UNIVERSIDADE FEDERAL DE SÃO CARLOS

CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Predicting the Direction, Maximum, Minimum and Closing Price of Daily/Intra-daily Bitcoin Exchange Rate Using Batch and Online Machine Learning Techniques

Dennys Ch. A. Mallqui

Orientador: Prof. Dr. Ricardo A. S. Fernandes

São Carlos - SP

Setembro/2018

Predicting the Direction, Maximum, Minimum and Closing Price of Daily/Intra-daily Bitcoin Exchange Rate Using Batch and Online Machine Learning Techniques

Qualificação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação, área de concentração: Inteligência Artificial.

Orientador: Prof. Dr. Ricardo A. S. Fernandes

São Carlos - SP
Setembro/2018

UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Ciência da Computação

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado do candidato Dennys Christian Mallqui Arguelles, realizada em 19/09/2018:

Prof. Dr. Ricardo Augusto Souza Fernandes
UFSCar

Prof. Dr. Marcelo Suetake
UFSCar

Prof. Dr. Danilo Hernane Spatti
USP

This thesis work is dedicated to my wife, who gave me the strength to pursue my dream of realize a master degree outside of my country. This work is also dedicated to my parents and daughter, whose are my source of inspiration and motivation.

Acknowledgements

I would like to express my deep gratitude to Professor Dr. Ricardo A. S. Fernandes, my research supervisor, for their patient guidance, enthusiastic encouragement and useful critiques of this research work. My special thanks are extended to scholarship program of the Organization of American States and Coimbra Group of Brazilian Universities for their generosity in funding of my mastering degree. I am very honored to be the recipient of this award.

"Experimentation is the least arrogant method of gaining knowledge.
The experimenter humbly asks a question of nature."
(Isaac Asimov)

Resumo

Bitcoin é a criptomoeda mais aceita no mundo, o que a torna atraente para investidores e comerciantes. No entanto, o grande desafio em prever a taxa de câmbio do Bitcoin é sua alta volatilidade. Portanto, a previsão de seu comportamento é de grande importância para os mercados financeiros. Desta forma, nos últimos anos, alguns estudos foram propostos com base no uso de técnicas de aprendizado de máquina para prever a direção de sua taxa de câmbio, embora com baixa precisão. Portanto, como primeira contribuição deste trabalho, pode-se destacar a análise e identificação de variáveis/atributos internos e externos considerados relevantes para a previsão da taxa de câmbio do Bitcoin em frequências diárias e intra-diárias. O aumento do uso de técnicas de aprendizado de máquina para prever séries temporais e a aceitação de criptomoedas como instrumentos financeiros motivaram o presente estudo a buscar previsões mais precisas para a taxa de câmbio do Bitcoin. Portanto, foram utilizadas diferentes técnicas de seleção de atributos para variáveis candidatas. Em relação às variáveis internas, propõe-se usar informação de Blockchain e gerar indicadores técnicos comumente utilizados pelos *traders*. Sobre variáveis externas é proposto o uso de índices econômicos internacionais e tendências sociais extraídos do Google e da Wikipedia. Como segunda contribuição, uma metodologia é proposta para prever a direção da taxa de câmbio do Bitcoin em relação ao dólar. Além disso, explorou-se a possibilidade de prever diretamente os preços máximo, mínimo e de fechamento, incluindo essas informações para predizer a tendência. Para isso, foram utilizadas redes neurais artificiais, redes neurais recorrentes, máquinas de vetores de suporte e modelos Ensemble (combinando regressão e clusterização). Como uma terceira contribuição, para frequência de tempo intra-diário, os métodos de aprendizado por fluxo de dados são explorados sob a hipótese de que o preço do Bitcoin apresenta um comportamento não-estacionário. Assim, observa-se que, no longo prazo, o Bitcoin se comporta mais como um instrumento tradicional e, portanto, é cada vez mais afetado pelo contexto internacional e fundamentos econômicos. Assim, os resultados obtidos mostraram que as variáveis/atributos selecionados e o melhor modelo de aprendizado de máquina obtêm uma melhoria de mais de 10% na precisão em relação aos últimos trabalhos da literatura correlata, usando o mesmo período de informação. Em relação à predição direta dos valores da taxa de câmbio do Bitcoin, foi possível obter Erros Absolutos Percentuais Médios entre 1% e 2%. Finalmente, na previsão do movimento de preços intra-diários, por meio do uso de técnicas de aprendizado de fluxo de dados, obteve-se uma melhora em mais de 6% de precisão em relação a estudos prévios.

Palavras-chaves: Taxa de câmbio de Bitcoin, previsão de séries temporais, previsão de preços de ações, métodos de seleção de atributos, tendências sociais, indicadores técnicos, aprendizado de máquina, aprendizado em fluxo de dados.

Abstract

Bitcoin is the most accepted cryptocurrency in the world, which makes it attractive for investors and traders. However, the great challenge in predicting the Bitcoin exchange rate is its high volatility. Therefore, the prediction of its behavior is of great importance for financial markets. In this way, in recent years, Few studies were proposed based on the use of machine learning techniques to predict the direction of their exchange rate, albeit with low precision. Therefore, as a first contribution of this paper, it can be highlighted the analysis and identification of internal and external variables/attributes considered as relevant for predicting the Bitcoin exchange rate in daily and intra-daily time frequencies. The increased use of machine learning techniques to predict time series and the acceptance of cryptocurrencies as financial instruments motivated the present study to seek more accurate predictions for the Bitcoin exchange rate. For this purpose, it was used different techniques of attribute selection to candidate variables. In relation of internal variables is proposed to use Blockchain information and generate technical indicators commonly used by traders. About external variables is proposed to use international economic indices and social trends extracted from Google and Wikipedia. As a second contribution, a methodology is proposed to predict the direction of the Bitcoin exchange rate against the dollar. In addition, it was explored the possibility of directly predict the maximum, minimum and closing prices, including these information to predict the trend. For this, Artificial Neural Networks, Recurrent Neural Networks, Support Vector Machines and Ensemble models (combining regression and clusterization) were used. As a third contribution for intra-daily time frequency, the data-stream learning methods are explored under the hypothesis that Bitcoin price presents a non-stationary behavior. Thus, it is observed that in long term, Bitcoin behaves more like a traditional instrument and, therefore, is increasingly affected by the international context and economic fundamentals. Likewise, the results showed that the selected attributes and the best machine learning model achieved an improvement of more than 10% in accuracy, for the price direction predictions with respect to the state-of-the-art papers, using the same period of information. In relation to the maximum, minimum and closing Bitcoin prices regressions, it was possible to obtain Mean Absolute Percentage Errors between 1% and 2%. Finally, in the prediction of intra-daily price movement, through the use of data-stream learning techniques, is obtained a result that improves more than 6% in accuracy to other previous studies.

Key-words: Bitcoin, prediction, direction, OHLC price, regression, attribute selection, social trends, technical indicators, data-stream learning, machine learning.

List of Figures

List of Tables

List of acronyms

ANN Artificial Neural Network

ARMA Autoregressive-Moving-Average

AUC Area Under ROC Curve

AUE Accuracy-Updated Ensemble

AWE Accuracy-Weighted Ensemble

BNN Bayesian Neural Network

CC Cryptocurrency

CNN Convolutional Neural Network

DC Digital Currency

GA Genetic Algorithm

GARCH Generalized AutoRegressive Conditional Heteroskedasticity

GLM Generalized Linear Model

HT Hoedffing Tree

LSTM Long Short-Term Memory

MAE Mean Absolute Error

MAPE Mean Absolute Percentage Error

MLP Multilayer Perceptron

MSE Mean Square Error

OHLC Open-High-Low-Close price

RC Real Currency

RF Random Forest

RMSE Root Mean Square Error

RNN Recurrent Neural Network

SMA Simple Moving Average

SVM Support Vector Machine

SVR Support Vector Regression

VC Virtual Currency

VFDT Very Fast Decision Tree

WEKA Waikato Environment for Knowledge Analysis

WMA Weighted Moving Average

Contents

1 Introduction

In recent years, the expansion of Internet and encryption technologies are generating disrupting changes in the valuation, accounting and exchange of economic assets and services. Thus, in this scenario, the Virtual Currencies (VCs) are becoming popular and used for financial transactions worldwide. In particular, the Cryptocurrencies (CCs) are the most representative (ABBOUSHI, 2016; PENG et al., 2018), because they have received much attention by the media and investors. This fact can be attributed to their innovative characteristics, transparency, simplicity and increasing acceptance (URQUHART, 2017).

Currently, Bitcoin is the most famous CC and, according to Kristoufek (2015), it is presented as a potential alternative to traditional currencies (e.g., US dollar, the Euro, Japanese Yen), because it has advantages such as low costs per transaction, a controlled and known algorithm for currency generation, and transparency. Thus, relating to the importance of the Bitcoin and its impact on the economy, it can be highlighted that, according to the website *https://coinmarketcap.com* accessed on March 3rd 2018, the CC market capitalization value represents approximately US$ 441 billions, where the Bitcoin represents more than 42%.

In accordance with Cuthbertson (2015), in February 2015, more than 100,000 businesses accept Bitcoins. The list includes famous companies like Amazon, CVS, Dell, Expedia, Home Depot, Pay Pal, Subway, Target, Victoria Secret, Gap, among others. Furthermore, the list continues to grow among all companies, small and large, including Fortune 500, such as presented by Chokun (2016). Currently, Moreau (2018) shows another list with retail companies as Overstock, eGifter, Newegg, Microsoft (funds for purchase movies, games and apps), Shopify stores and so on.

Although there are criticisms regarding security aspects of anonymity for the Bitcoin transactions, recent studies such as those made by Khalilov & Levi (2018) and Conti et al. (2018) suggest that there are opportunities to improve these aspects. However, it would be needed the adaptation of the current architecture of the Bitcoin to support the evolution of its demand and advances in the cryptographic and data security research fields.

Due to the above, the Bitcoin has experienced a rapid growth in visibility and interest from investors, financial press and financial regulatory agencies in the United States, Europe, Japan, and others (ABBOUSHI, 2016; MCINTYRE; HARJES, 2016). Thus, it is important to remember that accurate forecasts about trends and prices of any investment instrument can help investors to gain opportunities to make a profit (QIU;

SONG, 2016; GERLEIN et al., 2016).

1.1 Purpose and Problem Definition

For traders or general users of CCs, the greatest challenge is the Bitcoin exchange rate volatility, as will be seen later in the present study. Therefore, idealize a model that can explain the Bitcoin price behavior for this unsettled market is meaningful (MCINTYRE; HARJES, 2016). However, as mentioned by Alstyne (2014), the author states that the high volatility of thBitcoin can not be a factor that invalidates it as a currency, but it is a motivation for traders and the general public to seek solutions to reduce their risk. Therefore, in the financial world, the possibility to predict direction of assets is a practical matter that strongly influences a trader decision to buy or sell an instrument of investment (MONTGOMERY; JENNINGS; KULAHCI, 2015).

Thus, the number of studies about the time series of the Bitcoin exchange rate is increasing, but is relatively recent. Many of them try to identify factors or attributes that show more correlation with Bitcoin price variation (KRISTOUFEK, 2013; KRISTOUFEK, 2015; MATTA; LUNESU; MARCHESI, 2015; CIAIAN; RAJCANIOVA; KANCS, 2016; VASSILIADIS et al., 2017; BALCILAR et al., 2017; ZHU; DICKINSON; LI, 2017; LI; WANG, 2017). In addition, others previous studies try to make predictions for the Bitcoin exchange rate behavior, such as those made by Madan, Saluja & Zhao (2015), Greaves & Au (2015), Mcnally (2016), Kim et al. (2016), but they show many discrepancies or results with low precision, showing that there is still a long way to find better forecasting models.

1.2 Hypothesis

As a result of the particular nature of the Bitcoin (commodity and currency), there are studies that try to identify which is the best criterion for its treatment. Thus, the assumption of treating the Bitcoin as an instrument of investment will be used, according with previous studies presented by Glaser et al. (2014), Wu & Pandey (2014), Li & Wang (2017). Therefore, the main hypothesis of this study is to demonstrate that it is possible to generate consistent predictive models of the Bitcoin exchange rate, considering it as an investment instrument.

1.3 Goals

Based on the above context, this research seeks to contribute to the decision support literature, identifying relevant attributes and machine learning techniques to make predictions of the Bitcoin exchange rate (Bitcoin against US dollar), in order to obtain greater accuracy than recent studies and, consequently:

1. identify techniques of attribute selection that can obtain the most relevant attributes;

2. analyze the best attributes that explain the behavior of the variation of exchange rate in different frequencies;

3. get the configurations of machine learning techniques that obtain the best results.

Thus, the goal of this study is focused on improving the accuracy of forecasts of the daily exchange rate behavior of the Bitcoin considering the direction, maximum, minimum, and closing prices. Thus, the goal of this paper is to propose a methodology that can improve the decision making process for Bitcoin traders. For this purpose, it was employed algorithms to determine the most relevant attributes/variables and the combinations of machine learning algorithms were explored to predict the Bitcoin market behavior in daily and intra-daily time frequencies.

1.4 General Approach

Within the basic form of typification of scientific research work, according to Kothari (2004) and Silva & Menezes (2005), this study is framed in the following categories:

- *ex post facto research* – because this study uses historical information and attempts to discover the relationship between the variables selected and the Bitcoin's exchange rate behavior;

- *applied research* – related to time series analysis and focused on the Bitcoin exchange rate (trend prediction);

- *quantitative research* – where all information will be quantifiable and will be used statistical measures (e.g. means, standard deviation, kappa index);

- *experimental research* – in which, from the collected data, a set of techniques related to the analysis of time series and machine learning will be applied. In this way, it will be proved which information is relevant and which configuration of the prediction models obtain better results.

1.5 Success Criteria

The purpose of the present work is to generate a Bitcoin exchange rate prediction model that can be consistent and used as a decision support tool. For this, classification and regression metrics will be used to measure the prediction performance of the generated model. Moreover, these results will be compared with those obtained by similar studies.

2 Theory

2.1 Virtual Currencies

Virtual Currency (VC) is a different concept of Real Currency (RC), because RC like banknotes and coins are issued and controlled by sovereign government whereas VCs are not, as mentioned by Abboushi (2016). Thus, the authors detailed the main features of this type of currency:

- it is a type of Digital Currency (DC), which is considered as a digital representation of the measurement of the economic value of an asset or transaction;
- may or may not be exchangeable to real currency;
- it is issued by non-government party; and,
- it is used as a medium of exchange value similar to RC, but does not have supported by governments. In many countries, it is not illegal, but it also is not protected by them.

Thus, VCs cover a wide array of assets as coupon issued by a retailer and retrieved as digital code on mobile device, digital currencies backed by tangible economic assets such as gold or national currency, and are more sophisticated and popular (ABBOUSHI, 2016). Table 1 shows the different sorts of DCs and VCs.

Table 1 – Digital and virtual currency types.

Type of Digital Currency	Denomination	Key Feature
Not VC	Denominated in RC, e.g. US Dollar	Digital payment mechanism, e.g. Paypal, digital bank wallet and so on.
Non-convertible VC	Own units, e.g. Air miles points	Use only for restricted products and services. Non-convertible to RC or other VC
Convertible VC	Own units.	Centralized or decentralized VC system, convertible to RC, goods, and more.
Cryptocurrency (convertible VC sub-type)	Own units, e.g. Bitcoin unit BTC	Use only cryptography to validate value and transaction based on decentralized VC system.

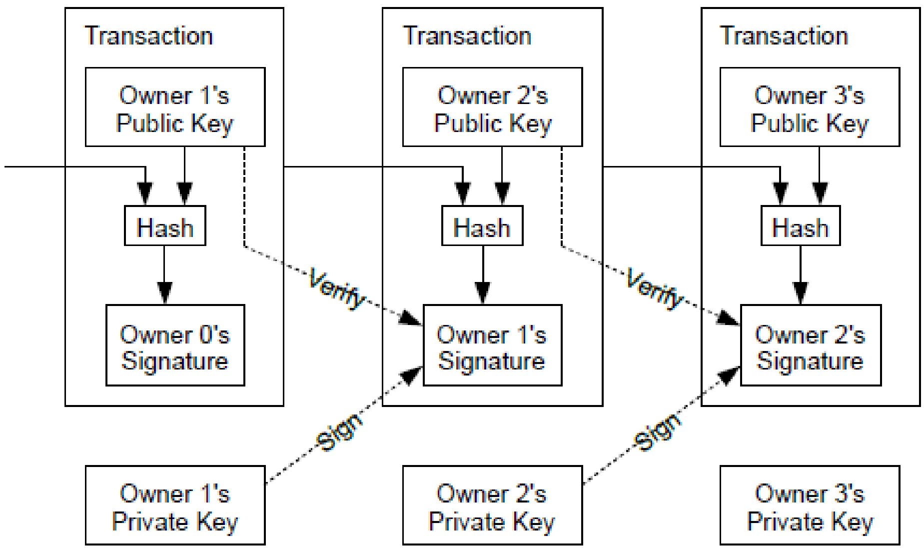
Source: Amended from Abboushi (2016).

2.2 Bitcoin

As seen in Table 1, the Bitcoin is considered as a sort of convertible VC, based on cryptography and a decentralized system. This category of currency is known as Cryptocurrency (CC). Thus, the Bitcoin works using open source peer-to-peer system that was created by a group or person under the pseudonym of “Satoshi Nakamoto” (NAKAMOTO, 2008).

According to Abboushi (2016), a disruptive characteristic of decentralized CCs is its accounting system. Because, while the participants in cryptocurrency are anonymous (or pseudonymous), their transactions are not. The transactions are registered in a distributed ledger and is transparent to all participants without revealing their identities. As detailed in Nakamoto (2008), each coin is defined as a chain of digital signatures, where each owner transfers the coins to the other using a digital signature which contain a hash of the previous transactions and the public key of the next owner, adding this information to the end of the coin. This signature can be verified by the holder to prove the chain of ownership.

Figure 1 – Blockchain structure.



Source: Nakamoto (2008).

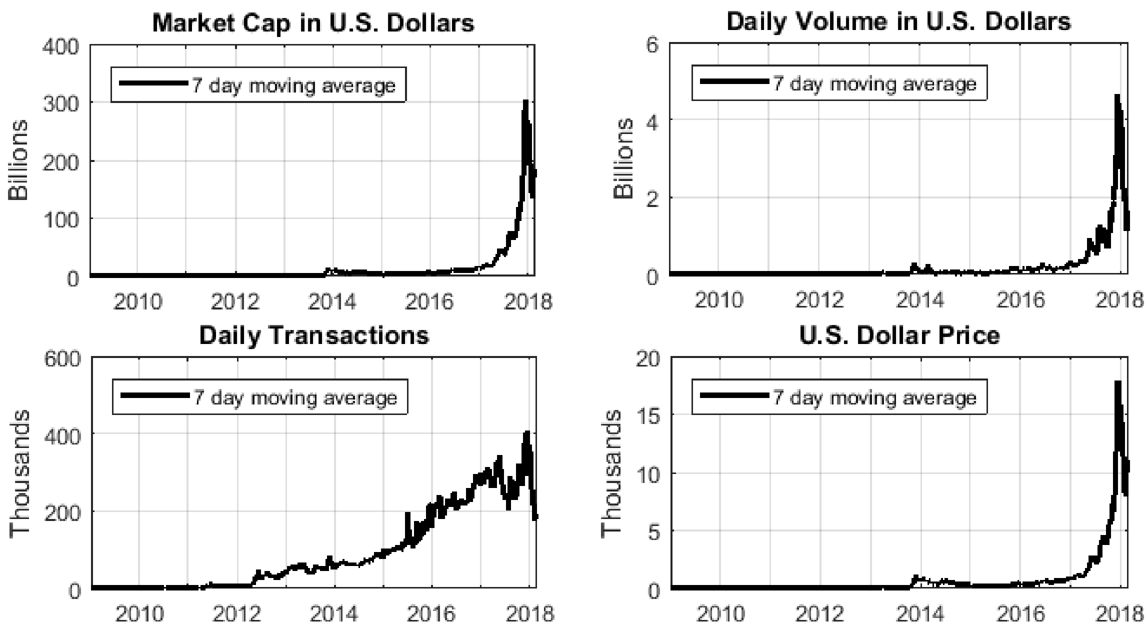
The Bitcoin network is composed by a high number of computers connected through the Internet. In order to avoid the need of a trusted party to validate transactions, it was implemented a proof-of-work mechanism where the nodes or “miners” perform complex mathematical procedures to verify the correctness and truthfulness of the transactions (COCCO; CONCAS; MARCHESI, 2017). Thus, the “miners” compete with others, through resolving complex mathematical problems to do the task of collecting new transactions, validate and group them into “blocks” of transactions, and assign a cryptographic hash to connect (chain) them to previous blocks. For this reason, the ledger is called Blockchain, which is a chain of blocks of verified transactions (ABBOUSHI, 2016). Moreover, it is important to emphasize that Bitcoin “blocks” or transactions are irreversible (COCCO; CONCAS; MARCHESI, 2017). In Figure 1, it is presented the schematic representation of the Blockchain structure.

As an incentive for the “miners”, the first transaction in a block is a special transaction that creates new coins (currently, 12.5 bitcoins) owned by the responsible for generating the block. This represents a similar scheme to the gold miners (NAKAMOTO, 2008). Besides, the system provides a limited total amount of money in circulation, equal to 21 million of Bitcoins. Consequently, this action avoids the risk of increasing the number of coins and generating inflation (COCCO; CONCAS; MARCHESI, 2017).

2.2.1 Market Evolution

Even the CCs market is small compared to traditional currencies (according to *http://money.visualcapitalist.com* until February 2018, the total value of currencies is USD 7 trillions). However, it can be mentioned that, in recent years, this market has grown significantly. In particular, the Bitcoin market has been an accelerated growth.

Figure 2 – Bitcoin market evolution (2009 - 2018).



Source: Amended from McIntyre & Harjes (2016) using data from *https://www.quandl.com*.

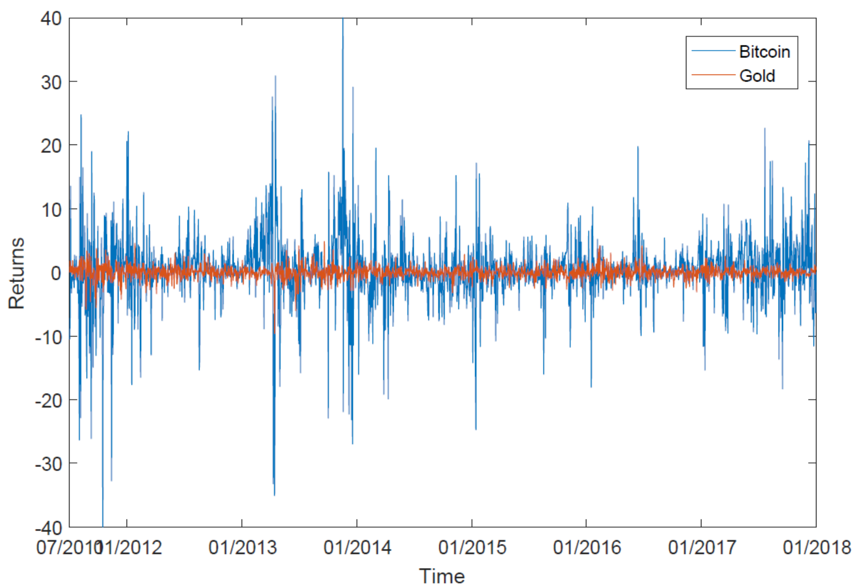
In Figure 2, it is presented the evolution of the Bitcoin market, where it can be observed its explosive acceptation between the years of 2013 and 2017. In Alstyne (2014), the author present four reasons for the Bitcoin success:

- avoid double-spending;
- Blockchain technology enables near friction-less;
- frauds are easily detected; and,
- the Bitcoin has value because people accept it.

