Association for Information Systems
AIS Electronic Library (AISeL)

Wirtschaftsinformatik 2021 Proceedings Track 3: Student Track

Predicting Hourly Bitcoin Prices Based on Long Short-term Memory Neural Networks

Maximilian Schulte
Fachhochschule Aachen

Mathias Eggert
Fachhochschule Aachen

Predicting Hourly Bitcoin Prices Based on Long Short-term Memory Neural Networks

Maximilian Schulte[1], Mathias Eggert[2]

[1] Fachhochschule Aachen, Department of Economics, Aachen, Germany
maxschulte@mailbox.org
[2] Fachhochschule Aachen, Department of Economics, Aachen, Germany
eggert@fh-aachen.de

Abstract. Bitcoin is a cryptocurrency and is considered a high-risk asset class whose price changes are difficult to predict. Current research focusses on daily price movements with a limited number of predictors. The paper at hand aims at identifying measurable indicators for Bitcoin price movements and the development of a suitable forecasting model for hourly changes. The paper provides three research contributions. First, a set of significant indicators for predicting the Bitcoin price is identified. Second, the results of a trained Long Short-term Memory (LSTM) neural network that predicts price changes on an hourly basis is presented and compared with other algorithms. Third, the results foster discussions of the applicability of neural nets for stock price predictions. In total, 47 input features for a period of over 10 months could be retrieved to train a neural net that predicts the Bitcoin price movements with an error rate of 3.52 %.

Keywords: bitcoin, neural nets, LSTM, data analysis, price prediction.

1 Introduction

In the past few years, the concept of cryptocurrencies made its way to the public with an open debate whether digital, decentralized currencies should be taken seriously or not. While many perceive Bitcoin and other cryptocurrencies as a pure speculative bubble, others see similarities with the early age of the internet, because of its underlying technology called Blockchain [1]. Certainly, Bitcoin is a highly volatile asset class. Solely in 2017, the price of Bitcoin rose by 2000 percent from under $1.000 in January, to almost $20.000 by the end of the year. In the following months of 2018, the price plummeted rapidly to under $7.000 in early February [1]. In December 2020, Bitcoin reached a new high of $23.000 and a renewed rapid increase is indicated[1]. These Bitcoin price changes are hard to predict due to the underlying high volatility. The general objective of this paper is to address this difficult task. Several researchers worked on predicting Bitcoin price changes based on twitter sentiments and blockchain information [e.g., 2, 3].

[1] https://www.nytimes.com/2020/11/30/technology/bitcoin-record-price.html.

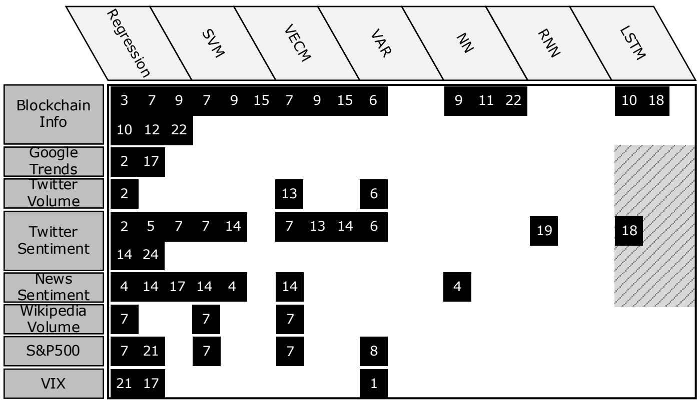
The majority of them focuses on Support Vector Machine (SVM) and regression models. So far, only Guo et. Al. [4] and Mohanty et. Al. [5] present results of an advanced neural network with a variety of input features as a predictive model for Bitcoin price changes. Against this background, the paper at hand answers the following research questions:

- What are significant indicators of Bitcoin price performance? (RQ1)
- How applicable are neural networks for predicting hourly Bitcoin prices? (RQ2)

The structure of the paper at hand is as follows. First, we provide related work on neural nets and a literature review about research works on the prediction of bitcoin price movements. Section 3 comprises the research design and a detailed description of the data collection, adjustment, and analysis. In Section 4, we present the results of our analysis, which contains the predictors found and the performance of the developed neural net. Section 5 contains the discussion of the results. The paper ends with a summary and an outlook on further research in the field of bitcoin price prediction.

2 Preliminary Study on Bitcoin Price Prediction

At the time of writing this paper, 26 research works investigating Bitcoin prediction models were found by using the IEEE Xplore Digital Library, the AIS eLibrary (AISEL), Google Scholar and the following search terms: *prediction, predict, bitcoin, stock market, time-series, regression, neural net, recurrent neural net, machine learning and LSTM*. We classify these papers by applying two dimensions: input features and applied analysis method. The results are depicted in Figure 1. Historical data on price or trade volumes as basic input parameters in time-series forecasting can be complemented by more indicators. For example, public interest or public opinions in a certain subject are influencing factors for the performance of a stock [6]. Data from Twitter, Facebook, Reddit, and popular News-Sites arguably represent a portion of the public opinion, while google search trends for example represent the public interest. Researchers and practitioners used such data of the past to address regression problems [e.g., 7, 8].



