UNIVERSITY OF LJUBLJANA
SCHOOL OF ECONOMICS AND BUSINESS

MASTER’S THESIS

# FORECASTING THE VALUE OF CRYPTO ASSETS BASED ON NEURAL NETWORKS AND SENTIMENT OF MARKET PARTICIPANTS

Ljubljana, April 2019 ANGELA BOGOEVSKA

## AUTHORSHIP STATEMENT

The undersigned Angela Bogoevska a student at the University of Ljubljana, School of Economics and Business, (hereafter: SEB LU), author of this written final work of studies with the title Forecasting the Value of Crypto Assets Based on Neural Networks and Sentiment of Market Participants, prepared under supervision of prof. dr. Aleš Berk Skok

D E C L A R E

1. this written final work of studies to be based on the results of my own research;
2. the printed form of this written final work of studies to be identical to its electronic form;
3. the text of this written final work of studies to be language-edited and technically in adherence with the SEB LU's Technical Guidelines for Written Works, which means that I cited and / or quoted works and opinions of other authors in this written final work of studies in accordance with the SEB LU's Technical Guidelines for Written Works;
4. to be aware of the fact that plagiarism (in written or graphical form) is a criminal offence and can be prosecuted in accordance with the Criminal Code of the Republic of Slovenia;
5. to be aware of the consequences a proven plagiarism charge based on the this written final work could have for my status at the SEB LU in accordance with the relevant SEB LU Rules;
6. to have obtained all the necessary permits to use the data and works of other authors which are (in written or graphical form) referred to in this written final work of studies and to have clearly marked them;
7. to have acted in accordance with ethical principles during the preparation of this written final work of studies and to have, where necessary, obtained permission of the Ethics Committee;
8. my consent to use the electronic form of this written final work of studies for the detection of content similarity with other written works, using similarity detection software that is connected with the SEB LU Study Information System;
9. to transfer to the University of Ljubljana free of charge, non-exclusively, geographically and time-wise unlimited the right of saving this written final work of studies in the electronic form, the right of its reproduction, as well as the right of making this written final work of studies available to the public on the World Wide Web via the Repository of the University of Ljubljana;
10. my consent to publication of my personal data that are included in this written final work of studies and in this declaration, when this written final work of studies is published.

Ljubljana, 24 April 2019 Author's signature: _________________________

# TABLE OF CONTENTS

# LIST OF FIGURES

## LIST OF TABLES



## LIST OF APPENDICES

## LIST OF ABBREVIATIONS

| | | |
|---|---|---|
| **sl.** | – | Slovene |
| **Crypto** | – | (sl. kriptografija); Cryptography/Cryptographic |
| **FinTech** | – | (sl. finančna tehnologija); Financial Technology |
| **Altcoins** | – | (sl. alternativni kovanci); Alternative coins |
| **IRS** | – | (sl. Ameriška zvezna davkarija); Internal Revenue Service |
| **NN** | – | (sl. nervronska mreža); Neural Network |
| **FF** | – | (sl. usmerjene nevronske mreže); Feedforward Neural Network |
| **RNN** | – | (sl. rekurzivne nevronske mreže); Recurrent Neural Network |
| **LSTM Algorithm** | – | (sl. algoritem dolgotrajne-kratkotrajne memorije); Long Term-ShortTerm Memory Algorithm |
| **PoW Algorithm** | – | (sl. dokaz za delo algoritem); Proof-of-Work algorithm |
| **ICO** | – | (sl. začetna ponudba kovancev); Initial Coin Offering |
| **ITS** | – | (sl. začetna prodaja žetonov); Initial Token Sale |
| **CAPM Model** | – | (sl. model določanja cen dolgoročnih naložb); Capital Asset Pricing Model |
| **DCF Model** | – | (sl. model diskontiranja denarnih tokov); Discounted Cash Flow Model |
| **Market Cap.** | – | (sl. tržna kapitalizacija); Market Capitalization |
| **AML Regulation** | – | (sl. zakon o preprečevanju pranja denarja); Anti-money Laundry Regulation |
| **CFT Regulation** | – | (sl. zakon o preprečevanju financiranje terorizma); Combating the Financing of Terrorism Regulation |
| **FINMA** | – | (sl. Švicarski komisiji za nadzor finančnih trgov); The Swiss Financial Market Supervisory Authority |
| **BTC** | – | (sl. bitcoin); Bitcoin |
| **LTC** | – | (sl. lajtkoin); Litecoin |
| **ETH** | – | (sl. eter); Ether |
| **SPX** | – | (sl. Standard & Poor’s 500 indeks); Standard & Poor’s 500 Index |
| **CBOE** | – | (sl. Ameriška opcijska borza v Chicagu); Chicago Board Options Exchange |

| | | |
|---|---|---|
| **VIX** | – | (sl. indeks volatilnosti); Chicago Board Options Exchange Volatility Index |
| **SPDR** | – | (sl. Standard & Poor's zlatni zaklad); Standard & Poor's Depository Receipts |
| **GLD** | – | (sl. SPDR zlatne delnice); SPDR Gold Shares |
| **TNX** | – | (sl. CBOE desetletna državna obveznica); CBOE 10-year Treasury-note Yield |
| **NVT Ratio** | – | (sl. razmerje omrežne vrednosti po transakciji); Network Value to Transaction Ratio |
| **NYSE** | – | (sl. Newyorška borza); New York Stock Exchange |
| **NASDAQ** | – | (sl. Nacionalno združenje trgovcev z vrednostnimi papirji); National Association of Securities Dealers Automated Quotations |
| **SA** | – | (sl. analiza sentimenta); Sentiment Analysis |
| **VADER** | – | (sl. slovar valenčnega ozaveščanja, pri analize sentimenta); Valence Aware Dictionary for Sentiment Reasoning |
| **API** | – | (sl. aplikacijski programski vmesnik); Application Programming Interface |
| **MA** | – | (sl. drseča sredina); Moving Average |
| **EMA** | – | (sl. eksponencialna drseča sredina); Exponential Moving Average |
| **MACD** | – | (sl. konvergenca divergenca drsečih sredin); Moving Average Convergence/Divergence |
| **RSI** | – | (sl. indeks relativne snage); Relative Strength Index |

# INTRODUCTION

Cryptocurrencies present encrypted digital currencies embodied to a fully distributed ledger technology, the Blockchain. Their price instability and the consequently increased attention towards the Blockchain, as the underline technology, questions the future of the present financial formation (Tapscott & Tapscott, 2016). The hype has initiated various considerations among the financial analysts and the investment community in general; the question is how to evaluate and model-in the cryptocurrency value? Due to the embodied complexity of crypto assets, the challenge of defining proper valuation model is yet outstanding (Hubrich, 2017).

This master thesis explores the predictive power of historical prices, crypto specific features, public sentiment and the neural network models for forecasting future movements on the cryptocurrency markets. Conclusively, the algorithmically-defined trading strategy performance will be compared with the basic Buy-and-Hold strategy.

# 1 RESEARCH OVERVIEW – CRYPTOCURRENCIES AND THE BLOCKCHAIN

Cryptocurrencies present purely digital assets, backed by complex cryptography and distributed hashing power used for solving individual blocks on the Blockchain (Tapscott & Tapscott, 2016). The Blockchain is a digital, public ledger, where cryptography is used for regulating the generation and verification of transactions (Crosby, Nachiappan, Pattanyak, Verma & Kalyanaraman, 2016). Decentralization is the key feature that differentiates blockchain systems from other networks. Unlike fiat currencies, crypto assets are not controlled by any central authority or legal entity – i.e. governments, central banks, public institutions (Peters, Panayl & Chapelle, 2015).

Cryptocurrency originated in the late 90s with the creation of digital money (Tapscott & Tapscott, 2016). However, the global surge of cryptocurrencies can be tracked since October 2008, when Bitcoin was launched as a project signed by Satoshi Nakamoto (Gantori et al., 2017). This invention has spurred the emergence of other altcoins which employ different algorithms but use the similar cryptographic technology.

According to CoinMarketCap.com, as of 1st of July 2018, the total value of the cryptocurrency market was U.S. $256.2 Billion. The ending of 2017 showed culmination of the optimism toward cryptocurrencies, with a 3,000 % observed growth in 2017. Starting at $17.7 Billion U.S. and reaching $598.5 Billion in market capitalization. However, after reaching their peak, the value has dropped by 57 % as of 1st of July 2018 (CoinMarketCap.com, 2018). This reveals the embodied high volatility of crypto assets as a new asset class.

