

Sentiment Analysis Based Direction Prediction in Bitcoin using Deep Learning Algorithms and Word Embedding Models

Zeynep Hilal Kilimci*¹

Submitted: 03/09/2019 Accepted: 17/06/2020

Abstract: Sentiment analysis is a considerable research field to investigate enormous quantity of knowledge and specify user opinions on many subjects and is resumed as the extraction of ideas from the textual data. Like sentiment analysis, Bitcoin which is a digital cryptocurrency also attracts the researchers considerably in the domain of cryptography, computer science, and economics. The objective of this work is to forecast the direction of Bitcoin price by analyzing user opinions in social media such as Twitter. To our knowledge, this is the very first attempt which estimates the direction of Bitcoin price fluctuations by using word embedding models in addition to deep learning techniques in the state-of-the-art studies. For the purpose of estimating the direction of Bitcoin, recurrent neural networks (RNNs), long-short term memory networks (LSTMs), and convolutional neural networks (CNNs) are used as deep learning architectures and Word2Vec, GloVe, and FastText are employed as word embedding models in the experiments. In order to demonstrate the contribution of our work, experiments are carried out on English Twitter dataset. Experiment results show that the usage of FastText model as a word embedding model outperforms other models with 89.13% accuracy value to estimate the direction of Bitcoin price.

Keywords: Bitcoin, deep learning, FastText, long short-term memory networks, sentiment analysis

1. Introduction

Social media is widely-used environment to interpret/understand thoughts of users about occurrences, products, supplies, demands. One of the well-known social media platforms is Twitter is employed by over the 100 million active users to explain ideas of them [1]. Because of expressing a huge amount of information by the users, Twitter encloses valuable knowledge which can be efficient for market dynamics. That is why the sentiment analysis is an important to understand and interpret the user requests in terms of negative and positive ways.

Decision tree (DT), naïve Bayes (NB), support vector machine (SVM), and so on, are conventional machine learning techniques employed in sentiment analysis field to determine the polarity of sentiment such as neutral, positive, or negative. On the other hand, deep learning algorithms are more popular and preferred by the researchers for the purpose of getting higher classification success. These popular models are used for both feature extraction and classification purposes in the literature. The main idea behind of the deep learning methodologies is to ensure the phase of feature extraction automatically thereby training complex features with least exterior reinforcement and come by more expressive demonstration of data. Deep belief networks (DBNs), convolutional neural networks (CNNs), recursive neural networks, recurrent neural networks (RNNs), are utilized for both feature extraction step and classification while word embedding models such as word2vec, Glove, and FastText are generally employed for feature extraction. Deep learning methodologies have been broadly performed by researchers in various fields such as natural language processing, computer vision, speech recognition, image/video processing, and so on.

Bitcoin is proposed by an mystery community or person employing the title Satoshi Nakamoto. In 2009, it is released as open-source software. Bitcoin is actually a form of electronic cash investment which is called a cryptocurrency. On the network of peer-to-peer bitcoin, Bitcoin can be handed on from a person to another one without any necessity of intermediaries. It is exchanged for other currencies, products, and services. In 2007, a cryptocurrency wallet is used by almost 5.8 million users according to the report in [2].

In this study, we propose sentiment analysis based direction prediction in Bitcoin employing deep learning methodologies, and word embedding models. This is the very first attempt, to the best of our knowledge, to estimate the direction of Bitcoin price using word embedding models in addition to deep learning techniques on the sentiment analysis of short texts. For this purpose, recurrent neural networks, long short-term memory networks, and convolutional neural networks as deep learning architectures and Word2Vec, GloVe, and FastText as word embedding models are employed on English Twitter dataset. Experiment results show that handling FastText model boosts the classification success of the proposed model.

The paper is designed as: Related works on direction or price prediction of Bitcoin are summarized in Section 2. In Section 3, the proposed framework is mentioned. Experiment setup and results are presented in Section 4. In Section 5, conclusion is given.

2. Related Work

This section maintains a brief summary of the literature studies on direction or price prediction of Bitcoin.