N°	Paper	N°	Paper
1	[9]	14	[10]
2	[7]	15	[11]
3	[12]	16	[13]
4	[14]	17	[15]
5	[8]	18	[4]
6	[2]	19	[16]
7	[3]	20	[17]
8	[18]	21	[19]
9	[20]	22	[21]
10	[5]	23	[22]
11	[23]	24	[24]
12	[25]		
13	[26]		

Figure 1. Bitcoin Price Prediction Approaches

In the reviewed papers, Blockchain information (e.g., blocks mined, number of transactions, total mining revenue, and cost per transaction or hash rates) is the most commonly used input parameter, supplementary to price data. The second mostly applied input feature is the Twitter sentiment. Nine research teams applied Twitter sentiments. Surprisingly, only three teams used Twitter volume data. Abraham et al. found out that Twitter volume and Google search trends are highly correlated with daily price changes [7]. Mc Wharter [15] found Google trends data to be the best predictor out of eight studied variables. Lamon et al. [10] used news headlines as text input feature. This method provides promising results for predicting general price trends, but struggles in accurate price predictions [10].

Out of the 26 considered works, 14 papers used regression models, either to forecast or to complement their approach by comparing it to other models. Their focus is on multivariate linear regression, logistic regression, and vector auto regression. The five works on SVM models are either early works in the field or showed that SVM performs worse than other approaches. Madan et al. [11] showed a decrease in accuracy when applying a SVM algorithm in contrast to binominal generalized linear models. In the reviewed papers, three approaches apply *Vector Autoregression* (VAR) [2, 9, 18]. *Vector error correction models* (VECM) are applied in an event study in order to find a connection between Twitter sentiment, Twitter volume and price reactions [26]. Three approaches that apply simple feedforward *neural networks* (NN) train the model solely with price data. Three out of six papers presenting *recurrent neural networks* (RNN) or LSTM models apply solely historical data on prices and trade volume [13, 27, 28]. Mc Nally et al. [13] for example build both a RNN and a LSTM network on daily price data. At the time of writing this paper and to our best knowledge, no investigations of applying neural networks, as predictive models of hourly Bitcoin price movements are available, which motivates the work at hand.

3 Research Design

3.1 Research Planning

In order to identify relevant predictors (RQ1) and to evaluate the performance of neural networks for hourly Bitcoin price predictions (RQ2), we conduct a four-step procedure (Figure 2). As the collection of the required data comes from different sources, we describe the *Data Collection* individually for each data source. The step *Data Adjustment* comprises data cleansing. *Sentiment Scoring* comprises the finding of sentiment polarity in both the collected tweets and news headlines. Afterwards, we merge the separate data sets and remove duplicates. Finally, *Model Development and Validation* comprises the development and validation of four different forecasting models.

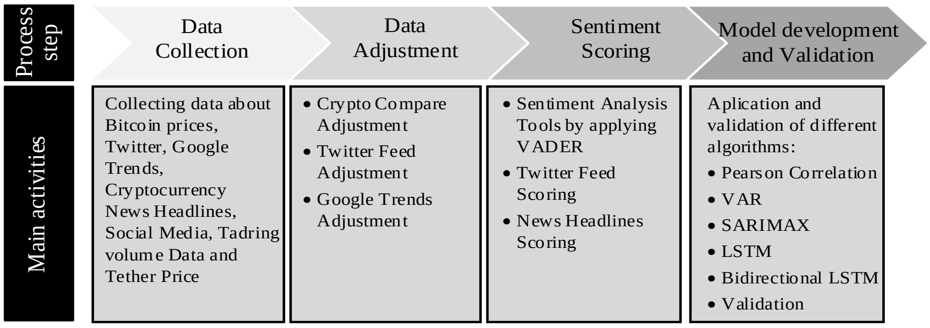


Figure 2. Data Preparation and Analysis Process

3.2 Data Collection

We list all data collected for the analysis in Table 1. Every data set is collected between September 2nd 2018 and July 26th 2019. In the following, we briefly describe all data sources and the retrieved data sets.

Data	Source	Method	Details
Price Data	Crypto Compare API	Python REST API	Bitcoin and Tether prices
Twitter Feed	Twitter API + kaggle.com	Python library "tweepy"	Represents public opinion
Google Trends	Google API	Python library "PyTrends"	Represents public interest
News Headlines	Crypto Compare API	Python REST API	Represents public opinion
Reddit/ Facebook/ GitHub Data	Crypto Compare API	Python REST API	Represents public opinion

Table 1. Data Sources

Bitcoin price data: Deviating from all related works, we did not collect the price changes of Bitcoin daily, but on hourly basis instead. A site, which provides global price data, is *cryptocompare.com*, which offers a free-to-use API to collect hourly aggregated price and volume data of over 70 cryptocurrency exchanges.