Since their emergence, many new cryptocurrencies have been created to address the shortcomings of Bitcoin, such as: improving transaction time, throughput, scalability, and required resources. Correspondingly, more than 1,500 diverse crypto assets are being traded worldwide, as of the 1st of July 2018. Blockchain technology is developing fast, and the best types and use cases of crypto assets are yet expected to be developed.

The question for the regulatory status of the novel crypto asset class is still opened. To obtain a proper understanding of the matter of crypto assets, it is crucial to look at what de facto, money is, and if the cryptocurrencies can be classified as money. Essentially, money presents an abstraction with a set of fundamental characteristics which justify its use. Fiat money has the following use cases: a store of value, unit of account and a medium of exchange. The Government plays a critical role in controlling the money supply and money circulation (Bell, 2001). Cryptocurrencies are the first independent global currencies since the gold and the silver standard. Cryptocurrencies could be considered as continuation of the historical trend to replace physical items with abstract (digital) items. They present an interesting departure protecting against government overreach. Crypto assets provide an alternative avenue to traditional financial systems applied for preservation, transfer and wealth management. A discerning look indicates new financial technology with the capacity of fundamental changes in the overall economic landscape.

In addition, cryptocurrencies exhibit low adoption costs, divisibility and have no minimum account requirements. These characteristics make crypto assets a suitable substitution for traditional bank accounts. Having in mind that currently, over two million individuals do not hold personal bank accounts, cryptocurrencies can help integrate this unbanked population. At the same time, synergies across the economy and new tax revenues can be generated (Saidov, 2018).

On the other hand, their yet low transaction volume limits their wider, global adoption and recognition. For instance, the Bitcoin network can process only five to seven transactions in a second. While Visa or Mastercard can deal with thousands of transactions in a second. Clearly, cryptocurrencies need many enhancements, and the crypto space is continually engaged in finding workable solutions.

By looking for secure and fast means of storing, spending and moving value, cryptocurrencies are challenging the traditional pillars of the existing financial system. Despite their tenuous nature, crypto assets have attracted broad interest and monetary momentum, in their attempt to change the way transactions and investments are done and to create economic value. Based on this knowledge, the study will challenge the contemporary understanding of the crypto markets while attempting to develop a proper model for explaining the development of the future crypto assets value.

## 1.1 Research Problem

Cryptocurrencies are novel, unique instruments with growing acceptance among the financial institutions and the general investment public. They share some of the characteristics of the traditional currencies and, as well, can serve as platforms for more sophisticated financial products (Tapscott & Tapscott, 2016).

Analyzing their historic price trend development and questionable offerings, crypto assets can be easily misclassified as fraudulent (Wilson & Ateniese, 2015). In 2017 the media was overwhelmed by cryptocurrency news. The majority of analysts and investors have a tendency of treating cryptocurrencies as assets, rather than to compare them with fiat currencies (Culpan, 2017). In fact, the valuation of cryptocurrencies should differ from that of fiat currencies and other traditional financial instruments. Since cryptocurrencies are required for the genuine functionality of the blockchain technology, their economic potential arises from the utility of the blockchain application, and not from the currency itself. The cryptocurrency is secondary to the blockchain. Still, the true nature of cryptocurrencies remains somewhat unclear. Currently, they are differently treated by regulators, bankers, and the general public. For instance, the U.S. Internal Revenue Service (hereinafter: IRS) categorizes cryptocurrencies as property, while the U.S. legal system considers them as a commodity (IRS, 2014).

The key difference between the equity and cryptocurrency market lays in the fact that the network value of the latter is not assigned to a centralized corporation, but to the users and participants in the network itself. The value is generated with the creation and exchange of the associated currency. It is the real-world connection lead by the real-world utility of blockchain applications that, once overlooked, can make cryptocurrencies seem like a pure mathematical invention or a “Ponzi scheme” (Hubrich, 2017).

Hence, the question that arises is: how to model in the intrinsic value of such complex holdings? So far, many analysts and institutions have been working on the challenge to develop optimal valuation approach for cryptocurrencies that will successfully capture all connected non-stationary and noisy signals (Bouoiyour, Tiwari, Selmi & Olayeni, 2016). The analysis done so far is mainly focused on technical or relative models, which compare cryptocurrencies with other traditional asset classes such as stocks, gold or bonds. The author finds these approaches as inapplicable since cryptocurrencies hold unique characteristics that further differentiate them as a distinctive and novel asset class.

## 1.2 Research Approach

The advanced developments by FinTech are disrupting and reshaping the financial markets worldwide. Cryptocurrencies present one of the FinTech forces that triggers numerous alterations in the financial systems (Chishti & Barberis, 2016). Touching upon wide financial fields, they challenge the existence of the present payment systems, investment models and government regulations.

The purpose of this master thesis is to construct an explanatory model for cryptocurrency value prediction by leveraging on the field of Feedforward and Recurrent neural networks. The field of stock market predictions based on neural networks has grown due to the development of high-frequency, low-latency trading hardware, enhanced with robust machine learning algorithms (Gomber & Haferkorn, 2015). Hence, there is a sense in replicating this prediction methodology for the cryptocurrency market, as the network gains greater liquidity and general public adoption.

The model will combine crypto specific metrics and sentiment derived indicators as two key parameters for the purpose of the price study. The initial goal is to perform a diligent study of all potential key value drivers, including the investor sentiment influence on the investment decision processes. VADER lexicon rule-based methodology will be applied for the purpose of capturing public opinion and sentiment that affects individuals when deciding to take a position in crypto assets. The focus is set on testing the optimal algorithmic model that will aim to explain the current and, to a certain degree, predict the future worth of cryptocurrencies.

In addition, the obtained data will be applied as an input to the developed neural network models (FF and RNN models). The purpose is to investigate how the crypto specific metrics and the public sentiment can be utilized for forecasting future movements on the cryptocurrency market. In conclusion, the algorithmically-defined trading strategy performance will be compared with the basic Buy-and-Hold strategy.

The development and the programming of the NN models was guided and supported by assistant Matej Dobrevski from the Faculty of Computer and Information Science at University of Ljubljana.

## 1.3 Research Questions

Due to their growing popular appeal and investor acceptance, it has become increasingly important to perceive the factors that affect the value formation of crypto assets. Given the internal complexity of the blockchain technology and the crypto market, methodological problems arise when one tries to assess the coin value - facing non-stationarity and non-linearity challenges.

Given the problem background and purpose described, this research study aims to identify the likely determinants of cryptocurrency value formation. The research questions are formulated to define the scope of the analysis of cryptocurrency value creation, investor decision making processes and construction of proper valuation methodology.

The first section of the study will be focused on testing the applicability of neural network methodology for making cryptocurrency value predictions. Based on the technical statistic-based features specific for cryptocurrency assets, a pricing model is constructed for three focus cryptocurrencies: Bitcoin, Litecoin and Ether. The model which can plausibly explain cryptocurrency value is constructed with the help of Feedforward and the Recurrent Neural Networks. The quantitative research study is performed based on the neural network algorithms applied as a tool in defining the optimal predictive model. With these models (FF and RNN) the author will try to explain the present value creation of cryptocurrencies. The end goal is to test the performance of the predictive models on predefined, 3-day trading strategy.

*RESEARCH QUESTION I: Can cryptocurrency value be predicted with the application of the Feedforward (FF) and Recurrent Neural Network (RNN – LSTM) algorithm? Can the algorithmic trading models improve cryptocurrency risk-adjusted returns?*

The reasoning behind the second section of this research study is to understand how public sentiment, once quantified, can contribute towards explaining and predicting cryptocurrency values. When the asset is new, new information has been released, or big players are making noticeably large bets, investing can result as a rational response to this present state of information. Moreover, collective optimism driven by peer influences among individual traders can create an endogenous effect relevant for understanding the investment dynamic. Thus, the author will test the research question:

*RESEARCH QUESTION II: Can the predictive power of the neural network models be improved by introducing public sentiment feature related to cryptocurrencies and estimated based on simple rule-based model (VADER)?*

The both research question will be tested with the neural network algorithm and cumulative results will be presented for all three focus crypto assets: Bitcoin, Litecoin and Ether.

## 1.4 Research Purpose

In the management portfolio, the alternative investments are known as commodities, hedge funds, real estate, and private equity. Traditionally, alternative investments have a lower historical correlation to the conventional asset classes – stocks, bonds and cash equivalents (Naik, Devarajan, Nowobilski, Page & Pedersen, 2016). This makes them a good variant for investment portfolio diversification.

