Forecasting Cryptocurrency Value by Sentiment Analysis: An HPC-Oriented Survey of the State-of-the-Art in the Cloud Era

Aleš Zamuda[1](✉), Vincenzo Crescimanna[2], Juan C. Burguillo[3], Joana Matos Dias[4], Katarzyna Wegrzyn-Wolska[5], Imen Rached[5], Horacio González-Vélez[6], Roman Senkerik[7], Claudia Pop[8], Tudor Cioara[8], Ioan Salomie[8], and Andrea Bracciali[2]

[1] Faculty of Electrical Engineering and Computer Science, University of Maribor, Koroška cesta 46, 2000 Maribor, Slovenia
ales.zamuda@um.si

[2] Department of Computing Science and Mathematics, University of Stirling, Stirling FK9 4LA, UK
vincenzo.crescimanna1@stir.ac.uk,abb@cs.stir.ac.uk

[3] Department of Telematic Engineering, University of Vigo, Campus Universitario, 36310 Vigo, Spain
j.c.burguillo@uvigo.es

[4] Faculty of Economics, University of Coimbra, Av. Dias da Silva 165, 3004-512 Coimbra, Portugal
joana@fe.uc.pt

[5] Allianstic Research Laboratory, EFREI Paris, Ecole d'ingénieurs généraliste Informatique et technologies du numérique, 30-32 Avenue de la République, 94800 Villejuif, France
{katarzyna.wegrzyn-wolska,imen.rached}@efrei.fr

[6] Cloud Competency Centre, National College of Ireland, Mayor Street Lower – IFSC, Dublin 1, Ireland
horacio@ncirl.ie

[7] Faculty of Applied Informatics, Tomas Bata University in Zlin, T. G. Masaryka 5555, 760 01 Zlin, Czech Republic
senkerik@utb.cz

[8] Computer Science Department, Technical University of Cluj-Napoca, 26-28 Baritiu Street, 400027 Cluj-Napoca, Romania
{claudia.pop,tudor.cioara,ioan.salomie}@cs.utcluj.ro

Abstract. This chapter surveys the state-of-the-art in forecasting cryptocurrency value by Sentiment Analysis. Key compounding perspectives of current challenges are addressed, including blockchains, data collection, annotation, and filtering, and sentiment analysis metrics using data streams and cloud platforms. We have explored the domain based on this problem-solving metric perspective, i.e., as technical analysis, forecasting, and estimation using a standardized ledger-based technology. The envisioned tools based on forecasting are then suggested, i.e., ranking

J. Kołodziej and H. González-Vélez (Eds.): cHiPSet, LNCS 11400, pp. 325–349, 2019.
https://doi.org/10.1007/978-3-030-16272-6_12

Initial Coin Offering (ICO) values for incoming cryptocurrencies, trading strategies employing the new Sentiment Analysis metrics, and risk aversion in cryptocurrencies trading through a multi-objective portfolio selection. Our perspective is rationalized on the perspective on elastic demand of computational resources for cloud infrastructures.

Keywords: Cryptocurrency · Blockchain · Sentiment Analysis · Forecasting · ICO · CSAI · Cloud computing

1 Introduction

This chapter presents a position survey on the overall objective and specific challenges encompassing the state of the art in forecasting cryptocurrency value by Sentiment Analysis. The compounding perspectives of current challenges are addressed, such as the blockchain technologies, underlying data collection of items from social media, the annotation, and filtering of such items, and the Sentiment Analysis as a resulting metric of the observed data streams.

We describe a Cryptocurrency Sentiment Analysis Indicator (CSAI) and identify its required inner workings compounding perspectives. Further possibilities are then explored, based on this new metric perspective, such as technical analysis, forecasting, and beyond. The envisioned tools based on forecasting are then suggested, i.e., ranking Initial Coin Offering (ICO) values for incoming cryptocurrencies, trading strategies employing the new Sentiment Analysis metrics, and risk aversion in cryptocurrencies' trading through multi-objective portfolio selection. Since the introduction of Bitcoin [94] and rise of blockchain-related algorithms [96] and technologies [10,74,98,130,137], there has been a significant increase in their recognition and analysis. In this chapter, we focus on the specific aspect of value for the blockchain projects related to cryptocurrencies.

While High-Performance Computing (HPC) and Cloud Computing are not *sine qua non* for cryptocurrencies, their use has become pervasive in their transaction verification ("mining"). Cryptocurrencies rely on powerful computational nodes to verify transactions, convert them into groups "blocks", and add them to the blockchain. Such verification is based on well-established complex cryptology algorithms [70] (ergo, the term cryptocurrencies) which assure user anonymity and payment untraceability. With the convergence of HPC and clouds [102], elastic computational resource utilization has become commonplace in different domains, including of course, Cryptocurrencies. Moreover, a recent survey on open challenges and trends in cloud computing [18] recognizes blockchain as a disruptive influencer of the field.

In the following section, related work is provided, introducing sentiment analysis, cryptocurrencies, and their value forecasting. Section 3 highlights for the interested reader, more about cryptocurrencies and blockchain, in general, and lists an example blockchain application for energy markets. After Sect. 3, specific challenges are addressed, listing the specific perspectives. Section 4 covers influence pertaining to social media, Sect. 5, social media data annotation and sentiment dictionary, Sect. 6, Filtering of Tweets, Sect. 7, the perspective on core resulting Sentiment Analysis for cryptocurrencies, Sect. 8, technical analysis, Sect. 9, ranking of ICOs, Sect. 10, portfolio selection using multi-objective optimization, and Sect. 11, investment approaches. Then, the Conclusion section summarizes the surveyed perspectives.

2 Related Work

As this chapter focuses on the objective of Sentiment Analysis for cryptocurrencies, the more recent related work leading to the formation of this objective is initially presented in this section. Namely, when analyzing the impact of cryptocurrencies, there are different possibilities on which to focus, like prediction of value or some other underlying principles and features of technologies enabling these cryptocurrencies.

We first focus on the Bitcoin cryptocurrency and the trading volumes of Bitcoin as introduced using Google Search Engine as the media feed [87]. In [126], search engine query trends are predicted for Bitcoin. A connection between Bitcoin search queries on Google Trends and Wikipedia is established in [68], where their relationship is also studied. An attempt to explain Bitcoin prices and adoption rates using Google Search is made in [109]. Returns and volatility perspective when using Bitcoin volume analysis is discussed in [9]. A text mining from an online forum which analyzes user opinions and aims to predict value fluctuations for Bitcoin is reported in [67]. A survey of methods and models for textual sentiment in finance is presented in [63]. Dynamic topic modeling for cryptocurrency community forums is introduced in [76]. A crowdsourced perspective is presented in [80], listing future directions in international financial integration research. For Bitcoin value formation, a model based on an empirical study is given in [57]. In [101], high frequency volatility is forecasted for cryptocurrencies. Some of the underlying determinants, including technology and economic factors influencing exchange rates for cryptocurrencies, are presented for the case of Bitcoin in [73]. A cryptocurrency price prediction using news and social media sentiment was first introduced in [69], followed shortly by the predicting of cryptocurrency price bubbles using social media data and epidemic modeling in [104]. The popular modern techniques from Artificial Intelligence and general softcomputing paradigms have been utilized in recent work [89], where price prediction based on historical price values was compared over machine learning with different types of neural networks, and further, in [72], where the sentiment-based prediction of alternative cryptocurrency price fluctuations using a gradient boosting tree model is given. In [105], a mutual-excitation of cryptocurrency market returns and social media topics is investigated, whereas the paper [84] reports a recent study considering the impact of social media on Bitcoin value. Finally, a study measuring the interaction between media sentiment, based on news articles as well as blog posts and the Bitcoin price, is given in [136]. Another broader time-series analysis used to study the general relationships between Bitcoin prices and fundamental economic variables, technological factors and measurements of collective mood derived from Twitter feeds was presented in [51].