Besides, it is important to mention that the acceptance of the Bitcoin by people is usually due to aspects related to the economic situation or the level of confidence with respect to traditional currencies.

Relating to the trading market composition, based on the historical data shown by McIntyre & Harjes (2016) and Kim (2017) and on what is observed until the end of February 2018, approximately 80% of all Bitcoin trading was realized between itself and the US dollar (US$). Additionally, it is possible to mention other important currencies such as the Euro (EUR) and the Chinese Yuan (CNY).

Figure 3 – Daily return series of Gold and Bitcoin (2012 - 2018).



Source: Klein, Pham Thu & Walther (2018).

In the financial world, one of the characteristics that worries operators is the volatility, which is understood as the standard deviation of the logarithmic returns of the price exhibited by an investment instrument. Thus, as mentioned by Yermack (2013) and

Klein, Pham Thu & Walther (2018), in comparison with the traditional currencies, the Bitcoin presents a volatility approximately twice greater referring to its value of exchange rate. For example, Figure 3 shows a comparison between the price volatility of the Bitcoin and the gold.

2.3 Forecasting and Time Series

According to Montgomery, Jennings & Kulahci (2015), the *forecasting* is an effort for predict some future event(s), and exist three type of forecasts, which are based on the prediction period: short-term (days to months), medium-term (1 to 2 years) and long-term (many years). Because the historical data usually has inertia, it is possible to identify, model and extrapolate patterns to perform predictions.

In order to realize short-term and medium-term forecasting, it is necessary to use time series data. The term *time-series* is defined as a chronological sequence of observations about an object or event of interest (MONTGOMERY; JENNINGS; KULAHCI, 2015). Thus, it can be identified the following elements (GEURTS, 2001):

- An universe U of objects or events, where each object or event o is observed in a frame of time $[0, t_f(o)[$;
- Exists candidate attributes (or relevant features) that they can describe the object or event o. Thus, the function $a_i(o, t)$ defines the value of an attribute i for the specific object or event o at the time t.

Based on the above definition, it is possible to define an object or event with one or more attributes. In the first case, it is known as an *univariate* problem and, in the last case, it is called as a *multivariate* problem. Thus, time series analysis can reveal patterns such as randomness, similarities, trends, level shifts, periods, non-common observations, or a combination of the above (MONTGOMERY; JENNINGS; KULAHCI, 2015; NANOPOULOS; ALCOCK; MANOLOPOULOS, 2001).

The *forecasting* knowledge is applied in many areas because the prediction of future events is a critical information for decision-making process (MONTGOMERY; JENNINGS; KULAHCI, 2015). In particular, for the financial area, it is used to minimize investment risks (LABOISSIERE; FERNANDES; LAGE, 2015).

2.3.1 Types of forecasting

Based on the definition given by Geurts (2001), an object or event can be described through a set of attributes. This description can be represented by $y_{t'}$, where t' represents the *forecast horizon* or future time. Thus, the time series problem can be defined as $y_{t'} = f(a_i(o,t))$, where f is a function that represents the pattern presented in the observations.

In general, the task of predicting time series can be classified according to the type of value that is intended to predict ($y_{t'}$). Therefore, two types of quantitative forecasts on time series can be identified:

- *Regression* – when $y_{t'}$ is real-valued. As mentioned by Murphy (2014), another sort, known as *ordinal regression* occurs, where label space Y ($y_{t'} \in Y$) has some logical ordering, such as grades A–F.
- *Classification* – When $y_{t'}$ is a categorical value, where $y_{t'} \in \{c_1, ..., c_M\}$ and M represents the number of classes (GEURTS, 2001).

In relation to the identification of patterns from the data, as mentioned by Nanopoulos, Alcock & Manolopoulos (2001), there are other categories, such as:

- *Generalization* – when it is necessary to generate a simple description and identify associations or rules from complex data;
- *Clustering* – when it is necessary to identify a set of subsets within the data that can be categorized.

2.4 Machine Learning

According to Murphy (2014), *machine learning* is defined as a set of techniques that allows the learning or identification of patterns that are present in the data; with this information it is possible to try to predict future events or carry out some other type of decision-making support in uncertain scenario. Thus, the methods proposed by this area of knowledge are very useful tools to solve the problems raised in the previous section. Moreover, as mentioned by Duda, Hart & Stork (2012) and Laboissiere, Fernandes & Lage (2015), it is possible to find many applications for machine learning, from speech recognition, fingerprint and face identification, optical character recognition, DNA sequence identification, market stock predictions and so on.

Thus, it can be mentioned the most known algorithms within this set of techniques, genetic algorithms, artificial neural networks (ADELI; HUNG, 1994), Support Vector Machine (DUDA; HART; STORK, 2012), etc.

2.4.1 Types of machine learning

Usually, machine learning problems can be divided into 3 categories based on the type of learning or adaptation (DUDA; HART; STORK, 2012; MURPHY, 2014):

1. *Supervised Learning* – it is also known as *predictive learning*. In this case, from training data (T) containing n samples of pairs of input (X_i) and output (y_i) of the form $T = \{(X_i, y_i)\}_{i=1}^{n}$, it is desired to extract the patterns that map the output;

2. *Unsupervised Learning* – in this case, there is no previous definition of classes that identify the category or type of pattern within the data. Thus, only the features are available $T = \{X_i\}_{i=1}^{n}$ and the goal is to find "interesting or natural patterns" in the data;

3. *Reinforcement Learning* – it is also known as *learning with a critic* and is based on reward or punishment signals for teaching feedback when the tentative category is right or wrong, respectively. Thus, no desired category signal is given.

Initially, the time series problems were addressed by the statistics area. However, in recent years, the Machine Learning approach has increased due to its success in the identification of patterns. Thus, in Table 2 it is possible to observe a comparison of some terminology related between *machine learning* and statistics areas.

Table 2 – Machine learning vs statistics terminology comparison.

Machine learning	Statistics
network, graphs	model
weights	parameters
earning	fitting
generalization	test set performance
supervised learning	regression/classification
unsupervised learning	density estimation, clustering

Source: Amended from Tibshirani (2011).

2.4.2 Batch and Online Approach

With the increase in speed, variability and volume of data, the challenge faced by machine learning techniques is the ability to process this type of data and identify their patterns, within the types of approach to processing information:

- *Batch Learning* – it is the most traditional approach and assumes that a finite data set (T) is available. This data set is used to train and test the models. Additionally, it assumes that these data present a stationary behavior or distribution, which means that the patterns found are usually uniform to the long of the time. Therefore, based on this approach, if it is necessary to learn from new data, in most cases, the models need to be retrained;

- *Online Approach* – unlike the previous approach, as mentioned by Gama (2010), it is assumed that the data are available in an unlimited streams form, with continuous flow and, in some cases, at high-speed. The distribution of the data over time is commonly non-stationary. Thus, an incremental learning approach is important, but not sufficient. It is necessary to incorporate the concept of drift, forgetting outdated data and fit with the most recent state. Thus, data stream solutions, according to Uriarte-Arcia et al. (2015), include algorithms which work with a limited amount of time and computational capacity and process instances one or only few times.

2.4.3 Ensemble Models

Until this point, it was mentioned models based on data. However, as in real life, in many cases, it is useful to consult different experts on the same subject in order to make a more precise decision. According to Dietterich (1997), the term *ensemble* or *multiple* models represent a set of models where their individual decisions or base learners are combined to predict or identify patterns in new samples. So, many methods to construct ensemble models have been developed, such as:

1. *Bayesian Voting (Enumerating the Hypotheses)* – the individual model error factor is used for construct a conditional probability. Thus, it is used the Bayes rule, where the posterior probability is proportional to the likelihood of the training data (S) times the prior probability of each individual model (h), i.e., $P(h|S) \approx P(S|h)P(h)$;

2. *Manipulating the Training Samples* – the learning process or training process is run several times, each time with a different subset of the training data. This is good for unstable learning algorithms like neural networks, decision trees or rule learning models. Thus, *Bagging* consists of a sample of m training data drawn randomly with replacement from the original training set of m items for training the base learners. Other similar technique is *AdaBoost*, where it maintains a set of weights over the training data and more weight is given for instances that were misclassified by earlier rounds. Finally, *Stacking* is a technique that creates base level models with the complete training set. Then, the final model is trained on the outputs of the base level model as features;

3. *Manipulating the Input Features* – in this case, each base learner is trained using different subsets of input attributes or features of the training data set. Thus, this technique works when the input features are highly redundant;

4. *Manipulating the Output Targets* – in case of classification, if the number of classes (K) is large, then it is possible to construct partitions of K classes into two subsets A_i and B_i. Thus, each input data is relabeled so that any of the original classes in

set A_i are located to the derived label 0 and the original classes in the set B_i are located to the derived label 1. This process is repeated L times and generates L models (h_i). Therefore, given a new data point x, if $h_i(x) = 0$, then each class in A_i receives a vote. Otherwise, the vote is given to B_i. Finally, after L models has voted, the class with the highest quantity of votes is selected as the prediction of the ensemble;

5. *Injecting Randomness* – this method consists on inject randomness into the learning algorithm of base learners. For example, considering the C4.5 algorithm, in the feature test process it is possible to select randomly (with equal probability) the feature of top n best tests.

2.4.4 Additional concepts

In addition, some concepts that will be used in later chapters of this work will be defined:

- *Parametric and non-parametric models* – the *Parametric* models have a fixed number of parameters. For this reason, the generated models have the advantage of often being faster to use, but the disadvantage of making stronger assumptions about the nature of the data distributions. On the other hand, the *Non-parametric* models have non-fixed number of parameters because these grow with the amount of training data. Thus, these models are more flexible, but often computationally intractable for large data sets (MURPHY, 2014);

- *The curse of dimensionality* – if the number of features or dimensions grows, the quantity of samples in the training data set need to grows exponentially for generalize accurately. For example, if 10 instances of an entry with one dimension are available and it is increased to two dimensions, then 100 samples are required, that is, 90 additional instances (MURPHY, 2014).

- *Feature Selection* – in order to reduce the high dimensionality, it was necessary to use methods to select the more relevant attributes. The best-known attributes selection techniques are described:

    1. correlation analysis (*corr*) – this function evaluates the value of an attribute by measuring the cross-correlation (Pearson's coefficient) between it and the class (MURPHY, 2014);

    2. relief technique (*relief*) – this method estimates the worth of an attribute by repeatedly sampling an instance and considering the value of the given attribute for the nearest instance of the same and different class (GAO et al., 2014);

3. information gain method (*info*) – this technique assess the value of an attribute by measuring the gain of information relative to the class using the concept of entropy (DAI; QING, 2013);
4. principal component analysis (*pca*) – reduces dimensionality by choosing sufficient eigenvectors to explain a percentage of the variance in the original data (95%). As a consequence, new attributes are calculated from the original (WANG et al., 2016; KIM; RATTAKORN, 2011);
5. correlation-based feature subset selection (*cfs*) – evaluates the value of a subset of attributes, considering the individual predictive capacity of each feature along with the degree of redundancy between them (HALL, 1999);

- *Validation Metrics* – to validate the predictive capacity of the models, metrics are used, which vary depending on whether it is a classification or a regression. The most used for each case are:

  1. Classification – basically, define the success of predictions counting number of success classification versus errors.

     – Accuracy is considered the more simple metric and is defined as:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\%. \tag{2.1}$$

where TP = True Positive; TN = True Negative; FP = False Positive; and FN = False Negative.

     – Area Under ROC Curve (AUC) metric tries to avoid bias of majority class. Thus, is composed by the sensitivity and specificity metrics.

$$Sensitivity = \frac{TP}{TP + FN}, \tag{2.2}$$

$$Specificity = \frac{TN}{TN + FP}, \tag{2.3}$$

Thus, it is calculated the area under the curve generated Figure 4. If its value is greater than 0.5, then it is considered a not randomly classification and the score 1.0 means a perfect accurate.

     – F1-Score is similar to AUC, because try to measure the quality of the classification, avoiding bias of majority class.

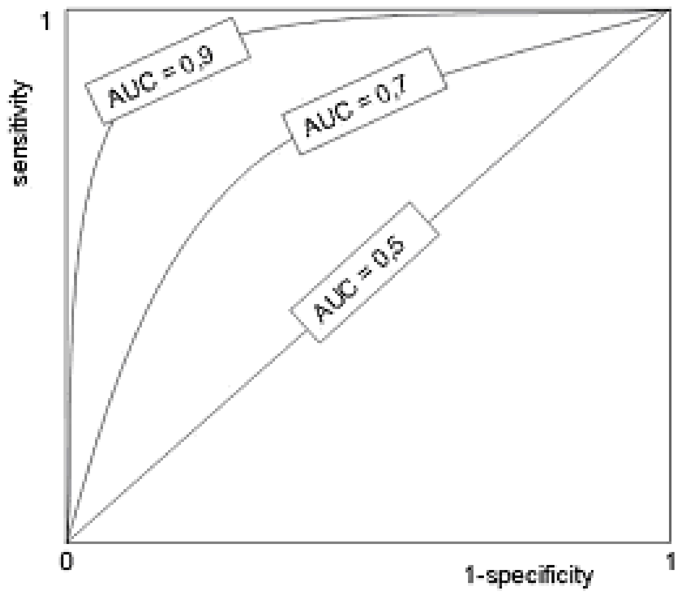
$$F_{score} = 2 \cdot \frac{Precision \times Recall}{Precision + Recall}, \tag{2.4}$$

where $Precision$ and $Recall$ metrics are defined as:

$$Precision = \frac{TP}{TP + FP}, \tag{2.5}$$

$$Recall = \frac{TP}{TP + FN}. \tag{2.6}$$

Figure 4 – Area under the ROC curve.



Source: https://acutecaretesting.org

– Cohen's kappa or *Kappa* is a statistical metric and define if the classification is randomly or not.

$$Kappa_{\kappa} = 1 - \frac{1-\rho_o}{1-\rho_e}, \tag{2.7}$$

where ρ_o is the observer level of agreement (empirical probability) to assign the label to any sample; and ρ_e is the expected/hypothetical probability of agreement to randomly assign the label.

2. Regression – try to measure mean error on the regression process.

– Mean Absolute Error (MAE):

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|Y_i - \overline{Y_i}\right|, \tag{2.8}$$

– Mean Absolute Percentage Error (MAPE):

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\frac{\left|Y_i - \overline{Y_i}\right|}{Y_i}, \tag{2.9}$$

– Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \overline{Y_i}\right)^2}, \tag{2.10}$$

- *Overfitting* – this phenomenon occurs when it is seek to model every minor variation in the input. Since, this is more likely to be noise than true signal. Thus, the fit of the model with the training data is high, but the fit or accuracy with unobserved data is low Murphy (2014);

- *No free lunch theorem* – in Wolpert (1996), this term is used for the first time. The author claims that not exists a universally best model, because the necessary set of assumptions that works well in one sort of problem may work poorly in another.

3 Literature Review

3.1 Research Methodology

A systematic review of the literature was made based on the steps detailed on Khan, Kunz & Klerjnen (2003). It is possible to summarize the steps carried out according to as follow:

1. First, the basic objectives of the research were established, which were: (i) understand the main components of the operation of the Bitcoin; (ii) identify the factors that affect the price of this Cryptocurrency (CC); (iii) identify methodologies/models used for the Bitcoin price prediction and other types of investment instruments;

2. After that, two search engines of information were used, the first was the Periodicos Capes Website and second one was the Google Scholar Website;

3. In addition, the papers were evaluated according to: (i) their contribution with the research questions mentioned above; (ii) if they had peer review; (iii) their year of publication; and (iv) the number of citations obtained;

4. Finally, each of them was organized extracting the most relevant information to design and compare the experiments to be carried out in the present study.

Therefore, the present study argues that the aforementioned literature offers possibilities to explore a new set of attributes and different configurations of machine learning techniques to improve both the direction prediction and the maximum, minimum and closing price prediction for the Bitcoin in daily and intra-daily time frequencies.

3.2 Related Studies

3.2.1 Bitcoin and its economic nature

Regarding to the nature of the Bitcoin, in Kristoufek (2013), the author argue that it presents characteristics as a standard financial asset and a speculative instrument at the same time. In Popper (2015), it is considered as a commodity or digital gold. Thus, in Iglesias (2015), the author highlights the technological advantages of the Bitcoin, converting it into a relevant alternative to credit cards and traditional bank transfers.

Unlike the previous study, which was focused on the advantages of the Bitcoin as a currency, in Glaser et al. (2014), the authors present solid indications that new users are not primarily interested in its transaction advantages, but to participate in a new investment vehicle. Similar to the previous study, in Wu & Pandey (2014) and Li & Wang (2017), the behavior of the Bitcoin is analyzed based on their capabilities as an investment instrument. Thus, this capability will be explored in the present study.

3.2.2 Relevant attributes for the Bitcoin price prediction

Relating to the identification of relevant attributes for the Bitcoin price and trend forecasting, there are a variety of studies (KRISTOUFEK, 2013; KRISTOUFEK, 2015; MATTA; LUNESU; MARCHESI, 2015; CIAIAN; RAJCANIOVA; KANCS, 2016; KIM et al., 2016; LI; WANG, 2017; BALCILAR et al., 2017; ZHU; DICKINSON; LI, 2017; VASSILIADIS et al., 2017; EROSS et al., 2017). In these papers, it is possible to identify the following kinds of relevant attributes:

- *Internal Features – Blockchain data* (e.g. OHLC prices, volumes, mining difficult and validation fees);
- *External Features – Economic fundamentals* or international indices (e.g., S&P500, NASDAQ, DAX, Dow Jones index, crude oil and gold prices) and public recognition or *social trends* (e.g., Google searches, Wikipedia searches and Twitter mentions).

In Kristoufek (2013), the relationship between the Bitcoin price and search queries on Google and Wikipedia is analyzed by the author for the period that starts on 2011 and ends on 2013. The results demonstrate that the volume of searches on Google and Wikipedia is statistical representative, specifically when the price of the Bitcoin has increased (i.e., a possible bubble behavior). The same author, in another paper Kristoufek (2015), analyzes long-term and short-term correlations with different sorts of factors (from 2011 to 2014). This study considered several aspects that could influence the Bitcoin price: economic or fundamental drivers, transaction drivers, technical drivers, public interest for the CCs and the effects produced by Chinese Bitcoin market. Thus, the author argues that the Bitcoin price is positively affected in long-term if it is used more for trade (e.g., non-exchange transactions). Moreover, when its price increases, then boosts the exchange transactions in the short-term. The first is a result from the economic theory and the second explains potential bubbles. Finally, the price level of dollar affects negatively the Bitcoin exchange rate in long-term. Therefore, the Bitcoin behavior does not contradict the standard monetary economics in the long-term.

In other research, Matta, Lunesu & Marchesi (2015) analyzed whether social media activity or information extracted from web search media could be helpful to predict the behavior of the Bitcoin price. As a result, Google Trends could be seen as a sort of predictors, because of its high cross-correlation.

In Ciaian, Rajcaniova & Kancs (2016), the authors present three factors that affect the price of the Bitcoin: (i) supply-demand interactions; (ii) attractiveness for investors; and (iii) global macroeconomic and financial indicators. The proposed experiments used daily data from 2009 to 2014. The authors showed that supply-demand factors have an important impact on the Bitcoin price, particularly the size of the Bitcoin economy and the velocity of circulation. In addition, they cannot reject the hypothesis that speculations are also influence the price. Meanwhile, the effect of macroeconomic indicators, compared with the first two factors, becomes statistically insignificant.

User comments obtained from CCs communities are analyzed by Kim et al. (2016). The authors predict the fluctuations in the prices of CCs and in the number of transactions. This way, it was possible to identify the types of comments most relevant for the predictions of the Bitcoin, Ethereum and Ripple. Furthermore, the simulated investment demonstrates that the proposed method is applicable to CC trading. In addition, based on the predictions, they made simulated investments achieving a higher return than a random investment. In the case of the Bitcoin, the analyzed data correspond to the period that starts on 2013 and ends on 2016.

In an experimental study, Li & Wang (2017) used daily data from 2011 to 2014, where the authors suggested that the determinant factors of the Bitcoin exchange rate are classified in technical (hash rate and public recognition) and economics (economic fundamentals and trading volume). Specifically, in long-term models, the exchange rate shows a significant reaction to economic fundamentals (including money supply, gross domestic product, inflation, and interest rate) and, in the short-term, it responds promptly to changes in hash rate and public recognition (Google searches and Twitter mentions).

In Balcilar et al. (2017), the authors focused on the relation of the Bitcoin price returns and volatility with the trading volume, considering a period from 2011 to 2016. The results show that trading volume can predict returns (when the market is operating around the median values), but not volatility. However, when the market is operating with strong highs or lows (potential bubbles), trading volume information is irrelevant.

Another experimental study, conducted by Zhu, Dickinson & Li (2017), analyzed how economics fundamentals influence the Bitcoin price using daily information from 2011 to 2016. The follow factors were examined: Custom price index, US dollar index, Dow Jones index, Federal Funds Rate and Gold price. Thus, the authors argued that all variables analyzed has a long-run influence, where the US dollar index has the most importance and gold price has the least relevance.

In Vassiliadis et al. (2017), it was collected data from 2013 to 2015, because the authors argue that earlier Bitcoin prices and transactions showed a high frequency. The authors show that the Bitcoin price has a strong cross-correlation with the number of transactions and transaction fees. In addition, in contrast to other studies, a good cross-correlation with gold and crude oil price and a moderated cross-correlation with contemporary stock market indices (such as NASDAQ, DAX and S&P500).

Finally, in Eross et al. (2017), it is examined the importance of intra-daily variables correlated with the Bitcoin exchange rate. In that study, the author suggests that volatility and the supply-demand differential are closely related, which is probably a result of the Bitcoin market being still immature. Likewise, these intra-daily variables are highly correlated, have significant delay relationships and high bilateral Granger causality.