Apart from the discussion on whether cryptocurrencies can become part of the mainstream financial system, billions of US dollars of cryptocurrencies are being traded worldwide (Filippi, 2014). There are several reasons for this. Firstly, crypto assets provide solid liquidity which makes them a good investment opportunity. Critics are addressing the limited number of coin creation, arguing that this can jeopardize the investing ability like with the limited gold supply. However, there are many promising altcoins in place and the number is still increasing, thus the whole crypto market could be seen as a good alternative to the traditional systems. Particularly it is a good alternative for diversification against mainstream assets. In addition, prices of cryptocurrencies do not fluctuate in the same direction as the marketplace, indicating the low correlation of overall returns (Hayes, 2017).

The valuation of cryptocurrencies is significantly different from the valuation of traditional financial instruments. A meaningful portion of cryptocurrencies have fixed supply, so evaluating them as money cannot be appropriate. Also, unlike equities or bonds, digital currencies do not generate any cash flow, making the discounted cash flow (hereinafter: DCF) valuation inapplicable. Furthermore, the performed studies are focusing more on the relative models, comparing altcoins to other asset classes such as gold, stocks or bonds. These approaches are less suitable since cryptocurrencies hold unique characteristics that further separates them as a crucially distinctive, diverse and novel asset class.

Crypto assets are given to investors as proof of future cash flow, payments, or possible future exchange, or the right to participate, vote, build blocks or purchase (Chuen, Guo & Wang, 2017). On top of the future cryptocurrency benefits, the network effect of cryptocurrency can be a crucial factor in its valuation, for the connected technology and perceived value of the cryptocurrency by the public.

Crypto assets exhibit some specific characteristics. Their prices are prone to constant fluctuations. Moreover, crypto assets are hardly taxable since there are various events related to them, each requiring arguably different methodological approach. Price volatility generates big issues regarding the valuation processes. Since monitoring a crypto coin purchase date is difficult and the currency is extremely volatile, the IRS defined detailed valuation formula to determine taxable gains and losses related to cryptocurrencies. Three different taxable events occur in the lifecycle of a crypto transaction, which results in a quite difficult valuation. Since the transactions can take the form of either receipt for mining, the sale of investment or use as currency, the right approach will incorporate these differences. While the sale as investments can be treated similarly in the form of a gain or loss, treatment of the mining gain can be approached as ordinary income for the performance of service. This can cause difficulties for the valuation if the source of the coins cannot be determined.

Cryptocurrencies have been the subject of many recent debates, and there are several important questions which require attention. Are cryptocurrencies a bubble? Are they used for illegal activities? Will governments around the globe try to regulate them? The core underlying certainty is that cryptocurrency future is unknown. Hence, it is uncertain whether they will turn out to be an invasive pest or a useful addition to the financial ecosystem. There is a well-documented efficacy of some factors across a wide range of asset classes that come close to being treated as a “law of nature” for the financial markets. Thus, this raises the question of whether the same pattern is present among cryptocurrencies, as well.

## 1.5 Expected Results

As previously mentioned, the aim of this master's degree study is to challenge the contemporary understanding of the crypto markets by developing a proper algorithmic model for explaining the future development of crypto assets values.

The expected outcome is a reasonable and justifiable accuracy of the tested neural network models. In addition, the aim is to achieve a proper understanding of the influence of diverse crypto-specific variables in explaining and inducing value movements on the cryptocurrency markets.

Moreover, the sentiment feature is introduced to the network models as an additional model variable for making future price predictions. The expected result is the enhancement of the predictive power and model performance by correcting for public opinion, subjectivity and, sentiment towards the novelty of cryptocurrencies.

As the overall research result, the author's expectations are that the tested models will outperform the basic Buy-and-Hold trading strategy, while achieving superior investment returns for the analyzed crypto assets.

## 1.6 Cryptocurrencies and the Blockchain

This chapter is an introduction to the purpose and application of crypto assets and the Blockchain technology. The main discussion revolves around the development of the cryptocurrency market and the Blockchain, with a special focus on its scope, dynamics, market share, and government regulation.

Cryptocurrencies are digital assets secured by cryptography and stored electronically on the Blockchain network (Hubrich, 2017). Crypto securities apply encryption techniques for controlling the creation of monetary units and verifying network transactions. The cryptography is used for regulating creation, verification, and record of transactions on the blockchain, eliminating the need for third-party interference (Tapscott & Tapscott, 2016).

The global surge for cryptocurrencies begun in 2008 with the introduction of the Bitcoin project signed by Satoshi Nakamoto. Even though digital currencies existed prior to Bitcoin, this creation marks an important milestone in the digital asset history, due to Bitcoin's distributed and decentralized nature. Back in 2008, Bitcoin was introduced and presented as "a purely peer-to-peer version of electronic cash that allows direct online transactions, neglecting the need of financial intermediation" (Nakamoto, 2008). The Bitcoin network is based on Proof-of-Work (PoW) algorithm which allows users to generate digital coins by performing meticulous computations. The PoW algorithm secures the system against potential fraud and counterevidence (Hayes, 2016).

The Bitcoin code is an open-source, and it is publicly available on the GitHub. Hence, developers around the world have used this code for creating thousands of modified crypto assets, known as alternative cryptocurrencies, or the “altcoins”. In its essence, altcoins are referred to as coins alternative to the Bitcoin, built on Bitcoin’s open-sourced protocol. The objective of this coin alterations is to create new assets which can potentially solve the existing issues: the high computational costs, the low scope of transactions per unit of time or increase of the block size. On the contrary, some of the cryptocurrencies provide novel and innovative features that offer substantially different functionality. Cryptocurrencies and Blockchain innovations can be grouped into two categories: new (public) blockchain systems that feature their own blockchain (e.g., Ethereum, Peercoin, Zcash), and dApps or other that exist on additional layers built on top of existing blockchain (Buterin, 2014).

Initial Coin Offerings (hereinafter: ICOs) have become a worldwide accepted approach for financing projects based on blockchain and so called “tokens”. The ICOs present a form of raising funds for financing project in the earliest stage in other words before the project has been launched. The purpose of these ICOs is the potential network development in the future. Tokens are a representation of asset or utility created on the top of the existing Blockchain. Tokens can take any form of tradable commodities which are generated based on smart contracts. The main difference between altcoins and tokens is in their structure. Altcoins are separate currencies with own separate blockchain. Tokens, on the opposite, exist and operate on top of a blockchain that facilitates the creation of decentralized applications (Buterin, 2014).

In this master thesis, the focus will be on analyzing and explaining the price fluctuations of the Bitcoin and its alternatives: Litecoin and Ethereum. At the present, the selected “in focus” cryptocurrencies hold more than 62 % of cryptocurrency market capitalization in total ($159.4 out of $256.2 Billion US – as of the 1st of July 2018).

### 1.6.1 Background of the Blockchain and Cryptocurrencies

The Blockchain is a distributed, shared, encrypted-database that serves as an irreversible and incorruptible public ledger of information. It is a public register where transactions are stored in a secure, verifiable and permanent system (Houben & Snyers, 2018). The term “blockchain” refers to blocks of transactions consolidated in a chain of accepted history of transactional data. The data encryption cancels out the need for third-party intermediation. Confirmation of transactions is done by unrelated nodes in the network. Nodes present an electronic device connected to the blockchain network which plays a supporting role by maintaining a copy of a blockchain and process transactions.

Distributed ledger technology incorporates pervasive, persistent and permanent data structures, which are replicated across numerous nodes and provide unique data representation. The technology works with double key encryption or public-key cryptography and peer-to-peer shared data storage. Consequently, the data source is “central”, while the technically “distributed” nodes have open access to the ledger with all the data transactions from the origin (“data genesis”). The further use of this technology was enabled with the massive increase of the computational processing speed and the drop in the data storage costs (Houben & Snyers, 2018).

Transactions on the Blockchain are conducted, verified, cleared and stored in a block on average every ten minutes. Each new block is linked to the preceding block, forming the chain of transactions – the Blockchain. The system structure permanently timestamps and stores the exchanges of values, and at the same time prevents further ledger alterations (Tapscott & Tapscott, 2016). This new digital ledger can be programmed to record virtually everything presenting value and importance, such as money, stocks, bonds, deeds, contracts and virtually all other kinds of assets.