An example type of media data feeds' content influencing cryptocurrency sentiment are social media channels' posts about ransomware and email-based hacker attacks, mostly influencing the sentiment as negative sentiment posts. The cybersecurity threads' (i.e., computer malware, ransomware, virus, worm, trojan horse, retrovirus, botnet, etc.) [129] development and their impact on society and financial markets are a significant topic today. Hackers enjoy anonymity under blockchain technology, that may have high social cost [82]. For example, the paper [6] shows that a substantial amount of Bitcoin blockchain operations and addresses are involved in ransomware money processing. Ransomware represents a type of software cyber-attack, where the hackers take advantage of victims' operating system vulnerabilities to deploy dedicated ransomware code. Once it has happened, the major harmful activity lies in the encryption of files with certain extensions that are expected to be highly personally important (documents, photos, etc.). The victim still can access (boot) the operating system, usually getting a message specifying the amount of ransom in cryptocurrency, the hacker's wallet address and the time left to pay the ransom in exchange to get the decryption key. On the other hand, the email-based hacker attacks are more frequent and easier to perform. The victim usually receives an email stating some mix of technical info and blackmailing based description of a situation that the hacker takes over the control of web browser, email, and social networks applications remotely, and will send some private information to all gathered contacts, unless a certain amount in cryptocurrency is paid within a given time slot. Both of the cybersecurity threads mentioned above are influencing the blockchain value and should be taken into consideration in any future developed forecasting sentiment aggregator to achieve better accuracy. The reasons and relations are given in an original work [122], where the author investigates how the cybersecurity shocks affect the demand for blockchain settlement, transaction fees, mining reward, and cryptocurrency exchanges. That paper also explores in detail the theory given in [44], and their findings show that sudden shocks to the exogenous demand for blockchain settlement are resulting in an increase in transaction fees and a reduction in the endogenous demand [122].

Following the ongoing research and published works listed in this section, the perspectives compounding the presented research topic of this chapter are identified in Sects. 4–11 as specific challenges in sentiment analysis for cryptocurrency value forecasting.

3 Background: Cryptocurrencies and Blockchain

Widely considered immutable time-stamped data structures, blockchains implement peer-to-peer networks where participants can verify interactions concurrently using decentralized peer-to-peer consensus protocols. As an emerging technology trend, different research [115] and industrial [29] perspectives are being assembled to document its potential disruptive impact [22].

Blockchains have five unique characteristics, namely:

1. Peer-to-peer communication without a central authority.
2. Transparent transaction processing with optionally-disclosed ownership.
3. Decentralized transaction history verifiable by all participants.
4. Immutability of records assuring chronological sequence and accessibility.
5. Logic-based processing to trigger algorithms and events.

The aforementioned characteristics have made blockchain particularly suitable to manage *cryptocurrencies*: Electronic cash systems administered via peer-to-peer consensus. Indeed, the most widely known for cryptocurrency, the Bitcoin [94], remains something like the gold Standard for financial blockchain applications. Nonetheless, while blockchains have been used extensively in financial entities, their decentralized immutability characteristics have made them particularly suitable for applications in other domains as diverse as Law [92], Food Traceability [50], and Open-source Software Management [42]. To highlight the importance of blockchain technologies evaluation, an example in surrounding blockchain technologies for Energy markets follows in the next subsection.

3.1 Blockchain Technologies for Energy Markets

The energy grid is moving to a new era, shifting from centralized broadcast-like energy systems to decentralized smart energy systems by incorporating a large number of small-scale Distributed Energy Prosumers (DEP). The advent of intermittent decentralized renewable energy sources is completely changing the way in which electricity grids are managed, supporting the shift to more decentralized smart energy systems. Variations in energy production, either surplus or deficit, may threaten the security of supply, leading to energy distribution systems' overload, and culminating in power outages or service disruptions, forcing the DEPs to shed or shift their energy demand to deal with peak load periods [34,108,131].

The centralized energy systems take limited account of local conditions and are difficult to be optimized, and offer no incentives for consumers to manage and adjust their consumption according the generation profiles. In this context, the H2020 eDREAM project [45] contributes to the transformation of the traditional energy systems into novel decentralized and community-driven ones, by leveraging on blockchain technology to exploit local capacities and constraints fully at micro-grid level to preserve the continuity and security of supply at affordable costs at smart grid level. The grid is modeled as a collection of DEPs' resources able to coordinate through a blockchain based infrastructure to support fully decentralized management, control and stable grid operation. The eDREAM project presents a blockchain decentralized management relying on the implementation of [34]:

- A distributed ledger for storing DEPs' energy data at the micro-grid level. All monitored energy data recorded at the level of a DEP are registered and stored as immutable transactions. Therefore, the individual energy production or

energy consumption values are aggregated in blocks which are then replicated in the ledger.
– A set of self-enforcing smart contracts for decentralized energy management and control. Through specific mechanisms, these contracts enable the peer-to-peer-trading of energy among DEPs and offer-demand matching and decentralized coordinated control for energy stakeholders, such as the DSO (Distribution System Operator). The contracts are able to assess and trace the share of contracted flexibility services actually activated in real-time by the aggregators (from their enrolled prosumers).
– Consensus based validation for transactions' validation and financial settlement. The project offers the solution of a novel blockchain-based validation that goes in the direction of increased reliability of the smart grid system operation, better energy incentives for DEPs, and increased usage of renewable energy.

Three types of smart management scenarios are supported with self-enforcing smart contracts [33]: (i) The provisioning of energy flexibility services by the DSO leveraging on aggregators, (ii) The implementation of a decentralized green energy market at the micro-grid level promoting the consumption of renewable energy where it is produced, and (iii) The creation of a community based coalition of prosumers, allowing them to participate in the national energy or capacity market.

The Provisioning of Flexibility Services supposes that prosumers are able to offer and trade their flexibility in terms of loads' modulation. They are involved via enabling aggregators, or directly with the DSO via direct Demand Response and control of DEP's energy assets. Using smart contracts, the DSO is able to assess and trace the share of contracted flexibility services, actually activated in real-time by the aggregators (from their enrolled prosumers) at the grid level. At the same time, the self-enforcing smart contracts act as a decentralized control mechanism used to manage the levels of energy flexibility from aggregators and enrolled prosumers on one side, and from aggregators to the DSO on the other side, associating incentive and penalties' rates. If relevant deviations between the expected energy flexibility request and the actual delivered flexibility are detected by smart contracts, specific actions are taken to rebalance the energy demand with the energy production.

The Decentralized Green Energy Market designed at the micro-grid level enacts any small-scale prosumer to participate and trade energy directly. The market acts as a management mechanism by rewarding the consumption of renewable energy when it is available, leveraging on green energy tokens, making sure that the potential energy imbalances at the micro-grid level are addressed locally and not exported to higher smart grid management levels. The non-fungible tokens are generated at a rate proportional with the forecast renewable energy production [103], transforming the energy in a transactable digital asset. The producers and consumers use the generated tokens to participate in the electricity market sessions and leverage on self-enforcing smart contracts to submit energy bids/offers and transact energy in a peer-to-peer fashion.

The Creation of Virtual Power Plants (VPP) addresses the increasing need to optimize the output from multiple local generation assets (i.e. wind-turbines, small hydro, photovoltaic, back-up generators, etc.) that serve primarily local communities and have export connections at the power distribution network. The benefits behind creating such coalitions are that a mix of different energy generation resources, which have different energy generation models and scale, may be interested to cooperate in a convenient way, with a view to achieving pre-defined smart grid sustainability objectives. The VPP aims at maximizing the utilization and revenues from RES and classic generation sources through accessing different markets as an aggregated portfolio, bringing its capacity to the optimum paying service at any time. The DEPs can ultimately participate on other higher-level energy markets, such as a flexibility service provider to a TSO (Transmission System Operator) or a wholesale capacity provider on the wholesale or capacity market.

The adoption of the above presented blockchain based management approaches will transform the smart grid into a democratic community that no longer relies on a central authority, but can take any decision through smart contracts rules, enforced and verified by each DEP of the grid. At the same time, it is in line with the European strategic vision of creating a secure and sustainable energy system by 2050, and reducing the greenhouse gas emissions by at least 80% [46].

4 Specific Challenge A: Influence on Social Media

Social networks are enabling people to interact, and are ever-changing their human relations to the virtual world. People utilize these media platforms for different activities; to express their opinions and their sentiments, to share their experiences, to react to another person. We can observe in this space different human interactions, and we can define many roles related to relations between different users.

We can observe, among others, influential, trusted, or popular individuals. These roles of users in social networks are significant and useful in various disciplines, such as Economy, Finance, Marketing, political and social campaigns, and recommendations. Defining, distinguishing, and measuring the strength of those relations becomes challenging, both in the theoretical and practical fields. The roles of trusted or influential users, users with high reputation, and popular users, can be used in various ways. An interesting work to define, classify, and present a hierarchy of all the terms was done by Rakoczy [111]. In this survey, we are interested in particular in *influence* approach, the relation which, as we suggest, has the strongest correlation between social media and the cryptocurrency tendency, and, in general, with the financial trends. The research involving influence and influential users is an important part of social network analysis.

The term *influence* is used widely and intuitively means the capacity to influence the character, development, or behavior of someone or something, or the effect itself. We name influencer the person exerting the influence action.