3.2.3 Bitcoin trend prediction

On works related to predicting the Bitcoin exchange rate direction, in an empirical study realized by Mcnally (2016), the author used Open-High-Low-Close price (OHLC) data from CoinDesk Website and the hash rate taken from Blockchain (1066 instances – using 80% as training data and the remaining 20% as test data). These data are normalized (mean equal to 0 and standard deviation equal to 1) and used to obtain the Simple Moving Average (SMA), which can improve the capacity of the model to recognize trends by smoothing the data. In addition, all extracted attributes were used as inputs of deep learning models: Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. It was observed that LSTM achieves the highest classification accuracy, about of 52%.

In addition, another similar study has been conducted on predicting fluctuations in the price and number of transactions of three relevant CCs (Bitcoin, Ethereum and Ripple) Kim et al. (2016). Thus, it was used comments (from people) in online CC communities (Bitcoin Forum, Ethereum Community Forum and Ripple Forum). In this study, a total of 793 instances were divided into 88% to train and 12% to test the model. These data were tagged in positive or negative, using VADER engine. However, it was determined 5 categories: very positive, positive, neutral, negative and very negative. Fluctuations in the Bitcoin price demonstrate to be significantly correlated with the number of positive/very positive comments and with positive replies. Granger causality test was used to get the maximum accuracy of 79.57%, an f1-score of 0.796 and Matthews correlation coefficient of 0.606.

Related to intra-daily trend prediction, in Madan, Saluja & Zhao (2015), the authors used 10 minute time intervals and Blockchain network-based features. In that work, it was modeled the price prediction problem as a binomial classification task. For this purpose, it was used Random Forests (RFs) and Generalized Linear Models (GLMs). The results had 57.40% of accuracy in predicting the sign of future price change, where RF performed better than the GLM. Similar to the above study, in Greaves & Au (2015), Blockchain network-based features are used and it is predicted the "up or down" Bitcoin price movement in hourly intervals with an accuracy of roughly 55%, where Artificial Neural Network (ANN) is performed better than linear models as Support Vector Machine (SVM), logistic regression and linear regression.

3.2.4 Machine learning applied to time series prediction

In terms of time series prediction (that is the context of this work), in Kara, Acar Boyacioglu & Baykan (2011), the authors predict the direction of movement for the National Index 100 of the Istanbul Stock Exchange (ISE) based on the use of technical features used by traders. In that experiment, two classification models are compared: ANNs and SVMs, where the average performance of the ANN was (75.74%), being better than the SVM model (71.52%).

In Patel et al. (2015), it was predicted the direction of daily movement of stock and the stock price index for Indian stock markets. The authors used two approaches for generate classification features: (i) computation of ten technical features using stock trading data (OHLCs); and (ii) represent these features as trend deterministic data, named as (*discretized* version). That study use ANNs, SVM, Naive Bayes and RF models. The experimental results show that the performance of all the prediction models improve with second approach, where Random Forest had the better results.

Related to the forecast of stock prices (*regression* problem), in Laboissiere, Fernandes & Lage (2015) it was used the OHLC and international economic indexes. The authors proposed a methodology for feature transformation and selection based on Weighted Moving Average (WMA) and correlation index, respectively. In that experimental study, the maximum and minimum stock daily prices for Brazilian distribution companies was predicted with a Mean Absolute Percentage Error (MAPE) between 0.6% and 2.1% using ANNs.

In Qiu & Song (2016), the authors compare two sort of technical indicators used by traders to predict the direction of the daily stock market index. Thus, in that experiment an optimized ANN model with Genetic Algorithms (GAs) is performed. Based on the scores obtained with traditional (*Type 1*) and non-traditional technical (*Type 2*) features, the results show that the *Type 2* input variables can generate a higher forecast accuracy.

One of the challenges in the prediction of time series related to stock prices is the non linear relation of its input variables. In this sense, Kocadağli & Aşikgil (2014) propose a sort of ANN model (Bayesian neural networks) for the time series forecasting. The authors argued that this approach provides a natural way to model the non linear relation as an ANNs. Likewise, in Rather, Agarwal & Sastry (2015) is used a hybrid model that combines Autoregressive-Moving-Average (ARMA) model, exponential smoothing model and a non linear model based on RNNs. In the same way, Cramer et al. (2017) show the benefit of machine learning algorithms

Ensemble models were explored, considering that the error of the ensemble decreases, respecting to each individual classifier or base learners, if and only if each individual classifier has a performance better than a random choice Gama (2010). Ensemble methods are attractive because they can often be more accurate than a single classifier alone Bifet et al. (2011), Das, Bisoi & Dash (2018), specially when exists the concept of drifts Wang (2003), Brzeziński & Stefanowski (2011). Likewise, it is possible to quote the work of Ballings et al. (2015), where the author conducted an experiment to create a benchmark of ensemble methods (Random Forest, AdaBoost and Kernel Factory) against single classifier models (Neural Networks, Logistic Regression, Support Vector Machines and K-Nearest Neighbor) for prediction of stock price direction.

In Jang & Lee (2018), the authors also select relevant attributes. However, the prediction was done by using a Bayesian Neural Network (BNN). Thus, the BNN results were compared to a Support Vector Regression (SVR) and linear models. The time series covers the daily Bitcoin data from Sep 11, 2011, to Aug 22, 2017. Based on this data set, the BNN was parameterized, trained and tested to predict the log price and the log volatility of the Bitcoin price. The results obtained present MAPEs equals to 0.0198 and 0.6302 for log price and log volatility, respectively.

The paper of Peng et al. (2018) combines a Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) model with the SVR. Thus, it was evaluated its performance for CCs (Bitcoin, Ethereum and Dash market price) and traditional currencies (Euro, British pound and Japanese yen). All of them were considered in US dollars. Moreover, it was used low (daily) and high (hourly) frequency data to predict the volatilities. The authors show that the errors, Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), obtained from high frequency data are much lower than for low frequency data.

About data-stream learning algorithms, Domingos & Hulten (2000) proposed the Very Fast Decision Tree (VFDT) based on Hoedffing Trees (HTs). It can execute thousand of instances using few computational resources with a performance similar to a batch decision tree given enough observations Gama, Fernandes & Rocha (2006). Recently, there are studies that use data-stream learning techniques to predict stock price direction Gao & Lei (2017), Lin, Chen & Chen (2017), because with this approach the model continuously evolves over time, is ideal for non-stationary time series and requires only a small retrain time per new data sample in comparison with batch learning Gao & Lei (2017).

4 Proposed Methodology

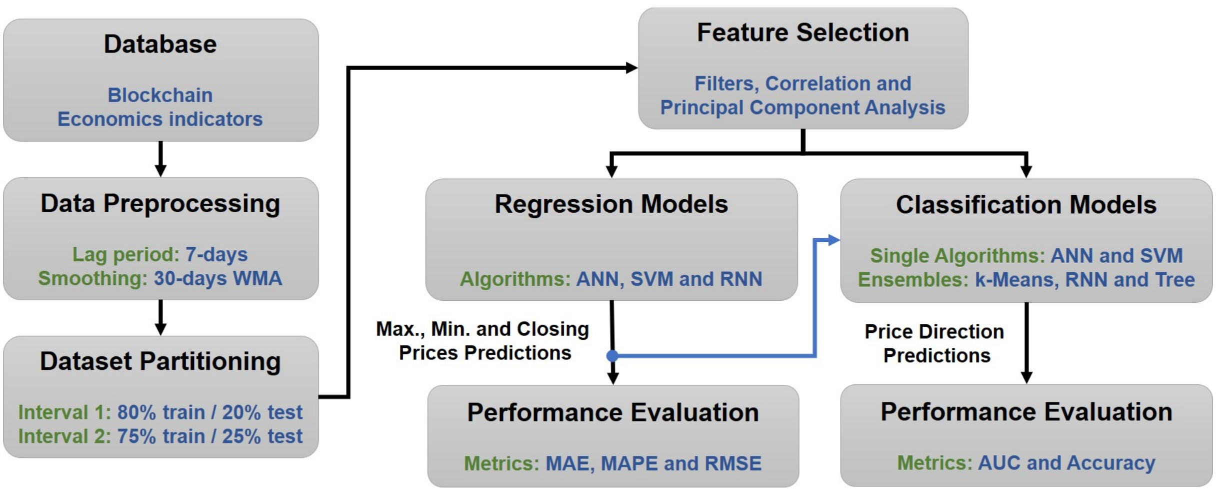
Two sorts of experiments were carried out to predict price movement of Bitcoin. First type considers a daily frequency of the price, meanwhile second performs intra-daily predictions (each 10 minutes).

4.1 Daily Prediction

4.1.1 Experiment 1 – Exploratory Approach (E1)

First, a exploratory methodology is conducted, where is focused on identify the best feature selection methods and evaluate the prediction capacity of Support Vector Machine (SVMs) and Artificial Neural Networks (ANNs) models with single and ensemble approaches. Thus, the methodology proposed can be visualized on Figure 5, which summarizes the: (1) the sources where the data were collected; (2) transformations used for data pre-processing; (3) distribution of the data partitioning for training and testing purposes; (4) sort of attribute selection methods applied; (5) application of machine learning techniques to classify/predict the price direction using single and ensemble approaches, including classification by regression to predict the maximum, minimum and closing prices; and (6) finally, the performance evaluation metrics used in each cases.

Figure 5 – Daily prediction E1 – Overview of the proposed methodology.



Source: Personal collection

A Input of data collected

The sources of information can be categorized into internal (the behavior of different parameters of Bitcoin) and external (the economic factors, external demand or information obtained from social networks or specialized forums, also named as public). In this sense, as an internal data source, the Blockchain information is considered in a similar way as suggested by the authors of Mcnally (2016), Balcilar et al. (2017). This information includes Open-High-Low-Close price (OHLC) of Bitcoin, the volume of trades, total transaction fees, number of transactions, cost per transaction and average hash rate.

As a contribution of the present study to the identification of relevant attributes for the prediction of the Bitcoin price trend, external information was considered and obtained from international economic indicators. These indicators were used due to the high correlation identified by Vassiliadis et al. (2017) and the good results obtained by Laboissiere, Fernandes & Lage (2015). Thus, the following indicators were used: crude oil future prices, gold future prices, S&P500 future, NASDAQ future and DAX index.

The OHLC exchange rates (Bitcoin against US dollar) were collected from the website BitcoinChart Website and the remaining internal data were obtained from the website Quandl Website. The external information was collected from Investing Website.

In order to compare the proposed methodology with the state-of-the-art, specifically with the models proposed by Mcnally (2016), a first interval was considered, ranging from August 19th, 2013 to July 19th, 2016. However, a second interval was considered, ranging from April 1st, 2013 to April 1st, 2017.

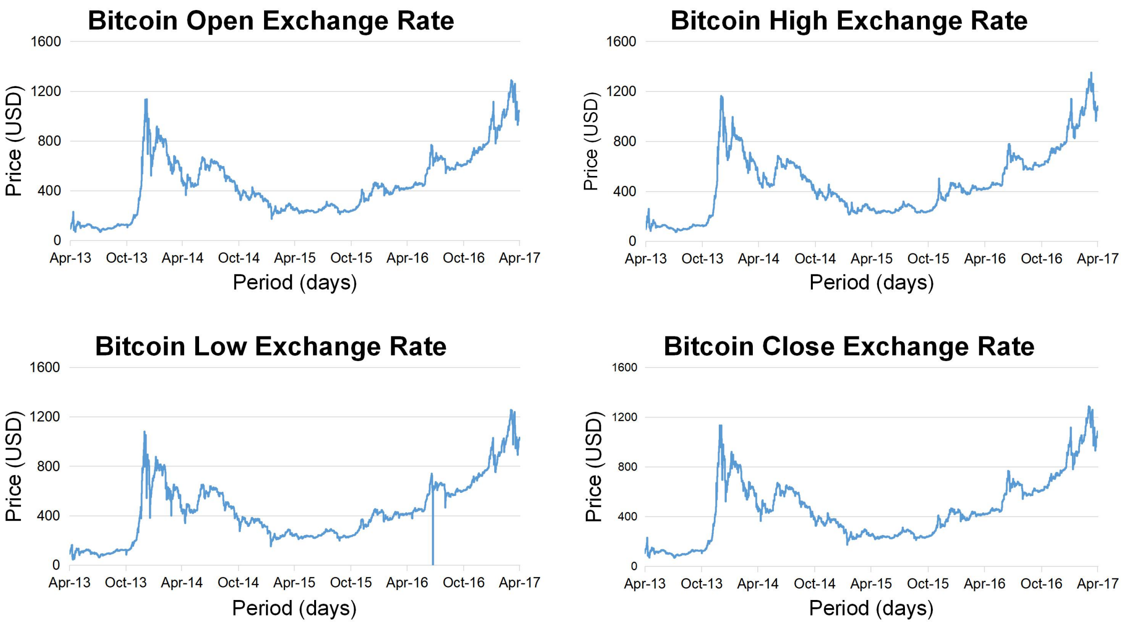
B Data pre-processing techniques

In Figure 6, it is presented the behavior of the Bitcoin exchange rate (OHLC). Thus, it can be highlighted the high volatility, especially for the minimum price (Low price) that presents a high fall in Jun 23th, 2016. This represents that Bitcoin market, in general, is immature yet.

However, it is possible to observe, in Figure 7, that the Blockchain data shows a similar volatility, which can be used to predict the exchange rate.

In this work, the data pre-processing stage suggested by Laboissiere, Fernandes & Lage (2015) was used, i.e., the lag period concept and the smoothing of the data. Thus, in the preprocessing stage, the value “1" was assigned to the class if the closing exchange rate of Bitcoin at a Day (D) is greater than or equal to the previous day ($D-1$). Otherwise, it was assigned the value “0". Unlike the case presented by Laboissiere, Fernandes & Lage (2015), the Bitcoin cryptocurrency is traded every hour and every day. For this reason, it was considered a lag period of 7 days. This way, for each class (“0" or “1"), at time D, it was considered historical data from the previous 7 days as input attributes.

Figure 6 – Daily prediction E1 – Daily Bitcoin exchange rate (OHLC).



Source: Personal collection using data described on Section A

As suggested by Jubert de Almeida, Ferreira Neves & Horta (2018), Patel et al. (2015) and Laboissiere, Fernandes & Lage (2015), a new variable is created from the Weighted Moving Average (WMA) transformation, and due to what is indicated by the previous paragraph, WMA variable was calculated for 30 days to all input attributes. The WMA calculation is used to identify possible trends in the exchange rate, which can be expressed as:

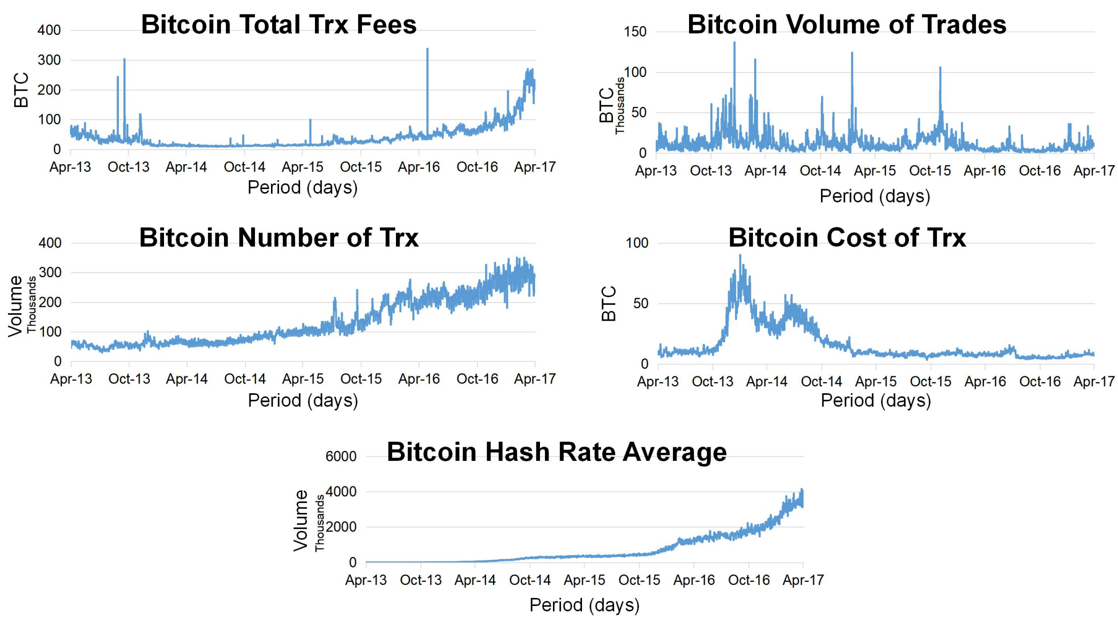
$$WMA_M = \frac{\sum_{n=1}^{M} np_n}{\sum_{n=1}^{M} n}, \tag{4.1}$$

where $M = 30$ due to the number of days considered in the WMA. So, p_n corresponds to the value $M - n$ days before the current day.

In the case of the economic indicators, only the 30-day WMA calculation was considered. Figure 8 shows the historical data for this sort of input data.

After the preprocessing stage, it was obtained a data set composed of the attributes shown in Table 3.

Figure 7 – Daily prediction E1 – Additional daily Blockchain information.



Source: Personal collection using data described on Section A

Table 3 – Daily prediction E1 – List of possible input attributes.

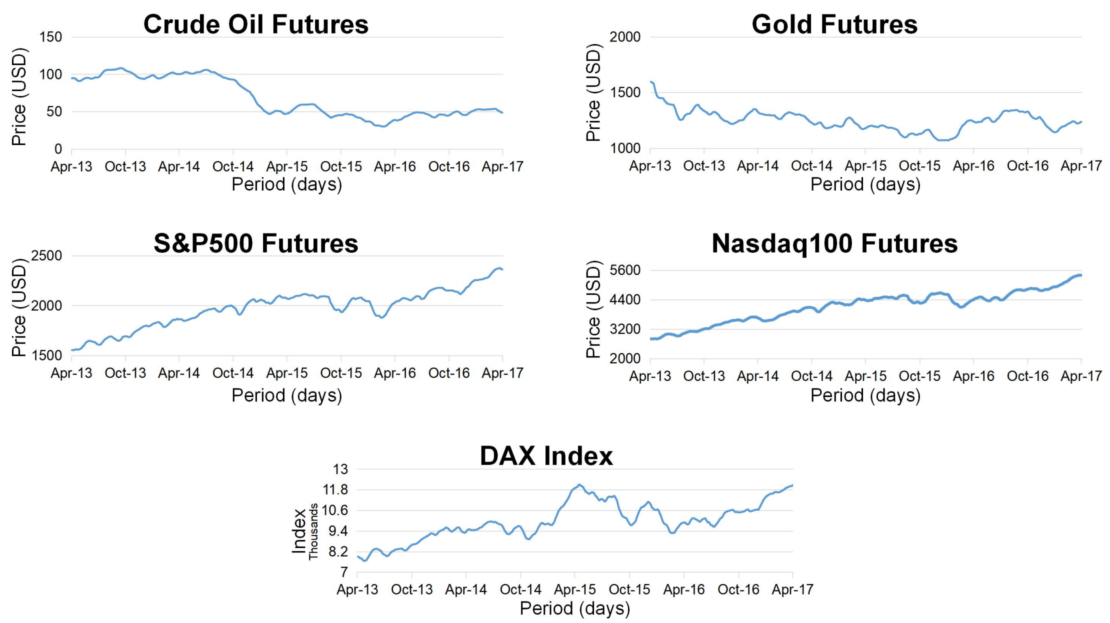
Day D	**Day $(D-i)$**	**30-day WMA**
Opening price	Price direction	Opening price
Timestamp	Opening price	Maximum price
	Maximum price	Minimum price
	Minimum price	Closing price
	Closing price	Volume of trades
	Volume of trades	Number of txn
	Number of txn	Transaction fees
	Transaction fees	Cost per txn
	Cost per txn	Hash rate avg
	Hash rate avg	Closing crude price
		Closing gold price
		Closing S&P500 price
		Closing Nasdaq price
		Closing DAX price

Source: Personal collection.

C Data Partitioning

In order to compare the obtained results with the methodology proposed by Mcnally (2016) and also as it is suggested by other studies as Wang et al. (2011), a data interval (named as *interval 1*) was considered, which considers the same data partitioning (80% of the data for training and the remaining 20% most recent data for validation/test). In addition, a larger data interval was also considered and used to generate a baseline for

Figure 8 – Daily prediction E1 – Daily economic indicators.



Source: Personal collection using data described on Section A

future researches. This larger interval was prepared (named as *interval 2*), which considers 75% of the data for training and the remaining 25% most recent data for validation as is suggested by Laboissiere, Fernandes & Lage (2015).

These data sets, *interval 1* and *interval 2*, were used in the training and validation/testing process of all machine learning algorithms that will be presented in Section E.

D Attribute Selection

As indicated in the previous section, it was considered up to 86 possible input attributes. Therefore, in order to reduce this high dimensionality, it was necessary to use methods to select the most relevant attributes. From the review of the literature, it was verified that the proposal of Laboissiere, Fernandes & Lage (2015) uses the degree of correlation to identify the most relevant attributes for the stock market. However, because the nature of Bitcoin is different from that of a stock market, as cited by Cocco, Concas & Marchesi (2017) and Li & Wang (2017), it was preferred to explore different selection and transformation techniques to reduce the dimensionality of the data set. Thus, five attribute selection techniques were considered, which are mentioned as follows: *correlation analysis*, *relief technique*, *information gain*, *principal component analysis* and *correlation-based feature subset selection*. These methods are explained in Section 2.4.4.

It is important to mention that for all of the five selectors, the 20 best attributes were selected. These algorithms were executed by means of Waikato Environment for

Knowledge Analysis (WEKA) version 3.8.1.

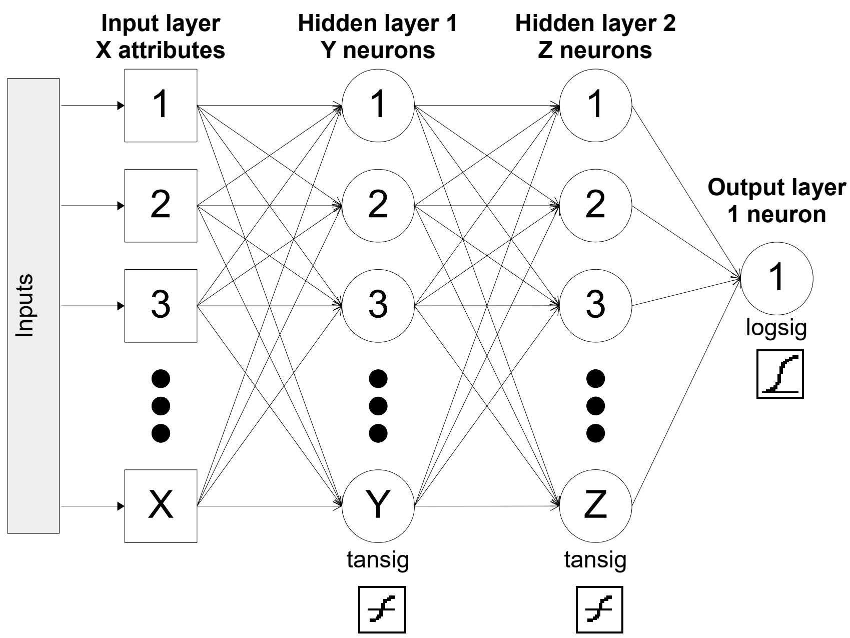
E Soft Computing Algorithms Applied to the Predictions

As can be seen in Section 3.2.3, some machine learning techniques such as ANN and SVM are widely used in stock market predictions. Thus, in this experiment, it was proposed a comparison between such techniques in relation to ensembles that combine regression models with classification and clustering algorithms.