Cryptocurrencies are required for the genuine functioning of the blockchain technology. Hence, their economic potential arises from the utility of the blockchain application, and not from the currency itself. The cryptocurrency is secondary to the blockchain in a comparable manner to the common stocks on the equity market. Common stocks do not exist as a purpose of itself. Stocks are created to finance real-world business activities of some associated corporations. The plurality of common stocks issued by all corporations generates what we know as the equity market. The key difference between the equity and cryptocurrency market is that the net value of the latter is not assigned to a centralized corporation, but to the users and participants in the network itself. Their value is generated with the creation and exchange of the associated currency.

Decentralization is the key feature that differentiates Blockchain systems and cryptocurrencies from traditional, fiat currencies. Cryptocurrencies are not controlled by any central authority nor legal entity (i.e. governments, central banks, institutions). They are designed to work as a medium of exchange, using cryptography to secure transactions and control the creation of additional units of these currencies. The inherent cryptographic security removes the need for centralized authorities that act as custodian (Chuen, Guo & Wang, 2017).

These decentralized currencies are produced by the entire cryptocurrency system collectively, at a rate predefined with the creation of the system. Within cryptocurrency systems, the safety and balance of ledgers are maintained by a community of mutually distrustful parties, the miners. Miners use their computing power to validate and timestamp transactions by adding them to the ledger in according to a particular timestamping scheme. The security of the crypto network is based on cryptography and a clever incentive system. The incentive system is created to compensate miners for their activity and to differentiate cryptocurrencies. The cryptography secures the transactions in the blockchain, manages the supply and prevents fraudulent activities.

What should be inferred as the real breakthrough is not solely the advanced cryptography, but rather the distributed ledger or blockchain technology, defined as the real "game changer". It reduces transaction costs among all participants in the economy and supports models of peer-to-peer collaboration, which all in all, makes the existing organizational forms redundant.

## 1.7 Cryptocurrencies and the Traditional Framework

OECD defines currency purchasing power parity (hereinafter: PPP) as "the rate of currency conversion that equalizes the purchasing power of different currencies by eliminating the differences in price levels between countries". The exchange rates between two currencies are equal to the ratio of the respective currency purchasing power (OECD, 2012). A decrease in purchasing power reduces the currency value on the foreign exchange market.

Traditional currency purchasing power and its exchange rate are closely dependent of the national GDP growth, levels of inflation and public debts, employment rates, productivity growth and various other macroeconomic factors (Dornbusch, 1985). Higher GDP growth and lower level of unemployment straightens up the value of the national currency. Furthermore, national central banks have the chance to influence currency value throughout inflation rates. Hence, higher inflations rates decrease the value of the national currencies. International trade can also influence currency values. For instance, current-account deficit forces currency depreciation. In addition, political instability can further decrease the value of the national currency.

Cryptocurrencies present a subset of the virtual, digital coins and a departure from the conventional currencies. According to the European Central Bank (2012), virtual currencies present unregulated, digital money issued and controlled by its developers, used and recognized among the members of a specific virtual community. Crypto assets share some common characteristics to both, foreign exchange and the security market and at the same time, incorporate their unique features of a completely new asset class. Hence, trying to explain the value of cryptocurrency trough out the perspective of the traditional currencies and the PPP theory can lead us to a systematically inconsistent result. Due to the novelty of the crypto assets and their specific nature, the same macroeconomic reasoning cannot apply for an understanding of the movements on the cryptocurrency market.

The same holds for applying the traditional investment theory and the conventional financial models for estimating cryptocurrency value. Building a portfolio based on the exposure to macroeconomic or statistical factors which explain the return differences between financial assets is known as factor investing (Naik, Devarajan, Nowobilski, Page & Pedersen, 2016). The capital asset pricing model (hereinafter: CAPM) is the primal factor investing model. CAPM is based on the market risk factor – the higher expected return due to the higher risk exposure to the equity market. Withal, the CAPM model lacks in explaining the cross-sectional differences in the asset returns. Fama and French (1993) integrate the market, size and value factor into one single model. In the following years, other factors such as momentum and different fixed-income effects have been introduced to the factor pricing models.

The Fama and French five-factor asset pricing model (2014) is used for assessing the average stock returns. The model grounds on the market, size, value-growth and the profitability factor:

$$R_{it} - R_{Ft} = a_i + b_i Mkt_t + s_i SMB_t + h_i HML_t + r_i RMW_t + c_i CMA_t + e_{It}, \qquad (1)$$

where $R_{it}$ is the risk-free rate, $Mkt_t$ is the value-weight market portfolio returns, $SMB_t$ – small minus big stocks, high and low Book-to-Market stocks $HML_t$, robust and weak profitability $RMW_t$ and stocks of low/high investment firms – $CMA_t$. The $e_{it}$is the zero-mean residual.

Factors are also relevant in the valuation of alternative, non-equity asset classes. Alternative investments are often considered as effective diversifiers that exhibit low correlation with the traditional assets and have fairly different volatility and risk profiles. Cryptocurrencies can be seen as the alternative to the traditional investment portfolio components, such as equity, government, corporate bonds or gold.

The main difficulty in including alternatives to the traditional asset portfolio is how to model the exposure of alternatives to the same (or similar) risk factors to which traditional assets are exposed to. Consequently, the risk exposure needs to be defined based on the selected variables which, according to the analyst's intuition, present the fundamental characteristics of the crypto assets (Naik, Devarajan, Nowobilski, Page & Pedersen, 2016). This approach relies on basic valuation principles and knowledge of the underlying assets, comparable to the commonly applied approach used in fundamental asset valuation models (ex. DCF model). Hence, if we assume that investors value alternative assets as DCF streams, then, the volatility of the assets is mostly determined by the asset growth perspectives and its discount rates.

The discount rate of one alternative asset should differ from the traditional assets discount rates. Liquidity beta presents an important component of the risk profile of the majority of alternative asset classes. The liquidity factor is applied as a discount rate for estimating the risk exposure toward the traditional alternative investments such as private equities, venture capital, and real estates. The Pastor-Stabaugh liquidity factor captures the excess returns of assets with large exposure to changes in their aggregate liquidity (Pastor & Stambaugh, 2001). The liquidity on the cryptocurrency market is closely related to the level of transactional fees established on the crypto-exchanges and the number of active addresses of the network participants. In addition, the mining profitability predefined in the network protocol can serve as an additional incentive for mining activity, and hence, it can contribute towards higher supply and increased liquidity.

On the other hand, we can define cryptocurrency future growth perspectives to be closely dependent on the network's ability for computing increased number of transactions, block size and transaction value and further on, reduction of the difficulty rates and the confirmation times required for the block unit creation.

So, in order to draw a line between cryptocurrencies and the established financial framework of asset valuation, the author introduces cryptocurrency metrics which can be used in broadcasting the cryptocurrency fundamental value: number of transactions in cryptocurrency, block size, sent transactions from single address, difficulty rate, hash rate, mining profitability, value sent in U.S. dollars, transaction fees, block time, transaction value and the number of cryptocurrency active addresses.

Due to the highly volatile and speculative nature of crypto assets, their prices are likely dependent on the sentiment and the market momentum. The market momentum motivates future investments and it further on, increases market speculations. Consequently, the task of constructing a solid model that will detect key features driving the cryptocurrency prices will be incomplete without the proper quantification of the market momentum and therefore estimation of the market sentiment. Therefore, the SA tool is applied for the purpose of extracting the general prevailing sentiment of investors, so to anticipate future price developments (Agaian & Kolm, 2017). The collective optimism driven by the peer influences among individual traders can create an endogenous effect which is relevant for understanding the investment dynamic of the cryptocurrency market. The aim is to provide a proper understanding of how the public sentiment can contribute towards explaining and predicting cryptocurrency values.

## 1.8 Cryptocurrency Market Analysis

Looking from a historical perspective, the year 2017 can be marked as the year when cryptocurrencies were first introduced to the broad world of finance. Over a one-year period, the total market capitalization of cryptocurrencies has grown from \$17.7 to \$610.2 Billion U.S. (as of 31st December 2017). The exponential growth of the cryptocurrency market has attracted considerable attention. According to CoinMarketCap.com, as of 1st of July 2018, the total value of the cryptocurrency market is \$256.2 Billion U.S. This value presents a 57 % decline, compared to the 2017-year ending (\$572.3 Billion U.S.). This evident price volatility inherent for the cryptocurrency assets presents a real challenge for investors in their attempt for understanding the fundament behind cryptocurrency value.