Although this intuitive meaning is well understood, in social network analysis, *influence* seems to be an ambiguous term that depends strictly on the presumed assumptions. For instance, Kempe et al. [64] focused on influence in the information flow and spread sense. On the other hand, other works, e.g. [56] and [140], targeted the quantity aspect of influence, targeting mainly users in central positions in a social network. The analysis of influence is an interdisciplinary domain, involving not only social network analysis, but also social sciences, e.g. Sociology, use of graph theory, statistics, and others [100].

The existing state-of-the-art methods considering influence regard this topic in two different approaches, which are: (i) Influential users' discovery and (ii) Influence diffusion information spread within the network with particular focus on the most optimized way to diffuse the information in a social network maximally.

4.1 Influential Users' Discovery

The methods for influential users discovery try to find users who have an impact on the network and the users who, in some way (structurally, by modifying the behavior, adding information to network, etc.), try to answer the following question: Which users are in such a position and are making such a difference to other users that the structure of the network, behavior, actions, or preferences of other users is changed because of this influential activity? Here, we can find two approaches: Based only on the central position of the user (topology-based approaches: Degree centrality, closeness centrality, and betweenness centrality [142]), or more complex, which takes into account more aspects, like a user's history, content, and other influence properties. Many methods which used topology criteria are based on PageRank [23,56,128,138].

Alternatively, there are also some works that have used additional information provided directly from users about the topic, namely hashtags. RetweetRank and MentionRank [140] are examples of such approaches that are using hashtag information in order to group users together via subject. There have been approaches to provision cloud resources elastically, based on social media, potentially for disaster situations [121]. These methods are similar, using as a base for the network either retweets or mentions. Additionally to the works presented above, there are obviously other works that also deal with the influence evaluation. In the famous work of Hirsh [58], the author presents a metric that aims to calculate the influence of researches, based on individual's publications.

4.2 Influence Diffusion and Influence Maximization

The issue of influence diffusion is a general problem of how the information spreads within a network, where users (nodes) are connected by edges which signify the influence. The best known methods are the *Linear Threshold* (LT) model [64], the modified LT named Delayed LT (DLT) model [93], or *Independent Cascade* (IC) model [64].

Influence maximization is a particular subproblem originating directly from the influence diffusion problem, that is particularly interesting and studied widely. Influence maximization is searching the answer to the following question: Which users to target for spreading some information in order to have maximum possible users in a social network talking/knowing about this information? Hence, such methods aim to find users who will share the information most widely (propagate it further). On the contrary to approaches for the influential users' discovery, these methods are connected strictly with diffusion of information.

4.3 General Model

A general and simple model for for evaluating the influence between a user's network platform is *Action-Reaction-Influence-Model* (ARIM). It is based on the users' proactive and reactive behaviors, that can be found on basically any social networking site, and can be used with different data sets. A few interesting approaches based on ARIM were proposed: e.g., method to calculate influence in time-dependent citation networks [113], and an algorithm to predict the reputation using the known influence value [114]. Finally, a very interesting and new approach of micro-influencer was defined [112].

5 Specific Challenge B: Social Media Feeds' Annotation and Dictionary Definition

A market sentiment dictionary for financial social media data applications is presented in [27], and then a fine-grained analysis of financial Tweets is described in [26]. The dictionary in [27] with 8,331 words, 112 hashtags, and 115 emojis is available publicly at http://nlg.csie.ntu.edu.tw/nlpresource/NTUSD-Fin/ under CC BY-NC-SA 4.0 license. It is built from posts in the StockTwits dataset [71], at that time providing 334,798 labeled posts from 13,059 users and crawled through StockTwits API (https://api.stocktwits.com/developers/docs). The dictionary stores unstemmed tokens appearing at least ten times and showning significant difference in chi-squared test (significance level 0.05) between expected and observed frequency. Also, stopwords, punctuations, digits, URLs, user ids, and tickers are removed from the input posts, while still specifically processing emojis. Input posts with less than two words are also removed. As the paper also analyses sentiment, it is discussed that based on this dictionary, the sentiment of investors may depend on the positions they hold (e.g. positive investor sentiment does not imply bullish market sentiment due to the target of the investor).

The FiQA 2018 Task 1 (https://sites.google.com/view/fiqa/home) is supported by Social Sentiment Index (SSIX) Horizon 2020 project (https://ssix-project.eu/), and it includes 675 training instances to predict continuous sentiment. In [26], utilizing the dictionary [27], the FiQA 2018 Task 1 is evaluated over different neural network models with Keras (https://github.com/keras-team/keras): Convolution Neural Network (CNN), Bidirectional Long Short-Term Memory (Bi-LSTM), and Convolution Recurrent Neural Network (CRNN).

6 Specific Challenge C: Filtering Tweets

As Sentiment Analysis (SA) usually works on top of Twitter feeds, it is necessary to pre-process the Tweets before SA, as described in the following. Preprocessing the data, as defined in [55], is the process of cleaning and preparing the text for classification. This phase has a relevant role in sentiment analysis tasks, indeed, reducing the redundancy and the noise intrinsic in the online data, allowing a better and fast classification [55]. The pre-processing method can be distinguished in two stages: *Transformation* and *filtering*. The first stage is quite standard, consisting in operations like white space removal, expanding abbreviation, stemming, stop words removal, and negation handling, while, instead, the second stage regards the choice and selection of the features, which, in Machine Learning, is called *feature selection* [145].

Such task can be described mathematically as the problem to find an optimal function $f : \mathcal{T} \to S$, where $\mathcal{T}$ is the (cleaned) text space and $S \subset \mathbf{R}^d$ the feature space; where the optimum is defined according to a certain metric that is defined a priori.

In a text mining context, the feature selection task is composed by three steps: Choice of the type of features, giving a weight to each feature, and selecting the relevant features [145]; i.e. scoring each potential feature according to a particular metric, and then taking the n best features. According to the mathematical description given above, the problem is equivalent to choosing the metric with which to evaluate the embedding map. We note that, differently from other fields like image or audio processing, when a combination of features does not have any meaning, then the type of features must be chosen by the user. In this section, we will focus on the third step: The possible methods to extract features specific for Sentiment Analysis, while the ranking of these features is addressed in Sect. 9. For a complete review of the possible type of features, we refer to [1] and [55], and for a description of the weight models and for a complete review on feature selection on text mining we refer to [25,49,54].

The most common metric in feature selection is the χ^2 (chi-squared), a measure expressing the divergence between the feature and the class. A drawback of this metric is that it works only when the dataset is big and there are no rare features. Another two popular metrics are the Accuracy, measuring the expected accuracy of a simple classifier built from the single feature, and the F_1, the harmonic mean of the precision and recall.

As observed in [49], all these metrics are equivalent, and one outperforms the others according to the dataset. In the review where twelve metrics are compared, Forman [49], proposed a new metric, the Bi-Normal Separation (BNS), an extension of the χ^2 where also the frequency of the appearance of the feature is considered. This metric is good in most common cases, but, when predictions move outside of this set, good results are not guaranteed. For example, in a highly skewed data context, the BNS does not work better and two information theoretic measures outperform it: Information Gain and $ITF \cdot IDF_2$, [77].

7 Specific Challenge D: Sentiment Analysis

Sentiment Analysis, also known as opinion mining, refers to the use of natural language, text analysis and computational linguistics to identify or extract subjective information from the attitude of a speaker/writer from a set of customer resources in order to classify the polarity. From the point of view of text mining, Sentiment Analysis is an automatic classification massive task as a function of the positive or negative emotions transmitted by the textual message.

In general, Sentiment Analysis tries to determine the opinion from a person with respect to a topic. Such opinion may involve an evaluation, the emotional state of the user at the time of writing, or the emotional communicative intention, i.e., how the customer tries to influence the reader or interlocutor.

Present approaches can be grouped into three main categories: Knowledge-based, statistical methods, and hybrid techniques [21].

- Knowledge-based techniques classify text into emotional categories based on the presence of unambiguous affective words like happy, sad, bored, etc. [97]. These methods also imply the use of lexical affinity to assign arbitrary words a certain affinity to particular emotions [125].
- Statistical methods take advantage of Machine Learning techniques like Support Vector Machines, mutual inference, semantic analysis, etc. [134]. More sophisticated methods try to detect the emotion and what is the target of such feeling [66]. Grammar dependent relations among words are usually applied to achieve such complex purpose, implying a deep grammar analysis of the message [41].
- Hybrid approaches leverage on both knowledge representation and Machine Learning techniques. These methods take advantage of knowledge representation models like ontologies and semantic nets, being able to extract implicit information [19].

On the one hand there are open source tools that use Machine Learning techniques, statistics, and natural language processing to automate the processing of huge amounts of data [41], including web pages, online news, discussion fora, social networks, web blogs, micro-blogging, etc. On the other hand, knowledge based systems use public access resources like SentiWordNet [7] or SenticNet [20] to extract semantic and affective information linked to natural language concepts. Sentiment Analysis can also be performed on visual content, i.e., images and videos (denoted as Multimodal sentiment analysis). One of the first approaches in this direction was SentiBank [14], utilizing an adjective noun pair representation of visual content.