Twitter data: To collect every tweet regarding Bitcoin, we apply the Python library *tweepy*. 2.8 million Tweets regarding Bitcoin have been collected in the period of 7th of May to 26th of July in 2019. We extracted the timestamp, the text, the number of likes and retweets and complement it by historical tweets regarding Bitcoin, uploaded by a user on Kaggle.com, which led to over 6 million tweets in the period of 2nd September 2018 to 26th July 2019.

Google Trends data: In addition to sentiment scores as a representation of the public opinion, Google Trends data is going to represent overall public interest. To pull trend data from Google's API, the Python library *PyTrends* was used. We pulled Google Trends data between September 2nd 2018 and July 26th 2019 and scaled the results.

Cryptocurrency news data: To study the influence of cryptocurrency-related news sites on price movement and volatility of Bitcoin, we collect related headlines through Crypto Compare's API. Crypto Compare tracks the 40 biggest cryptocurrency-related news sites including: *CoinDesk, TodayOnChain, CoinTelegraph, CCN* and *NullTx*. According to crypto compare, headlines regarding Bitcoin include the phrases *"BTC", "BITCOIN"* and *"SATOSHI"*, while headlines with the phrase *"BITCOIN CASH"* are excluded. In total, 35.000 headlines for the timespan of 2nd of September 2018 to 26th of July 2019 were retrieved.

Social media volume data: Besides providing data on price changes and news feed, Crypto Compare also offers hourly data on social media platforms regarding certain cryptocurrencies. For Reddit, the *subreddit "r/Bitcoin"* is tracked on the number of subscribers, active users, posts, and comments per hour, as well as posts and comments per day. For Facebook, the page *"@Bitcoin"* is tracked on the number of likes and the total number of users that are talking about the page. The Twitter account "@Bitcoin" is tracked on the number of followers, favorites, statuses, the number of users wo are followed and the number of lists that the account is part of. The data set also contains seven data points on Bitcoins repository on *GitHub*. The platform offers the number of repository stars, forks, open and closed pulls, as well as open and closed bugs.

Tether price data: Tether is a special form of cryptocurrency, a so-called *stable coin*. The increasing amount of trading in Tether has a big impact on Bitcoin prices and according to Griffin and Shams [29] should be investigated for price manipulation. In order to receive hourly price and volume data on Tether, we apply the Crypto Compare API once again.

3.3 Data Adjustment

Crypto Compare data: By applying the python library pandas, we convert the column containing the hourly timestamp from a regular string object into a *datetime object*, in order to sort it in chronological order. The datetime object was then reduced to hours, by strafing the datetime object to format "%y-%m-%d %H", in order to merge it with other data sets later on.

Twitter data: To obtain a clean data set, we drop all duplicates and all data sets containing no data (empty containers). Finally, we reduce the original data to keep the timestamp and the text rows for the sentiment analysis.

Google Trends data: To get the correct scaling of the Google Trends data, we apply a Python script to overlap the weekly data points and calculate a ratio, in which the scales are adjusted. The script takes a start- and endpoint as a datetime object and a keyword input as a string object. Afterwards, we create a list, in which a datetime object of the starting point represents every new week. A for-loop now iterates over the range of the weekly list, starts downloading data, creates a Pandas data frame out of it and appends the data frames to another list. We fill the data frames with hourly timestamps of the weekly timeframe and obtain the weekly scaling between 0 and 100. By overlapping the weekly timeframes with one datapoint, a recalling of the weekly data would be possible. For that purpose, we use a third list to store the ratio between the score of the last hour of week 1 and the first hour of week 2 and so on. A second for-loop now iterates through all weekly data and applies this ratio as a correction parameter to the list of weekly Pandas data frames, except the first element (because there is no last hour of week zero). The weekly data frames are now merged into a single data frame representing the complete time period requested. Even though the correction parameter led to accurate values, the overall scaling needs to be fixed once again, since it is not in a range between 0 and 100.

3.4 Sentiment Scoring

In order to receive sentiment scores, we apply the python library VADER on both the collected tweets and collected news headlines. All hyperlinks included in the tweets were deleted. Due to the optimization of VADER to social media texts including emojis and special characters, we decided to keep them, as they probably provide more accurate results regarding the polarization of the given text. Since every tweet with the hashtag Bitcoin were collected, no additional filtering on languages was done. In order to handle non-english tweets, we apply the G*oogle Translate* Python library. The Google Translate script checks whether the given text is written in English and thus translates the text if necessary.

VADER delivers a score that indicates the polatization of the text (positive: 1, neutral: 0, negative: -1) We solely append a score of a text if it is either above 0.3 or below -0.3, otherwise a 0 is added in the final data set. Applying this approach brings 3.648.079 individual scored tweets for the timespan of 2nd of September 2018 to 26th of July 2019. The final Twitter dataset consists of hourly timestamps with the average of the sentiment values for the respective hour. We apply the same approach on the data set of 35.000 collected news headlines.