For instance, we can compare cryptocurrencies with different marketable asset value versus the cryptocurrency value. As of 2017 year-ending, the global money supply has reached up to \$89.6 Trillion US, whereas the global stock market amounted \$66.8 Trillion U.S., and the Gold Market capitalization was \$1.5 Trillion U.S. On the opposite, the total value of the crypto asset market is much lower, reaching \$572.3 Billion U.S. by the end of the year 2017 (\$256.2 Billion U.S. as of 1st of July 2018). However, we can still note that, although the cryptocurrency market is currently in the early stage of its development, the trend indicates that there is a serious potential for positive future development.

*Table 1: Cryptocurrency Market – top 10 cryptocurrencies as of 1st of July 2018*

| # | Name | Market Cap. | Price | Circulating Supply | Volume (24h) |
|---|---|---|---|---|---|
| 1 | Bitcoin | $ 109.221.170.204 | $ 6.377,9 | 17.125.022 BTC | $ 4.329.018.274 |
| 2 | Ethereum | $ 45.584.874.181 | $ 454,0 | 100.412.052 ETH | $ 1.401.705.283 |
| 3 | XRP | $ 18.005.373.483 | $ 0,5 | 39.262.097.329 XRP | $ 304.091.528 |
| 4 | Bitcoin Cash | $ 12.708.687.096 | $ 738,3 | 17.212.897 BCH | $ 607.524.980 |
| 5 | EOS | $ 7.211.307.163 | $ 8,1 | 896.149.492 EOS | $ 720.214.320 |
| 6 | Litecoin | $ 4.594.033.393 | $ 80,3 | 57.212.111 LTC | $ 279.443.639 |
| 7 | Stellar | $ 3.645.118.635 | $ 0,2 | 18.760.438.885 XLM | $ 43.538.174 |
| 8 | Cardano | $ 3.545.505.830 | $ 0,1 | 25.927.070.538 ADA | $ 85.034.081 |
| 9 | IOTA | $ 2.833.538.257 | $ 1,0 | 2.779.530.283 MIOTA | $ 44.779.143 |
| 10 | Tether | $ 2.705.091.364 | $ 1,0 | 2.707.140.346 USDT | $ 2.639.539.372 |

*Source: CoinMarketCap (2018).*

## 1.9 Cryptocurrency Market Characteristics

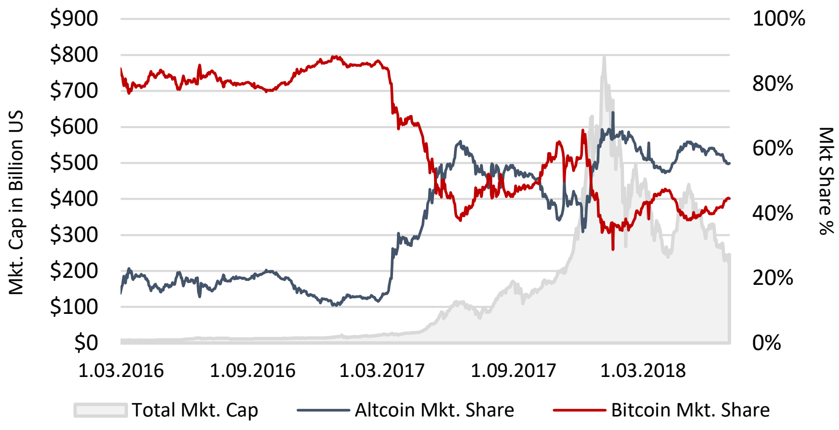
Despite the increasing relevance of the cryptocurrency market into the financial world, a comprehensive analysis of the whole system is still lacking, as most researchers focus exclusively on the behavior of selected cryptocurrencies. The following chapter presents an an overview of the main characteristics which define cryptocurrency market in total. This chapter includes overview of the market shares of diverse altcoins, compared against the Bitcoin's market share, review of the regulative changes, their impact and the overview of the market momentum effects.

### 1.9.1 Bitcoin Market Share Dominance

The analysis focuses on the market share of cryptocurrencies for the period between 19th of February 2016 and 1st of July 2018. The dataset includes 1.567 cryptocurrencies, of which around 500 were active during this period.

Figure 1 shows the evolution of the market capitalization over the last two years (starting from 19th of February 2016) for all altcoins and the Bitcoin. The grey area shows the development of the cryptocurrency market overall. For the period of 2016–2018, the total market capitalization of cryptocurrencies has increased exponentially (CoinMarketCap, 2018). Since its introduction in 2009, Bitcoin had the first-mover advantage of being the most well-known and widely adopted cryptocurrency. In the beginning of 2017, the dominance of Bitcoin started a downside trend. Bitcoin market share has been steadily decreasing over the past two years.

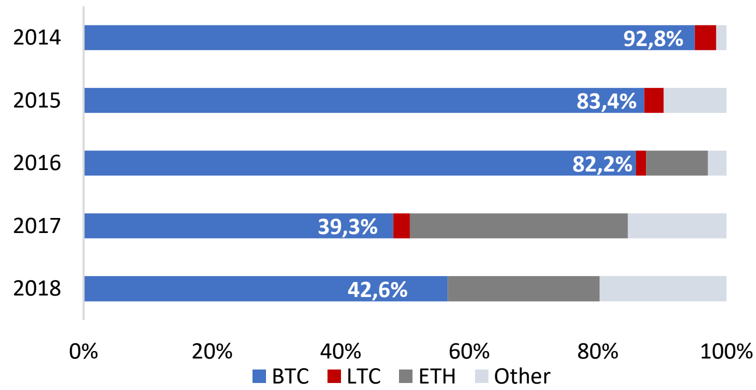
*Figure 1: Cryptocurrency Market Share Trend 2016–2018 – Bitcoin vs. Altcoins*




*Adapted from Bitinfocharts (2018).*

On the 15th of May 2017, for the first time in the cryptocurrency history, the Altcoin market share has overpassed Bitcoin by reaching 50,4 % or in value $28.3 Billion US. Consequently, it can be noticed that Altcoins have continued to gain significant market share, cumulatively accounting for 55,4 % of the market as of 1st of July 2018.

*Figure 2: Top 5 Cryptocurrency Market Share as of 1st of July 2018*




*Adapted from CoinMarketCap (2018).*

Although Bitcoin remains the dominant cryptocurrency in terms of market capitalization, other cryptocurrencies are increasingly overtaking the share of the cryptocurrency market. While Bitcoin's market capitalization accounted for 94,8 % of the total market in June 2013, it has dropped to 42,6 % as of June 2018. Litecoin (LTC) in 2018 lost its position among the top five cryptocurrencies. On the opposite, Ether (ETH), the native cryptocurrency of the Ethereum network, has established itself as the second-largest cryptocurrency, accounting for 17,8 % market share as of June 2018 (CoinMarketCap, 2018).

### 1.9.2 Cryptocurrency Market Regulation

Government regulation has a sore impact on the prices of cryptocurrencies. The cryptocurrency regulative landscape changes dynamically with time, and consequent price reactions can be tracked. The impact of regulation can affect both positive and negative price movements. Regulative, like taxes and asset laws, can change the perceived value of altcoins, hence, move cryptocurrency market prices.

In one-year virtual currency, capitalization has increased from $17.7 Billion U.S. to $573.0 Billion U.S. For the same period, the cryptocurrency frame has been widened with the invention and spread of ICOs. Both private and institutional investors became interested in the worth of these assets. Authorities, on the other hand, are concerned about the associated risk. Token valuations exhibit enormous volatility, while several ICOs resulted to be a scam and some crypto-exchanges were hacked (Howell, Niessner & Yermack, 2018).

The limited ability to control the cryptocurrency is due to the open network platforms which are often seen as an opportunity for criminal activities and illegal transactions. Hence, authorities around the world have raised their concerns regarding these developments. South Korea and China, both considered among the largest cryptocurrency markets in the world, have banned the ICOs (Howell, Niessner & Yermack, 2018). This action has initiated discussions for potential changes in the U.S. and EU regulations which will continuously impact crypto asset values. An overview is following of the regulative changes on the biggest crypto markets and how these changes have affected the cryptocurrency prices.

The changes which the United States implied towards the crypto asset regulations have been in focus since the creation of these assets and in fact, any change can have an immense knock-on effect on the global markets. The U.S. Securities and Exchange Commission (SEC) is targeting the change in the treatment of the majority of ICO tokens to securities. In addition, the U.S. Congress in their 2018 Annual Economic report (February 2018) accentuates the need for incorporating AML and CFT regulation to the cryptocurrency market.

The Chinese authorities have implemented a blanket ban on all cryptocurrency trading and domestic exchanges back in September 2017. An observation of the cryptocurrency market movements from that time shows short-term downside movements. Despite the fear that this regulation can have a detrimental effect on a global level, the market has bounced back very fast.