In order to measure the precision and recall of a Sentiment Analysis tool, it is usually compared with human judgments about others' opinions [59]. However, humans when evaluated in the same way only show an 80% of precision on average, and this means that a program with a 75% of precision works almost as well as humans [99].

8 Specific Challenge E: Technical Analysis and Forecasting

SA has a wide range of applications [48], for example monitoring the review of consumer products, evaluating the popularity of public people or discovering a fraud; but, the SA can also be applied in prediction settings. Indeed, SA was used with good results to predict: Elections [132], football matches [110], stock prices [13,85,147], and cryptocurrency fluctuations [2,68,124]. In this survey, we are interested in the latter application, but in all the different contexts described in this paragraph, the prediction problem, from a mathematical point of view, can be described in the same way: Find a function $f : \mathcal{S} \rightarrow \mathbf{R}$, such that given the sentiment s_t (called the dependant or the explanatory variable), the associated value is the prediction (unknown a priori) of (the independent or response variable) x_{t+1}, i.e. $f(s_t) = x_{t+1}$. A simple linear regression model could be summarized by the equation:

$$x_{t+1} = \beta s_t + \epsilon,$$

where other unobserved factors determining x_{t+1} are captured by the error term ϵ. The dependent variable x_{t+1} in the textual sentiment finance literature is typically Bitcoin price and/or trade volume [2,37,51,61,67,136] or Bitcoin return and volatility, as in [30].

Logistic regression is used if the response variable is dichotomous. Here, the dependent variable is a binary variable, whose value is equal to 1 if a certain event has happened, and 0 otherwise. The idea is to examine if Sentiment Analysis is associated significantly with a certain event.

In almost all the examples listed above in this section, the models used to estimate the function f are restricted to the linear ones: Linear or logistic regression in case the features are correlated, or Naive Bayes when the dataset is relatively small [141]. The only exception is represented by [13], where the authors used Fuzzy Neural Networks, a hybrid system that combines the theories of Fuzzy Logic and Neural Networks, [95]. Before proceeding with the description of recent forecasting models for cryptocurrency, it is worth underlining two points: First, such simple models are used because the feature space S is, in general, low dimensional; indeed, more complex models like Support Vector Machines or Neural Network architectures suffer from overfitting. Second, although many works show the relationship between the market and the social sentiments [85,117,147] as highlighted by [148], if the financial instruments attract sufficient messages, in general, the sentiments do not lead financial markets in a statistically significant way. The latter observation can explain the reason why most of the work on cryptocurrency forecasting based on SA are tested with the well-known cryptocurrency Bitcoin, see e.g. [35], or with a specific cryptocurrency, like *ZClassic*, [72]. For financial time series analysis, in order to overcome the problem of the incomplete reliability of sentiments, it is usual to combine SA with auto-regressive models (see e.g. [53] and reference therein, with such approach being suggested also for cryptocurrency), indeed as shown in [12], where the traded volume is driven primarily by past returns. This idea considers multimodal architectures, and although it is suggested in many works [12,24], it is still not applied.

When forecasting cryptocurrencies' values, a key topic is to consider if they should be classed as currencies, assets, or investment vehicles. If they are traded for investment purposes, like hedging or pricing instruments, investigating the volatility of these cryptocurrencies becomes important, and could help others make better informed decisions in terms of portfolio and risk management. Specifically, cryptocurrencies' volatility levels are usually much higher than traditional currencies [39]. Volatility models were used in [139] to test the effects of sentiment or sentiment-related variables on the second moment of stock returns. GARCH-type models and the "realized volatility" approach were employed. A stock volatility prediction using Recurrent Neural Networks with SA was presented in [79]. In [62], the ability of several GARCH models to explain the Bitcoin price volatility was studied, and the optimal model was presented in terms of goodness-of-fit to the data. GARCH modeling of 7 other Cryptocurrencies was presented in [32]. A study on forecasting high frequency volatility for cryptocurrencies and traditional currencies with Support Vector Regression was presented in [101].

As another approach to forecasting, in [133], a fuzzy inference system for forecasting cryptocurrency price variation is developed, with trading rules being optimized through the use of a Genetic Algorithm (the use of Genetic Algorithms to determine trading rules has also been used before in Foreign Exchange markets, FX [90]). One of the advantages of this type of methodology compared with black box machine learning models, is its interpretability.

9 Specific Challenge F: Ranking ICOs

After the introduction of a new measurement indicator like CSAI, such introduction brings with it a new challenge of how to include the new indicator within existing schemes that fuse different indicators to then support decision-making.

In fusing the different data streams together, one approach is to assign weights or ranks to indicators. The Machine Learning algorithm designs operating on Big Data, such as stability selection [145], might be used on top of indicator values when tackling the ranking of ICOs. As stability selection also highlights outliers, it might also be helpful in optimizing [144] real-world value chains like energy efficiency and sustainability [52], given a proper access to blockchain data including cryptocurrencies, as well as possible Big Data from the SA streams.

10 Specific Challenge H: Multi-objective Cryptocurrency Portfolio Selection

As cryptocurrency evaluation usually includes taking into account risk [47], several cryptocurrencies might be needed as a set to balance this risk. Namely, a portfolio selection of cryptocurrencies might increase robustness when trading. In portfolio selection, multi-objective optimization algorithms are applied, and are gaining importance [4,91,107,149], as their applicability over global optimization problems contributes towards better multiple criteria decision-making. As a perspective therefore, this is an important challenge to be listed among future directions connected to cryptocurrency analysis.

In [16] it is stated that it is possible to reduce investment risk substantially by adding several different cryptocurrencies to a portfolio, using the traditional Markowitz model. In a similar manner in [106], performance of an equally weighted portfolio and an optimal Markowitz portfolio considering four popular cryptocurrencies is compared, considering different levels of risk-aversion, and reaching the conclusion that there are no significant differences between the two portfolios. In [36], the diversification advantages of including cryptocurrencies in short-term investment portfolios is analyzed, justified by the fact that cryptocurrencies are highly connected to each other but not to other financial assets. In [78], data from ten cryptocurrencies are used, and the conclusion is reached that, through diversification, investment results can be reached.

Portfolio selection is inherently a multi-objective problem, even when using mono-objective models like the Markowitz mean-variance framework. The optimal portfolio will always be a compromise solution between risk and return. By changing parameters in the Markowitz model, it is possible to define an efficient frontier that will exhibit the existing risk-return compromises. There is evidence that cryptocurrencies can be seen more as a speculative asset than as a currency [11,17]. These findings justify the use of optimization models for determining optimal portfolios that also consider investments in cryptocurrencies. If there are metrics capable of calculating expected risk and return for cryptocurrencies' investments, then most of the available methodologies can be used in this setting.

In [88], a copula-based approach is presented considering the multiperiod portfolio optimization, using Differential Evolution (DE). DE is a metaheuristic for global optimization [38,127,135,143] that is also applied successfully to many other challenges, like [52,146], therefore improvements in DE can also have potential impact on portfolio selection and, hence, cryptocurrency analysis. In a multiperiod problem [88], investors have the possibility of changing the allocation of their investments during the considered horizon. In [88] it is stated that this flexibility of rebalancing the portfolio is advantageous, especially due to the high volatility that cryptocurrencies present. Portfolio rebalancing is also considered in [65], where information about the market psychology is included, evaluated using fuzzy reasoning processes that are able to be determined under overevaluation possibilities. The investor profile will, hence, determine the efficient solution that is most adequate to his/her risk aversion preferences.

10.1 Optimization Algorithms

Metaheuristics have long been used for portfolio optimization. In [91] a review is presented of Evolutionary Algorithms used for multi-objective portfolio management. As a recent contribution, a review of Swarm Intelligence methods applied to portfolio optimization can be found in [47].

A Genetic Algorithm that takes into consideration transaction costs is introduced in [5]. A Genetic Algorithm is also used in [120]. The authors consider cardinality constraints, floor constraints, and also round-lot constraints. The objective functions consider the maximization of return and minimization of risk, along with the maximization of the sampling distance. This last objective is only used to improve the behavior of the optimization algorithm. A multi-objective evolutionary approach that is able to produce an efficient frontier of portfolios is introduced in [31], taking into account some constraints, like lower and upper limits to the investment that can be made in each asset, or the maximum and minimum number of assets that the portfolio should consider. This type of cardinality constraints is also considered in the work of [118], where the authors describe a hybrid algorithm combining Evolutionary Algorithms, quadratic programming, and a pruning heuristic. Five different multi-objective Evolutionary Algorithms are compared in [3] for the mean-variance cardinality constrained portfolio optimization problem. The authors conclude that SPEA2 seems to be better, but all multi-objective approaches perform better than single-objective ones. A mean-semivariance framework, taking into account adverse return variations only, is described in [83]. Two different Genetic Algorithms are compared (NSGAII and SPEA2), embedding the use of technical analysis indicators. Other applications of Genetic Algorithms can be found in [75,123].