3.5 Model Development

In order to find linear relationships between input features and the Bitcoin close price, we apply the *Pearson correlation* analysis. In addition, we compare the results and the predictive power of four approaches: VAR, SARIMAX, LSTM and BiLSTM**.** A Vector auto regression (VAR) model is a multivariate linear time-series model and is considered as a simple and flexible alternative to the traditional multiple-equations models. VAR uses linear relations between variables, a trend component, constant intercepts and uncorrelated errors (Garcia and Schweitzer 2015). The definition of a lagging parameter and a minimum of two endogenous variables are needed, in order to fit the model. We train the VAR on Bitcoin close, low and high prices as exogenous variables and since VAR models do not have any hyperparameters, we do not need to tune these models.

A Seasonal Autoregressive Integrated Moving-Average with Exogenous Regressors (SARIMAX) model is an ARIMA model that can also handle seasonal components (S) and includes the modeling of exogenous variables (X). We configure the Bitcoin close price as endogenous variable and the collected input features as exogenous variables.

A *Neural Network* (NN) is an information-processing mechanism that is inspired by the human brain. NN learn from "observational data, figuring out its own solution to the problem at hand" [30]. By receiving a set of inputs (also called features) and performing increasingly complex calculations, the network outputs a predictive value or class assignment. A NN consists of a web of nodes called *neurons*, which are grouped up in layers and linked to each other through connectors.

Unlike conventional NN, also called feed forward networks, Recurrent Neural Networks (RNN) can receive a sequence of values as input. To make predictions on statistical data, a time series can be implemented as a sequence, where an output can be the next value in that sequence. In a RNN, the output of the layer is added to the next input and fed back into the same layer, which is typically the only layer in the entire network [31]. The problem of *vanishing gradient*, already known from feedforward networks, is further reinforced by the architecture of RNNs, because each time step is the equivalent of an entire layer of a feed-forward network. This leads to even smaller gradients and to a loss of information over time [32]. To address that problem, so-called gates were introduced to RNNs to forget or remember the current input, if the network decides the information is required for future time steps [31] (Chung et al. 2014). An often-used gate architecture today is called Long Short-Term Memory (LSTM). LSTM Networks were proposed in 1997 by Hochreiter and Schmidhuber and were designed to soften the vanishing or exploding gradient problem [33].

We build the LSTM models with the Python libraries *TensorFlow* and *Keras* [34], while the training was done with a GPU (GeForce GTX 1660Ti). We prepare the data set by framing it as a supervised learning problem and normalizing the input variables. Next, we split the given data into a training set and a test set, in order to test the model on data that is unknown. Therefore, we use a ratio of 90% training data and 10% test data.

We improve the training of the model by changing the hyperparameters, such as the number of layers or neurons or Epochs. For implementing a *Bidirectional LSTM* (BiLSTM) network, which is an advanced version of a regular (or unidirectional) LSTM, we use the same procedure as for the regular LSTM.

A commonly used technique for validating prediction models is cross-validation, which tests how the results of a model generalize to an independent data set, by estimating how accurate a predictive model performs outside the training set [35]. Therefore, the data set must be split into a training set and test set. Cross-validation also works for tuning hyperparameters. To use hyperparameter tuning, the training set needs to be split again into a validation set and a training subset. The model is trained on the training subset, while the parameters are chosen in a way that minimizes the error for the validation set. By using the selected parameters, we train the model on the full training set and test it on the test data set. In order to validate the performance of the VAR and SARIMAX models, a 5-fold cross-validation was performed. In order to tune the hyperparameters and validate the performance of the LSTM and BiLSTM, we apply a 5-fold nested cross-validation.

The final step comprises the tuning of hyper parameters. SARIMAX models have two sets of parameters, the order parameters (p, q, d) and the seasonal parameters (P, Q, D). We evaluate the performance of the model by the Akaike Information Criteria (AIC) score. While the number of layers is fixed to one and the number of epochs to 30, the nested cross-validation process tries to find the best combination of an Optimizer (Adam, RMSprop or SGD), 64 or 128 neurons and a batch size of either 32 or 72. We evaluate the results by applying the loss functions Mean Square Error (MSE) and Mean Absolute Percentage Error (MAPE). The used Python script reveals the best working combination of parameters for every test set in the cross-validation process. The BiLSTM and LSTM parameters are also tuned by nested cross-validation. The number of layers is fixed to three and the number of epochs to 30. Due to unconvincing results of our factor analysis, we decided to integrate all collected features into our training. In order to compare the performance of the five models, we use the results of 5-fold cross-validation for all models for the same time span. To quantify the results, we calculate a MAPE score for each model and for each fold, as well as an average of the MAPE for the complete time span.

4 Results

4.1 Influencing Factors

In total, we analyze the power of 47 influencing factors to predict the Bitcoin price movements. Table 2 provides the results of the Pearson correlation analysis. 2 out of 47 possible factors are statistically non-significant (News sentiment score and Tether price high). In contrast to the non-significant factors, we identify eight factors that strongly correlate with the Bitcoin price development, i.e. these factors have a correlation coefficient above 0.7 or less than -0.7. In line with other studies about the predictive power of Twitter sentiment [e.g. 7, 26], we also identify a strong correlation for the hourly Bitcoin price movements. In addition, we confirm the close relationship between the cryptocurrencies Bitcoin and Tether. All Tether prices that are measured in Bitcoin provide a correlation coefficient above -0.9. Thus, the data reveals that the higher the Tether price, the lower the Bitcoin price and vice versa. All Bitcoin price data have a significant influence on the Bitcoin hourly price, which is not surprising because these values (BTC price high, low, and open) relate to the hourly price directly. Besides factors that have a strong correlation with the Bitcoin price, we also identify factors with a medium correlation. Among others, the page views of Crypto Compare receive correlation coefficients between 0.4 and 0.5. Less important indicators are Facebook likes as well as GitHub code pulls.