In April 2018, the government of Japan has announced its plans to legalize ICOs and utilize token sales for national gains. The Russian main legislative body, the State Duma in May 2018 has released a report which is addressing the regulation of digital financial assets. In this report, it is stated there is a need for implementation of AML and CFT regulations and for placing an unprecedented set of ICO regulations. The regulations have promoted further development of the Russian-based ICOs (Button, 2018).

*Figure 3: Effects on regulative changes on cryptocurrency market*


*Adapted from Bitinfocharts (2018).*

The Swiss regulator, FINMA in February 2018, shared its strategy to treat tokens based on their function. Different treatment will be placed for tokens used for payments, utility tokens, and tokens which serve as an investment. Hence, token classification will change during the time (Zetzsche, Buckley, Armer & Föhr, 2018). The tendency to regulate crypto assets comes from the money-laundering and consumer protection considerations. According to FINMA, crypto exchanges should undergo similar standardization to those of banks. This requires customer identification and record keeping of the unusual transactions. This already became a practice in Australia and South Korea, while the EU Commission in April 2018 passed a directive that is stipulating the same outcome.

At the moment of writing, regulators from different countries and comities have a diverse treatment towards the cryptocurrencies. Some treat them as asset commodities, while others as digital currencies. The lack of a unique and structured legal framework in many countries is still a hurdle, while the legal precedent for cryptocurrency is still being settled.

### 1.9.3 Cryptocurrency Market Momentum

In comparison to the traditional financial markets, cryptocurrency markets have been historically characterized as speculative investments, attracting momentum and speculative traders. Cryptocurrency markets are still undeveloped and highly influenced by limited groups of investors, companies, and institutions. Investors can speculate, and in that manner, influence the value of the less-known coins and inadvertently affect price movements. With a large amount of capital at their disposal, they can hold a large percentage of the coin supply, and later, attempt to promote positive information about the coin, in order to surge prices.

Momentum and speculative investors can commonly overwhelm value investors, hence create an environment where bubbles flourish. Furthermore, assets can be priced by applying their recent past values as the only indicator for the future. In this way, a significant feedback loop is generated, amplifying the value of ups and downs, as opposed to more balanced, stable financial markets.

The value of the cryptocurrency is highly dependent on the perception of the majority. The main factor affecting the value of cryptocurrency is what people are willing to invest in a unit of cryptocurrency: money and time (Naik, Devarajan, Nowobilski, Page & Pedersen, 2016). Any asset can have value to someone based on their beliefs. So, creating a positive perception of a certain cryptocurrency is crucial for maintaining the value of the cryptocurrency or commodity.

In addition, the media has a meaningful influence on crypto asset prices, both positively and negatively. The momentum is enhanced by news and media announcements which, later on, create a major vector for potential price manipulations. By focusing on certain positive or negative aspects of the cryptocurrency, the media can play a guiding role in the asset price fluctuations (Cheah & Fry, 2015). In a playing field of massive social media coverage combined with rapid technology changes, altcoin prices swing up and down in a matter of a day. Fast information diffusion causes rapid coin price changes. These changes lead to a volatile market, which has been a common characteristic for the overall cryptocurrency market.

*Table 2: Social Media Coverage – top 5 cryptocurrencies as of 1st of July 2018*

| # | Name | Average Users Online | Reddit Subscribers | Twitter Followers |
|---|---|---|---|---|
| **1** | **Bitcoin** | 10.973 | 910.495 | 62.319 |
| **2** | **Ethereum** | 5.627 | 375.362 | 416.681 |
| **3** | **XRP** | 4.319 | 190.156 | 884.465 |
| **4** | **Bitcoin Cash** | 3.641 | 35.832 | 106.127 |
| **5** | **EOS** | 4.300 | 56.041 | 184.994 |

*Source: CoinGecko (2018).*

# 2 CRYPTOCURRENCIES IN RESEARCH FOCUS

The purpose of this research is to define a proper, cohesive approach for estimating and predicting the range of cryptocurrency value. The initial objective is to define a trading strategy applicable for diverse crypto assets. Since 2009, various cryptocurrencies have been developed, with 1.567 in existence, as of 1st of July 2018. In this master's study research, the scope of the analysis will be focused on three major cryptocurrencies networks with the most extensive data history: Bitcoin, Litecoin and Ether. These cryptocurrencies were selected based on their relevance in terms of diffusion and market capitalization (presenting 62 % of total cryptocurrency market capitalization, as of 1st of July 2018) and historic data availability. The following chapter presents an introduction towards these three cryptocurrencies with a special focus on their fundamental specifics.

*Table 3: BTC, LTC & ETH Data overview as of 1st of July 2018*

| | **Bitcoin (BTC)** | **Litecoin (LTC)** | **Ether (ETH)** |
|---|---|---|---|
| **Market Price** | $6.378 US | $80 US | $454 US |
| **Market Capitalization** | $109,2 Billion US | $4.6 Billion US | $45,6 Billion US |
| **Avg. Transactions Per Hour** | 6.991 | 1.442 | 27.430 |
| **Avg. Sent Per Hour** | 23,482 BTC *($176.7 million US)* | 104.453 LTC *($8.7 million US)* | 41.218 ETH *($19.1 million US)* |
| **Avg. Transaction Volume** | 3.27 BTC *($24.599 US)* | 127,03 LTC *($10.606 USD)* | 1.50 ETH *($697 USD)* |
| **Block Time** | 8m 31s | 2m 51s | 14,5s |
| **Avg. Blocks Per Hour** | 7 | 21 | 248 |
| **Reward per Block** | 2.100 +14,93 BTC | 25+0.06436 LTC | 3+0.09+0.027+0.66 ETH |
| **Block count** | 533,149 | 1.461.076 | 6.011.806 |
| **Hash rate** | 37,485 Ehash/s | 264,402 Thash/s | 273,061 Thash/s |
| **Tweets Per Day** | 33.267 | 2.287 | 12.959 |

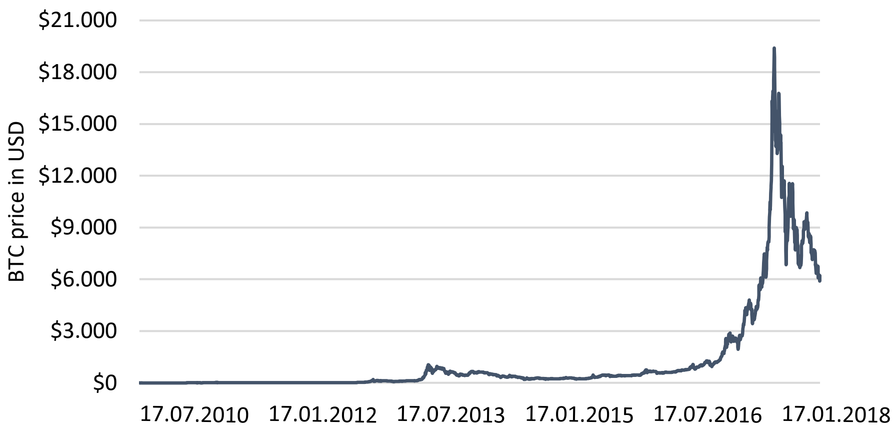
*Source: Bitinfocharts (2018).*

## 2.1 Bitcoin

Bitcoin is the first decentralized cryptocurrency built upon the use of Blockchain technology. The virtual currency was conceptualized in the 2008 whitepaper, written by a pseudonymous author sign by the name of Satoshi Nakamoto. The initial idea behind Bitcoin was to create "a new electronic cash system completely decentralized and with no server or central authority" (Nakamoto, 2008). Bitcoin's reputation as the oldest digital currency has been the subject of mainstream media coverage due to its innovative technical concept and rapid value fluctuations. It is conceived as the "gold standard" of cryptocurrencies since all alternative cryptocurrency market prices are matched to the price of the Bitcoin (Tapscott & Tapscott, 2016).

Bitcoin transactions are encrypted, verified by network nodes and recorded on the publicly available distributed ledger. Bitcoins are created with the process of mining and the use of SHA-256 algorithm as cryptographic proof of transactions. Transactions are recorded in "blocks" per unit of time, collectively forming the Blockchain. The term "mining" is refereeing to the process of performing advanced mathematical computation and record-keeping of transactions, performed by the nodes in the network. The mining process requires special software packages and energy-intensive hardware. In this way, miners manage to convert blocks into a sequence of code, known as the hashes. New hashes are placed at the end of the blockchain and the miners are awarded with Bitcoins. The number of awarded Bitcoins decreases over time, as pre-defined in the Bitcoin protocol (Nakano & Takahashi, 2018). A single Bitcoin can be spent in fractional increments, as small as 0.00000001 BTC per transaction. Bitcoin total supply is limited up to 21 million Bitcoins. This feature makes Bitcoin comparable to gold and other precious metals.