Ant Colony Optimization is the methodology chosen in [43] to tackle multi-objective portfolio selection, and the authors realize that there are some efficient portfolios that are extremely difficult to find. The authors report better results than the ones obtained with simulated annealing and NSGA. Particle Swarm Optimization is used by [40]. The method seems to be particularly well suited to be used for low-risk investment portfolios. An Artificial Bee Colony metaheuristic is described in [28] for portfolio optimization, interpreting returns as fuzzy numbers. Varying the risk tolerance parameter will lead to different efficient portfolios. The authors acknowledge that real-world constraints have an important impact on the investment strategy. Fuzzy numbers are also used by [119] to represent the uncertainty in future returns. Three objectives are considered there simultaneously: The maximization of the possibilistic expected returns, the minimization of the downside risk (absolute semi-deviation below mean), and the minimization of the skewness of every portfolio. Cardinality constraints are also included.

The multi-objective portfolio optimization is transformed into a mono-objective problem using fuzzy normalization and uniform design method in [116]. The resulting model is then solved using an invasive weed optimization algorithm. The original model considers two more objectives in addition to the Markowitz mean-variance model: the stock profit gained relative to its market price, and the representation of experts' recommendations. Fuzzy numbers are also used in [86], to describe asset returns and develop a multi-objective approach based on a genetic algorithm. In [81] an approach is presented that uses a self-organizing Fuzzy Neural Network embedded into an heuristic that explores simplex local searches. Finally, [60] use machine learning tools, namely, deep reinforcement learning, to define cryptocurrencies' portfolios based on technical aspects only (price and movement).

11 Specific Challenge G: Investment Approaches with the New ICO Sentiment Indicator

The technical analysis covered in this chapter is the SA (Sect. 7) with forecasting (Sect. 8) and indicators fusion (Sect. 9), followed by optimization algorithms' supported multi-objective portfolio selection (Sect. 10). Besides technical analysis however, investment approaches with cryptocurrencies must take into account particular characteristics, like the herding effect, as identified in [15]. Risks from these characteristics can be caused through a lack of diversification of cryptocurrencies' portfolios, making investors more exposed to investment risk, as explained in the previous section. These characteristics also include long memory, stochastic volatility, and leverage [104]. Keeping pertained characteristics in mind and together through technical analysis, a trader might then evaluate a cryptocurrency as a whole and trade the cryptocurrency as a financial asset [8].

12 Conclusion

This chapter covered the different aspects of necessary perspectives needed when preparing forecasting and investment, supported by cryptocurrency social media Sentiment Analysis. Eight specific perspectives were identified, and the current state-of-the-art was covered in separate sections of the chapter, all focusing around a new type of indicator, the CSAI. In the following work, some more specific implementations and experimental results could be presented, based on the guidelines, outlines, and integration possibilities presented in this chapter.

Acknowledgements. This chapter is based upon work from COST Action IC1406 High-Performance Modelling and Simulation for Big Data Applications (cHiPSet), supported by COST (European Cooperation in Science and Technology). The author AZ acknowledges the financial support from the Slovenian Research Agency (Research Core Funding No. P2-0041). AZ also acknowledges EU support under Project No. 5442-24/2017/6 (HPC – RIVR). This chapter is also based upon work from COST Action CA15140 "Improving Applicability of Nature-Inspired Optimisation by Joining Theory and Practice (ImAppNIO)" supported by COST. The authors CP, TC and IS also acknowledge that the work presented in this chapter is supported financially by the eDREAM Project Grant number 774478, co-funded by the European Commission as part of the H2020 Framework Programme. This work was supported by the Fundacao para a Ciencia e a Tecnologia (FCT) under Project Grant UID/Multi/00308/2019. Author RS also acknowledges that work was supported by the Ministry of Education, Youth and Sports of the Czech Republic within the National Sustainability Programme Project No. LO1303 (MSMT-7778/2014), further supported by the European Regional Development Fund under the Project CEBIA-Tech no. CZ.1.05/2.1.00/03.0089.

References

1. Abbasi, A., Chen, H., Salem, A.: Sentiment analysis in multiple languages: feature selection for opinion classification in Web forums. ACM Trans. Inf. Syst. (TOIS) **26**(3), 12 (2008)
2. Abraham, J., Higdon, D., Nelson, J., Ibarra, J.: Cryptocurrency price prediction using tweet volumes and sentiment analysis. SMU Data Sci. Rev. **1**(3), 1 (2018)
3. Anagnostopoulos, K.P., Mamanis, G.: The mean-variance cardinality constrained portfolio optimization problem: an experimental evaluation of five multiobjective evolutionary algorithms. Expert Syst. Appl. **38**(11), 14208–14217 (2011)
4. Aouni, B., Doumpos, M., Pérez-Gladish, B., Steuer, R.E.: On the increasing importance of multiple criteria decision aid methods for portfolio selection. J. Oper. Res. Soc. **69**, 1525–1542 (2018)
5. Aranha, C., Iba, H.: Modelling cost into a genetic algorithm-based portfolio optimization system by seeding and objective sharing. In: 2007 IEEE Congress on Evolutionary Computation, pp. 196–203. IEEE, Singapore (2007)
6. Athey, S., Parashkevov, I., Sarukkai, V., Xia, J.: Bitcoin pricing, adoption, and usage: theory and evidence. Stanford University Graduate School of Business Research Paper No. 16–42 (2016). https://ssrn.com/abstract=2826674
7. Baccianella, S., Esuli, A., Sebastiani, F.: SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: Seventh International Conference on Language Resources and Evaluation (LREC), vol. 10, pp. 2200–2204 (2010)
8. Bagheri, A., Peyhani, H.M., Akbari, M.: Financial forecasting using ANFIS networks with quantum-behaved particle swarm optimization. Expert Syst. Appl. **41**(14), 6235–6250 (2014)
9. Balcilar, M., Bouri, E., Gupta, R., Roubaud, D.: Can volume predict Bitcoin returns and volatility? A quantiles-based approach. Econ. Model. **64**, 74–81 (2017)
10. Bartoletti, M., Bracciali, A., Lande, S., Pompianu, L.: A general framework for Bitcoin analytics. arXiv preprint arXiv:1707.01021 (2017)
11. Baur, D.G., Hong, K., Lee, A.D.: Bitcoin: medium of exchange or speculative assets? J. Int. Fin. Markets Inst. Money **54**, 177–189 (2018)
12. Bianchi, D.: Cryptocurrencies as an asset class? An empirical assessment (6 June, 2018). WBS Finance Group Research Paper (2017)
13. Bollen, J., Mao, H., Zeng, X.: Twitter mood predicts the stock market. J. Comput. Sci. **2**(1), 1–8 (2011)
14. Borth, D., Ji, R., Chen, T., Breuel, T., Chang, S.F.: Large-scale visual sentiment ontology and detectors using adjective noun pairs. In: Proceedings of the 21st ACM International Conference on Multimedia, pp. 223–232. ACM (2013)
15. Bouri, E., Gupta, R., Roubaud, D.: Herding behaviour in cryptocurrencies. Fin. Res. Lett. (2018). https://doi.org/10.1016/j.frl.2018.07.008
16. Brauneis, A., Mestel, R.: Cryptocurrency-portfolios in a mean-variance framework. Fin. Res. Lett. (2018). https://doi.org/10.1016/j.frl.2018.05.008
17. Brauneis, A., Mestel, R.: Price discovery of cryptocurrencies: Bitcoin and beyond. Econ. Lett. **165**, 58–61 (2018)