- *Artificial Neural Network*

  As mentioned in the literature review, the ANNs have been widely applied in the forecasting and prediction direction of stock values. Thus, it was used the Multilayer Perceptron (MLP) architecture due to its flexibility and good results presented in other studies Kara, Acar Boyacioglu & Baykan (2011), Kocadağli & Aşikgil (2014), Patel et al. (2015), Ballings et al. (2015), Rather, Agarwal & Sastry (2015), Laboissiere, Fernandes & Lage (2015), Qiu & Song (2016), Cramer et al. (2017). In this exploratory experiment, the hidden layers use the hyperbolic tangent transfer function and, for the output layer, it was used the logistic transfer function, such as presented in Figure 9.

Figure 9 – Daily prediction E1 – ANN/MLP architecture employed.



Source: Personal collection

The scaled conjugate gradient was employed as learning method and *crossentropy* as performance metric (recommended for classification purposes). Thus, several configurations with one and two hidden layers were tested, with combinations of 5, 10, 15, 20, 25, 30 and 35 neurons with a number of epochs ranging from 20 to 500. This algorithm was implemented and parameterized in Matlab® platform.

- *Support Vector Machine*

  The SVM algorithm is based on the principle of minimization of structural risk. Moreover, it estimates a function that reduces the generalization error, demonstrating a resistance to the problem of overfitting. It is important to mention that the SVM is not a stochastic technique. Therefore, if the dataset is not changed, the same result will be always obtained (Huang, Nakamori & Wang (2005)).

  The basic idea is to create a hyperplane that can separate the classes of the problem (Kara, Acar Boyacioglu & Baykan (2011)). Since each sample in each side of the hyperplane have a distance to it, the smallest distance is called the separation margin. The hyperplane is optimal, if the margin is maximized. Therefore, the training process of the SVM consists of finding the optimal hyperplane, that is the one with the maximum distance from the nearest training samples (Duda, Hart & Stork (2012)). In order to avoid the excessive computational cost for calculating the optimal hyperplane, the concept of “soft margin" is used, which establishes a tolerance level (C) to accept samples that are not within the limit established by the hyperplane Bishop (2012).

  In cases where the data are not linearly separable, Cover’s theorem is used, which suggests raising the dimensionality to achieve a linear separation. In this way, the SVM makes use of *kernels* that allow to raise the dimensionality of the data and, thus, achieve to separate them linearly (Bishop (2012)). The *kernels* used in this paper are described in Equations 4.2 and 4.3:

$$Polynomial = (x^T x')^d, \tag{4.2}$$

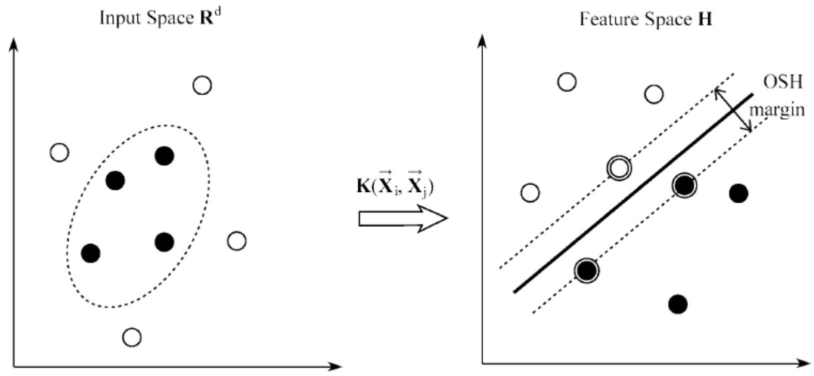
$$Gaussian = \exp\left(-\frac{x - x'^2}{2\sigma^2}\right), \tag{4.3}$$

  where d is the degree and σ is the gamma parameters.

  Thus, when using a polynomial kernel, it will be necessary to define the parameter d, that represents the degree of the polynomial expressed in Equation 4.2. On the other hand, if the model uses a radial kernel, then the standard deviation (σ) must be defined in Equation 4.3.

  It is important to mention that the classical SVM algorithm requires to solve a quadratic optimization problem. In order to avoid the amount of memory needed, it was used the Sequential Minimal Optimization algorithm, described in Platt (1998).

Figure 10 – Daily prediction E1 – SVM hyperplane concept.

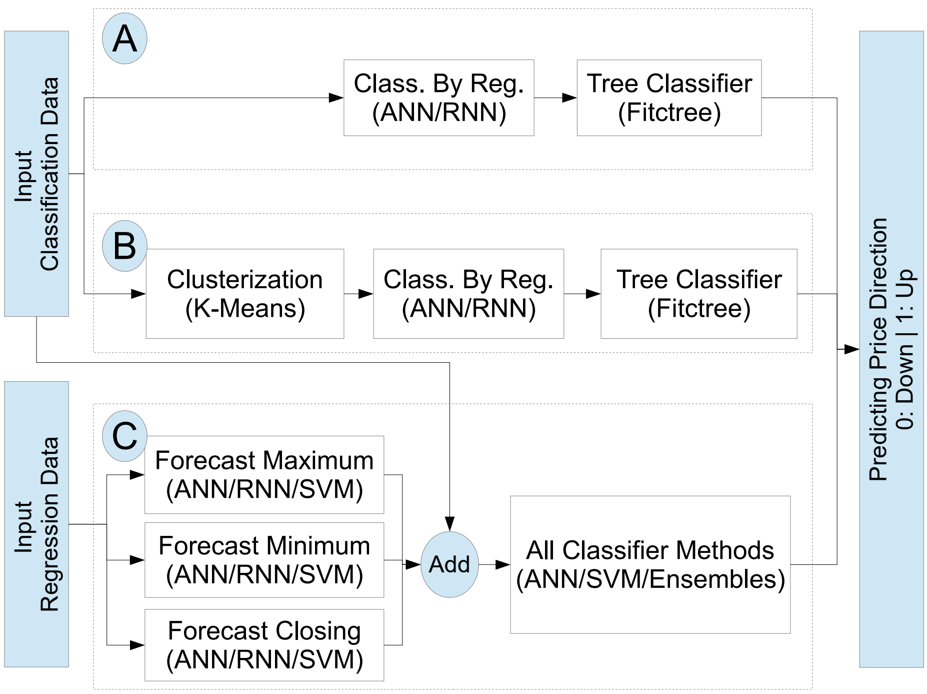


Source: Hua & Sun (2001)

- *Ensembles*

  Three ensembles of machine learning techniques were used to compare their prediction performance with the classifiers (ANN and SVM) mentioned above. Figure 11 presents an overview of ensembles A, B and C proposed in this experiment.

Figure 11 – Daily prediction E1 – Ensemble of machine learning techniques.



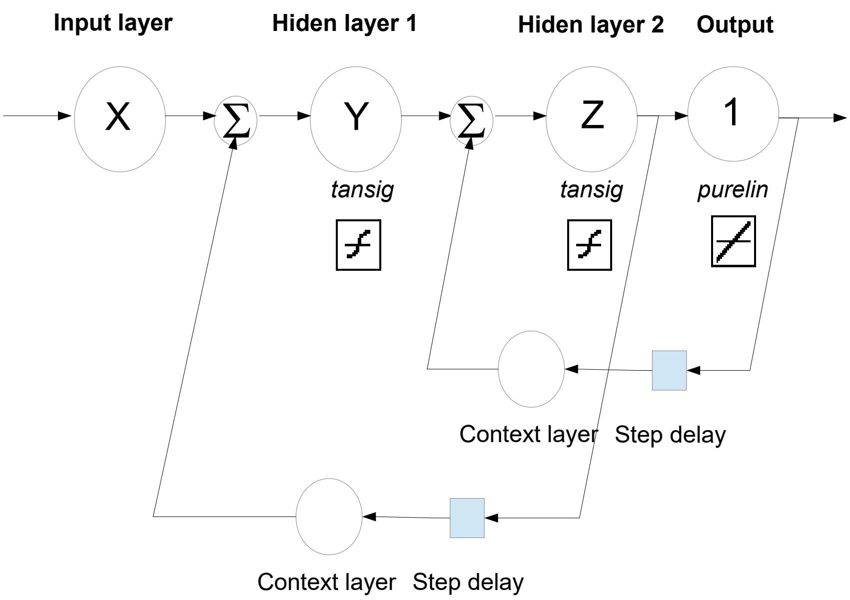
Source: Personal collection

  – *Ensemble A*

First, with the input data prepared in Section B, it is executed a regression model based on RNN (Jordan architecture) to classify the Bitcoin price direction. After that, the result is used as a input of a tree classifier model that predicts the Bitcoin price direction.

For classification by regression, Recurrent Neural Networks (RNNs) were explored through a Jordan architecture inspired by Mcnally (2016). For the hidden layers were used the hyperbolic tangent transfer function and, for the output layer, it was used the linear transfer function, such as presented in Figure 12.

Figure 12 – Daily Prediction E1 – RNN architecture.



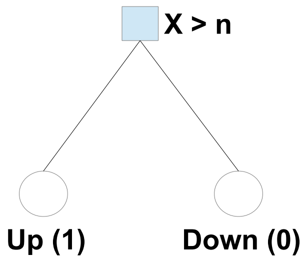
Source: Personal collection

As a learning method, a function based on the gradient descent with momentum and an adaptive learning rate was employed. Mean Square Error (MSE) metric was used to evaluate the performance during the training stage. Thus, several configurations with one and two hidden layers were tested, with combinations of 5, 10, 15, 20, 25, 30 and 35 neurons with a number of epochs ranging from 20 to 500. This method was also implemented in Matlab® platform.

The tree classifier model (Figure 13) was configured with only one decision rule that is $X > n$, where X is the input value and n is a threshold adjusted during the training process. Thus, it is worth to mention that, after train all the models generated, n ranged from 0.45 to 0.55.

– *Ensemble B*

Figure 13 – Daily prediction E1 – Decision tree classifier.



Source: Personal collection

In this ensemble model a prior clustering is executed previously to apply the RNN (classification by regression) and the tree classifier used in the Ensemble A. A clustering method based on k-Means algorithm was implemented, where k represents the number of clusters (in this case, 2 clusters).

According to Jain (2010), the k-Means works by stages: (a) two-dimensional input data with three clusters; (b) three seed points selected to generate k cluster centroids and initial classification of the data points to these clusters; (c) intermediate iterations x updating cluster labels and their centroids; (d) final clustering obtained after the convergence. For implementation of this technique was considered two clusters and used the *city block* distance metric.

– *Ensemble C*

The last ensemble is responsible for executing two tasks: (1) the first task is the forecasting of Maximum, Minimum and Closing Bitcoin exchange rates, using ANN, RNN and SVM in their regression versions; and (2) use the outputs of the forecasting process as inputs to classifiers (ANN, SVM, and Ensembles A and B) in order to predict the Bitcoin price direction. Thus, as inputs for each forecasting method, it was used the most relevant attributes selected by each attribute selection technique mentioned in Table 3.

The ANNs were used with a similar architecture at presented in Figure 9, changing only the transfer function of the output layer by a linear function. In addition, it was used the Levenberg-Marquardt learning method and the MSE as performance metric. The same combinations of the neurons, layers and epochs presented in previous part were used. Each model generated was evaluated by regression performance metrics presented in the following equations Equation 2.8, Equation 2.9 and Equation 2.10.

Finally, the maximum, minimum and closing Bitcoin exchange rates predicted values were added to the input of each classifier (ANN, SVM, and Ensembles A and B), which are responsible for predicting the Bitcoin price direction.

F Performance Metrics used to Evaluate the Price Direction Prediction

To compare the performance of each individual classifier and ensemble used for the purpose of predicting the Bitcoin price direction, it was used Area Under ROC Curve (AUC). Thus, it was calculated *sensitivity* and *specificity* metrics, such as described on Equation 2.2 and Equation 2.3, respectively. The accuracy (Equation 2.1) is considered to compare the models proposed in this paper (with the best performances) with the state-of-the-art results.

It can be highlighted that each generated model was evaluated (trained and validated/tested) 50 times in order to obtain values statistically significant for each considered performance metric.

4.1.2 Experiment 2 – Technical Indicators and Social Trends Approach (E2)

Unlike the previous experiment, this part will focus on identifying the most relevant input data transformations. In particular, the use of technical indicators commonly used by traders will be explored. Likewise, once the best set of technical attributes has been identified, information regarding the acceptance or public recognition of Bitcoin will be added based on the volume of searches on Google and Wikipedia.

In addition to the previous experiment, information on international economic indicators will be added. Also, for the selection of attributes, one of the techniques with the best results from the previous experiment will be used. Finally, from the objective that is to identify the prediction capacity of the attributes and transformations studied, different configurations of models based on ANN and SVM will be used, due to their flexibility and complementarity identified in the previous experiment.

An overview of the methodology proposed in this paper can visualized by means of Figure 14.

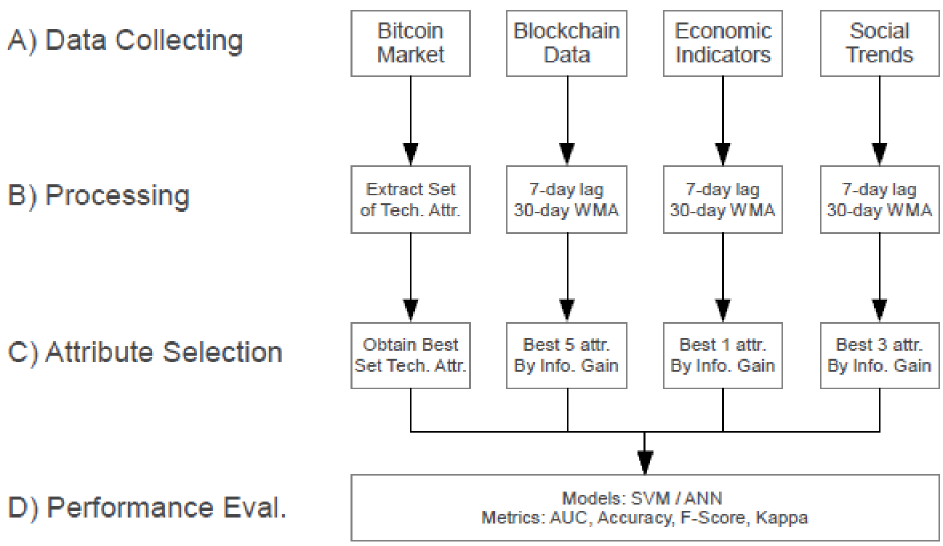
A Input of data collected

In addition to the information considered in Section 4.1.1, is considered Social Trends Information, on the period from 2013 to 2017. Thus, this data is extracted from Google Trends and Wikipedia Searches, such as used by Kristoufek (2013), Ciaian, Rajcaniova & Kancs (2016) and Li & Wang (2017). It is important to mention that Wikipedia information was obtained using a R script shared by Kim et al. (2016).

B Data pre-processing techniques

Based on the database prepared in Section 4.1.1, the information from Social Trends Information is added, where for this information is considered 7-day (Wikipedia

Figure 14 – Daily prediction E2 – Overview of the proposed methodology.



Source: Personal collection

Searches) and 4-week (Google Trends) lag periods and 30-day WMA. This way, the database was processed and its attributes can be seen in Table 4.

Table 4 – Daily prediction E2 – List of all attributes.

Day (*D*)	**Day (*D-i*)**	**30-Day (*WMA*)**
Opening price	Opening price	Opening price
	Maximum price	Maximum price
	Minimum price	Minimum price
	Closing price	Closing price
	Volume of trades	Volume of trades
	Number of transactions	Number of transactions
	Transaction fees	Transaction fees
	Cost per Transaction	Cost per Transaction
	Hash rate average	Hash rate average
	Wikipedia trends	crude oil price
		gold price
		S&P500 index
		Nasdaq index
		DAX index
		Wikipedia trends
	Week (*W-j*)	**30-Day (*WMA*)**
	Google trends	Google trends

Source: Personal collection.

Moreover, it was extracted/calculated features from the Bitcoin exchange rates, generating two sets of trading technical indicators commonly used for stock price prediction models. The first one was composed by the attributes proposed by Kara, Acar Boyacioglu & Baykan (2011). Thus, it was used the Opening Price (OP) for the day (D) as the only raw exchange rate data. It was considered the calculation of Moving Average (MA_{10}) and the Weighted Moving Average (WMA_{10}) for 10 days, such as presented in Equation 4.4 and Equation 4.5:

$$MA_{10} = \frac{C_t + C_{t-1} + ... + C_{t-10}}{10}, \tag{4.4}$$

$$WMA_{10} = \frac{((n_{wma}) \times C_t + (n-1) \times C_{t-1} + ... + C_{t-10}}{(n_{wma} + (n_{wma} - 1) + ... + 1)}. \tag{4.5}$$

Also, it was calculated the following features: Momentum – M (Equation 4.6), Relative Strength Index – RSI (Equation 4.7), momentum index created by Larry Williams – %R (Equation 4.8), Commodity Channel Index – CCI (Equation 4.9), and Accumulation/Distribution – A/D (Equation 4.10):

$$M = C_t - C_{t-n_M}, \tag{4.6}$$

$$RSI = 100 - \frac{100}{1 + (\sum_{i=0}^{n-1} Up_{t-i}/n) / \sum_{i=0}^{n-1} Dw_{t-i}/n)}, \tag{4.7}$$

$$\%R = \frac{HH_n - C_t}{HH_n - LL_n} \times 100, \tag{4.8}$$

$$CCI = \frac{M_t - SM_t}{0.015 D_t}, \tag{4.9}$$

$$A/D = \frac{H_t - C_{t-1}}{H_t - L_t}. \tag{4.10}$$

Finally, it was calculated two stochastic oscillators namely %K and %D (respectively presented in Equation 4.11 and Equation 4.12) and the Moving Average Convergence/Divergence oscillator – MACD (Equation 4.13):

$$\%K = \frac{C_t - LL_{t-n}}{HH_{t-n} - LL_{t-n}} \times 100, \tag{4.11}$$

$$\%D = \frac{\sum_{i=0}^{n-1} K_{t-i}\%}{n}, \tag{4.12}$$

$$MACD = MACD(n_M)_{t-1} + 2/n_M + 1 \times (DIFF_t - MACD(n_M)_{t-1}). \tag{4.13}$$

Based on these equations, C_t represents the closing price at the day t; n_{wma} is the WMA period equal to 10; n_M is equal to 9 and represents the period for momentum and MACD indicators; n represents the period of indicators RSI, %R, %K and %D and is equal to 14; LL_t and HH_t are the lowest minimum and highest maximum at the day t, respectively; L_t is the minimum exchange rate and H_t is the maximum exchange rate at time t; $DIFF = EMA(12)_t - EMA(26)_t$, where EMA is the exponential moving average for 12 and 26 days; $M_t = \frac{H_t+L_t+C_t}{3}$; $SM_t = \frac{\sum_{i=1}^{n} M_{t-i+1}}{n}$; $D_t = \frac{\sum_{i=1}^{n} |M_{t-i+1} - SM_t|}{n}$; Up_t and Dw_t means the upward and downward exchange rate changes at time t, respectively.

The second set of trading technical indicators was extracted/calculated in accordance with Qiu & Song (2016). Again, the only raw exchange rate maintained was the Opening Price for the day (D). It was considered the calculation of On Balance Volume – OBV (Equation 4.14):

$$OBV_t = OBV_{t-1} + \theta \times V_t, \tag{4.14}$$

where V_t is the volume of trade of the Bitcoin at time t and $\theta = \begin{cases} +1, & C_t \geq C_{t-1} \\ -1, & C_t < C_{t-1} \end{cases}$.

Moreover, it was calculated the Moving Average (MA_5) and the proportional deviation respect to the mean ($BIAS_6$), such as presented in Equation 4.15 and Equation 4.16, respectively:

$$MA_5 = (\sum_{i=1}^{5} C_{t-i+1}/5), \tag{4.15}$$

$$BIAS_6 = (\frac{C_t - MA_6}{MA_6}) \times 100. \tag{4.16}$$

In addition, information was provided on the proportion of times that the Bitcoin price increased for a period of 12 days. This information is detailed in the calculation of the Psychological Line – PSY_{12} (Equation 4.17):

$$PSY_{12} = (A/12) \times 100, \tag{4.17}$$

where A is the number of rising days in the last n days.

Finally, it was calculated the difference of return for Bitcoin exchange rate between two days. Therefore, five formulas with the form ASY were considered and are detailed in the Equation 4.18 to Equation 4.22:

$$ASY_5 = (\sum_{i=1}^{5} ASY_{t-i+1})/5, \tag{4.18}$$

$$ASY_4 = (\sum_{i=1}^{4} ASY_{t-i+1})/4, \tag{4.19}$$

$$ASY_3 = (\sum_{i=1}^{3} ASY_{t-i+1})/3, \tag{4.20}$$

$$ASY_2 = (\sum_{i=1}^{2} ASY_{t-i+1})/2, \tag{4.21}$$

$$ASY_1 = ASY_{t-1}, \tag{4.22}$$

where ASY_n is the average return in the last n days.

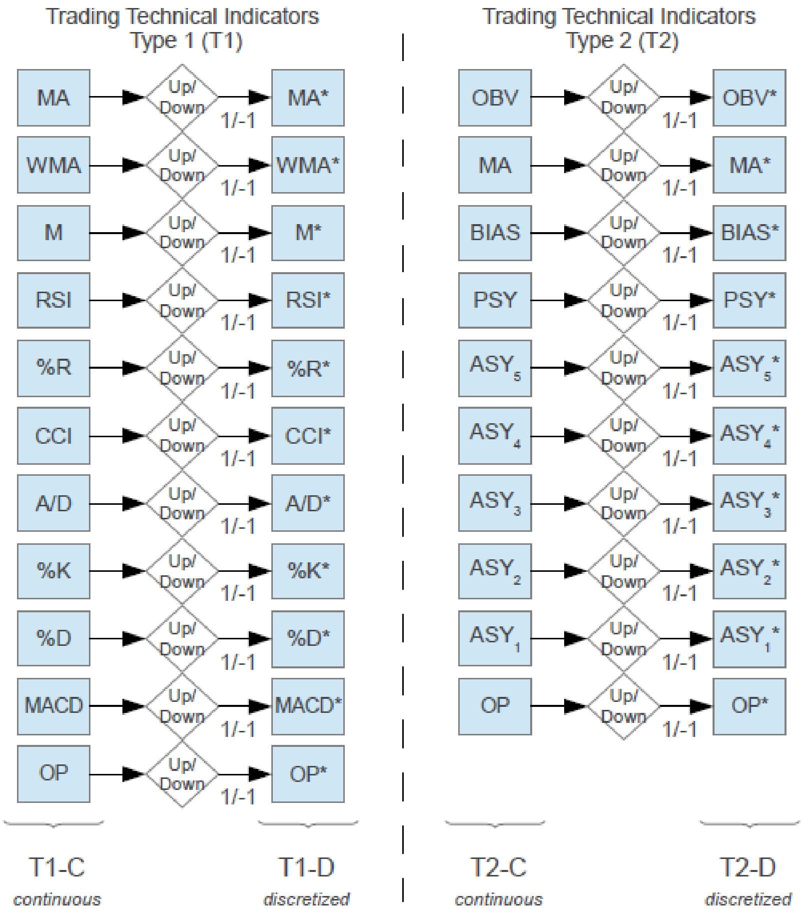
It is important to mention that for both trading technical indicators were considered their continuous and discretized versions (+1 if increase or -1 if decrease), such as proposed by Patel et al. (2015). These versions are explained on Figure 15.