In [3], Isaac et al. aim to estimate the Bitcoin price employing machine learning models. For this purpose, they focus on daily trends in Bitcoin market and optimal features surrounding Bitcoin price by before gathering the dataset. Then, the dataset is constructed daily with more than 25 features which are related to

Information Systems Engineering, Kocaeli University, Kocaeli – 41001, TURKEY ORCID ID :0000-0003-1497-305X

* Corresponding Author Email: zeynep.kilimci@kocaeli.edu.tr

the Bitcoin price. After obtaining 98.7% of accuracy for the indicator in changes of Bitcoin price daily, they try to analyze leveraged Bitcoin data at 10-second time and 10-minute intervals to evaluate the opportunity purchasing in Bitcoin. Finally, the study is concluded that 50-55% accuracy performance is acquired in forecasting the signal of future changes in Bitcoin price utilizing 10 minute time intervals. In another study [4], they investigate the effect of Bayesian regression method for forecasting the price variation of Bitcoin with real-valued quantity. Based on this method, a basic trading strategy in Bitcoin is designed. They claim that the strategy is able to almost double investment less than 60-day when employ against real-time data performance.

In [5], Jang and Lee propose to forecast prices in Bitcoin by modeling with bayesian neural networks. First, authors investigate the impact of Bayesian neural network models by focusing on the time series analysis of Bitcoin data. Second, the most appropriate information on features are gathered from Blockchain platform. Features are trained with the proposed models to advance the system performance in Bitcoin. Experiments are performed by comparing the proposed model with non-linear and linear conventional methods on estimating the Bitcoin price. In conclusion, experiment results display that the utilization of BNN exhibits remarkable results in forecasting price of Bitcoin.

In other work [6], McNally et al. propose to estimate the Bitcoin price in USD using deep learning techniques. For this purpose, a long short-term memory network, and bayesian optimised recurrent neural network are employed. The LSTM exhibit the highest classification accuracy of 52% and a RMSE of 8%. Moreover, ARIMA as a time series model is applied to compare with the performance of deep learning methods. The experiment results show that deep learning models perform better than ARIMA whose performance is poor. In an interesting work [7], Sin and Wang propose to forecast the price of Bitcoin employing neural network ensembles. Authors focus on to investigate the correlation between changes in Bitcoin price in the next day and the characteristics of Bitcoin and utilizing ensembles of neural networks. To construct the neural network ensembles, multi-layered perceptron is employed as individual learner of the ensemble system. In other words, proposed ensemble model is constructed for the purpose of predicting direction of Bitcoin price of the next day implemented nearly 200 features of the cryptocurrency more than 2 years time period. They conclude the study that the previous approach produces just about 85% revenues, surpassing the “prior-day trend surveillance” approach which generates nearly 38% revenues and a trading model that observes the best performing MLP method by using ensemble methodology that generates almost 53% in revenues.

In [8], Matta et al. focus on the Bitcoin spread prediction using social media. They compare trends in price employing Google Trends dataset. They claim that there is a significant cross relation values, especially between the price of Bitcoin and the dataset gathered from Google trends. In [9], Garcia and Schweitzer propose to demonstrate the effect of social signs and trading strategy for Bitcoin by employing economic signs of transaction volume and price in exchange for USD. They analyze social signs considering the search of information, negative and positive aspects from tweets, so on additionally to the volume of transactions in Bitcoin. They claim that rises in analysis of sentiments and exchange volume take precedence of compared to the rising Bitcoin prices. They also verify that having great performance in terms of profitability with sturdy statistical methods that take into account costs of trading and risk analysis. In [10], Galeshchuk et al. focus on behavioral signals for the

purpose of predicting Bitcoin fluctuations by analyzing Twitter comments of users. Actually, they both evaluate the Bitcoin exchange rate dataset that is based on numeric values and text-based Twitter dataset. Random Walk (RW) and Integrated Moving Average (ARIMA) models are employed for the prediction of Bitcoin exchange rate while Multilayer Perceptron (MLP) and Convolutional Neural Networks (CNNs) are used to analyze sentiments of Twitter users. Authors conclude the study that the usage of CNNs is more advantageous compared to the others. In [11], the author propose to assess the effect of user comments on Twitter for predicting the direction of Bitcoin by only considering sentiment score of each user comment. In [12], the impact of Twitter sentiments on Bitcoin price prediction is investigated by assessing Convolutional Neural Networks (CNNs), Long Short Term Memory (LSTM), and Gated Recurrent Unit (GRU). Aggarwal et al. observe that deep learning models are obviously efficient for the estimation of Bitcoin price.