18. Buyya, R., et al.: A manifesto for future generation cloud computing: research directions for the next decade. ACM Comput. Surv. **51**(5), 105:1–105:38 (2018)
19. Cambria, E., Hussain, A.: Sentic Computing: Techniques, Tools, and Applications, vol. 2. Springer, Dordrecht (2012). https://doi.org/10.1007/978-94-007-5070-8
20. Cambria, E., Olsher, D., Rajagopal, D.: SenticNet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. In: Twenty-Eighth AAAI Conference on Artificial Intelligence (2014)
21. Cambria, E., Schuller, B., Xia, Y., Havasi, C.: New avenues in opinion mining and sentiment analysis. IEEE Intell. Syst. **28**(2), 15–21 (2013)
22. Casino, F., Dasaklis, T.K., Patsakis, C.: A systematic literature review of blockchain-based applications: current status, classification and open issues. Telematics Inform. (2018, in Press) https://doi.org/10.1016/j.tele.2018.11.006
23. Cataldi, M., Aufaure, M.A.: The 10 million follower fallacy: audience size does not prove domain-influence on Twitter. Knowl. Inf. Syst. **44**(3), 559–580 (2015)
24. Catania, L., Grassi, S., Ravazzolo, F.: Forecasting cryptocurrencies financial time series. CAMP Working Paper Series No. 5 (2018)
25. Chandrashekar, G., Sahin, F.: A survey on feature selection methods. Comput. Electr. Eng. **40**(1), 16–28 (2014)
26. Chen, C.C., Huang, H.H., Chen, H.H.: Fine-grained analysis of financial Tweets. In: Companion of the Web Conference 2018 on the Web Conference 2018, pp. 1943–1949. International World Wide Web Conferences Steering Committee (2018)
27. Chen, C.C., Huang, H.H., Chen, H.H.: NTUSD-Fin: a market sentiment dictionary for financial social media data applications. In: Proceedings of the 1st Financial Narrative Processing Workshop (FNP 2018) (2018)
28. Chen, W.: Artificial bee colony algorithm for constrained possibilistic portfolio optimization problem. Phys. A: Stat. Mech. Appl. **429**, 125–139 (2015)
29. Chen, Y.: Blockchain tokens and the potential democratization of entrepreneurship and innovation. Bus. Horiz. **61**(4), 567–575 (2018)
30. Cheoljun, E., Kaizoji, T., Kang, S.H., Pichl, L.: Bitcoin and investor sentiment: statistical characteristics and predictability. Physica A **514**, 511–521 (2019)
31. Chiam, S.C., Al Mamun, A., Low, Y.L.: A realistic approach to evolutionary multiobjective portfolio optimization. In: 2007 IEEE Congress on Evolutionary Computation, pp. 204–211. IEEE (2007)
32. Chu, J., Stephen, C., Saralees, N., Joerg, O.: GARCH modeling of cryptocurrencies. J. Risk Fin. Manag. **10**(4), 1–15 (2017)
33. Cioara, T., et al.: Enabling new technologies for demand response decentralized validation using blockchain. In: 2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe, pp. 1–4. IEEE (2018)
34. Claudia, P., Tudor, C., Marcel, A., Ionut, A., Ioan, S., Massimo, B.: Blockchain based decentralized management of demand response programs in smart energy grids. Sensors **18**, 162 (2018)
35. Colianni, S., Rosales, S., Signorotti, M.: Algorithmic trading of cryptocurrency based on Twitter sentiment analysis. CS229 Project (2015)
36. Corbet, S., Meegan, A., Larkin, C., Lucey, B., Yarovaya, L.: Exploring the dynamic relationships between cryptocurrencies and other financial assets. Econ. Lett. **165**, 28–34 (2018)
37. Cretarola, A., Figà-Talamanca, G., Patacca, M.: A sentiment-based model for the BitCoin: theory, estimation and option pricing (2017). https://doi.org/10.2139/ssrn.3042029. https://ssrn.com/abstract=3042029

38. Das, S., Mullick, S.S., Suganthan, P.: Recent advances in differential evolution - an updated survey. Swarm Evol. Comput. **27**, 1–30 (2016)
39. David, Y.: Handbook of Digital Currency Bitcoin, Innovation, Financial Instruments, and Big Data. Elsevier, Amsterdam (2015). https://doi.org/10.1016/B978-0-12-802117-0.00002-3
40. Deng, G.F., Lin, W.T., Lo, C.C.: Markowitz-based portfolio selection with cardinality constraints using improved particle swarm optimization. Expert Syst. Appl. **39**(4), 4558–4566 (2012)
41. Dey, L., Haque, S.M.: Opinion mining from noisy text data. Int. J. Doc. Anal. Recogn. (IJDAR) **12**(3), 205–226 (2009)
42. D'mello, G., González-Vélez, H.: Distributed software dependency management using blockchain. In: 27th Euromicro International Conference on Parallel, Distributed and Network-Based Processing, PDP 2019, pp. 1–8. IEEE, Pavia (2019, to appear)
43. Doerner, K., Gutjahr, W.J., Hartl, R.F., Strauss, C., Stummer, C.: Pareto ant colony optimization: a metaheuristic approach to multiobjective portfolio selection. Ann. Oper. Res. **131**(1–4), 79–99 (2004)
44. Easley, D., O'Hara, M., Basu, S.: From Mining to Markets: The Evolution of Bitcoin Transaction Fees. SSRN 3055380 (2018)
45. eDREAM: H2020 eDREAM Project. http://edream-h2020.eu/en/
46. Energy4Europe, European Commission: Energy Strategy and Energy Union, Secure, Competitive, and sustainable Energy. https://ec.europa.eu/energy/en/topics/energy-strategy-and-energy-union
47. Ertenlice, O., Kalayci, C.B.: A survey of swarm intelligence for portfolio optimization: algorithms and applications. Swarm Evol. Comput. **39**, 36–52 (2018)
48. Feldman, R.: Techniques and applications for sentiment analysis. Commun. ACM **56**(4), 82–89 (2013)
49. Forman, G.: An extensive empirical study of feature selection metrics for text classification. J. Mach. Learn. Res. **3**(Mar), 1289–1305 (2003)
50. Galvez, J.F., Mejuto, J., Simal-Gandara, J.: Future challenges on the use of blockchain for food traceability analysis. TrAC Trends Anal. Chem. **107**, 222–232 (2018)
51. 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. In: MCIS 2015 Proceedings 20 (2015). https://aisel.aisnet.org/mcis2015/20/
52. Glotić, A., Zamuda, A.: Short-term combined economic and emission hydrothermal optimization by surrogate differential evolution. Appl. Energy **141**, 42–56 (2015)
53. Guidolin, M., Hyde, S., McMillan, D., Ono, S.: Non-linear predictability in stock and bond returns: when and where is it exploitable? Int. J. Forecast. **25**(2), 373–399 (2009)
54. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. J. Mach. Learn. Res. **3**(Mar), 1157–1182 (2003)
55. Haddi, E., Liu, X., Shi, Y.: The role of text pre-processing in sentiment analysis. Procedia Comput. Sci. **17**, 26–32 (2013)
56. Haveliwala, T.H.: Topic-sensitive PageRank. In: Proceedings of the 11th International Conference on World Wide Web, pp. 517–526. ACM (2002)
57. Hayes, A.S.: Cryptocurrency value formation: an empirical analysis leading to a cost of production model for valuing Bitcoin. Telematics Inform. **34**(7), 1308–1321 (2017)

58. Hirsch, J.E.: An index to quantify an individual's scientific research output. Proc. Nat. Acad. Sci. U.S.A. **102**(46), 16569 (2005)
59. Hussein, D.M.E.D.M.: A survey on sentiment analysis challenges. J. King Saud Univ.-Eng. Sci. **30**(4), 330–338 (2018)
60. Jiang, Z., Liang, J.: Cryptocurrency portfolio management with deep reinforcement learning. In: 2017 Intelligent Systems Conference (IntelliSys), pp. 905–913. IEEE (2017)
61. Kaminski, J.C.: Nowcasting the Bitcoin market with Twitter signals. arXiv preprint arXiv:1406.7577v3 (2016). https://arxiv.org/pdf/1406.7577v3.pdf
62. Katsiampa, P.: Volatility estimation for Bitcoin: a comparison of GARCH models. Econ. Lett. **158**, 3–6 (2017)
63. Kearney, C., Liu, S.: Textual sentiment in finance: a survey of methods and models. Int. Rev. Fin. Anal. **33**, 171–185 (2014)
64. Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 137–146. ACM (2003)
65. Khayamim, A., Mirzazadeh, A., Naderi, B.: Portfolio rebalancing with respect to market psychology in a fuzzy environment: a case study in Tehran Stock Exchange. Appl. Soft Comput. **64**, 244–259 (2018)
66. Kim, S.M., Hovy, E.: Identifying and analyzing judgment opinions. In: Proceedings of the Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, pp. 200–207. Association for Computational Linguistics (2006)
67. Kim, Y.B., Lee, J., Park, N., Choo, J., Kim, J.H., Kim, C.H.: When Bitcoin encounters information in an online forum: using text mining to analyse user opinions and predict value fluctuation. PloS One **12**(5), e0177630 (2017)
68. Ladislav, K.: BitCoin meets Google Trends and Wikipedia: quantifying the relationship between phenomena of the Internet era (2013). https://www.nature.com/srep/
69. Lamon, C., Nielsen, E., Redondo, E.: Cryptocurrency Price Prediction Using News and Social Media Sentiment (2016). http://cs229.stanford.edu/proj2017/final-reports/5237280.pdf
70. Law, L., Sabett, S., Solinas, J.: How to make a mint: the cryptography of anonymous electronic cash. Technical report, National Security Agency, Office of Information Security Research and Technology, Cryptology Division, 1001 Pennsylvania Avenue, DC, USA, June 1996
71. Li, Q., Shah, S.: Learning stock market sentiment lexicon and sentiment-oriented word vector from StockTwits. In: Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017), pp. 301–310 (2017)
72. Li, T.R., Chamrajnagar, A.S., Fong, X.R., Rizik, N.R., Fu, F.: Sentiment-based prediction of alternative cryptocurrency price fluctuations using gradient boosting tree model. arXiv preprint arXiv:1805.00558 (2018)
73. Li, X., Wang, C.A.: The technology and economic determinants of cryptocurrency exchange rates: the case of Bitcoin. Decis. Support Syst. **95**, 49–60 (2017)
74. Lim, S.Y., et al.: Blockchain technology the identity management and authentication service disruptor: a survey. Int. J. Adv. Sci. Eng. Inf. Technol. **8**(4–2), 1735–1745 (2018)
75. Lin, C.C., Liu, Y.T.: Genetic algorithms for portfolio selection problems with minimum transaction lots. Eur. J. Oper. Res. **185**(1), 393–404 (2008)