In Table 5 are presented the basic statistics of Bitcoin exchange rates, OHLC, for the two datasets created, where the high volatility is evident. For example, the minimum price (Low price) presents a fall in Jun 23th, 2016 (1.50 USD).

Similarly, Table 6 shows the basic statistics of Blockchain data for these two datasets, which demonstrate even more volatility than Bitcoin exchange rates.

Table 7 presents basic statistics for economic indices data, considering the two intervals. However, in this case, it can be observed a lower volatility. And, the statistics

Figure 15 – Daily prediction E2 – Discretization process.



Source: Personal collection

Table 5 – Daily prediction E2 – Statistics of Bitcoin exchange rate.

Interval 1 - from Aug. 19th, 2013 to Jul. 19th, 2016				
Attributes	Max.	Min.	Mean	Std
Open	1135.00	99.32	410.17	187.91
High	1163.00	99.99	421.66	195.84
Low	1080.00	**1.50**	396.06	177.83
Close	1132.01	99.30	410.68	187.91
Interval 2 - from Apr. 1st, 2013 to Apr. 1st, 2017				
Attributes	Max.	Min.	Mean	Std
Open	1287.38	66.34	449.02	257.93
High	1350.00	72.88	460.60	265.03
Low	1255.00	**1.50**	434.99	249.61
Close	1285.33	66.34	449.68	258.34

Source: Personal collection.

Table 6 – Daily prediction E2 – Statistics of Blockchain data.

Interval 1 - from Aug. 19th, 2013 to Jul. 19th, 2016				
Attributes	Max.	Min.	Mean	Std
Volume of trades	137070.18	**0.00**	13964.01	13638.31
Total transaction fees	337.76	8.00	25.74	21.69
Cost per transaction	90.20	3.44	19.25	16.16
Number of transactions	276448.00	35815.00	113434.03	58828.98
Hash rate avg.	1776788.55	449.59	422210.50	444679.91
Interval 2 - from Apr. 1st, 2013 to Apr. 1st, 2017				
Attributes	Max.	Min.	Mean	Std
Volume of trades	137070.18	**0.00**	12322.58	12450.27
Total transaction fees	337.76	8.00	41.88	43.30
Cost per transaction	90.20	3.44	16.04	14.82
Number of transactions	350751.00	28865.00	132905.79	80846.22
Hash rate avg.	4161948.39	52.26	710026.71	885313.13

Source: Personal collection.

Table 7 – Daily prediction E2 – Statistics of economic indices.

Interval 1 - from Aug. 19th, 2013 to Jul. 19th, 2016				
Attributes	Max.	Min.	Mean	Std
Crude oil futures	108.24	30.05	69.46	26.65
Gold futures	1389.96	1069.11	1227.38	76.06
S&P 500 futures	2115.11	1647.66	1960.38	123.98
Nasdaq 100 futures	4673.42	3090.93	4064.26	434.72
DAX index	12106.08	8263.26	9981.29	883.85
Interval 2 - from Apr. 1st, 2013 to Apr. 1st, 2017				
Attributes	Max.	Min.	Mean	Std
Crude oil futures	108.24	30.05	68.49	25.89
Gold futures	1597.68	1069.11	1247.84	90.54
S&P 500 futures	2375.48	1553.58	1972.77	187.03
Nasdaq 100 futures	5396.27	2789.91	4107.05	629.99
DAX index	12106.08	7647.83	9972.34	1078.98

Source: Personal collection.

for social trends information are showed in Table 8, where Wikipedia searches presents greater volatility than Google popularity index.

In addition, Table 9 and Table 10 show the summary of the main statistics for the two sets of technical indicators extracted/calculated for the both intervals. Analyzing these tables, it is possible to notice the high variability of the data that is reflected in the statistics, especially for the first set of technical indicators.

Table 8 – Daily prediction E2 – Statistics of social media trends.

Interval 1 - from Aug. 19th, 2013 to Jul. 19th, 2016				
Attributes	Max.	Min.	Mean	Std
Google popularity index	65.00	6.00	14.43	9.99
Wikipedia searches	847614.00	**0.00**	13117.77	**36843.84**
Interval 2 - from Apr. 1st, 2013 to Apr. 1st, 2017				
Attributes	Max.	Min.	Mean	Std
Google popularity index	65.00	6.00	14.83	9.57
Wikipedia searches	847614.00	**0.00**	12619.23	**32263.13**

Source: Personal collection.

Table 9 – Daily prediction E2 – Statistics of the first set of technical indicators.

Interval 1 - from Aug. 19th, 2013 to Jul. 19th, 2016				
Name of indicators	Max.	Min.	Mean	Std
Simple MA	1035.81	96.79	408.28	185.84
Weighted MA	1053.42	98.03	409.08	186.12
Momentum	532.99	-379.90	4.77	72.24
Stochastic K%	100.00	0.00	54.85	27.71
Stochastic D%	98.17	6.14	54.84	26.19
RSI	94.75	17.51	52.55	14.31
MACD	180.47	-47.96	3.65	27.06
LW R%	100.00	0.00	45.15	27.71
A/D%	1.05	-0.05	0.48	0.32
CCI	452.84	-568.14	6.11	97.45
Open. Price	1135.00	99.32	410.70	187.85
Interval 2 - from Apr. 1st, 2013 to Apr. 1st, 2017				
Name of indicators	Max.	Min.	Mean	Std
Simple MA	1237.88	77.17	446.78	255.39
Weighted MA	1252.52	74.63	447.73	256.03
Momentum	532.99	-379.90	5.73	70.12
Stochastic K%	100.00	0.00	56.76	27.47
Stochastic D%	98.17	6.14	56.77	25.97
RSI	96.99	17.51	53.35	14.33
MACD	180.47	-47.96	5.14	24.99
LW R%	100.00	0.00	43.24	27.47
A/D%	1.21	-0.05	0.48	0.32
CCI	452.84	-568.14	10.52	96.41
Open. Price	1287.38	66.34	449.69	258.27

Source: Personal collection.

C Data Partitioning

Is considered similar partition detailed on Section C.

Table 10 – Daily prediction E2 – Statistics of the second set of technical indicators.

Interval 1 - from Aug. 19th, 2013 to Jul. 19th, 2016				
Indicators	Max.	Min.	Mean	Std
OBV	924475.69	-141105.19	297055.72	258748.53
SMA_5	1061.80	98.91	409.61	186.85
$BIAS_6$	1073.50	98.78	409.34	186.66
PSY_{12}	91.67	8.33	52.45	16.34
ASY_5	10.61	-10.79	0.18	1.94
ASY_4	11.82	-12.31	0.18	2.13
ASY_3	16.16	-16.85	0.18	2.48
ASY_2	21.58	-22.24	0.18	3.10
ASY_1	33.75	-28.09	0.18	4.46
Open. Price	1135.00	99.32	410.70	187.85
Interval 2 - from Apr. 1st, 2013 to Apr. 1st, 2017				
Indicators	Max.	Min.	Mean	Std
OBV	924475.69	-141105.19	350098.79	298288.58
SMA_5	1270.88	71.33	448.37	256.84
$BIAS_6$	1264.58	72.67	448.04	256.53
PSY_{12}	91.67	8.33	53.97	15.86
ASY_5	11.13	-18.43	0.17	2.14
ASY_4	15.50	-22.90	0.17	2.39
ASY_3	18.25	-26.66	0.17	2.81
ASY_2	23.66	-50.50	0.17	3.50
ASY_1	33.75	-66.39	0.17	4.94
Open. Price	1287.38	66.34	449.69	258.27

Source: Personal collection.

D Attribute Selection

In order to select the most relevant attributes for Blockchain, Economic indices and Social trends, the measure of information gain (Equation 4.23) was used. This measure is based on the amount of entropy (Equation 4.24) provided by each attribute (X) in relation to the class (Y), such that:

$$I(X, Y) = H(X) - H(X|Y), \tag{4.23}$$

$$H(X) = -\sum p(X) \log p(X), \tag{4.24}$$

where X is the vector of input attributes and Y is the class vector.

In the case of the Blockchain data, the best five attributes selected are presented in Table 11.

One interesting fact is the differences of attributes selected for each interval, because only the Minimum Price of $D-5$ remains in both periods. Thus, this may be an

Table 11 – Daily prediction E2 – Best five Blockchain attributes by $Info.\ Gain$.

Interval 1	**Interval 2**
Transaction fees $D-2$	Maximum Price $D-5$
Hash rate average $D-2$	Minimum Price $D-5$
Minimum Price $D-5$	Closing Price $D-5$
Hash rate average $D-7$	Volume of trades $D-5$
Number of trx $30-Day\ WMA$	Hash rate avg $30-Day\ WMA$

Source: Personal collection.

indication of the difference in the price behavior of Bitcoin. In addition, a summary of the main statistics (Table 12) for these attributes are presented for both intervals analyzed.

Table 12 – Daily prediction E2 – Statistics of the best five Blockchain attributes.

Interval 1 - from Aug. 19th, 2013 to Jul. 19th, 2016				
Indicators	Max.	Min.	Mean	Std
Trx. fees $D-2$	337.76	8.00	25.72	21.68
Hash rate $D-2$	1776788.55	449.59	420728.20	443446.77
Min. Price $D-5$	1080.00	1.50	393.94	178.02
Hash rate $D-7$	1776788.55	388.97	413694.63	438077.88
# of trx WMA	235960.72	50558.52	112069.48	56561.62
Interval 2 - from Apr. 1st, 2013 to Apr. 1st, 2017				
Indicators	Max.	Min.	Mean	Std
Max. Price $D-5$	1350.00	72.88	457.94	263.82
Min. Price $D-5$	1255.00	1.50	432.43	248.41
Close Price $D-5$	1285.33	66.34	447.07	257.12
Volume $D-5$	137070.18	0.00	12312.43	12454.55
Hash rate WMA	3550041.98	48.68	686356.25	850098.29

Source: Personal collection.

Also, it can be seen that the Weighted Moving Average operation on "Hash rate" and "Number of trx" reduces their original volatility. As far as economic indicators is concerned, only the best global economic index was selected, being the same for both intervals, such as presented in Table 13.

Table 13 – Daily prediction E2 – Best economic attribute selected by the $Info.\ Gain$.

Interval 1	**Interval 2**
DAX index $30-Day\ WMA$	DAX index $30-Day\ WMA$

Source: Personal collection.

In Table 14, the main statistics of this attribute for both intervals are presented.

In this case, the Weighted Moving Average over DAX index was considered as the best predictive attribute, which reduces its original volatility.

Table 14 – Daily prediction E2 – Statistics of the best economic attribute.

Interval 1 - from Aug. 19th, 2013 to Jul. 19th, 2016				
Indicators	Max.	Min.	Mean	Std
DAX index 30 – *Day* WMA	12106.08	8263.26	9981.29	883.85
Interval 2 - from Apr. 1st, 2013 to Apr. 1st, 2017				
Indicators	Max.	Min.	Mean	Std
DAX index 30 – *Day* WMA	12106.08	7647.83	9972.34	1078.98

Source: Personal collection.

Finally, the Table 15 presents the three best social trends attributes for both intervals, highlighting that the same values were obtained. In addition, the main statistics of these attributes are presented in Table 16.

Table 15 – Daily prediction E2 – Best three social attributes selected by the $Info.\ Gain$.

Interval 1	**Interval 2**
Google trends $W-4$	Google trends $W-4$
Wikipedia trends $D-1$	Wikipedia trends $D-1$
Wikipedia trends 30 – *Day* WMA	Wikipedia trends 30 – *Day* WMA

Source: Personal collection.

Table 16 – Daily prediction E2 – Statistics of the three best social attributes.

Interval 1				
Indicators	Max.	Min.	Mean	Std
Google trends $W-4$	65.00	6.00	14.23	10.03
Wikipedia trends $D-1$	847614.00	0.00	13117.77	36843.84
Wikipedia trends WMA	1283953529	5761	2260342	40885351
Interval 2				
Indicators	Max.	Min.	Mean	Std
Google trends $W-4$	65.00	6.00	14.55	9.47
Wikipedia trends $D-1$	847614.00	0.00	12619.23	32263.13
Wikipedia trends WMA	1283953529	5761	1839714	34947161

Source: Personal collection.

Finally, it is possible to observe that, as in the previous cases, in this case an attribute was obtained with the calculation of Weighted Moving Average over the "Wikipedia trends" variable, which reduces its volatility and rescues information about its trend.

E Soft Computing Algorithms Applied to the Predictions

Because the present experiment will focus on the predictability of the selected attributes, the two models that stand out for their flexibility (ANN and SVM) will be used.

- *Artificial Neural Network*

  Similar with Section E is used a MLP architecture with two hidden layers. For both hidden layers, it was used the rectified linear unit function and the logistic sigmoid function was used for the output layer. Back-propagation algorithm was used as learning method and *log loss* function as performance metric. Thus, several configurations were tested, 400 combinations, and each of them executed 5 times, which are presented in Table 17.

  Table 17 – Daily prediction E2 – ANN parameter combinations tested.

Parameters	**Level(s)**
Number of neurons by hidden layers ($arch$)	[10, 20, ..., 100]-[10, 20, ..., 100]
Epochs (ep)	50, 100, 500, 1000
Momentum constant (mc)	.1
Learning rate (lr)	.1

  Source: Personal collection.

- *Support Vector Machine*

  Several configurations with polynomial and radial basis kernel were tested (681 combinations), which are presented in Table 18.

  Table 18 – Daily prediction E2 – SVM parameter combinations tested.

Parameters	***Polynomial* (p)**	***Radial basis* (r)**
Degree of kernel (d)	1,2,3,4	-
Gamma in kernel (γ)	1/n, .1, .2, ..., 1.0	1/n, .1, .2, ..., 10.0
Regularization (c)	.5, 1, 5, 10	.5, 1, 5, 10, 100
where, n is the number of input attributes.		

  Source: Personal collection.

Unlike the previous experiment, both models were implemented in *Python* from the use of the *Sklearn* library.

F Classification performance metrics

In addition to the metrics used in Section F, it was calculated the F-score (or F-measure), in accordance with Equation 2.4.

Finally, the Cohen's kappa measure (*Kappa*) was used as a secondary comparison metric (Equation 2.7).

To compare AUC values, the statistical significance test was used in this part. The null hypothesis considers that the means of AUC values when compared are equal. In cases where the probability value of this statement is not statistically significant (p-value), the null hypothesis is rejected. Thus, in order to consider value of AUC, obtained in the experiments, as smaller or greater than another, it is necessary to reject the null hypothesis.

For the calculation, a Student's t-distribution was used considering a significance value of 95% and two tails (Equation 4.25):

$$\begin{cases} \text{reject null hypothesis}, & p\text{-value} < 0.05 \\ \text{accept null hypothesis}, & \text{otherwise}. \end{cases} \tag{4.25}$$

4.2 Intra-daily Prediction

4.2.1 Data-stream/Online Learning Approach

As is mentioned in Gao & Lei (2017), the main advantage of data stream learning approach is that the prediction model can capture the changing pattern of Bitcoin price since the model is continuously updated whenever new data are available. Thus, this experiment seeks to make intra-daily predictions, but unlike Madan, Saluja & Zhao (2015), Greaves & Au (2015), an approach is proposed using the aforementioned advantages of stream learning approach.

A Data Collected

In this part was collected GMT-stamped tick data, from Bitstamp(USD) exchange market, and aggregate it to the 10-minutely frequency to analyze intra-daily behavior. It was computed OHLC exchange rate and volume of trades realized in this interval, from April 1, 2013 to April 01, 2017 (around of 209K instances). In addition, it was included data with best daily aggregation that they used in the previous section.

B Data Pre-processing

Firstly, it was computed a sort of technical indicators commonly used for stock price predictions models, proposed by Qiu & Song (2016). In the Table 19 is showed the attributes and formulas considered.

Secondly, it was evaluated the change of the classification performance adding daily frequency data, proposed by in previous section and described in the Table 20.

Table 19 – Intra-daily prediction – Input data - 10 *minutely* frequency.

Technical Indicators				
Attribute	Max.	Min.	Mean	Std
On-balance volume (OBV_t)	270914.51	-115571.73	19546.44	82974.30
Simple Moving Average SMA_5	1291.33	54.21	450.39	258.02
$BIAS_6$	51.21	-30.8	0.00	0.63
PSY_{12}	100	0	48.75	13.09
ASY_5	8.54	-11.97	0.00	0.22
ASY_4	14.10	-13.49	0.00	0.26
ASY_3	19.94	-14.76	0.00	0.31
ASY_2	36.22	-18.53	0.00	0.39
ASY_1	73.69	-29.73	0.00	0.60

Source: Personal collection.

Table 20 – Intra-daily prediction – Input data - 1 *daily* frequency.

Technical Indicators				
Attribute	Max.	Min.	Mean	Std
OBV	924475.69	-141105.19	350098.79	298288.58
SMA_5	1270.88	71.33	449.53	256.92
$BIAS_6$	1264.58	72.67	449.19	256.61
PSY_{12}	91.67	8.33	54.02	15.86
ASY_5	11.13	-18.43	0.17	2.14
ASY_4	15.50	-22.90	0.17	2.39
ASY_3	18.25	-26.67	0.17	2.81
ASY_2	23.66	-50.50	0.16	3.50
ASY_1	33.75	-66.39	0.16	4.94
Open. Price	1287.38	66.34	450.85	258.35
Blockchain attributes				
Attribute	Max.	Min.	Mean	Std
Max. Price $D-5$	1350.00	72.88	459.07	263.95
Min. Price $D-5$	1255.00	1.50	433.49	248.49
Close Price $D-5$	1285.33	66.34	448.19	257.22
Volume $D-5$	137070.18	0.00	12308.62	12459.47
Hash rate WMA	3550041.98	48.68	688621.85	851472.46
International economic trend				
Attribute	Max.	Min.	Mean	Std
DAX index $30 - Day\ WMA$	12106.08	7647.83	9975.81	1078.60
Social popularity				
Attribute	Max.	Min.	Mean	Std
Google trends $W-4$	65.00	6.00	14.57	9.50
Wikipedia trends $D-1$	847614.00	0.00	12633.64	32274.30
Wikipedia trends WMA	1283953529	5761	1851343	35075353

Source: Personal collection.

Furthermore, in order to select the best combination of 10-minutely and daily frequency data, it was used information gain score, computing the entropy value provided by each feature in relation to the class. The attributes selected by this method is described in Table 21.

Table 21 – Intra-daily prediction – Input data - information gain filtered.

10 *minutely* frequency				
Attribute	Max.	Min.	Mean	Std
$BIAS_6$	51.21	-30.8	0.00	0.63
PSY_{12}	100	0	48.75	13.09
ASY_5	8.54	-11.97	0.00	0.22
ASY_4	14.10	-13.49	0.00	0.26
ASY_3	19.94	-14.76	0.00	0.31
ASY_2	36.22	-18.53	0.00	0.39
ASY_1	73.69	-29.73	0.00	0.60
1 *daily* frequency				
Attribute	Max.	Min.	Mean	Std
ASY_5	11.13	-18.43	0.17	2.14
ASY_4	15.50	-22.90	0.17	2.39
ASY_3	18.25	-26.67	0.17	2.81
ASY_2	23.66	-50.50	0.16	3.50
ASY_1	33.75	-66.39	0.16	4.94

Source: Personal collection.

Finally, the value "1" was assigned to the class if the closing exchange rate of Bitcoin at a 10 min (t) is greater than or equal to the previous period $(t-1)$, otherwise was assigned the value "0".

C Soft Computing Algorithms Applied to the Predictions

Massive Online Analysis (MOA) Holmes, Kirkby & Pfahringer (2007) is a software platform selected for perform all algorithms using a laptop computer with an Intel Core i5-3320M processor running Windows 7 64-bit operating system and 8GB of RAM.

- *Model based on Hoedffing Tree (HT)*

  Is explored a technique with single classifier approach, was proposed by Domingos & Hulten (2000), Very Fast Decision Tree (VFDT) algorithm based on HT. One important property of this algorithm is that the trees it produces are asymptotically arbitrarily close to the ones produced by a batch learner. This is possible thanks to the employ *Hoeffding bound*. This bound states that with probability 1 - δ, the true mean of a random variable of range R will not differ from the estimated mean after n independent observations by more than Bifet et al. (2011):

$$\epsilon = \sqrt{\frac{R^2 \ln(1/\delta)}{2n}} \tag{4.26}$$

where ϵ is a limit and R is the base 2 logarithm of the number of possible class labels.

Algoritmo 1: Hoeffding tree induction algorithm.

Let HT be a tree with a single leaf (the root);
forall *training examples* **do**
  Sort example into leaf l using HT;
  Update sufficient statistics in l;
  Increment n_l , the number of examples seen at l;
  if $n_l \mod n_{min} = 0$ ***and*** *examples seen at l not all of same class* **then**
    Compute $\overline{G}_l(X_i)$ for each attribute;
    Let X_a be attribute with highest $\overline{G}_l$;
    Let X_b be attribute with second-highest $\overline{G}_l$;
    Compute Hoeffding bound (4.26);
    if $X_a \neq X_\emptyset$ ***and*** $(\overline{G}_l(X_a) - \overline{G}_l(X_b) > \epsilon$ ***or*** $\epsilon < \tau$ **then**
      Replace l with an internal node that splits on X_a;
      forall *branches of the split* **do**
        Add a new leaf with initialized sufficient statistics;
      end
    end
  end
end

The complete basic algorithm is presented in the Algorithm 1 (Bifet et al. (2011)) where $\overline{G}$ is a split function, n_{min} is a number of instances a leaf should observe between split attempt, and τ is a threshold below which a split will be forced to break ties (innovation realized by Domingos & Hulten (2000)).

For experiments was considered $\overline{G}$ as $InformationGainSplit$ function, $n_{min} = 200$, $\delta = 1\text{e}{-7}$ and $\tau = 0.05$.