Our work is different from the above aforementioned literature studies in that this is the first study for employing word embedding models in addition to deep learning techniques for forecasting the direction of Bitcoin price. In Section 3, proposed model is introduced in details.

3. Proposed Framework

A summary of the methods, materials, and proposed method are introduced in this section.

3.1. Dataset Preparation and Proposed Framework

In this study, we focus on to forecast the direction of Bitcoin employing deep learning algorithms and word embedding models by evaluating sentiment analysis of Bitcoin related tweets.

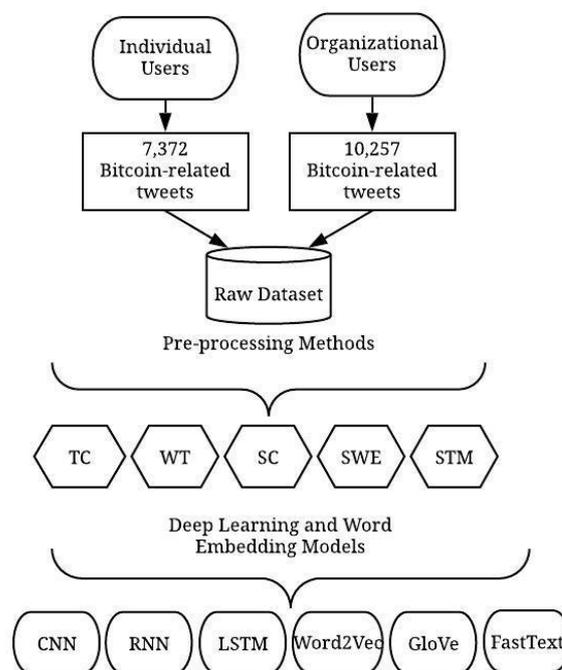


Fig. 1. Flowchart of the system.

For this purpose, “BitcoinDollar”, “BitcoinUSD”, “BTCDollar”, “BTCUSD” related tweets are gathered from both individual and organizational user accounts in English between 05/01/2019 and 08/01/2019 time interval with our own crawler to eliminate the

limitation of Twitter API.

In this study, individual accounts with public tweets are taken a consideration because of the protected tweets of some individual accounts. 10,257 for organizations and 7,372 for individual account totally; 17629 of tweets are gathered and labelled as positive, or negative through TextBlob [13] that uses naive Bayes classifier to determine the sentiment of tweets and produces the class probability of each tweet. The sentiment of tweets is assigned as 82.5% average classification accuracy through TextBlob. Because the raw dataset collected from each account is quite noisy in social media platforms, different pre-processing techniques are applied. In this study, stop-word elimination, removing hashtags, removing URLs, removing punctuations, word tokenization, stemming, and spell correction techniques are implemented. In this work, Word2Vec, GloVe, and FastText are employed as word embedding models for the purpose of enriching tweets in terms of meaning, context and syntax. Then, instead of using conventional machine learning algorithms, three different deep learning architectures such as long short-term memory networks, recurrent neural networks, and convolutional neural networks are utilized. Flowchart of the system is given in Fig. 1.

3.2. Deep Learning Architectures

In this study, we focus on the widely-applied three deep learning algorithms namely, long short-term memory networks (LSTMs), recurrent neural networks (RNNs), and convolutional neural networks (CNNs).

Convolutional Neural Networks (CNNs): CNN is widely known and used deep learning model in image recognition, image classification, video classification, visual recognition, natural language processing [14-22]. CNN is also a feedforward neural network that covers an hidden layers, output and input layers. Hidden layers occurs from convolutional layers blending by pooling layers. The convolution layer is the most significant part among the blocks of CNN. A convolution filter is performed in convolutional layer to generate a feature map for the purpose of consolidating information and data that is located on filter. Multiple filters are implemented on inputs in order to obtain a bulk of feature maps. The final output is transformed by these feature maps. Local dependencies are obtained with convolution operation on the original dataset. Moreover, rectified linear unit as a supplementary activation function is implemented to feature maps for the purpose of associating non-linearity to CNN. Next, the number of samples in each feature map is diminished by a pooling layer, which holds the most significant information. In this way, both dimensionality and over-fitting process is reduced with the usage of pooling layer and the training time decreases. CNN architecture is based on a combination of convolutional and pooling layers, followed by a sequences of fully connected layers.