Variable	Coefficient	Variable	Coefficient
Twitter Sentiment Score	0.7534131***	CryptoC. Forum Page Views	0.4774623***
News Sentiment Score	-0.0038534	CryptoC. Influence Page Views	0.4720214***
Google Trends	0.3141094***	CryptoC. Markets Page Views	0.4587880***
Reddit Subscribers	0.3656004***	CryptoC. Overview Page Views	0.5380816***
Reddit Comments per Hour	0.2833215***	CryptoC. Points	0.4829218***
Reddit Comments per Day	0.2833221***	CryptoC. Posts	0.4701490***
Reddit Active Users	0.0502295**	Bitcoin Price High	0.9998439***
Reddit Posts per Day	0.2574809***	Bitcoin Price Low	0.9998195***
Reddit Posts per Hour	0.2574877***	Bitcoin Price Open	0.9996947***
Facebook likes	0.1302711***	Bitcoin Volume from	0.0634226***
Facebook talked about	-0.0401814**	Bitcoin Volume to	0.4098491***
GitHub Rep. Open Pull Issues	0.3614214***	Tether Price Close (BTC)	-0.9380909***
GitHub Rep. Closed Issues	0.2054122***	Tether Price High (BTC)	-0.9375232***
GitHub Rep. Closed Pull Issues	0.2345545***	Tether Price Low (BTC)	-0.9384082***
GitHub Rep. Forks	0.2298345***	Tether Price Open (BTC)	-0.9378622***
GitHub Rep. Open Issues	0.2407842***	Tether Volume from (BTC)	0.6001629***
GitHub Rep. Stars	0.1874888***	Tether Volume to (BTC)	0.0765503***
GitHub Rep. Subscribers	-0.0261862*	Tether Price Close (USD)	0.0562241***
CryptoC. Total Page Views	0.5044371***	Tether Price High (USD)	0.0117722
CryptoC. Trades Page Views	0.5120901***	Tether Price Low (USD)	0.0909971***
CryptoC. Forum Comments	0.4724169***	Tether Price Open (USD)	0.0564081***
CryptoC. Analysis Page Views	0.4554351***	Tether Volume from (USD)	0.2937562***
CryptoC. Charts Page Views	0.4373161***	Tether Volume to (USD)	0.2944756***
CryptoC. Followers	0.5362737***	*** p ≤ 0.001, ** p ≤ 0.01, * p ≤ 0.05	

Table 2. Bitcoin Price Correlation Coefficients

4.2 Model Results

The SARIMAX model showed the best results with an order of (2, 1, 0) and a seasonal order of (0, 0, 0) 6000. The best working hyperparameters for LSTM and the Bidirectional LSTM network, found in nested cross-validation, are listed in Table 3. The comparison of the predictive power of the models for the fifth cross-validation is depicted in Figure 3.

The regular LSTM seems to deliver the best results judging by the graphical comparison. The BiLSTM shows a curve with big spikes. The SARIMAX model shows a good coverage of the original curve, but does not react to the price drop in mid-July. The VAR model curve appears to be a relatively solid representation of the original curve but is for most of the time below the real price level.

The corresponding MAPE scores of all five validations and averages are compared in Table 4. It should be noted that both LSTM and BiLSTM achieve their best performance in CV2, as the market was not very volatile at the time. The 3.52 % error in CV5 of LSTM Network is particularly good as the market was volatile at that time and can be viewed as the best performance of all models and validations. The VAR receives a MAPE of 6.36 % in CV5 although its average error rate is much higher. However, the average value is increased heavily by the outlier value of the second cross-validation. Apparently, the VAR model needs more data to function properly in comparison to the neural networks. Similarly, the SARIMAX model behaves worse in CV2 than in CV1, but improves significantly.

Network	loss function	Cross-Validation	Optimizer	Learn Rate	Batch Size	Neurons	Layers
LSTM	MSE	CV1	SGD	0.2	72	64	2
		CV2	SGD	0.2	32	64	2
		CV3	SGD	0.2	32	128	2
		CV4	Adam	0.2	72	128	2
		CV5	RMSProp	0.2	72	128	2
	MAPE	CV1	Adam	0.2	32	64	2
		CV2	RMSprop	0.2	32	64	2
		CV3	RMSprop	0.2	32	64	2
		CV4	Adam	0.2	32	64	2
		CV5	Adam	0.2	72	64	2
BiLSTM	MSE	CV1	Adam	0.2	72	64	3
		CV2	Adam	0.2	72	128	2
		CV3	Adam	0.2	72	128	3
		CV4	SGD	0.2	72	128	2
		CV5	SGD	0.2	72	64	2
	MAPE	CV1	Adam	0.2	72	64	3
		CV2	RMSprop	0.2	72	128	2
		CV3	RMSprop	0.2	72	64	3
		CV4	Adam	0.2	72	64	2
		CV5	RMSProp	0.2	72	64	2