*Figure 4: Bitcoin (BTC) price in U.S. dollars*




*Adapted from Bitinfocharts (2018).*

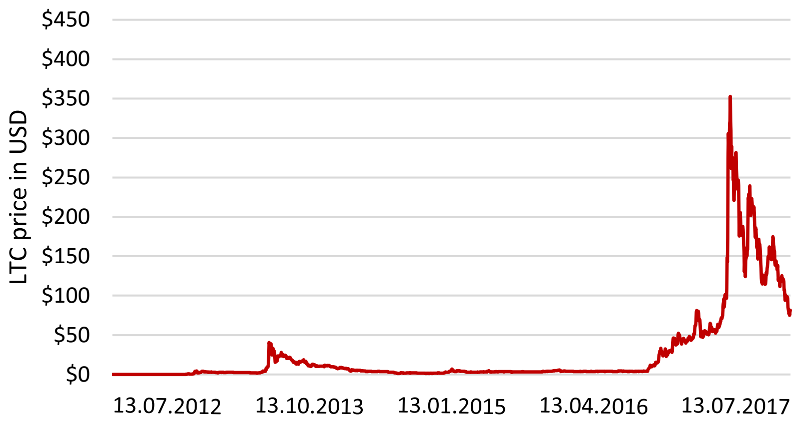
Over the course of its first eight years, the market price of a single Bitcoin has fluctuated from below $0,01 U.S. to above $19.000 U.S. The highly volatile price has attracted traders seeking for profits from market speculation, while at the same time, it has made long term investors and daily users hesitant to consider Bitcoin as an alternative investment. As of 1$^{st}$ July 2018, 17.1 million BTC were in circulation with a total market value of $109.2 Billion. The price per unit of Bitcoin is $6.378 U.S.

## 2.2 Litecoin

Litecoin can be considered as the "silver standard" of cryptocurrencies, as it has been the second most adopted cryptocurrency by both miners and crypto exchanges. It was created in 2011 by Charles Lee and the support of the Bitcoin community. Litecoin grounds on the same peer-to-peer protocol as Bitcoin, and it was treated as its leading competitor, up until 2016.

Litecoin uses Scrypt algorithm, as opposite of the SHA-256 algorithm. Scrypt enables standard computational hardware to be used in verifying transactions. It is a sequential memory-hard function which requires asymptotically more memory, compared to other algorithms. The Litecoin network utilizes FPGA and ASIC accelerated hardware technologies use for mining purposes. In addition, it has a faster time for transaction confirmation, which makes the coin especially attractive in time-critical situations (Litecoin Project, 2017). The time needed for the Litecoin network to process a block is 2.5 minutes, unlike the Bitcoin network which requires 10 minutes.

*Figure 5: Litecoin (LTC) price in U.S. dollars*



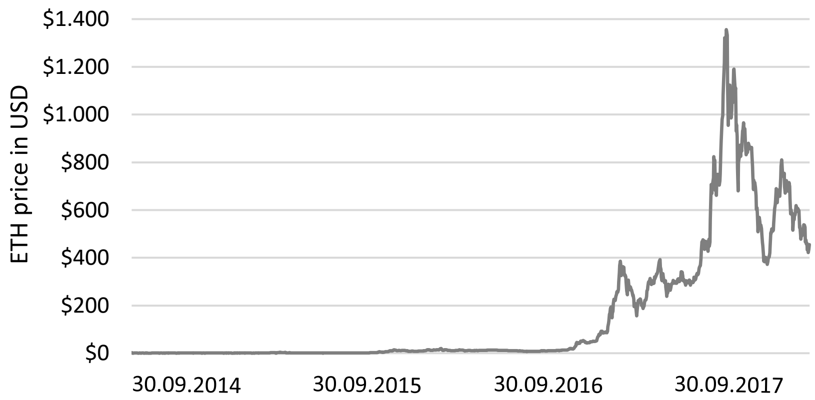

*Adapted from Bitinfocharts (2018).*

The total amount of Litecoin available for mining and circulation is four times the amount of Bitcoin. Litecoin protocol limits the coin total supply up to 84 million LTCs. As of 1st July 2018, 57.5 million LTC were in circulation with a total market value of $4.6 Billion U.S. and price of per LTC of $80.3 U.S.

## 2.3 Ether (ETH)

Ethereum presents a distributed open-source, public platform which enables the functioning of smart contracts. It was created in 2013 by Vitalik Buterin. Ethereum is a flexible, open software platform based on the Blockchain technology that enables developers to build and deploy decentralized applications (hereinafter: dApps).

*Figure 6: Ether (ETH) price in U.S. dollars*




*Adapted from Bitinfocharts (2018).*

The Ethereum Virtual Machine (EVM) is a software that runs on the Ethereum network. It enables anyone to run any program, regardless of the programming language. The EVM makes the process of creating block application much easier and efficient, by enabling development of unlimited applications on the same platform (Buterin, 2014).

The main leverage of Ethereum is the application of so-called “smart contracts”. Smart contracts present computer codes that facilitate the exchange of anything that involves a certain value. When running on the Blockchain, smart contracts become self-operating computer programs that automatically execute when contract terms are fulfilled. This feature serves as a guarantee against potential fraud or third-party interference.

In general, the value is driven by the increased utility and the possibility of third-party elimination by introducing contractual obligations. Ethereum leverages on its Turning-Complete language for solving computational problems. This allows the Ethereum network to facilitate the exchange of data and information. Ether is the cryptocurrency generated by the Ethereum platform. The idea behind Ether is not to be utilized for the purpose of purchasing and storing value, but rather to power the applications or computational services on the Ethereum blockchain.  Ether is also used for paying of transactional fees and services on the Ethereum network.

As of 1$^{st}$ July 2018, 100.4 million Ethers are in circulation, with the total market capitalization of $45.6 Billion U.S., with the market price of Ether of $454.0 U.S. Ethereum and Bitcoin differ substantially in their purpose and capabilities. Bitcoin is a peer-to-peer electronic cash system that enables online payments in BTC. On the opposite, the Ethereum blockchain focuses on running the programming code of any decentralized application.

# 3 CRYPTO ASSETS - PRICE ANALYSIS

Traditional financial models used to guide investment decisions are inadequate for constructing properly based investment strategy. Cryptocurrencies differ from the traditional financial assets in a way that they have no cash flows or historical dividend payouts, and their growth perspectives are rather uncertain and prone to subjectivity.

Investment strategies based on asset fundamentals have longer predictive horizons on factors influencing the price. Their infrequent nature challenges their application on a short run. Hence, professional traders tend to look beyond fundamentals when making trading decisions. The absence of the traditional fundament study leads us towards the analysis of the price data and asset statistics.

The following chapter presents the quantitative analysis of the three focus cryptocurrencies. The analysis is based on the Pearson correlation, Sharpe ratio, and the Network-Value-to-Transactions ratio. These three indicators proved to be widely adopted techniques utilized by the cryptocurrency daily traders. Crypto traders apply them in combination with chart trend analysis of historic coin prices when making their investment decisions. Hence, the following section will analyze the informative nature of these quantitative crypto-specific metrics.

### 3.1 Cryptocurrency Price Correlation

Correlation as a statistical measure can be applied for analyzing how asset prices move in relation to each other. Correlation statistics provide information of whether two currency pairs move in same, opposite or in an unrelated, random direction.

The correlation level is presented with the correlation coefficients, which have a range of +1/-1. Perfect positive correlation of +1 implies that two currency pair prices move in the same direction, while the perfect negative correlation of -1 means that price pair move in completely opposite direction. The correlation coefficient equals zero when prices are completely independent and random to each other.