Recurrent Neural Networks (RNNs): RNN is a kind of neural network in which output from the processing step pretends as input in the current stage. RNN is proposed due to the requirement of remembering the words. This problem is solved with the aid of hidden layer [20, 23-24]. Having a secret state for RNN model ensures to recollect information about a series. RNN is able to call up all calculations with the usage of memory. Same procedure is implemented on overall hidden layers or inputs for the purpose of generating output. Using the same parameters for each input reduces the complexity of the parameters.

Long Short-Term Memory Networks (LSTMs): LSTM is another popular technique is to employ which are designed as a specific type of recurrent neural networks to solve the gradient disappearing issues of RNNs [27-28]. LSTM preserves the error to

back-propagate through deeper layers and to proceeds to learn over many time steps. Actually, LSTMs are improved in order to obtain long-distance dependences by using sequences. The contextual semantics of information are retained and stored for the purpose of obtaining long dependencies between data with LSTM approach. For this purpose, special memory units are exploited to stock information for dependencies in long range context in LSTM. Each LSTM unit contains input, forget, and output gates to check which fragment of information to remember, pass and forget to the next step. In this way, The LSTM gains a capability to make decisions about what to store, and when to permit reads, writes and deletions by via gates that pass or block information through a LSTM unit.

3.3. Word Embedding Models

Word embeddings are intense vector representation of words that discover similarity of semantics or relationships among words which appear in nearby locations in a text. The relationships are unsupervisedly learned from a collection of documents by using neural networks. Word2vec, Glove and FastText are used as three of the most popular word embedding models.

Word2Vec: Word2vec is probably the most embedding important model that started a new trend in distributed semantics. It is proposed by Mikolov et al. [29-31] by the introduction of a toolkit that enables trainings of word embeddings from raw text and the use of pre-trained embeddings. By using neural networks, the toolkit enables learned representation of words as dense vectors which encode many linguistic regularities and patterns among words. Mikolov et al presented two different approaches for learning word embeddings: Continuous bag-of-words (CBOW) and Skip-gram models. In Skip-gram approach that is used in this work, the objective is to estimate the surrounded words of a dedicated word in a sentence or document. On the other hand, a target word is forecasted by surrounded words in CBOW model.

GloVe: Glove, signifies "Global Vectors" is another word embedding method proposed by Pennington et al. [32]. Word2vec method learns semantics of words by passing a local content window over the training data line-by-line to predict a word from its surroundings or surroundings of a given word. Pennington at al argue that local content window models are enough to extract semantics among words and do not exploit the count-based statistical information regarding word co-occurrences. Local content window and count-based matrix factorization approaches are consolidated in GloVe to acquire a better representation. Glove uses matrix factorization to get an accumulative global co-occurrence statistics of word-word from a dataset.

FastText: Bojanowski et al. [33-34] proposed FastText, a word embedding method enhanced with subword (character n-gram) embeddings. While GloVe and Word2Vec uses entire words as the smallest component, FastText employs character of n-grams to represent words as the smallest element. Each word is divided into character n-grams where $3 \leq n \leq 6$. For example, with $n=3$, the word "there" is divided into <th, the, her, ere, and re>, plus <there>. FastText also utilizes the skip-gram approach with negative sampling proposed for Word2Vec with a modified skip-gram loss function.

4. Experiments

In this chapter, experiment setup and results are presented in details.

4.1. Experiment Setup

In this study, the extensive experiments are performed to understand direction prediction in Bitcoin using word embedding models and deep learning algorithms. Accuracy is appraised as an evaluation metric in the experiments to display the success of each model and the contribution of our study. Experiments are carried out by varying the training set sizes as 5%, 10%, 30%, 50%, 80%. The percentages are represented with the prefix “ts” in order to forbid confusion with percentage of accuracies. The broadly-applied 5x2 cross validation method that is also performed in the literature studies [35-43] is implemented. Abbreviations are used for the preprocessing methods, and deep learning algorithms as follows: TC: Tweet cleaning that includes removal of hashtags, URL’s, and punctuations, WT: Word tokenization, SC: Spell correction, STM: Stemming, SWE: Stop-word elimination, AOT: All of them, CNN: Convolutional neural network, RNN: Recurrent neural network, LSTM: Long short-term memory network. The best accuracy results are obtained is indicated with boldface. For all processes including crawler, pre-processing steps, the implementation of word embedding models and deep learning algorithms with Keras deep learning library, Python programming language is employed.