Table 3. Best BiLSTM and LSTM Hyperparameters

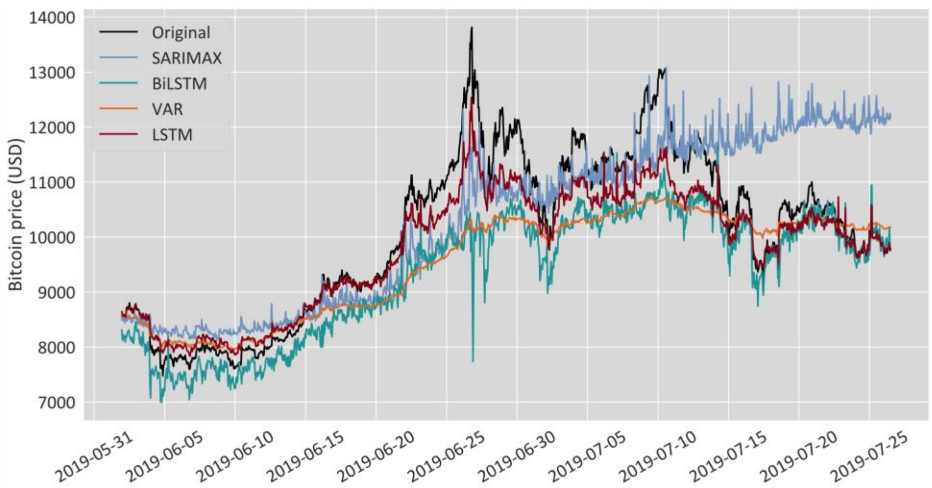


Figure 3. Result Comparison of the fifth Cross-Validation Prediction

Model	Cross-Validation	Percentage of absolute error CV5	Error average for all CV
	CV1	27.29 %	
	CV2	1.18 %	
LSTM	CV3	3.73 %	8.93 %
	CV4	8.95 %	
	CV5	3.52 %	
	CV1	63.39 %	
	CV2	407.7 %	
VAR	CV3	31.86 %	104.20 %
	CV4	11.55 %	
	CV5	6.36 %	
	CV1	31.68 %	
	CV2	1.93 %	
BiLSTM	CV3	7.55 %	13.03 %
	CV4	16.93 %	
	CV5	7.05 %	
	CV1	37.61 %	
	CV2	69.24 %	
SARIMAX	CV3	9.68 %	26.06 %
	CV4	5.66 %	
	CV5	8.10 %	

Table 4. Error Comparison

5 Discussion and Outlook

The results of the Pearson Correlation Analysis suggest that the public opinion is a measurable indicator of Bitcoin price changes. On the one hand, the very high correlation between Twitter sentiment scores and the close price shows the importance of the public opinion for the price change of Bitcoin and supports earlier work on daily data. Cryptocurrency news headlines on the other hand do not show any correlation. This is surprising since news outlets specialized on cryptocurrency have arguably a more direct contact to the industry than Twitter. This finding could implicate that the Twitter sentiment score also represents the Twitter community's reaction to certain price changes in the respective time span, rather than Bitcoin's price changing according to the general Twitter polarity. Nevertheless, the results confirm the influence of public opinion on Bitcoin's hourly closing prices.

Measurements of public interest in Bitcoin show a similar picture. The Google Trends scale data correlates moderately and therefore shows a relationship between the amount of Google search queries and the price trend. While the active users on Reddit per hour do not correlate with close prices, the track of Bitcoin's Subreddit shows a weak correlation. However, the values for Facebook activities do not seem to be associated with close prices, since Facebook's *likes* and Facebook's *talked about* show no linear relationship. In summary, public interest in Bitcoin has a measurable impact on hourly price movements.

Surprisingly, every data point of the internal data pulled from Crypto Compare correlates with the Bitcoin hourly close prices. The data points are a track of the usage of their site and forum. Even though not directly related to Bitcoin, the total page views of cryptocompare.com show a high correlation with close prices. This could be due to the fact, that Bitcoin remains a pseudonym for cryptocurrencies and people possibly get involved by hearing of Bitcoin first. Another surprising medium correlated input data point is the number of open pull issues on GitHub, which is the notification on changes being pushed to a repository, which is then being discussed and reviewed by collaborators. This predictor may represent the disagreement of the mining community about changes in the underlying Blockchain implementation of the Bitcoin network. Supporting the paper of Griffin and Shams, the hourly data of all Tether price measurements are very negatively correlated with Bitcoin close prices, as well as the volume of traded tether coins [29].