- *Accuracy-Weighted Ensemble (AWE) model*

This algorithm was explored because has been shown to be an efficient way for mining concept-drifting data streams Wang (2003). This is a horizontal ensemble method where is calculated the weight (w_i) of each base learner (C_i) should be inversely proportional to their mean square error in classifying calibration data of current user (MSE_i).

$$MSE_i = \frac{1}{|S_i|} \sum_{(x,c)\in S_i} (1 - f_c^i(x))^2; MSE_r = \sum_c p(c)(1 - p(c))^2 \tag{4.27}$$

$$w_i = argmax(-(MSE_i - MSE_r), 0) \tag{4.28}$$

$$P(x_{new}) = \frac{\sum_{i=1}^{k} w_i \times f_c^i}{\sum_{i=1}^{k} w_i} \tag{4.29}$$

where x is a input features, c is a class, S_i is a chunk i, f_c^i is a probability to classify x as c by base learner (C_i), $p(c)$ is a random probability to classify x as c using cross-validation technique with n folds, x_{new} is a unseen data and k is number of classifier in the ensemble.

For experiments was considered a Hoeffding Tree as a base learner (C), chunk size = 500, number folds = 10, k=15 and considered 30 maximum number of classifiers to store and choose from when creating an ensemble.

- *Accuracy-Updated Ensemble (AUE) model*

  This algorithm was inspired by AWE method. This method has two mainly innovations, first the weight calculation for each classifier is as follow Brzeziński & Stefanowski (2011):

$$w_i = \frac{1}{MSE_i + \epsilon} \tag{4.30}$$

where ϵ is a very short number because MSE_i maybe zero. Thus is not necessary calculate the random error for each classifier.

And the second change is that AUE update base classifiers rather than only adjust their weights, when $w_i > 1/MSE_r$.

For experiments was considered a Hoeffding Tree as a base learner (C), chunk size = 500, number folds = 10 and k=10 (number of members).

D Performance Metrics used to Evaluate the Price Direction Prediction

One of the goals of this experimental study is to perform a consistent comparison between the three different of feature groups detailed in Table 19, Table 20, Table 21 and evaluate the performance of different types of data stream learning algorithms described above. As is claimed in Uriarte-Arcia et al. (2015) the evaluation of data stream techniques presents a different approach that batch learning evaluation. Thus in this work was performed two sorts approaches.

- *Interleaved Test-Then-Train*

  Each instance can be used to test the model before it is used for training, and from this the classification performance can be incrementally updated Uriarte-Arcia et al. (2015).

- *Prequential with Sliding Window*

  Prequential evaluation provides a learning curve that monitors the evolution of learning as a process similar to interleaved method, but compute the error using using a time window of the most recent observed errors Gama (2010)

For experiments the sample frequency used was 1000 instances and for prequential evaluation was considered 1000 instances as size of the time window. Thus, for both evaluation approaches is used AUC metric to compare the classifiers. For compare the best models generated with other studies was added accuracy metric. Finally, the Cohen's kappa measure (kappa) was used as a complementary comparison metric (Equation 2.7).

5 Results and Discussion

5.1 Daily Prediction

5.1.1 Experiment 1 – Exploratory Approach (E1)

- *Prediction of Price Direction*

  The classification strategies described on Section 4.1.1 were evaluated and the best results of each of them are presented, in Table 22 and Table 23, for the first and second intervals described on Section A and Section C, respectively.

  In the column "Algorithm:Arch:fs" column is indicated machine learning technique used, its architecture and, in parenthesis, the attribute selection method applied to the data set (where, *all* represents that the better results were obtained using all attributes). In case of Artificial Neural Network (ANN), Ensemble A and B models, architecture used is described as: "h1"-"h2"-"e", where "h1" is the number of neurons used in the first hidden layer, "h2" is the number of neurons considered in the second hidden layer and "e" represents the number of epochs used in the training stage. For Support Vector Machine (SVM), architecture is described as "c"-"d", where "c" represents cost parameter and "d" is degree of the kernel polynomial function.

  The column "Individually" means that the classifiers where individually employed, i.e., the Ensemble C was not considered. While the results presented in the column "Ensemble C" means that the Ensemble C was taken into consideration, i.e., the forecasting of Maximum, Minimum and Closing Bitcoin exchange rates was used as inputs for each classifier.

  Analyzing the results presented in the Table 22 and Table 23, it can be observed that the Ensemble C did not demonstrate good performance for both data sets (intervals 1 and 2). Therefore, in the sequence, the Figure 16 and Figure 17 show the Area Under ROC Curve (AUC) score for different periods of validation (in days). In these graphs, the interval areas are highlighted with a statistical confidence of 95% (±2 times the standard deviation).

  In the first interval (Table 22), the best result was obtained by Ensemble A that has the greatest value of AUC (0.58) and an accuracy of 62.91%. It was used the correlation analysis technique as attribute selection method, without including the predicted values of maximum, minimum and closing Bitcoin exchange rates. Table 24 shows the attributes selected by the *Corr* method, which were used to obtain this

Table 22 – Daily prediction E1 – Best performances (Interval 1).

Individually			Ensemble C		
ANN/Ens.: *ep|arch* SVM: *C|d* (fs)	**AUC**	**Acc.**	**ANN/Ens.: *ep|arch* SVM: *C|d* (fs)**	**AUC**	**Acc.**
ANN:100|20-0 (Corr)	0.56 $\pm$*0.03*	58.84% $\pm$*7.25%*	ANN:500|25-30 (Corr)	0.51 $\pm$*0.02*	46.10% $\pm$*4.62%*
SVM:1|1 (CFS)	0.52 $\pm$*0.00*	56.81% $\pm$*0.00%*	SVM:1|1 (CFS)	0.51 $\pm$*0.00*	56.34% $\pm$*0.00%*
Ens. A:500|5-10 (Corr)	**0.58** $\pm$*0.00*	**62.91%** $\pm$*0.00%*	Ens. A:20|25-5 (corr)	0.51 $\pm$*0.00*	42.72% $\pm$*0.00%*
Ens. B:20|5-15 (all)	0.56 $\pm$*0.03*	61.31% $\pm$*3.89%*	Ens. B:1000|20-5 (all)	0.54 $\pm$*0.01*	60.83% $\pm$*0.48%*

Source: Personal collection.

Table 23 – Daily prediction E1 – Best performances (Interval 2).

Individually			Ensemble C		
ANN/Ens.: *ep|arch* SVM: *C|d* (fs)	**AUC**	**Acc.**	**ANN/Ens.: *ep|arch* SVM: *C|d* (fs)**	**AUC**	**Acc.**
ANN:100|25-0 (InfoGain)	0.54 $\pm$*0.03*	53.40% $\pm$*5.40%*	ANN:20|15-30 (InfoGain)	0.51 $\pm$*0.02*	46.11% $\pm$*5.78%*
SVM:1|1 (all)	**0.58** $\pm$*0.00*	**59.45%** $\pm$*0.00%*	SVM:1|1 (all)	0.55 $\pm$*0.00*	56.44% $\pm$*0.00%*
Ens. A:500|25-0 (InfoGain)	0.54 $\pm$*0.00*	48.85% $\pm$*0.00%*	Ens. A:20|20-15 (InfoGain)	0.50 $\pm$*0.00*	60.50% $\pm$*1.67%*
Ens. B:500|25-20 (Corr)	0.55 $\pm$*0.02*	58.19% $\pm$*2.37%*	Ens. B:1000|5-25 (Corr)	0.52 $\pm$*0.00*	42.16% $\pm$*0.75%*

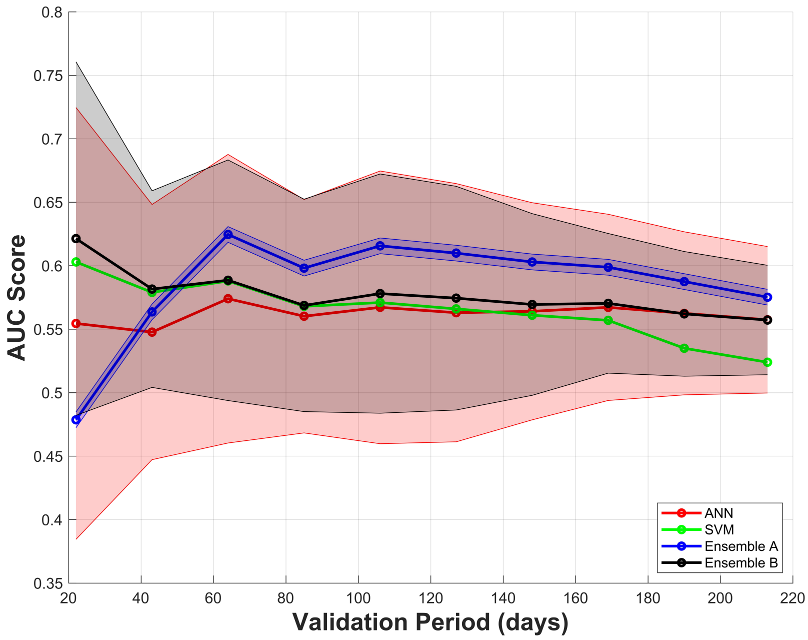
Source: Personal collection.

result. It is important to observe that only information from the Blockchain was used.

Analyzing the results obtained for second interval (Table 23), SVM was the algorithm with the best performance (with 0.58 of AUC and 59.45% of accuracy). The data set used is composed of all attributes described in Table 3, but without including predicted values of maximum, minimum and closing Bitcoin exchange rates.

In addition, the best result obtained by the Ensemble A was compared with those

Figure 16 – Daily prediction E1 – Performance validation for the first data set (interval 1), considering 95% of confidence.



Source: Personal collection

Table 24 – Daily prediction E1 – Attributes selected by the *Corr* method for interval 1.

Day *D*	**Day *(D − i)***	**30-day WMA**
Open. price	Open. price (i:1,5,6,7) Max. price (i:6,7) Min. price (i:1,2,4,6,7) Closing price (i: 1,6,7)	Opening price Maximum price Minimum price Closing price Transaction fees

Source: Personal collection.

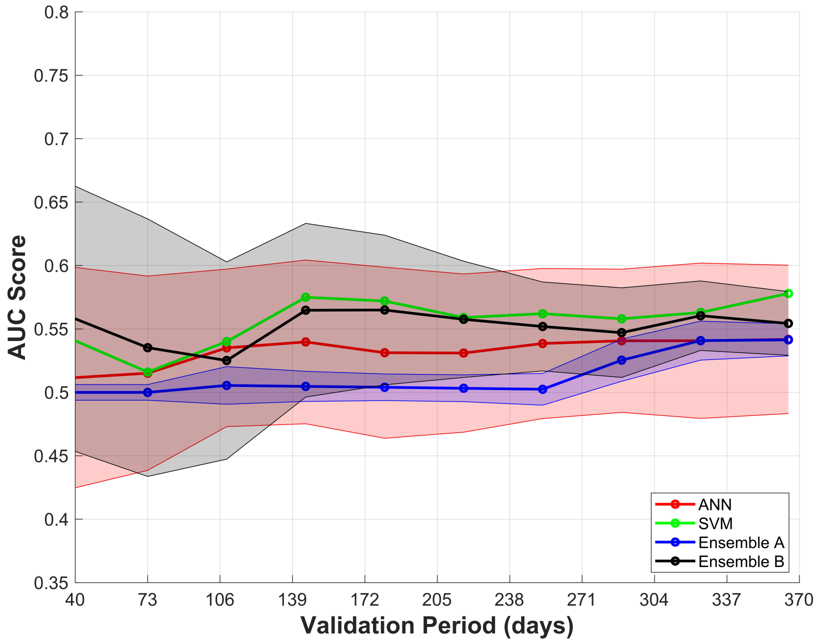
presented by Mcnally (2016), considering the same range of data (interval 1). For this comparison, only the accuracy metric could be used.

Comparing the results obtained by the LSTM algorithm proposed by Mcnally (2016), it was possible to note that the performances, in terms of accuracy, of all the individual algorithms (shown in Table 22, column “Individually") proposed in this paper are better. In addition, is presented in Figure 18 the confusion matrix where is observed an equilibrium between TP and TN.

- *Forecasting of Maximum, Minimum and Closing Bitcoin Exchange Rates*

  In the Table 26 and Table 27 are presented the best results for regression (fore-

Figure 17 – Daily prediction E1 – Performance validation for the first data set (interval 2), considering 95% of confidence.



Source: Personal collection

Table 25 – Daily prediction E1 – Comparison of accuracy with the models proposed by Mcnally (2016).

Model	Accuracy
Ensemble A:500|5-10 (Corr)	**62.91%**
LSTM Mcnally (2016)	52.78%
RNN Mcnally (2016)	50.25%
ARIMA Mcnally (2016)	50.05%

Source: Personal collection.

casting) experiments considering first and second intervals, respectively. For both intervals, the best results was obtained by SVM algorithm (regression version) using attributes selected by Relief method.

For both intervals, the best results were obtained by the SVM algorithm (in its regression version) using attributes selected by the Relief technique. Moreover, the SVM obtained the best results to forecast the maximum, minimum and closing Bitcoin exchange rates.

In the case of minimum exchange rate, Mean Absolute Percentage Error (MAPE) metric is very high because in Jun 23th, 2016 its value decreases from $588.03 to

Figure 18 – Daily prediction E1 – Confusion Matrix of Ensemble A (interval 1).

		Actual	
		1	0
Predicted	1	30.92%	19.08%
	0	18.01%	31.99%

Source: Personal collection

Table 26 – Daily prediction E1 – Best performances (Interval 1).

Value	ANN/RNN: *ep*|*arch* or SVM: *C*|*d* and (fs)	MAE	MAPE	RMSE
Max.	ANN:100|5-0(Relief)	27.02 ±*37.58*	4.94% ±*6.72%*	65.29 ±*110.7*
	RNN:500|10-10(InfoGain)	19.97 ±*0.00*	3.80% ±*0.00%*	32.16 ±*0.00*
	SVM:0.8|1(Relief)	**6.70** ±***0.00***	1.28% ±*0.00%*	12.12 ±*0.00*
Min.	ANN:20|5-5(CFS)	14.58 ±*3.52*	183.9% ±*1.37%*	45.90 ±*4.39*
	RNN:500|30-10(CFS)	13.51 ±*0.00*	183.2% ±*0.00%*	42.48 ±*0.00*
	SVM:0.8|1(Relief)	**10.08** ±***0.00***	183.7% ±*0.00%*	42.66 ±*0.00*
Close	ANN:20|5-0(CFS)	19.06 ±*10.06*	3.86% ±*2.02%*	25.85 ±*13.83*
	RNN:500|30-25(CFS)	14.54 ±*0.00*	3.08% ±*0.00%*	18.56 ±*0.00*
	SVM:1|1(Relief)	**9.63** ±***0.00***	1.91% ±*0.00%*	15.92 ±*0.00*

Source: Personal collection.

Table 27 – Daily prediction E1 – Best performances (Interval 2).

Value	ANN/RNN: *ep*|*arch* or SVM: *C*|*d* and (fs)	MAE	MAPE	RMSE
Max.	ANN:20|5-5(CFS)	55.03 ±*73.48*	6.51% ±*9.40%*	83.42 ±*95.86*
	RNN:500|10-35(CFS)	14.04 ±*0.00*	2.03% ±*0.00%*	20.38 ±*0.00*
	SVM:0.9|1(Relief)	**9.23** ±***0.00***	1.14% ±*0.00%*	17.17 ±*0.00*
Min.	ANN:20|5-0(CFS)	44.40 ±*31.20*	112.3% ±*4.22%*	79.99 ±*52.96*
	RNN:500|10-35(CFS)	32.65 ±*0.00*	113.6% ±*0.00%*	51.97 ±*0.00*
	SVM:1|1(Relief)	**13.26** ±***0.00***	107.8% ±*0.00%*	41.08 ±*0.00*
Close	ANN:20|5-0(CFS)	26.14 ±*5.18*	3.06% ±*0.60%*	41.62 ±*7.22*
	RNN:500|30-30(CFS)	27.85 ±*0.00*	3.36% ±*0.00%*	42.34 ±*0.00*
	SVM:1.3|1(Relief)	**14.32** ±***0.00***	1.81% ±*0.00%*	25.47 ±*0.00*

Source: Personal collection.

$1.5. However, if this date is not considered, the MAPE obtained using the SVM regression model decreases from 183.7% and 107.8% to 1.52% and 1.58% for the first and second intervals, respectively.

Table 28, Table 29 and Table 30 show the most effective attributes selected by the

Relief method to forecast the maximum, minimum and closing Bitcoin exchange rates. It can be noted that only the information from Blockchain was considered as relevant.

Table 28 – Daily Prediction E1 – Attr. Selected by Relief (Max. Price Reg.).

Interval *I*	**Day *D***	**Day *(D − i)***	**30-day WMA**
Interval 1	Open. price	Open. price (i:1,2) Max. price (i:1-3) Min. price (i:1,2) Closing price (i:1-3) Vol. of trades (i:1-4) **Cost Trx. (i:2)**	**Min. price** Vol. of trades Cost Trx. Trx. Fee
Interval 2	Open. price	Open. price (i:1,2) Max. price (i:1,2) Min. price (i:1,2) Closing price (i:1-3) **Vol. of trades (i:1-5)** **Cost Trx. (i:1,2)**	Vol. of trades Cost Trx. Trx. Fee

Source: Personal collection.

Table 29 – Daily Prediction E1 – Attr. Selected by Relief (Max. Price Reg.).

Interval *I*	**Day *D***	**Day *(D − i)***	**30-day WMA**
Interval 1	Open. price	**Open. price (i:1,2)** Max. price (i:1,2) Min. price (i:1,2) Closing price (i:1-3) Vol. of trades (i:1-3) Cost Trx. (i:1-4,6,7)	Vol. of trades
Interval 2	Open. price	Open. price (i:1) Max. price (i:1,2) Min. price (i:1,2) Closing price (i:1-3) Vol. of trades (i:1-3) **Cost Trx. (i:1-7)**	Vol. of trades

Source: Personal collection.

Analyzing the Table 28, the second interval presents some attributes equal to the first interval. The exceptions were the maximum price of the D-3 and the minimum price 30-day Weighted Moving Average (WMA). Moreover, the volume of trades D-5 and the cost of transaction D-1 were added.

Comparing the attributes selected for the first and second intervals to forecast the minimum price (Table 29), only the opening price of the D-2 day was excluded. On the other hand, the cost of transaction of the D-5 day was added.

Table 30 – Daily Prediction E1 – Attr. Selected by Relief (Max. Price Reg.).

Interval *I*	**Day *D***	**Day *(D − i)***	**30-day WMA**
Interval 1	Open. price	Open. price (i:1) Max. price (i:1,2) Min. price (i:1,2) Closing price (i:1,2) Vol. of trades (i:1-3) **Cost Trx. (i:1-7)**	Vol. of trades Cost Trx.
Interval 2	Open. price	Open. price (i:1) Max. price (i:1,2) Min. price (i:1,2) Closing price (i:1,2) Vol. of trades (i:1-3) Cost Trx. (i:1-6)	Vol. of trades Cost Trx. **Trx. Fee**

Source: Personal collection.

In addition, in Table 30, it can be observed that, for closing prices, only the cost transaction of the *D*-7 day was not considered for the second interval. However, the transaction fee 30-day WMA was considered as relevant for this forecasting.

Analyzing the data of Figure 16 and Figure 17, it is observed that the ANNs show a greater variation within whole validation period. Also, it is possible to observe that the Ensemble A has less variation in all cases, because it is based on Recurrent Neural Networks (RNNs). However, in all cases, the Ensemble A do not achieve better results than traditional ANNs, this is shown in Figure 17. In the case of SVM algorithm, this variability can not be noticed because it is not a stochastic model.

Regarding the attribute selection methods (dimensionality reduction), the most successful were: correlation score (*Corr*), *InfoGain* and *CFS*. However, for the second data set (interval 2), it was necessary to use all the attributes (including international economic indicators). This means that the attributes in this interval show a similar importance.

Although the Ensemble B did not improve the classification results, for both intervals it shows better results for the first 50 days of prediction. In addition, it shows a lower variability than the ANN algorithm.

Similar to the prediction of Bitcoin price direction, stochastic models based on RNN show less variability than those based on ANN. However, here it can be mentioned that the SVM models show better performances in all cases and for both intervals.

In terms of dimensionality reduction, the best methods were: *Relief*, *CFS* and *InfoGain*. In both intervals, it was possible to reduce the number of attributes with an improvement in the performance of each forecasting algorithm.

Thus, regarding the regression experiments, the SVM algorithm obtained the best results for all predictions (maximum, minimum and closing prices) and for both intervals. In terms of maximum price prediction, it was obtained low MAPE (1.28% and 1.14% for *intervals 1* and *2*, respectively). The same occurs to forecast the closing price, where the SVM presents 1.91% and 1.81% of MAPE for *intervals 1* and *2*, respectively. The worst results were obtained for the minimum price (183.7% and 107.8% of MAPE). However, these results were a consequence of an abrupt decrease of the Bitcoin in Jun 23th, 2016. Thus, by disregarding this date, the SVM obtains 1.52% and 1.58% of MAPE, respectively, demonstrating its potential to predict the Bitcoin exchange rates.

5.1.2 Experiment 2 – Technical Indicators and Social Trends Approach (E2)

First, it is compared performance of each classifier using different versions of technical indicators. In Table 31 and Table 32 are compared performance of SVM classifiers for intervals 1 and 2, respectively. In addition, type of technical indicators and the algorithm's configuration are presented. Relating to the datasets, T1 and T2 represent the first and second sort of technical indicators, respectively. Likewise, type of technical indicators on discretized (D) and continuos (C) version are considered. For describe algorithm configuration, p and r represent polynomial and radial-basis function kernels, respectively. If kernel function employed is polynomial, then next three values are degree, gamma and regularization parameters; else, next two values are gamma and regularization parameters.

Table 31 – Daily Prediction E2 – SVM Results Technical Ind. (Interval 1).

	Parameter combination (*function*;*d*; γ; *c*) (*attr.*)			
	p;4;.6;5 (*T1-C*)	p;2;.5;5 (*T1-D*)	p;4;1.0;.5 (*T2-C*)	p;3;0.1;.5 (*T2-D*)
metric	mean ±*std*	mean ±*std*	mean ±*std*	mean ±*std*
auc	0.5480 ±*0.00*	**0.5568** ±*0.00*	0.5490 ±*0.00*	0.5240 ±*0.00*
acc	55.87% ±*0.00%*	56.34% ±*0.00%*	57.28% ±*0.00%*	53.99% ±*0.00%*
f-score	61.79% ±*0.00%*	61.41% ±*0.00%*	65.50% ±*0.00%*	61.11% ±*0.00%*
kappa	9.57% ±*0.00%*	11.24% ±*0.00%*	10.04% ±*0.00%*	4.82% ±*0.00%*

Source: Personal collection.