4.2. Experiment Results

First, we analyze the impact of pre-processing methods on deep learning models and word embedding methods in Table 1. For all pre-processing method, FastText exhibits the best classification success at ts80. It is followed by LSTM algorithm with almost 2%-3% less classification performance. The performance order can generally be summarized as: FastText > LSTM > CNN > RNN ~ = GloVe > Word2Vec. Moreover, the usage of TC as a pre-processing model boosts the classification performance of the system compared to the other pre-processing models. On the other hand, it is clearly seen that WT and SWE demonstrate the similar performances to TC. For Word2vec model, TC performs 75.04% of accuracy when WT and SWE exhibit 74.80% of accuracy and 74.17% of accuracy, respectively. Thus, we consolidate these three pre-processing models, namely TC, WT, and SWE, in the proposed system. The combination of all pre-processing models (AOT) affects the success of the system negatively with approximately 4%-5% decrease in accuracy.

Table 1. The classification accuracies of deep learning models and word embedding methods in terms of pre-processing methods at ts80.

Models	Pre-processing Methods					
	TC	WT	SC	STM	SWE	AOT
Word2Vec	75.04	74.80	70.15	72.80	74.17	71.10
GloVe	76.35	75.48	70.42	73.15	76.05	70.00
FastText	84.27	83.65	78.27	81.36	83.96	79.08
CNN	78.80	78.06	73.60	76.91	77.58	74.35
RNN	76.15	75.70	71.34	73.11	76.09	70.27
LSTM	81.53	81.23	75.55	78.65	80.85	76.22

Second, we concentrate on the effectiveness of different training set percentages on word embedding models and deep learning algorithms when the combination of pre-preprocessing models is set to TC+WT+SWE in Table 2. At ts80, each word embedding

Table 2. The classification accuracies of deep learning models and word embedding methods at each training set percentage.

Models	Training Set Percentages				
	80	50	30	10	5
Word2Vec	81.95	78.17	64.09	59.20	46.14
GloVe	82.01	80.42	65.28	59.65	49.29
FastText	89.13	87.25	71.36	65.10	52.63
CNN	84.30	82.09	69.80	62.81	49.78
RNN	83.77	80.96	67.44	59.05	46.55
LSTM	87.45	85.78	70.95	65.51	53.00

model and deep learning algorithm demonstrates the best performance compared to the other training set sizes while the success of the system reflects the worst accuracy values for all models at ts5. The success of FastText is clearly observed in higher training set percentages as ts80, ts50, and ts30 while the classification success of LSTM exceeds FastText when training set sizes are set to 10 and 5. As a result, the use of LSTM instead of FastText is observed to be nearly 1% better at lower learning percentages in terms of system performance.

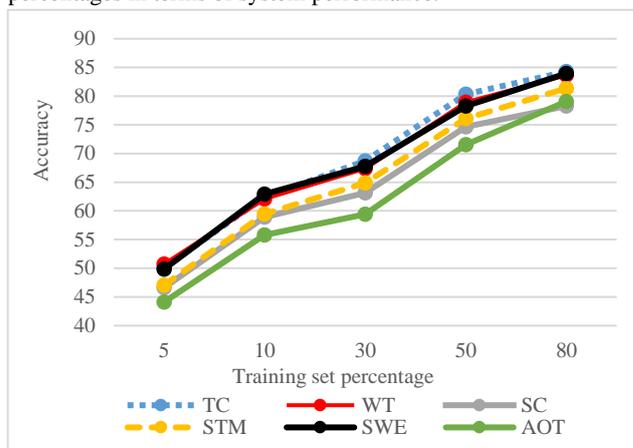


Fig. 2. The classification performances of pre-processing methods in terms of training set percentages when FastText is set to system.