The comparison of the forecasting results reveals that Long Short-Term Memory networks are best suited for Bitcoin price prediction out of the five models that we considered in the analysis. We cannot identify a particular optimizer to work better than another since the three optimizers are distributed equally in the best performing combinations of hyperparameters. The error rate of 3.52 percent on unseen data in the last validation process is a good result, since it outperforms the second-best models by 40 percent in the forecasting error. Even when comparing the average error rates, LSTM receives the best average error rate with a value of 8.993 percent, which outperforms the second-best model by 30 %. Guo et. al. achieved comparable results, as they stated that their study showed a 50% more accurate performance of LSTM networks compared to foundational statistical indicators [5].

In each validation process, the regular Long Short-Term Memory network outperforms the bidirectional LSTM, which is surprising since BiLSTM is a more advanced version of neural net models. Perhaps the network architecture does not fit this particular forecasting problem. The VAR model performs much better if it gets all input features than if it gets only a few. It also reaches the highest average error rate of the five models, but adapts convincingly over the course of the cross-validation, scoring the second lowest error on the last test set. The SARIMAX performance is the second worst performing model, according to the average error rate and the error rate of the fifth cross-validation.

Against the background of these results, the paper at hand contributes to research in three ways. First, 45 significant indicators on the Bitcoin price are identified and discussed. Second, we confirm that the usage of a trained long short-term memory (LSTM) neural network produces the best Bitcoin price predictions on an hourly basis. Third, the results provide a basis for a fruitful discussion of the applicability of neural nets for stock price predictions.

From a practical perspective, the tracking of hourly data points might be used as a trading strategy, as the model performs well on unknown data. The results could encourage asset managers to test the model in practice. All input features that the model was trained on could be streamed live for the respective hour and fed into a trading bot system. Another application is the use as a forecasting model for the highest price of the next hour, in order to follow the trading strategy of selling at the highest price in the respective hour.

A general limitation of this work is the restraint in data availability. This paper has no claim to completeness, as there could exist more predictors. Hourly data on Blockchain information, the Standard & Poor's 500 and the CBOE Volatility Index are not included in the data set due to missing data availability. The inclusion of such data or the increase of the analyzed time period might improve the forecasting model. Furthermore, solely Twitter feeds and news headlines represent the relationship between public opinion and Bitcoin prices. Even though these are arguably solid measurements, the public opinion on certain topics is obviously represented by more than just two data points and suggest integrating other data points. The same applies to Google Trends as a representation of public interest. Applying more advanced tuning techniques such as grid-search, especially for the SARIMAX model might improve the hyperparameter tuning of the considered models.

The developed models in this paper are a starting point for a more precise Bitcoin price forecasting and lead to more research on hourly forecasting with different models and different input variables. Further research should focus on the application and evaluation of such models in practice. We suggest conducting a case study together with asset managers in order to verify the applicability of LSTM for Bitcoin price prediction in asset management.

References

1. Popper, N.: After the Bust, Are Bitcoins More Like Tulip Mania or the Internet? New York Times (2019)
2. Garcia, D., Schweitzer, F.: Social signals and algorithmic trading of Bitcoin. Royal Society open science 2, 1–13 (2015)
3. Georgoula, I., Pournarakis, D., Bilanakos, C., Sotiropoulos, D.N., Giaglis, G.M.: Using Time-Series and Sentiment Analysis to Detect the Determinants of Bitcoin Prices. SSRN Journal (2015)
4. Mohanty, P., Patel, D., Patel, P., Roy, S.: Predicting Fluctuations in Cryptocurrencies' Price using users' Comments and Real-time Prices. In: 2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), pp. 477–482. IEEE (2018)
5. Guo, T., Bifet, A., Antulov-Fantulin, N.: Bitcoin Volatility Forecasting with a Glimpse into Buy and Sell Orders. In: 2018 IEEE International Conference on Data Mining Workshops (ICDMW), pp. 989–994. IEEE (2018)
6. Tetlock, P.C.: Giving Content to Investor Sentiment: The Role of Media in the Stock Market. THE JOURNAL OF FINANCE (2007)
7. Abraham, J., Higdon, D., Nelson, J., Ibarra, J.: Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis. SMU Data Science Review 1 (2018)
8. Galeshchuk, S., Vasylchyshyn, O., Krysovatyy, A.: Bitcoin Response to Twitter Sentiments. In: 6th International Workshop on Information Technologies in Economic Research (ITER 2018) (2018)