76. Linton, M., Teo, E.G.S., Bommes, E., Chen, C.Y., Härdle, W.K.: Dynamic topic modelling for cryptocurrency community forums. In: Härdle, W.K., Chen, C.Y.H., Overbeck, L. (eds.) Applied Quantitative Finance. SC, pp. 355–372. Springer, Heidelberg (2017). https://doi.org/10.1007/978-3-662-54486-0_18
77. Liu, C., Kübler, S., Yu, N.: Feature selection for highly skewed sentiment analysis tasks. In: Proceedings of the Second Workshop on Natural Language Processing for Social Media (SocialNLP), pp. 2–11 (2014)
78. Liu, W.: Portfolio diversification across cryptocurrencies. Fin. Res. Lett. (2018). https://doi.org/10.1016/j.frl.2018.07.010
79. Liu, Y., Qin, Z., Li, P., Wan, T.: Stock volatility prediction using recurrent neural networks with sentiment analysis. In: Benferhat, S., Tabia, K., Ali, M. (eds.) IEA/AIE 2017. LNCS (LNAI), vol. 10350, pp. 192–201. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-60042-0_22
80. Lucey, B.M., et al.: Future directions in international financial integration research - A crowdsourced perspective. Int. Rev. Fin. Anal. **55**, 35–49 (2018)
81. Lumanpauw, E., Pasquier, M., Quek, C.: MNFS-FPM: a novel memetic neuro-fuzzy system based financial portfolio management. In: 2007 IEEE Congress on Evolutionary Computation, pp. 2554–2561. IEEE, Singapore (2007)
82. Ma, J., Gans, J.S., Tourky, R.: Market structure in Bitcoin mining. Technical report, National Bureau of Economic Research (2018)
83. Macedo, L.L., Godinho, P., Alves, M.J.: Mean-semivariance portfolio optimization with multiobjective evolutionary algorithms and technical analysis rules. Expert Syst. Appl. **79**, 33–43 (2017)
84. Mai, F., Shan, Z., Bai, Q., Wang, X., Chiang, R.H.: How does social media impact Bitcoin value? a test of the silent majority hypothesis. J. Manag. Inf. Syst. **35**(1), 19–52 (2018)
85. Mao, Y., Wei, W., Wang, B., Liu, B.: Correlating S&P 500 stocks with Twitter data. In: Proceedings of the First ACM International Workshop on Hot Topics on Interdisciplinary Social Networks Research, pp. 69–72. ACM (2012)
86. Mashayekhi, Z., Omrani, H.: An integrated multi-objective Markowitz-DEA cross-efficiency model with fuzzy returns for portfolio selection problem. Appl. Soft Comput. **38**, 1–9 (2016)
87. Matta, M., Lunesu, I., Marchesi, M.: The predictor impact of Web search media on Bitcoin trading volumes. In: 2015 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K), vol. 1, pp. 620–626. IEEE (2015)
88. Mba, J.C., Pindza, E., Koumba, U.: A differential evolution copula-based approach for a multi-period cryptocurrency portfolio optimization. Fin. Markets Portfolio Manag. **32**(4), 399–418 (2018)
89. 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)
90. Mendes, L., Godinho, P., Dias, J.: A forex trading system based on a genetic algorithm. J. Heuristics **18**(4), 627–656 (2012)
91. Metaxiotis, K., Liagkouras, K.: Multiobjective evolutionary algorithms for portfolio management: a comprehensive literature review. Expert Syst. Appl. **39**(14), 11685–11698 (2012)
92. Millard, C.: Blockchain and law: incompatible codes? Comput. Law Secur. Rev. **34**(4), 843–846 (2018)
93. Mohammadi, A., Saraee, M., Mirzaei, A.: Time-sensitive influence maximization in social networks. J. Inf. Sci. **41**(6), 765–778 (2015)

94. Nakamoto, S.: Bitcoin: a peer-to-peer electronic cash system. Bitcoin.org (2008)
95. Nauck, D., Klawonn, F., Kruse, R.: Foundations of Neuro-Fuzzy Systems. Wiley, Hoboken (1997)
96. Nguyen, G.T., Kim, K.: A survey about consensus algorithms used in blockchain. J. Inf. Process. Syst. **14**(1), 101–128 (2018)
97. Ortony, A., Clore, G.L., Collins, A.: The Cognitive Structure of Emotions. Cambridge University Press, Cambridge (1990)
98. Panarello, A., Tapas, N., Merlino, G., Longo, F., Puliafito, A.: Blockchain and IoT integration: a systematic survey. Sensors **18**(8), 2575 (2018)
99. Pang, B., Lee, L.: Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales. In: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, pp. 115–124. Association for Computational Linguistics (2005)
100. Peng, S., Zhou, Y., Cao, L., Yu, S., Niu, J., Jia, W.: Influence analysis in social networks: a survey. J. Netw. Comput. Appl. **106**, 17–32 (2018)
101. Peng, Y., Albuquerque, P.H.M., de Sá, J.M.C., Padula, A.J.A., Montenegro, M.R.: The best of two worlds: forecasting high frequency volatility for cryptocurrencies and traditional currencies with Support Vector Regression. Expert Syst. Appl. **97**, 177–192 (2018)
102. Petcu, D., González–Vélez, H., Nicolae, B., García–Gómez, J.M., Fuster–Garcia, E., Sheridan, C.: Next generation HPC clouds: a view for large-scale scientific and data-intensive applications. In: Lopes, L., et al. (eds.) Euro-Par 2014. LNCS, vol. 8806, pp. 26–37. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-14313-2_3
103. Petrican, T., et al.: Evaluating forecasting techniques for integrating household energy prosumers into smart grids. In: 14th International Conference on Intelligent Computer Communication and Processing, pp. 79–85. IEEE (2018)
104. Phillips, R.C., Gorse, D.: Predicting cryptocurrency price bubbles using social media data and epidemic modelling. In: 2017 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1–7. IEEE (2017)
105. Phillips, R.C., Gorse, D.: Mutual-excitation of cryptocurrency market returns and social media topics. In: Proceedings of the 4th International Conference on Frontiers of Educational Technologies, pp. 80–86. ACM (2018)
106. Platanakis, E., Sutcliffe, C., Urquhart, A.: Optimal vs naïve diversification in cryptocurrencies. Econ. Lett. **171**, 93–96 (2018)
107. Ponsich, A., Jaimes, A.L., Coello, C.A.C.: A survey on multiobjective evolutionary algorithms for the solution of the portfolio optimization problem and other finance and economics applications. IEEE Trans. Evol. Comput. **17**(3), 321–344 (2013)
108. Pop, C., et al.: Decentralizing the stock exchange using blockchain an ethereum-based implementation of the Bucharest Stock Exchange. In: 14th International Conference on Intelligent Computer Communication and Processing, pp. 459–466. IEEE (2018)
109. Puri, V.: Decrypting Bitcoin prices and adoption rates using Google search. CMC Senior Theses **1418** (2016)
110. Radosavljevic, V., Grbovic, M., Djuric, N., Bhamidipati, N.: Large-scale World Cup 2014 outcome prediction based on Tumblr posts. In: KDD Workshop on Large-Scale Sports Analytics (2014)
111. Rakoczy, M., Bouzeghoub, A., Wegrzyn-Wolska, K., Gancarski, L.A.: Users views on others - analysis of confused relation-based terms in social network. In: Debruyne, C., et al. (eds.) OTM 2016. LNCS, vol. 10033, pp. 155–174. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-48472-3_9