Table 32 – Daily Prediction E2 – SVM Results Technical Ind. (Interval 2).

	Parameter combination (*function*;*d*; γ; *c*) (*attr.*)			
	p;3;.2;10 (*T1-C*)	r;-;.9;.5 (*T1-D*)	r;-;2.2;10 (*T2-C*)	p;1;.1;.5 (*T2-D*)
metric	mean ±*std*	mean ±*std*	mean ±*std*	mean ±*std*
auc	0.5670 ±*0.00*	0.5321 ±*0.00*	**0.5766** ±*0.00*	0.5152 ±*0.00*
acc	55.89% ±*0.00%*	59.18% ±*0.00%*	58.36% ±*0.00%*	51.78% ±*0.00%*
f-score	59.65% ±*0.00%*	70.50% ±*0.00%*	64.15% ±*0.00%*	57.28% ±*0.00%*
kappa	12.60% ±*0.00%*	6.93% ±*0.00%*	14.86% ±*0.00%*	2.90% ±*0.00%*

Source: Personal collection.

According to the results shown in Table 31, the best configuration of SVM has polynomial kernel (2 of degree, 0.5 of gamma and 5 of regularization parameter), using the first set of technical indicators in its discretized version (T1-D).

In Table 32, it is observed that the best configuration of SVM has radial basis kernel (2.2 of gamma and 10 of regularization parameter), using the second set of technical indicators in its continuous version (T2-C).

In Table 33 and Table 34, performance of ANN for interval 1 and 2 are evaluated, respectively. Thus, it was used a statistical test to compare the obtained results, because it is known that ANNs present variability in their results due to the random initialization of the synaptic weights. Thus, p-value is considered to compare the AUC scores obtained by ANN models.

Table 33 – Daily Prediction E2 – ANN Results Technical Ind. (Interval 1).

	Parameter combination (*ep*; *mc*; *arch*) (*attr.*)			
	500;.1;40-20 (*T1-C*)	50;.1;10-50 (*T1-D*)	500;.1;60-90 (*T2-C*)	100;.1;90-30 (*T2-D*)
metric	mean ±*std*	mean ±*std*	mean ±*std*	mean ±*std*
auc	**0.5291** ±*0.01938*	0.4960 ±*0.0139*	**0.5375** ±*0.0276*	0.5033 ±*0.0076*
acc	52.30% ±*1.84%*	50.33% ±*1.67%*	53.62% ±*4.03%*	51.17% ±*0.57%*
f-score	54.29% ±*4.72%*	55.81% ±*2.76%*	56.22% ±*8.98%*	56.91% ±*1.58%*
kappa	5.58% ±*3.80%*	-0.78% ±*2.72%*	7.38% ±*5.38%*	0.64% ±*1.52%*

Source: Personal collection.

In Table 33 it is compared if the difference between the values obtained using T1-C and T1-D is statistically significant. Thus, a p-value of 0.0157 is obtained, therefore, T1-C presents the best mean performance. However, when compared T1-C with T2-C, it was not identified a statistical difference, because, the p-values were greater than 0.05.

Table 34 – Daily Prediction E2 – ANN Results Technical Ind. (Interval 2).

	Parameter combination (*ep*; *mc*; *arch*) (*attr.*)			
	1000;.1;80-30 (*T1-C*)	1000;.1;100-30 (*T1-D*)	1000;.1;30-70 (*T2-C*)	1000;.1;10-10 (*T2-D*)
metric	mean ±*std*	mean ±*std*	mean ±*std*	mean ±*std*
auc	**0.5276** ±*0.0191*	0.5160 ±*0.0121*	**0.5381** ±*0.0190*	0.4889 ±*0.0053*
acc	48.99% ±*2.64%*	54.30% ±*2.21%*	49.21% ±*3.18%*	51.23% ±*1.04%*
f-score	46.23% ±*5.77%*	62.91% ±*3.42%*	44.35% ±*7.56%*	59.80% ±*1.77%*
kappa	4.91% ±*3.42%*	3.25% ±*2.50%*	6.70% ±*3.48%*	-2.20% ±*1.05%*

Source: Personal collection.

In Table 34, the comparison of scores obtained between the datasets T2-C and T2-D has a p-value of 0.0005. This result is lower than 0.05 and, for this reason, T2-C is considered better than T2-D. However, when compared T1-C with T2-C, it was not identified a statistical difference, because the p-values were greater than 0.05.

Therefore, is selected the second set of technical indicators in its continuous version (T2-C) because presents good results in both intervals for ANN and SVM methods. After that, is tested the addition of others selected attributes from Blockchain, Economic Indices and Social Trend Information (detailed in Section D). Thus, in Table 35 and Table 36 are presented the performances of SVM models for intervals 1 and 2, respectively.

Table 35 – Daily Prediction E2 – SVM T2-C vs *B*-*E*-*S* (Interval 1).

	Parameter combination (*function*;*d*; γ; *c*) (*attr.*)		
	p;3;.07;10 (**w/** *B*)	p;1;.1;.5 (**w/** *E*)	p;1;.08;.5 (**w/** *S*)
metric	mean ±*std*	mean ±*std*	mean ±*std*
auc	**0.5807** ±*0.00*	0.5679 ±*0.00*	0.5735 ±*0.00*
acc	62.44% ±*0.00%*	61.50% ±*0.00%*	61.97% ±*0.00%*
f-score	72.41% ±*0.00%*	72.11% ±*0.00%*	72.35% ±*0.00%*
kappa	17.29% ±*0.00%*	14.64% ±*0.00%*	15.83% ±*0.00%*

Source: Personal collection.

Table 36 – Daily Prediction E2 – SVM T2-C vs *B*-*E*-*S* (Interval 2).

	Parameter combination (*function*;*d*; γ; *c*) (*attr.*)		
	p;4;.7;.5 (*w/* *B*)	r;-;1.7;5 (*w/* *E*)	r;-;2.5;100 (*w/* *S*)
metric	mean ±*std*	mean ±*std*	mean ±*std*
auc	0.5652 ±*0.00*	**0.5897** ±*0.00*	0.5732 ±*0.00*
acc	52.60% ±*0.00%*	61.10% ±*0.00%*	61.64% ±*0.00%*
f-score	50.43% ±*0.00%*	68.30% ±*0.00%*	70.95% ±*0.00%*
kappa	11.54% ±*0.00%*	17.95% ±*0.00%*	15.32% ±*0.00%*

Source: Personal collection.

For first interval (Table 35), the best result is obtained with the addition of the Blockchain attributes. Moreover, this result shows greater value of AUC in comparison with the results presented in Table 31.

Relating to second interval (Table 36), with addition of the Economic attribute is obtained better performance. Moreover, this result presents greater value of AUC when compared with Table 32. Similarly, Table 37 and Table 38 compare the performances of ANN models for intervals 1 and 2, respectively; but there is no statistically significant difference with the previously obtained results.

In addition, the performances of models are compared using all attributes selected, that is, the second set of technical indicators (T2-C) together with the attributes selected for Blockchain, Economic indices and Social Trends Information (Section D). Thus, the performance of SVM and ANN are compared together in order to identify the outperform model. Thus, in Table 39 and Table 40 are presented comparisons between SVM and ANN models for intervals 1 and 2, respectively. In interval 1, SVM outperformed than ANN

Table 37 – Daily Prediction E2 – ANN T2-C vs *B*-*E*-*S* (Interval 1).

	Parameter combination (*ep*; *mc*; *arch*) (*attr.*)		
	100;.1;50-60 (**w/** *B*)	100;.1;40-50 (**w/** *E*)	50;.1;50-20 (**w/** *S*)
metric	mean ±*std*	mean ±*std*	mean ±*std*
auc	0.5565 ±*0.0198*	0.5397 ±*0.0136*	0.5430 ±*0.0155*
acc	59.44% ±*2.36%*	57.18% ±*2.11%*	55.21% ±*1.95%*
f-score	69.17% ±*3.35%*	66.54% ±*3.18%*	60.79% ±*3.03%*
kappa	11.94% ±*4.27%*	8.30% ±*2.94%*	8.56% ±*3.10%*

Source: Personal collection.

Table 38 – Daily Prediction E2 – ANN T2-C vs *B*-*E*-*S* (Interval 2).

	Parameter combination (*ep*; *mc*; *arch*) (*attr.*)		
	500;.1;40-80 (**w/** *B*)	50;.1;80-60 (**w/** *E*)	100;.1;80-80 (**w/** *S*)
metric	mean ±*std*	mean ±*std*	mean ±*std*
auc	0.5524 ±*0.0084*	0.5576 ±*0.0301*	0.5515 ±*0.0347*
acc	50.36% ±*4.07%*	52.44% ±*4.33%*	48.99% ±*5.98%*
f-score	43.57% ±*12.70%*	50.88% ±*9.38%*	38.73% ±*13.77%*
kappa	9.15% ±*1.65%*	10.40% ±*5.49%*	8.99% ±*6.60%*

Source: Personal collection.

(p-value equal 0.0385). However, in interval 2 the difference between SVM and ANN is not significant (p-value equal 0.0776). Thus, Figure 19 and Figure 20 show the confusion matrix in *interval 1* and *interval 2*, respectively.

Table 39 – Daily Prediction E2 – SVM vs ANN All Attr. (Interval 1).

	SVM Parameter combination (*function*;*d*; γ; *c*) ANN Parameter combination (*ep*; *mc*; *arch*)	
	p;3;.7;10 *SVM*	50;.1;90-60 *ANN*
metric	mean ±*std*	mean ±*std*
auc	**0.5902** ±*0.00*	0.5607 ±*0.0267*
acc	62.44% ±*0.00%*	56.90% ±*4.00%*
f-score	71.22% ±*0.00%*	61.21% ±*8.61%*
kappa	18.94% ±*0.00%*	12.10% ±*5.36%*

Source: Personal collection.

Therefore, in Table 41 and Table 42 are showed the results obtained in the present experiment in comparison with Section 5.1.1 and, if is possible, with previous studies. Thus, it is observed that the technical indicators and the selection of attributes proposed in this experiment obtain a better results in both intervals comparing the values of AUC with the previous experiment or its accuracy with results of other authors.

Figure 19 – Daily prediction E2 – Confusion Matrix of SVM (interval 1).

		Actual 1	Actual 0
Predicted	1	46.48%	11.74%
	0	25.82%	15.96%

Source: Personal collection

Table 40 – Daily Prediction E2 – SVM vs ANN All Attr. (Interval 2).

SVM Parameter combination ($function$;d; γ; c) ANN Parameter combination (ep; mc; $arch$)		
	r;-;.05;1 *SVM*	100;.1;90-70 *ANN*
metric	mean $\pm std$	mean $\pm std$
auc	**0.5910** $\pm$*0.00*	**0.5638** $\pm$***0.0301***
acc	63.84% $\pm$*0.00%*	51.07% $\pm$*5.92%*
f-score	73.06% $\pm$*0.00%*	43.46% $\pm$*14.84%*
kappa	19.27% $\pm$*0.00%*	11.25% $\pm$*5.90%*

Source: Personal collection.

Figure 20 – Daily prediction E2 – Confusion Matrix of SVM (interval 2).

		Actual 1	Actual 0
Predicted	1	46.44%	14.93%
	0	21.23%	17.40%

Source: Personal collection

Finally, as a summary of the experience in the use of the models of both the E1 and E2 experiments, it is possible to point out the following:

- The results obtained through the best ANN models were taken as a basis to explore the use of other models that presented a similar or better level of results;
- From the above, it was identified that SVM had similar or better results. However, it could be detected that when the attributes were increased, especially in the E2 experiment, the performance of this model degraded exponentially;

Table 41 – Daily Prediction – Comparison of best results (Interval 1).

Model	**AUC**	**Accuracy**
Ensemble A (E1)	0.58	62.91%
SVM (E2)	**0.59**	**62.44%**
LSTM *		52.78%

Source: Personal collection.

(*) Mcnally (2016).

Table 42 – Daily Prediction – Comparison of best results (Interval 2).

Model	**AUC**	**Accuracy**
SVM (E1)	0.58	59.45%
SVM (E2)	**0.59**	**63.84%**

Source: Personal collection.

- One of the benefits of using SVM is that it is not a stochastic model, which guarantees that its results do not vary randomly, as in the case of ANN;
- In order to reduce the randomness of the results of models based on ANN, it was decided to develop Ensemble models that use functions of linear outputs (classification through regression).

5.2 Intra-daily Prediction

5.2.1 Data-stream/Online Learning Approach

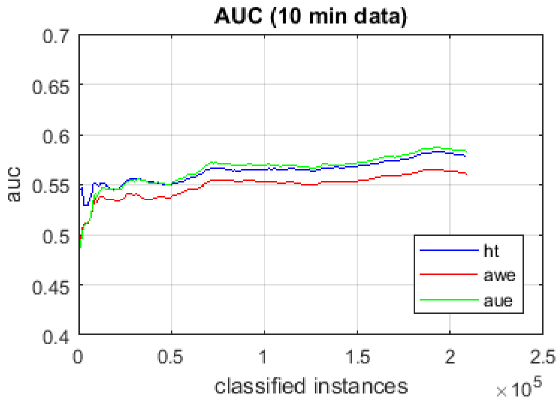
In the Figure 21 is presented the AUC values of the algorithms tested, comparing each of them in each of the scenarios proposed in the study.

Is possible to see the difference between interleaved and prequential evaluation method. In the sub-figures (a), (c) and (e) the AUC metric has a stable evolution, because interleaved method gathers all results over time; meanwhile in the sub-figures (b), (d) and (f) the AUC presents a volatile behavior, because prequential method reset all values each slide window.

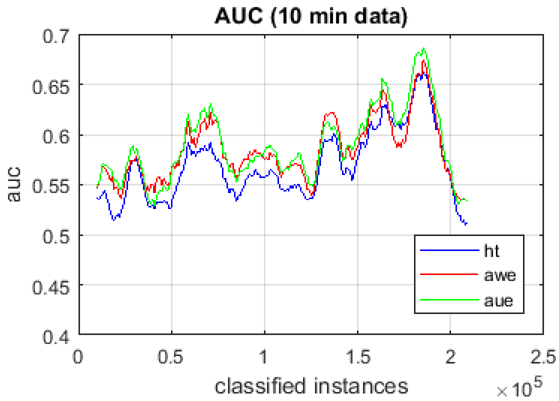
The sub-figures (a) and (b) show the classification performance obtained by dataset that it considered only 10-minutely frequency data (Table 19). In (a) the Hoedffing Tree (HT) model starts with a higher AUC but after 50,000 instances, the Accuracy-Updated Ensemble (AUE) model presents a greater classification performance. The sub-figure (b) presents a behavior more heterogeneous, but in general with best performance of AUE. The results obtained through the dataset formed with the best attributes according infor-

Figure 21 – Intra-daily Prediction – Interleaved Test-Train / Prequential Results.

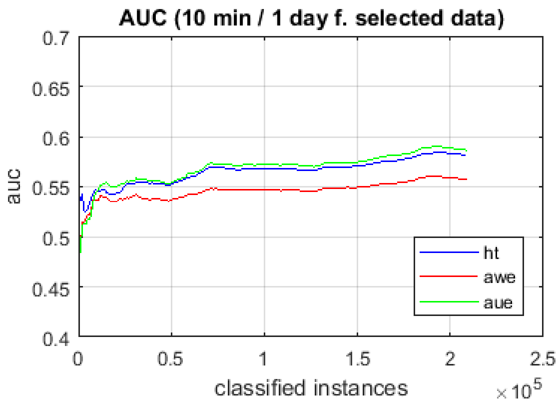
(a) using 10-min features (Interleaved)



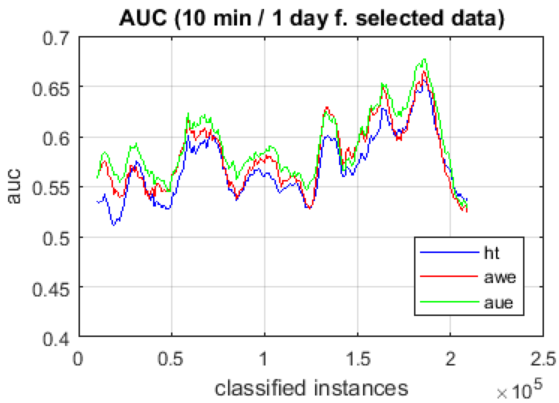
(b) using 10-min features (Prequential)



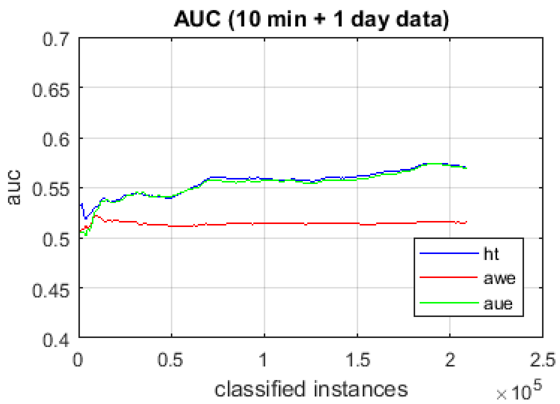
(c) 10-min/daily f. selection (Interleaved)



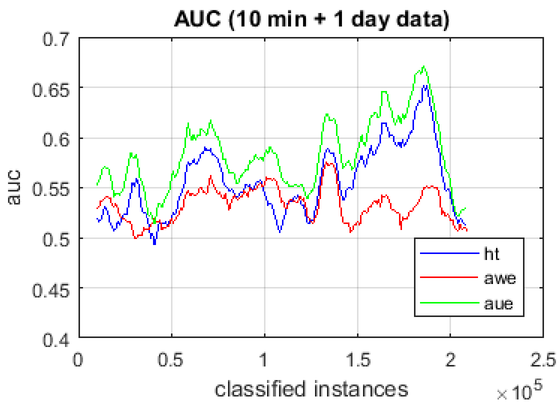
(d) 10-min/daily f. selection (Prequential)



(e) 10-min + daily features (Interleaved)



(f) 10-min + daily features (Prequential)

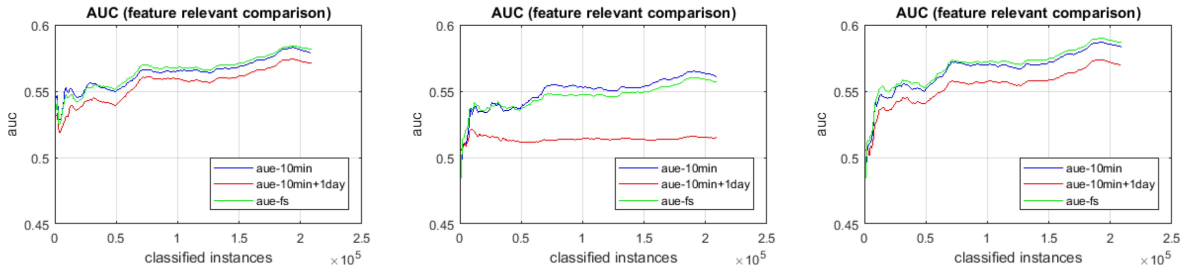


Source: Personal collection.

mation gain feature selection technique, is presented on the sub-figure (c) and (d) where these attributes are detailed in the Table 21. Similar to the previous graphics, AUE obtains the best results followed very closely by HT. Finally, in the sub-figures (d) and (e) is presented the results obtained by the dataset formed by all 10-minutely (Table 19) and daily frequency attributes (Table 20), where again Accuracy-Weighted Ensemble (AWE) presents the worst results.

Figure 22 – Intra-daily Prediction – Comparison Results (Interleaved Test-Train).

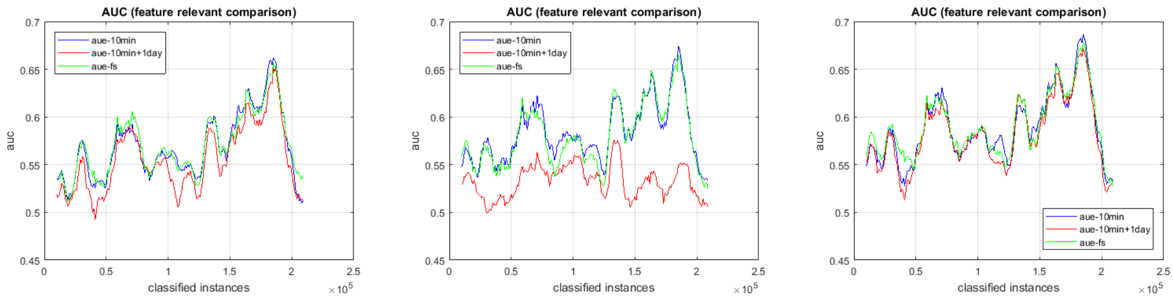
(a) HT (b) AWE (c) AUE



Source: Personal collection.

Figure 23 – Intra-daily Prediction – Comparison Results (Prequential).

(a) HT (b) AWE (c) AUE



Source: Personal collection.

Regarding to evaluate the impact of the classification performance on each algorithm through of the use different sorts of dataset, they were elaborate the Figure 22 and Figure 23 for interleaved and prequential, respectively.

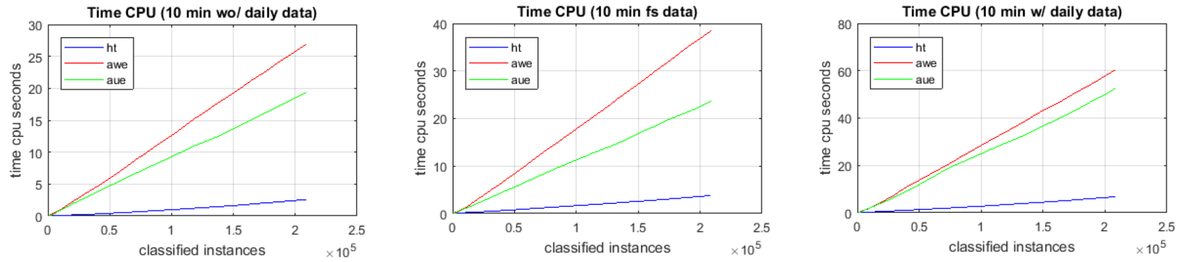
In Figure 22 is possible to observe that performance obtained by 10-minutely (Table 19) and feature selected data (Table 21) presents the best values. In Figure 23 similar to previous Figure 22 the best AUC scores were obtained by datasets detailed in Table 19 and Table 21, but the algorithm AUE has the least variability. Conversely, AWE presents the higher volatility results with a strong negative impact, when using all the daily frequency data.

Although the difference of results in prequential evaluation are less than in interleaved test-train evaluation, in all of cases the effect to include all daily information is negative and with exception to AWE, the use of feature selected data (Table 21) presents the best classification performance results.