9. Aalborg, H.A., Molnár, P., Vries, J.E. de: What can explain the price, volatility and trading volume of Bitcoin? Finance Research Letters 29, 255–265 (2019)
10. Lamon, C., Nielsen, E. and Redondo, E.: Cryptocurrency Price Prediction Using News and Social Media Sentiment, https://www.semanticscholar.org/paper/Cryptocurrency-Price-Prediction-Using-News-and-Lamon-Nielsen/c3b80de058596cee95beb20a2d087dbcf8be01ea (Accessed: 22.11.2019)
11. Madan, I., Saluja, S. and Zhao, A.: Automated Bitcoin Trading via Machine Learning Algorithms, https://www.semanticscholar.org/paper/Automated-Bitcoin-Trading-via-Machine-Learning-Madan/e0653631b4a476abf5276a264f6bbff40b132061 (Accessed: 22.11.2019)
12. Akcora, C.G., Dey, A.K., Gel, Y.R., Kantarcioglu, M.: Forecasting Bitcoin Price with Graph Chainlets. In: Phung, D., Tseng, V.S., Webb, G.I., Bo, B., Ganji, M., Rashidi, L. (eds.) Advances in knowledge discovery and data mining. 22nd Pacific-Asia conference, PAKDD 2018, Melbourne, VIC, Australia, June 3-6, 2018 : proceedings, 10939, pp. 765–776. Springer, Cham (2018)
13. McNally, S., Roche, J., Caton, S.: Predicting the Price of Bitcoin Using Machine Learning. In: 2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP), pp. 339–343. IEEE (2018)
14. Ceyhan, K., Kurtulmaz, E., Sert, O.C., Ozyer, T.: Bitcoin movement prediction with text mining. In: 2018 26th Signal Processing and Communications Applications Conference (SIU), pp. 1–4. IEEE (2018)
15. McWharter, N.: Bitcoin and Volatility: Does the Media Play a Role?, https://creativematter.skidmore.edu/econ_studt_schol/82
16. Pant, D.R., Neupane, P., Poudel, A., Pokhrel, A.K., Lama, B.K.: Proceedings on 2018 IEEE 3rd International Conference on Computing, Communication and Security (ICCCS). October 25th - 27th, 2018, Kathmandu, Nepal : an IEEE Nepal Sub Section Conference. IEEE, Piscataway, NJ (2018)
17. Radityo, A., Munajat, Q., Budi, I.: Prediction of Bitcoin Exchange Rate to American Dollar Using Artifical Neural Network Methods. In: 2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS) (2017)
18. Giudici, P., Abu-Hashish, I.: What determines bitcoin exchange prices? A network VAR approach. Finance Research Letters 28, 309–318 (2019)
19. Rodricks, Matthew, R.: Statistical Determinants of Bitcoin, http://www.wpi.edu/Pubs/E-project/Available/E-project-042518-122138/unrestricted/MQP_Final.pdf (Accessed: 22.11.2019)
20. Greaves, A. and Au, B.: Using the Bitcoin Transaction Graph to Predict the Price of Bitcoin, https://www.semanticscholar.org/paper/Using-the-Bitcoin-Transaction-Graph-to-Predict-the-Greaves-Au/a0ce864663c100582805ffa88918910da89add47 (Accessed: 22.11.2019)
21. Saad, M., Mohaisen, A.: Towards Characterizing Blockchain-based Cryptocurrencies for Highly-Accurate Predictions. In: IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), pp. 704–709. IEEE (2018)
22. Shah, D., Zhang, K.: Bayesian regression and Bitcoin. In: 2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton), pp. 409–414. IEEE (2014)
23. Indera, N.I., Yassin, I.M., Zabidi, A., Rizman, Z.I.: Non-linear Autoregressive with Exogeneous input (narx) bitcoin price prediction model using PSO-optimized

parameters and moving average technical indicators. J. Fundam and Appl Sci. 9, 791 (2018)
24. Stenqvist, E. and Lönnö, J.: Predicting Bitcoin price fluctuation with Twitter sentiment analysis, https://www.semanticscholar.org/paper/Predicting-Bitcoin-price-fluctuation-with-Twitter-Stenqvist-L%C3%B6nn%C3%B6/0954565aebae3590e6ef654fd03410c3bdd7d15a (Accessed: 22.11.2019)
25. Jang, H., Lee, J.: An Empirical Study on Modeling and Prediction of Bitcoin Prices With Bayesian Neural Networks Based on Blockchain Information. IEEE Access 6, 5427–5437 (2018)
26. Kremser, T., Radszuwill, S., Schweizer, A., Steffek, B.: How Do Large Stakes Influence Bitcoin Performance? Evidence from the Mt.Gox Liquidation Case. In: 27th European Conference on Information Systems (2019)
27. Wu, C.-H., Lu, C.-C., Ma, Y.-F., Lu, R.-S.: A New Forecasting Framework for Bitcoin Price with LSTM. In: 2018 IEEE International Conference on Data Mining Workshops (ICDMW), pp. 168–175. IEEE (2018)
28. Yiying, W., Yeze, Z.: Cryptocurrency Price Analysis with Artificial Intelligence. In: 5th International Conference on Information Management (2019)
29. Griffin, J.M., Shams, A.: Is Bitcoin Really Un-Tethered? SSRN Journal (2018)
30. Nielsen, M.: Neural Networks and Deep Learning, http://neuralnetworksanddeeplearning.com (Accessed: 22.11.2019)
31. Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. In: NIPS 2014 Deep Learning and Representation Learning Workshop (2014)
32. Hochreiter, S.: The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions. Int. J. Unc. Fuzz. Knowl. Based Syst. 06, 107–116 (1998)
33. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation 9, 1735–1780 (1997)
34. Abadi, M., Agarwal, A. and et al.: TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems, http://arxiv.org/pdf/1603.04467v2
35. Vanderplas, J.T.: Python Data Science Handbook. O'Reilly (2017)