112. Rakoczy, M.E., Bouzeghoub, A., Lopes, G.A., Wegrzyn-Wolska, K.: In the search of quality influence on a small scale – micro-influencers discovery. In: Panetto, H., Debruyne, C., Proper, H., Ardagna, C., Roman, D., Meersman, R. (eds.) OTM 2018. LNCS, vol. 11230. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-02671-4_8
113. Rakoczy, M.E., Bouzeghoub, A., Gançarski, A.L., Wegrzyn-Wolska, K.: Influence in time-dependent citation networks. In: 12th International Conference on Research Challenges in Information Science, RCIS 2018, 29–31 May 2018, Nantes, France, pp. 1–11 (2018)
114. Rakoczy, M.E., Bouzeghoub, A., Gançarski, A.L., Wegrzyn-Wolska, K.: Reputation prediction using influence conversion. In: 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/12th IEEE International Conference on Big Data Science and Engineering, TrustCom/BigDataSE 2018, 1–3 August 2018, New York, NY, USA, pp. 43–48 (2018)
115. Reyna, A., Martín, C., Chen, J., Soler, E., Daíz, M.: On blockchain and its integration with IoT. Challenges and opportunities. Future Gener. Comput. Syst. **88**, 173–190 (2018)
116. Rezaei Pouya, A., Solimanpur, M., Jahangoshai Rezaee, M.: Solving multi-objective portfolio optimization problem using invasive weed optimization. Swarm Evol. Comput. **28**, 42–57 (2016)
117. Ruiz, E.J., Hristidis, V., Castillo, C., Gionis, A., Jaimes, A.: Correlating financial time series with micro-blogging activity. In: Proceedings of the Fifth ACM International Conference on Web Search and Data Mining, pp. 513–522. ACM (2012)
118. Ruiz-Torrubiano, R., Suarez, A.: Use of heuristic rules in evolutionary methods for the selection of optimal investment portfolios. In: 2007 IEEE Congress on Evolutionary Computation, pp. 212–219. IEEE (2007)
119. Saborido, R., Ruiz, A.B., Bermúdez, J.D., Vercher, E., Luque, M.: Evolutionary multi-objective optimization algorithms for fuzzy portfolio selection. Appl. Soft Comput. **39**, 48–63 (2016)
120. Skolpadungket, P., Dahal, K., Harnpornchai, N.: Portfolio optimization using multi-objective genetic algorithms. In: 2007 IEEE Congress on Evolutionary Computation, pp. 516–523. IEEE (2007)
121. Smith, P., González-Vélez, H., Caton, S.: Social auto-scaling. In: 26th Euromicro International Conference on Parallel, Distributed and Network-Based Processing, PDP 2018, pp. 186–195. IEEE, Cambridge (2018)
122. Sokolov, K.: Ransomware Activity, Demand for Blockchain Settlement and the Upper Bound of Congestion-Induced Mining Reward. SSRN Working Paper (2018). https://ssrn.com/abstract=3175986
123. Soleimani, H., Golmakani, H.R., Salimi, M.H.: Markowitz-based portfolio selection with minimum transaction lots, cardinality constraints and regarding sector capitalization using genetic algorithm. Expert Syst. Appl. **36**(3), 5058–5063 (2009)
124. Stenqvist, E., Lönnö, J.: Predicting Bitcoin price fluctuation with Twitter sentiment analysis (2017)
125. Stevenson, R.A., Mikels, J.A., James, T.W.: Characterization of the affective norms for english words by discrete emotional categories. Behav. Res. Methods **39**(4), 1020–1024 (2007)
126. Stocchi, M., Lunesu, I., Ibba, S., Baralla, G., Marchesi, M.: The future of Bitcoin: a synchrosqueezing wavelet transform to predict search engine query trends. In: KDWeb (2016)

127. Storn, R., Price, K.: Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. J. Global Optim. **11**, 341–359 (1997)
128. Sung, J., Moon, S., Lee, J.-G.: The influence in Twitter: are they really influenced? In: Cao, L., et al. (eds.) BSI/BSIC 2013. LNCS (LNAI), vol. 8178, pp. 95–105. Springer, Cham (2013). https://doi.org/10.1007/978-3-319-04048-6_9
129. Szor, P.: The Art of Computer Virus Research and Defense. Pearson Education, London (2005)
130. Tonelli, R., Ducasse, S., Fenu, G., Bracciali, A.: 2018 IEEE 1st International Workshop on Blockchain Oriented Software Engineering (IWBOSE), 20 March 2018, Campobasso, Italy. IEEE (2018)
131. Tudor, C., et al.: Optimized flexibility management enacting data centres participation in smart demand response programs. Future Gener. Comput. Syst. **78**, 330–342 (2018)
132. Tumasjan, A., Sprenger, T.O., Sandner, P.G., Welpe, I.M.: Predicting elections with Twitter: what 140 characters reveal about political sentiment. In: International AAAI Conference on Web and Social Media, Fourth International AAAI Conference on Weblogs and Social Media, vol. 10, no. 1, pp. 178–185 (2010)
133. Tupinambas, T.M., Leao, R.A., Lemos, A.P.: Cryptocurrencies transactions advisor using a genetic Mamdani-type fuzzy rules based system. In: IEEE International Conference on Fuzzy Systems, p. 7 (2018)
134. Turney, P.D.: Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pp. 417–424. Association for Computational Linguistics (2002)
135. Viktorin, A., Senkerik, R., Pluhacek, M., Kadavy, T., Zamuda, A.: Distance based parameter adaptation for success-history based differential evolution. Swarm Evol. Comput., 12 November 2018. https://doi.org/10.1016/j.swevo.2018.10.013
136. Vytautas, K., Niels, D., Jochen, D.W.: Using sentiment analysis to predict interday Bitcoin price movements. J. Risk Fin. **19**(1), 56–75 (2018)
137. Wang, W., et al.: A survey on consensus mechanisms and mining management in blockchain networks. arXiv preprint arXiv:1805.02707 (2018)
138. Weng, J., Lim, E.P., Jiang, J., He, Q.: TwitterRank: finding topic-sensitive influential Twitterers. In: Proceedings of the Third ACM International Conference on Web Search and Data Mining, pp. 261–270. ACM (2010)
139. Werner, A., Murray, Z.F.: Is all that talk just noise? The information content of internet stock message boards. J. Fin. **59**, 1259–1294 (2004)
140. Xiao, F., Noro, T., Tokuda, T.: Finding news-topic oriented influential Twitter users based on topic related hashtag community detection. J. Web Eng. **13**(5&6), 405–429 (2014)
141. Yogatama, D., Dyer, C., Ling, W., Blunsom, P.: Generative and discriminative text classification with recurrent neural networks. In: Thirty-Fourth International Conference on Machine Learning (ICML 2017) (2017)
142. Zafarani, R., Abbasi, M.A., Liu, H.: Social Media Mining: An Introduction. Cambridge University Press, New York (2014)
143. Zamuda, A., Brest, J.: Self-adaptive control parameters' randomization frequency and propagations in differential evolution. Swarm Evol. Comput. **25**, 72–99 (2015)
144. Zamuda, A., Nicolau, M., Zarges, C.: A black-box discrete optimization benchmarking (BB-DOB) pipeline survey: taxonomy, evaluation, and ranking. In: Proceedings of the Genetic and Evolutionary Computation Conference Companion (GECCO 2018), pp. 1777–1782 (2018)

145. Zamuda, A., Zarges, C., Stiglic, G., Hrovat, G.: Stability selection using a genetic algorithm and logistic linear regression on healthcare records. In: Proceedings of the Genetic and Evolutionary Computation Conference Companion (GECCO 2017), pp. 143–144 (2017)
146. Zamuda, A., Sosa, J.D.H.: Success history applied to expert system for underwater glider path planning using differential evolution. Expert Syst. Appl. **119**(1 April 2019), 155–170 (2019)
147. Zhang, X., Fuehres, H., Gloor, P.A.: Predicting stock market indicators through Twitter "I hope it is not as bad as I fear". Procedia-Soc. Behav. Sci. **26**, 55–62 (2011)
148. Zheludev, I., Smith, R., Aste, T.: When can social media lead financial markets? Sci. Rep. **4**, 4213 (2014)
149. Zopounidis, C., Galariotis, E., Doumpos, M., Sarri, S., AndriosopouloS, K.: Multiple criteria decision aiding for finance: an updated bibliographic survey. Eur. J. Oper. Res. **247**(2), 339–348 (2015)