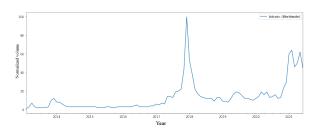


Figure 9: Illustration of the search volume of the query "Bitcoin" from the year 2013 until the year 2021.

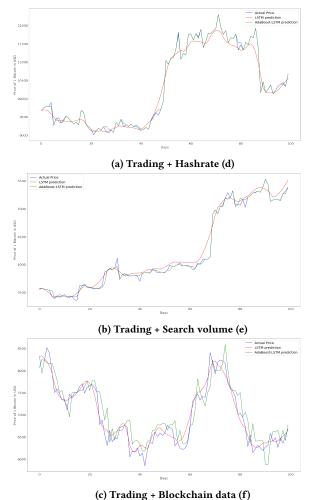


Figure 10: Qualitative prediction results of *LSTM* vs. *AdaBoost-LSTM* in different 100-days time window on trading data combined with (d) the hash rate, (e) search volume and (f) blockchain data

Figure 11 illustrate qualitative results of *AdaBoost-LSTM* and *LSTM* applied on all modality combinations for randomly sampled testing interval of 100 days.