In Fig.2, the classification success of each pre-processing method is presented in terms of training set sizes. These results are obtained when FastText is chosen as word embedding model because of its superior performance as shown in Table 1. It is clearly observed that TC, WT, and SWE exhibits similar classification performances especially at lower training set levels. Between ts50 and ts80, unlike the previous pattern observed in ts5, ts10, and ts30, TC model present better classification success with nearly 1% enhancement compared to WT and SWE. Due to these similar performances, TC, WT, and SWE methods are consolidated for the purpose of improving system performance. Moreover, the combination of each pre-processing model which is abbreviated as AOT exhibits the worst accuracy values at each training set levels. The selection of AOT is not suitable because of poor classification success of it. Furthermore, the performance of each model is ordered at ts80, ts50, and ts30 as: FastText> LSTM> CNN> RNN> GloVe> Word2Vec. On the other hand, this order is changed at ts10 and ts5 as: LSTM> FastText> CNN> GloVe> RNN> Word2Vec. As a consequence of Table 2, we perform all experiments at ts80 because of the superior performance of it for each model.

In Table 3, accuracy, f-measure, precision, and recall results are given for each model to predict the direction of Bitcoin. The experiment results show that the proposed system is able to predict the direction of Bitcoin in USD with mean accuracy of 84.77 % by evaluating sentiment analysis of both individual and organization

accounts.

It is hard to perform a fair comparison the success of proposed system with the literature works due to the deficiency of studies with the same datasets as textual data especially from social media platforms, word embedding models, and deep learning approaches. The comparison of experiment results between our system and the state-of-art studies is performed in terms of evaluation metrics considering the usage of different datasets in Table 4.

Table 3. Prediction results with different evaluation metrics.

Models	Evaluation Metrics			
	Accuracy	F-measure	Precision	Recall
Word2Vec	81.95	76.27	76.10	85.18
GloVe	82.01	76.96	77.85	86.44
FastText	89.13	84.06	86.97	90.28
CNN	84.30	78.05	79.36	88.55
RNN	83.77	77.32	78.15	87.30
LSTM	87.45	80.60	84.39	89.05

In study [6], author proposes to forecast the Bitcoin price by exploring the effects of machine learning models. The dataset is comprised from numeric data which is gathered from the Index of Bitcoin Price. To estimate the Bitcoin price, three models, namely LSTM, RNN, and ARIMA are compared in terms of classification success. They report that LSTM performs better classification accuracy with 2% improvement compared to the others. In our study, sentiment analysis based Bitcoin prediction system with FastText model performs 89.13% accuracy result. Moreover, LSTM in our model outperforms study [6] with 87.45 accuracy value.

Table 4. Comparison with state-of-the-art studies on the direction prediction of Bitcoin.

Study	Model	Accuracy	F-measure	Precision	Recall
[6]	LSTM	52.78	35.50	-	-
[7]	Ensemble	63.00	-	-	-
[44]	VADER	83.33	80.00	100	88.88
Our study	FastText	89.13	84.06	86.97	90.28

In another study [7], Sin et al. focus on Bitcoin price prediction using genetic algorithm based selective neural network ensemble on numeric price data. The ensemble system in [7] demonstrates 63.00% classification success while even son our worst model exhibits 81.95% accuracy value. It is meaningful to compare the study [44] with our work because of the usage of sentiment analysis. In [44], Stenqvist and Lönnö concentrate on to predict the fluctuation of Bitcoin prices by analyzing sentiments of tweets. For this purpose, VADER which is a software that consolidates sentiment analysis in terms of both lexicon and rule-based, is employed to detect polarity and the density of sentiments in the text. Although this study is parallel to our study in terms of the usage of sentiment analysis, our model exhibits 6% better classification performance by using preprocessing methods, deep learning models, and word embedding methods.

5. Conclusions

In this work, we propose sentiment analysis based direction prediction in Bitcoin using deep learning models, and word embedding methods. To our knowledge, this is the very first attempt which forecasts the direction of Bitcoin price in USD by investigating the impact of word embedding models in addition to the usage of pre-processing methods, and deep learning architectures in the state-of-the-art studies. To predict the direction of Bitcoin in USD, long-short term memory networks (LSTMs),

recurrent neural networks (RNNs), and convolutional neural networks (CNNs) are used as deep learning architectures and Word2Vec, GloVe, and FastText are employed as word embedding models in the experiments. In order to demonstrate the contribution of our study, experiments are carried out on English Twitter dataset. Experiment results show that the usage of FastText model as a word embedding model outperforms others with 89.13% accuracy result to forecast the direction of Bitcoin price in USD. In the future, we plan to improve a hybrid model includes both textual and numeric datasets for the purpose of empowering the Bitcoin direction prediction system.