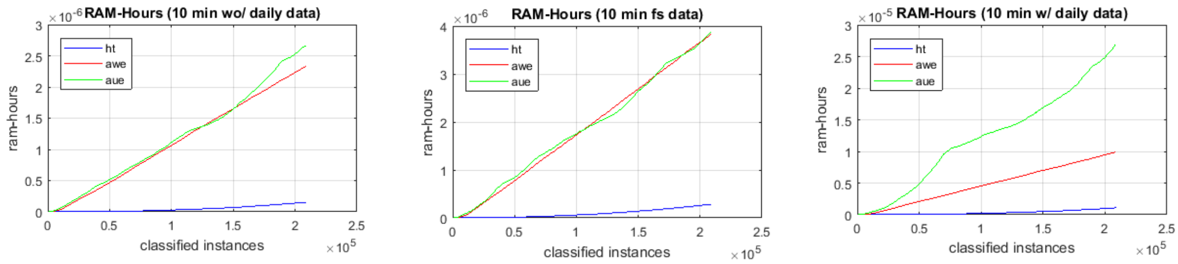
In addition, it was realized a computational performance test, comparing the amounts of time and memory consuming by each algorithms. Thus, the Figure 24 shows the evolution of time-memory consuming as the number of attributes increases, because (a) and (d) consider the computational performance using 10-minutely dataset that it

Figure 24 – Intra-daily Prediction – Time-Memory Performance.

(a) 10 min attr. (b) 10 min/1 day fs. attr. (c) 10 min + 1 day attr.



(d) 10 min attr. (e) 10 min/1 day fs. attr. (f) 10 min + 1 day attr.



Source: Personal collection.

has 9 attributes; (b) and (e) show the results considering attributes selected by information gain method that it has 12 attributes; and (c) and (f) show the computational performance considering all available attributes with 29 attributes. In general, although the difference in classification capacity between HT and AUE is small in the majority of cases, it is possible to observe that the computational cost of HT is significantly lower.

Finally, in Table 43 is presented the consolidated results where is calculated the average of prequential classification performance results obtained in the one year more recent data (∼52K instances), where the best results was obtained by AUE model with better values than obtained by Madan, Saluja & Zhao (2015).

Table 43 – Intra-daily Prediction – Prequential Avg. (1*year* of recent data).

Dataset	Algorithm	*AUC*	Accuracy	Kappa
10 min attr.	HT	0.5937	58.93%	0.17
10 min attr.	AWE	0.5982	56.79%	0.13
10 min attr.	AUE	0.6111	58.47%	0.16
10 min/1 day fs. attr.	HT	0.5974	59.16%	0.17
10 min/1 day fs. attr.	AWE	0.5974	57.04%	0.13
10 min/1 day fs. attr.	AUE	**0.6113**	**58.66%**	0.16
10 min + 1 day attr.	HT	0.5818	59.18%	0.17
10 min + 1 day attr.	AWE	0.5274	52.62%	0.04
10 min + 1 day attr.	AUE	0.6043	58.62%	0.16
Random Forest *			57.40%	–

Source: Personal collection.

(*) Madan, Saluja & Zhao (2015).

6 Conclusions and Future Works

6.1 Daily Prediction

For predicting the Bitcoin price direction, it is possible to highlight the selection of attributes by correlation analysis (*Corr*) and information gain analysis (*InfoGain*) as the techniques with the highest effectiveness rate. However, for the larger interval, the best result was obtained through the use of all attributes. This implies that it is still necessary to look for other data pre-processing in order to effectively select the attributes necessary for prediction. Also, for the regression experiment performed to forecast the maximum, minimum and closing Bitcoin exchange rates, it was observed that *Relief* is the technique that obtained the best results for all scenarios.

For first experiment, in terms of attribute analysis, in first interval (from August 19th, 2013 to July 19th, 2016), it was noticed a better classification performance with the internal attributes (from Blockchain). On the other hand, for the second interval, the best result was obtained with a combination of internal and external attributes. Regarding the regression experiment, in both intervals, the best results were obtained with the internal attributes. Similar to above, in second experiment, it was observed that for the second interval, incorporation of information on the international economic index DAX allows to improve the yield of prediction. In contrast to first interval, information coming from Blockchain demonstrated to be more relevant. This could be an indication that in long term Bitcoin behaves more like a traditional instrument and therefore is increasingly affected by international context and economic fundamentals, similar to that indicated by Li & Wang (2017). In addition, inclusion of information on trends in social media improves the predictability.

In first experiment, proposed Ensemble A model (based on Recurrent Neural Networks (RNNs)) obtains the best results for the first interval, and in comparison with the previous work of Mcnally (2016), a considerable improvement of 10% in precision can be appreciated. In second experiment, it was performed tests with different groups of attributes in order to identify which are the most relevant to make predictions about the Bitcoin price direction. Thus, the set of technical indicators proposed in previous study on stock price direction prediction Qiu & Song (2016) and detailed in Section B presents a higher prediction capacity than the commonly used technical indicators. Thus, the results of the second experiment using techniques proposed by Qiu & Song (2016) as the data of trends in social media manages to improve the prediction performance obtained in the first experiment, where for first interval it can be obtained an Area Under ROC Curve (AUC) of 59.10% and an accuracy of 63.84%; and for second interval, it was reached an

AUC of 59.02% and an accuracy of 62.24%. In both cases, these results are obtained by the algorithm Support Vector Machine (SVM).

The previous results (Gao & Lei (2017)) become relevant if we observe that in case of prediction of movement of more traditional instruments such as oil, the approximate accuracy is 70% for similar periods. The foregoing implies that even though Bitcoin is much more volatile than oil, the predictions are relatively close.

Regarding the regression experiments, the SVM algorithm obtained the best results for all predictions (maximum, minimum and closing prices) and for both intervals with a MAPE between 1.28% and 1.91%.

6.2 Intra-daily Prediction

In this experimental study, it is concluded that the addition of daily frequency information and later the selection of the most relevant attributes according to the information gain technique improves the intra-daily classification performance. In addition, the classification performance of Accuracy-Updated Ensemble (AUE) algorithm outperformed than Hoedffing Tree (HT) and Accuracy-Weighted Ensemble (AWE) techniques. About the computational cost HT presents the best results, where AUE algorithm obtain a better time-cost than AWE but worse memory-cost.

Furthermore, prequential is better than interleaved evaluation for understand changes about performance classifications in different time windows. Thus, considering the average of classification performance for the most recent data (one year: ∼52 K instances) the following best results were obtained: 0.6113 of AUC, 58.66% of accuracy and 0.16 of kappa index, being these results better than presented in Madan, Saluja & Zhao (2015), demonstrating the usefulness and feasibility of using data stream learning algorithms for time series predictions.

6.3 Future Works

In future work, in order to improve non-linear patterns identification, application of recurrence plot generation will be explored from input data (Romano et al. (2004), Hatami, Gavet & Debayle (2017)) and after that, use image classification techniques based on deep learning as Convolutional Neural Network (CNN) (Hatami, Gavet & Debayle (2018)).

Bibliography

ABBOUSHI, S. Global Virtual Currency - Brief Overview. *Competition Forum*, v. 14, n. 2, p. 230–236, 2016. Quoted 4 times on pages 21, 25, 26, and 27.

ADELI, H.; HUNG, S.-L. *Machine learning: neural networks, genetic algorithms, and fuzzy systems*. [S.l.]: John Wiley & Sons, Inc., 1994. Cited on page 30.

ALSTYNE, M. van. Why Bitcoin has value. *Communications of the ACM*, v. 57, n. 5, p. 30–32, 2014. Quoted 2 times on pages 22 and 28.

BALCILAR, M. et al. Can volume predict bitcoin returns and volatility? A quantiles-based approach. *Economic Modelling*, v. 64, p. 74–81, 2017. Quoted 4 times on pages 22, 38, 39, and 44.

BALLINGS, M. et al. Evaluating multiple classifiers for stock price direction prediction. *Expert Systems with Applications*, v. 42, n. 20, p. 7046–7056, 2015. Quoted 2 times on pages 42 and 48.

BIFET, A. et al. Data Stream Mining. *Methodology*, v. 8, n. May, p. 127–141, 2011. Quoted 3 times on pages 42, 67, and 68.

BISHOP, C. M. Pattern recognition and machine learning, 2006. *Korean Society of Civil Engineers*, v. 60, n. 1, p. 78–78, 2012. Cited on page 49.

BRZEZIŃSKI, D.; STEFANOWSKI, J. Accuracy Updated Ensemble for Data Streams with Concept Drift. *Hybrid artificial intelligent systems*, p. 155–163, 2011. Quoted 2 times on pages 42 and 69.

CHOKUN, J. *Who accepts Bitcoins as payment? List of companies, stores, shops.* 2016. Retrieved June 11, 2016 from http://www.bitcoinvalues.net/who-accepts-bitcoins-payment-companies-stores-take-bitcoins.html. Cited on page 21.

CIAIAN, P.; RAJCANIOVA, M.; KANCS d'Artis. The economics of bitcoin price formation. *Applied Economics*, v. 48, n. 19, p. 1799–1815, 2016. Quoted 4 times on pages 22, 38, 39, and 53.

COCCO, L.; CONCAS, G.; MARCHESI, M. Using an artificial financial market for studying a cryptocurrency market. *Journal of Economic Interaction and Coordination*, v. 12, n. 2, p. 345–365, 2017. Quoted 3 times on pages 26, 27, and 47.

CONTI, M. et al. A survey on security and privacy issues of bitcoin. *IEEE Communications Surveys & Tutorials*, (Early Access), p. 1–39, 2018. Cited on page 21.

CRAMER, S. et al. An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives. *Expert Systems with Applications*, v. 85, p. 169–181, 2017. Quoted 2 times on pages 41 and 48.

CUTHBERTSON, A. *Bitcoin now accepted by 100,000 merchants worldwide. International Business Times*. 2015. Retrieved April 25, 2016 from www.venturebeat.com. Cited on page 21.

DAI, J.; QING, X. Attribute selection based on information gain ratio in fuzzy rough set theory with application to tumor classification. *Applied Soft Computing*, v. 13, n. 1, p. 211–221, 2013. Cited on page 34.

DAS, P.; BISOI, R.; DASH, P. Data decomposition based fast reduced kernel extreme learning machine for currency exchange rate forecasting and trend analysis. *Expert Systems with Applications*, Elsevier, v. 96, p. 427–449, 2018. Cited on page 42.

DIETTERICH, T. G. Machine-learning research. *AI magazine*, v. 18, n. 4, p. 97, 1997. Cited on page 32.

DOMINGOS, P.; HULTEN, G. Mining High-Speed Data Streams. *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, p. 71–80, 2000. Quoted 3 times on pages 42, 67, and 68.

DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern classification*. New York: John Wiley & Sons, 2012. Quoted 2 times on pages 30 and 49.

EROSS, A. et al. The Intraday Dynamics of Bitcoin. n. August, 2017. Quoted 2 times on pages 38 and 40.

GAMA, J. *Knowledge Discovery from Data Streams*. Lisbon: CRC Press, 2010. 255 p. Quoted 3 times on pages 32, 42, and 70.

GAMA, J.; FERNANDES, R.; ROCHA, R. Decision trees for mining data streams. *Intelligent Data Analysis*, v. 10, p. 23–45, 2006. Cited on page 42.

GAO, L. et al. Research and application of data mining feature selection based on relief algorithm. *Journal of Software*, v. 9, n. 2, p. 515–522, 2014. Cited on page 33.

GAO, S.; LEI, Y. A new approach for crude oil price prediction based on stream learning. *Geoscience Frontiers*, Elsevier Ltd, v. 8, n. 1, p. 183–187, 2017. Quoted 3 times on pages 42, 65, and 88.

GERLEIN, E. A. et al. Evaluating machine learning classification for financial trading: An empirical approach. *Expert Systems with Applications*, v. 54, p. 193–207, 2016. Cited on page 22.

GEURTS, P. Pattern extraction for time series classification. In: *European Conference on Principles of Data Mining and Knowledge Discovery*. Berlin, Heidelberg: Springer, 2001. p. 115—-127. Quoted 2 times on pages 29 and 30.

GLASER, F. et al. Bitcoin-asset or currency? revealing users' hidden intentions. 2014. Quoted 2 times on pages 22 and 37.

GREAVES, A.; AU, B. Using the bitcoin transaction graph to predict the price of bitcoin. 2015. Quoted 3 times on pages 22, 40, and 65.

HALL, M. A. *Correlation-based feature selection for machine learning*. Thesis (PhD.) — The University of Waikato, 1999. Cited on page 34.

HATAMI, N.; GAVET, Y.; DEBAYLE, J. Bag of recurrence patterns representation for time-series classification. *Pattern Analysis and Applications*, Springer, p. 1–11, 2017. Cited on page 88.

HATAMI, N.; GAVET, Y.; DEBAYLE, J. Classification of time-series images using deep convolutional neural networks. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *Tenth International Conference on Machine Vision (ICMV 2017)*. [S.l.], 2018. v. 10696, p. 106960Y. Cited on page 88.

HOLMES, G.; KIRKBY, R.; PFAHRINGER, B. *MOA: massive online analysis*. 2007. Cited on page 67.

HUA, S.; SUN, Z. Support vector machine approach for protein subcellular localization prediction. *Bioinformatics*, v. 17, n. 8, p. 721–728, 2001. Cited on page 50.

HUANG, W.; NAKAMORI, Y.; WANG, S.-Y. Forecasting stock market movement direction with support vector machine. *Computers & Operations Research*, Elsevier, v. 32, n. 10, p. 2513–2522, 2005. Cited on page 49.

IGLESIAS, J. *Bitcoin: a new way to understand Payment Systems*. Thesis (PhD.) — Massachusetts Institute of Technology, 2015. Cited on page 37.

JAIN, A. K. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, v. 31, n. 8, p. 651–666, 2010. Cited on page 52.

JANG, H.; LEE, J. An empirical study on modeling and prediction of bitcoin prices with bayesian neural networks based on blockchain information. *IEEE Access*, v. 6, p. 5427–5437, 2018. Cited on page 42.

Jubert de Almeida, B.; Ferreira Neves, R.; HORTA, N. Combining Support Vector Machine with Genetic Algorithms to optimize investments in Forex markets with high leverage. *Applied Soft Computing*, Elsevier B.V., v. 64, p. 596–613, 2018. Cited on page 45.

KARA, Y.; Acar Boyacioglu, M.; BAYKAN, Ö. K. Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. *Expert Systems with Applications*, v. 38, n. 5, p. 5311–5319, 2011. Quoted 4 times on pages 41, 48, 49, and 55.

KHALILOV, M. C. K.; LEVI, A. A survey on anonymity and privacy in bitcoin-like digital cash systems. *IEEE Communications Surveys & Tutorials*, v. 20, n. 3, p. 2543–2585, 2018. Cited on page 21.

KHAN, K. S.; KUNZ, R.; KLERJNEN, J. Five steps to conducting a systematic review. *Journal of the royal society of medicine*, v. 96, 2003. Cited on page 37.

KIM, S. B.; RATTAKORN, P. Unsupervised feature selection using weighted principal components. *Expert Systems with Applications*, v. 38, n. 5, p. 5704–5710, 2011. Cited on page 34.

KIM, T. On the transaction cost of Bitcoin. *Finance Research Letters*, v. 0, p. 1–6, 2017. Cited on page 28.

KIM, Y. B. et al. Predicting fluctuations in cryptocurrency transactions based on user comments and replies. *PLoS ONE*, v. 11, n. 8, p. 1–18, 2016. Quoted 5 times on pages 22, 38, 39, 40, and 53.

KLEIN, T.; Pham Thu, H.; WALTHER, T. Bitcoin is not the new gold – a comparison of volatility, correlation, and portfolio performance. *International Review of Financial Analysis*, v. 59, p. 105–116, 2018. Quoted 2 times on pages 28 and 29.

KOCADAğLI, O.; AşIKGIL, B. Nonlinear time series forecasting with Bayesian neural networks. *Expert Systems with Applications*, v. 41, n. 15, p. 6596–6610, 2014. Quoted 2 times on pages 41 and 48.

KOTHARI, C. *Research methodology: methods and techniques*. New Delhi: New Age International, 2004. 1–418 p. Cited on page 23.

KRISTOUFEK, L. BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. *Scientific Reports*, v. 3, p. 1–7, 2013. Quoted 4 times on pages 22, 37, 38, and 53.

KRISTOUFEK, L. What are the main drivers of the bitcoin price? Evidence from wavelet coherence analysis. *PLoS ONE*, v. 10, n. 4, p. 1–15, 2015. Quoted 3 times on pages 21, 22, and 38.

LABOISSIERE, L. A.; FERNANDES, R. A.; LAGE, G. G. Maximum and minimum stock price forecasting of Brazilian power distribution companies based on artificial neural networks. *Applied Soft Computing*, v. 35, p. 66–74, 2015. Quoted 7 times on pages 29, 30, 41, 44, 45, 47, and 48.

LI, X.; WANG, C. A. The technology and economic determinants of cryptocurrency exchange rates: The case of bitcoin. *Decision Support Systems*, v. 95, p. 49–60, 2017. Quoted 6 times on pages 22, 38, 39, 47, 53, and 87.

LIN, C. C.; CHEN, C. S.; CHEN, A. P. Using intelligent computing and data stream mining for behavioral finance associated with market profile and financial physics. *Applied Soft Computing Journal*, Elsevier B.V., 2017. Cited on page 42.

MADAN, I.; SALUJA, S.; ZHAO, A. Automated bitcoin trading via machine learning algorithms. v. 20, p. 1–5, 2015. Quoted 5 times on pages 22, 40, 65, 86, and 88.

MATTA, M.; LUNESU, I.; MARCHESI, M. Bitcoin Spread Prediction Using Social And Web Search Media. *UMAP Workshops*, Dublin, p. 1–10, 2015. Quoted 2 times on pages 22 and 38.

MCINTYRE, K. H.; HARJES, K. Order Flow and the Bitcoin Spot Rate. *Applied Economics and Finance*, v. 3, n. 3, p. 136–147, 2016. Quoted 4 times on pages 21, 22, 27, and 28.

MCNALLY, S. *Predicting the price of Bitcoin using Machine Learning*. Thesis (PhD.) — National College of Ireland, 2016. Quoted 10 times on pages 15, 22, 40, 44, 46, 51, 73, 74, 83, and 87.

MONTGOMERY, D. C.; JENNINGS, C. L.; KULAHCI, M. *Introduction to time series analysis and forecasting*. Second edition. New Jersey: John Wiley & Sons, 2015. Quoted 2 times on pages 22 and 29.

MOREAU, E. *13 Major Retailers and Services That Accept Bitcoin*. 2018. Retrieved April, 2018 from www.lifewire.com. Cited on page 21.

MURPHY, K. P. *Machine Learning: A Probabilistic Perspective. Adaptive Computation and Machine Learning*. Massachusetts: MIT press, 2014. 1067 p. Quoted 3 times on pages 30, 33, and 35.

NAKAMOTO, S. Bitcoin: A peer-to-peer electronic cash system. *www.bitcoin.org*, p. 9, 2008. Quoted 2 times on pages 26 and 27.

NANOPOULOS, A.; ALCOCK, R. O. B.; MANOLOPOULOS, Y. Feature-based Classication of Time-series Data. *International Journal of Computer Research*, v. 10, n. 3, p. 49–61, 2001. Quoted 2 times on pages 29 and 30.

PATEL, J. et al. Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques. *Expert Systems with Applications*, v. 42, n. 1, p. 259–268, 2015. Quoted 4 times on pages 41, 45, 48, and 57.

PENG, Y. et al. The best of two worlds: Forecasting high frequency volatility for cryptocurrencies and traditional currencies with support vector regression. *Expert Systems with Applications*, v. 97, n. 1, p. 177–192, 2018. Quoted 2 times on pages 21 and 42.

PLATT, J. Sequential minimal optimization: A fast algorithm for training support vector machines. *Advances in Kernel Methods – Support Vector Learning*, p. 1–21, 1998. Cited on page 49.

POPPER, N. *Digital gold: The untold story of Bitcoin*. [S.l.]: Penguin UK, 2015. Cited on page 37.

QIU, M.; SONG, Y. Predicting the direction of stock market index movement using an optimized artificial neural network model. *PLoS ONE*, v. 11, n. 5, p. 1–11, 2016. Quoted 6 times on pages 22, 41, 48, 56, 65, and 87.

RATHER, A. M.; AGARWAL, A.; SASTRY, V. N. Recurrent neural network and a hybrid model for prediction of stock returns. *Expert Systems with Applications*, v. 42, n. 6, p. 3234–3241, 2015. Quoted 2 times on pages 41 and 48.

ROMANO, M. C. et al. Multivariate recurrence plots. *Physics letters A*, Elsevier, v. 330, n. 3-4, p. 214–223, 2004. Cited on page 88.

SILVA, E. L. d.; MENEZES, E. M. Metodologia da pesquisa e elaboração de dissertação. UFSC, Florianópolis, v. 4. ed. rev. atual, 2005. Cited on page 23.

TIBSHIRANI, R. *Glossary: machine learning vs statistics, Statistics 315a, Stanford University*. 2011. Retrieved March 21, 2018 from www-stat.stanford.edu. Cited on page 31.

URIARTE-ARCIA, A. V. et al. Data stream classification based on the gamma classifier. *Mathematical Problems in Engineering*, v. 2015, 2015. Quoted 3 times on pages 32, 69, and 70.

URQUHART, A. Price clustering in bitcoin. *Economics Letters*, v. 159, p. 145–148, 2017. Cited on page 21.

VASSILIADIS, S. et al. Bitcoin Value Analysis Based On Cross-Correlations. *Journal of Internet Banking and Commerce*, v. 22, n. S7, p. 1, 2017. Quoted 4 times on pages 22, 38, 39, and 44.

WANG, H. Mining Concept-Drifting Data Streams using Ensemble Classifiers. *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, p. 226–235, 2003. Quoted 2 times on pages 42 and 68.

WANG, J. et al. Improved v-support vector regression model based on variable selection and brain storm optimization for stock price forecasting. *Applied Soft Computing*, v. 49, p. 164–178, 2016. Cited on page 34.

WANG, J. Z. et al. Forecasting stock indices with back propagation neural network. *Expert Systems with Applications*, Elsevier Ltd, v. 38, n. 11, p. 14346–14355, 2011. Cited on page 46.

WOLPERT, D. H. The lack of a priori distinctions between learning algorithms. *Neural computation*, MIT Press, v. 8, n. 7, p. 1341–1390, 1996. Cited on page 36.

WU, C. Y.; PANDEY, V. K. The value of Bitcoin in enhancing the efficiency of an investor ' s portfolio. *Journal of Financial Planning*, v. 27, n. 9, p. 44–52, 2014. Quoted 2 times on pages 22 and 38.

YERMACK, D. Is Bitcoin a real currency? An economic appraisal. *National Bureau of Economic Research (NBER) Working Papers*, v. 19747, 2013. Cited on page 28.

ZHU, Y.; DICKINSON, D.; LI, J. Analysis on the influence factors of bitcoin's price based on VEC model. *Financial Innovation*, v. 3, n. 1, p. 3, 2017. Quoted 3 times on pages 22, 38, and 39.