References

- [1] V. M. Prieto, S. Matos, M. Alvarez, F. Cacheda, and J. L. Oliveira, "Twitter: a good place to detect health conditions", PLoS one, vol. 9, no. 1., pp. 1-11, Jan. 2014.
- [2] HILEMAN, Garrick; RAUCHS, Michel. Global cryptocurrency benchmarking study. Cambridge Centre for Alternative Finance, 2017, 33.
- [3] Madan, Isaac, Shaurya Saluja, and Aojia Zhao. "Automated bitcoin trading via machine learning algorithms." URL: <http://cs229.stanford.edu/proj2014/Isaac%20Madan20> (2015).
- [4] SHAH, Devavrat; ZHANG, Kang. Bayesian regression and Bitcoin. In: 2014 52nd annual Allerton conference on communication, control, and computing (Allerton). IEEE, 2014. p. 409-414.
- [5] JANG, Huisu; LEE, Jaewook. An empirical study on modeling and prediction of bitcoin prices with bayesian neural networks based on blockchain information. Ieee Access, 2017, 6: 5427-5437.
- [6] MCNALLY, Sean; ROCHE, Jason; CATON, Simon. Predicting the price of Bitcoin using Machine Learning. In: 2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP). IEEE, 2018. p. 339-343.
- [7] SIN, Edwin; WANG, Lipo. Bitcoin price prediction using ensembles of neural networks. In: 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD). IEEE, 2017. p. 666-671.
- [8] MATTA, Martina; LUNESU, Ilaria; MARCHESI, Michele. Bitcoin Spread Prediction Using Social and Web Search Media. In: UMAP Workshops. 2015. p. 1-10.
- [9] GARCIA, David; SCHWEITZER, Frank. Social signals and algorithmic trading of Bitcoin. Royal Society open science, 2015, 2.9: 150288.
- [10] S. Galeshchuk, O. Vasylyshyn and A. Krysovaty, "Bitcoin Response to Twitter Sentiments", in Proc. ICTERI, Kyiv, Ukraine, 2018, pp. 160-168.
- [11] J. B. Ramos. "¿Podemos comerciar Bitcoin usando análisis de sentimiento sobre Twitter?", Trabajo Fin de Grado, Universidad Pontificia de Comillas, Madrid, Spain, 2019.
- [12] A. Aggarwal, I. Gupta, N. Garg and A. Goel, "Deep Learning Approach to Determine the Impact of Socio Economic Factors on Bitcoin Price Prediction," 2019 Twelfth International Conference on Contemporary Computing (IC3), Noida, India, 2019, pp. 1-5, doi: 10.1109/IC3.2019.8844928.
- [13] Loria, S. (2018). Textblob Documentation (pp. 1-73). Technical report.
- [14] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, L. Fei-Fei, "Large-scale Video Classification with Convolutional Neural Networks", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2014.
- [15] A. Karpathy, L. Fei-Fei, "Deep Visual-Semantic Alignments for Generating Image Descriptions", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2015, pp. 3128-3137.
- [16] Andrej Karpathy, Connecting Images and Natural Language, PhD

- thesis, Stanford University, 2016.
- [17] Jiajun Sun, Jing Wang, Ting-chun Yeh, Video Understanding: From Video Classification to Captioning, Stanford University, 2017.
- [18] Y. LeCun, Y. Bengio, G. Hinton, "Deep learning", 2015.
- [19] B. Ginzburg, "Introduction: Convolutional Neural Networks for Visual Recognition", Intel, 2013
- [20] Kilimci, Z. H., & Akyokus, S. (2018). Deep Learning-and Word Embedding-Based Heterogeneous Classifier Ensembles for Text Classification. Complexity, 2018.
- [21] Zuxuan Wu, Ting Yao, Yanwei Fu, Yu-Gang Jiang, Deep Learning for Video Classification and Captioning, Fudan University, Microsoft Research Asia, University of Maryland, 2016.
- [22] Zuxuan Wu, Ting Yao, Yanwei Fu, Yu-Gang Jiang, Deep Learning for Video Classification and Captioning, University of Maryland, College Park, Microsoft Research Asia, Fudan University, 2018.
- [23] Z. C. Lipton, J. Berkowitz, and C. Elkan, "A Critical Review of Recurrent Neural Networks for Sequence Learning," May 2015.
- [24] J. L. Elman, "Finding structure in time," Cogn. Sci., 1990.
- [25] Z. C. Lipton, J. Berkowitz, and C. Elkan, "A Critical Review of Recurrent Neural Networks for Sequence Learning," May 2015.
- [26] J. L. Elman, "Finding structure in time," Cogn. Sci., 1990.
- [27] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A Search Space Odyssey," IEEE Trans. Neural Networks Learn. Syst., 2017.
- [28] D. Kent and F. M. Salem, "Performance of Three Slim Variants of The Long Short-Term Memory {(LSTM)} Layer," CoRR, vol. abs/1901.00525, 2019.
- [29] Q. V. Le and T. Mikolov, "Distributed Representations of Sentences and Documents," vol. 32, 2014.
- [30] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," pp. 1–12, 2013.
- [31] T. Mikolov et al., "Distributed Representations of Words and Phrases and their Compositionality arXiv : 1310 . 4546v1 [cs . CL] 16 Oct 2013," Adv. Neural Inf. Process. Syst., 2013.
- [32] J. Pennington, R. Socher, and C. Manning, "Glove: Global Vectors for Word Representation," in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014.
- [33] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of Tricks for Efficient Text Classification," 2016.
- [34] T. Mikolov, E. Grave, P. Bojanowski, C. Puhrsch, and A. Joulin, "Advances in Pre-Training Distributed Word Representations", arXiv:1712.09405 , 2017.
- [35] Kilimci ZH, Akyokus S. N-Gram Pattern Recognition using Multivariate Bernoulli Model with Smoothing Methods for Text Classification. 24th IEEE Signal Processing and Communications Applications Conference; 2016; Zonguldak, Turkey.
- [36] McCallum A, Nigam KA. Comparison of Event Models for Naive Bayes Text Classification. In: AAAI-98 Workshop on Learning for Text Categorization; 1998; Wisconsin, USA: pp. 41-48.
- [37] Kilimci, Z. H., & Omurca, S. I. (2018). The Impact of Enhanced Space Forests with Classifier Ensembles on Biomedical Dataset Classification. International Journal of Intelligent Systems and Applications in Engineering, 6(2), 144-150.
- [38] Rennie JDM, Shih L, Teevan J, Karger DR. Tackling the Poor Assumptions of Naive Bayes Text Classifiers. In: 20th International Conference on Machine Learning; 2003; Washington, USA: pp. 616-623.
- [39] Kilimci ZH, Ganiz MC. Evaluation of classification models for language processing. In: 10th International Symposium on INnovations in Intelligent SysTems and Applications; 2015; Madrid, Spain: pp. 1-8.
- [40] Amasyalı MF, Ersoy OK. Classifier Ensembles with the Extended Space Forest. IEEE Transactions on Knowledge and Data Engineering 2013; 26: 549-562.
- [41] Kilimci, Z. H., & Omurca, S. İ. (2018). Extended feature spaces based classifier ensembles for sentiment analysis of short texts. Information Technology and Control, Vol:47, no:3.
- [42] Adnan MN, Islam MZ, Kwan PWH. Extended Space Decision Tree. 13th International Conference on Machine Learning and Cybernetics; 2014; Lanzhou, China: pp. 219-230.
- [43] Kilimci, Z. H., & Akyokus, S. (2019, July). The Analysis of Text Categorization Represented With Word Embeddings Using Homogeneous Classifiers. In 2019 IEEE International Symposium on INnovations in Intelligent SysTems and Applications (INISTA) (pp. 1-6). IEEE.
- [44] Stenqvist, Evita, and Jacob Lönnö. "Predicting Bitcoin price fluctuation with Twitter sentiment analysis.", INDEGREE PROJECT TECHNOLOGY, FIRST CYCLE, 15 CREDITS, STOCKHOLM SWEDEN, 2017.