*Table 4: Cryptocurrency Correlation Matrix*

| | | BTC | ETH | LTC | ^SPX | ^VIX | ^GLD | ^TNX |
|---|---|---|---|---|---|---|---|---|
| Bitcoin | *BTC* | 1 | 0.64 | 0.63 | 0.03 | -0.11 | 0.02 | -0.01 |
| Ether | *ETH* | 0.64 | 1 | 0.67 | 0.07 | -0.14 | 0.11 | 0.01 |
| Litecoin | *LTC* | 0.63 | 0.67 | 1 | 0.1 | -0.14 | 0.02 | 0.06 |
| S&P 500 | *^SPX* | 0.03 | 0.07 | 0.1 | 1 | -0.8 | -0.07 | 0.22 |
| CBOE Index | *^VIX* | -0.11 | -0.14 | -0.14 | -0.8 | 1 | 0.08 | -0,24 |
| Gold Exchange | *^GLD* | 0.02 | 0.11 | 0.02 | -0.07 | 0.08 | 1 | -0.43 |
| 10yTreasury Note | *^TNX* | -0.01 | 0.01 | 0.06 | 0.22 | -0.24 | -0.43 | 1 |

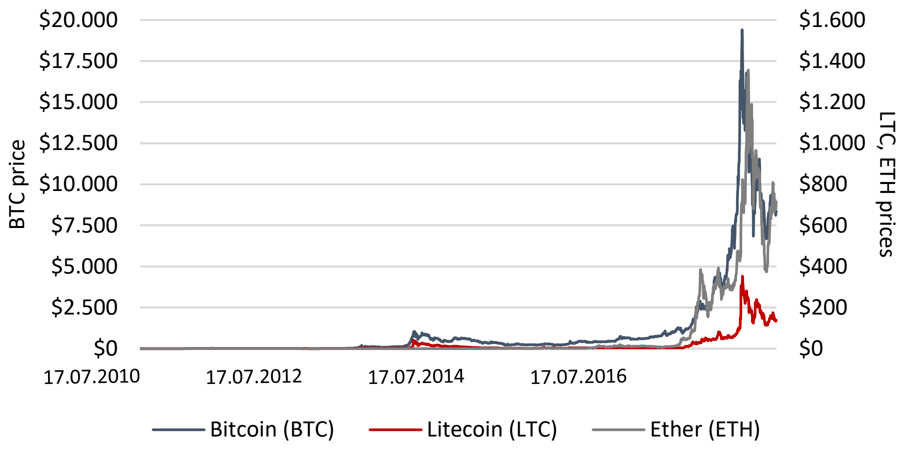
*Adapted from Sifirdata (2018).*

Table 4 shows the Pearson correlation indexes of the three focus cryptocurrencies among themselves and with other traditional financial instruments: Standard & Poor's 500 Index (^SPX), CBOE Volatility Index (^VIX), SPDR Gold Trust Shares (^GLD), and CBOE 10-year treasury-note yield (^TNX). Pearson correlation results are computed as log-returns of volume-weighted average daily prices over 365 days.

It can be observed that the three cryptocurrency prices exhibit a positive correlation among each other (>0,63). Hence, it can be expected the trend of collective price spikes and crashes (Figure 7). One reason for this may be the cryptocurrency network similarity. Overall, crypto asset values are influenced by the same market forces, such as the regulative changes, momentum investors, media hype and technology advancements. That inclines significant levels of positive price intercorrelation.

The positive correlation among crypto assets can be used for diversification purposes. For instance, Bitcoin and Ether have price correlation of 0.64 – statistically significant ($p<0.05$) positive correlation. Hence, by investing partially in both currencies, the investor can diversify its risk while still maintaining core directional view, trusting the future growth of the cryptocurrency market. In addition, the imperfect correlation between the two assets allows further diversification and marginally lower risk exposure.

*Figure 7: Bitcoin (BTC), Litecoin (LTC) and Ether (ETH) prices*




*Adapted from Bitinfocharts (2018).*

Since one conventional investment portfolio accounts for diverse assets: equities, hedging instruments, commodities and bonds/treasuries following is an analysis of how these standard investment portfolio components are correlated to the crypto assets. Hence, we examine how crypto asset prices are related to S&P 500 index (^SPX), CBOE Volatility Index (^VIX), SPDR Gold Shares (^GLD) and CBOE 10-year treasury-note yield (^TNX).

Standard & Poor's 500 (^SPX) is a U.S. stock market index based on market capitalizations of the 500 largest companies having common stock listed on the NYSE or NASDAQ (Chuen, Guo & Wang, 2017). It is designed to present a leading indicator of U.S. equities and is meant to reflect the risk/return characteristics of the largest companies. We note that the relationship between BTC, LTC and ETH and S&P 500 is weak-positive. Corporate stock prices depend on reaching financial targets (i.e. corporate margin targets), company fundamentals and wide macroeconomic factors. On the opposite are cryptocurrencies, where the majority of prices are inflated by hope and aspirations raised upon the Blockchain technology (Filippi, 2014). Stock markets, however, can be quite informative when it comes to the overall economic conditions and performance of specific industries regarding the Blockchain technology adoption.

VIX is the ticker symbol for the Chicago Board Options Exchange (hereinafter: CBOE) Volatility Index. It shows the market expectations of 30-day volatility. It is computed from the implied volatilities of a wide range of S&P 500 index options, both calls, and puts. VIX indicates the level of risk that is currently present in the markets for a given time frame. Correlation between the three focus currencies and VIX is weak-negative. Hence, there is an insignificantly low inverse dependency between these crypto assets and VIX Volatility index (Chuen, Guo & Wang, 2017).

SPDR Gold Trust Shares (^GLD) is part of the SPDR unity of exchange-traded funds (ETFs) managed by State Street Global Advisors (SSGA). The Trust holds gold bars and issues Baskets in exchange for deposits of gold. In addition, the fund distributes gold in connection with the redemptions of Baskets. The shared objective is to reflect the performance of the price of gold bullion, less than Trust expenses. The correlation of the focused currencies with GLD shares is weak-positive, indicating low interdependency.

CBOE 10-year treasury-note yield (^TNX) is based on 10 times the yield-to-maturity on the most recently auctioned 10-year Treasury note. The notes are usually auctioned quarterly. Bitcoin has weak negative, while Ether and Litecoin show a weak positive correlation to the CBOE 10-year treasury note. Thus, we can once again conclude insignificant interdependency between our focus coin prices and government security yields.

In conclusion, cryptocurrencies proved to have an insignificant (mainly negative) correlation with traditional asset classes: stocks, gold, and treasury notes. Further research can be placed on how much of the assumed correlation among cryptocurrencies is caused by sentiment rather than empirical dependency. Cryptocurrencies have crossed the threshold into the mainstream market. Traditional investors are starting to transfer risk from relatively stable equity markets into cryptocurrencies and vice versa. Bitcoin is considered to be a hedge against all cryptocurrencies, resembling the role of the gold in the traditional finance asymmetry.

With the potential increase in demand, recognition, and acceptance of crypto assets, there is an opportunity to take a similar position to gold and the stock market. The investor's sentiment can carry over from the stock market to the cryptocurrency market. Due to the mainstream adoption and acceptance of cryptocurrencies as investable assets, both markets start to crossover. However, none of the presented relationships can serve as a help in making educated future market predictions. The examined correlation only proves that investment fear and sentiment are not just isolated to certain cryptocurrency but are rather interrelated.

## 3.2 Cryptocurrency Sharpe Ratio ($S_h$-ratio)

According to Markowitz's modern portfolio theory (1952), a rational investor diversifies its portfolio in the way that there is no other portfolio that can offer him a superior risk-return analogy. Since Markowitz introduced his modern portfolio theory back in the year 1952, several metrics have been developed and utilized to determine an asset's required returns, given the level of risk, and vice versa. The Sharpe ratio is one of these metrics.

The Sharpe ratio (hereinafter: $S_h$ ratio) presents a score of the asset return on investment over asset volatility (Chiou & Lee, 2010). In fact, the ratio is measuring the investor's risk-adjusted expected return, given the risk level associated with the investment asset. When the cryptocurrency investor wants to invest in a portfolio of such risky assets as an alternative for a riskless asset, then the $S_h$ ratio gives an indication of the likelihood that the portfolio will out-perform the riskless asset. Hence, he/she will prefer a higher $S_h$ ratio as it represents the highest probability that his/her portfolio will be outperforming the risk-free assets. The $S_h$ ratio measures the excess return of an asset ($\mu$) per unit of risk ($\sigma$), and it can be defined as:

$$S_h = \frac{\mu}{\sigma} \tag{2}$$

The excess return ($\mu$) is calculated as the return estimated as surplus, above the given benchmark:

$$\mu = \int_{-\infty}^{+\infty} (r_c - r_b) p(r_c) dr_c \tag{3}$$

and risk ($\sigma$) is defined as the standard deviation of returns:

$$A\sigma = [\int_{-\infty}^{+\infty} (r_c - r_b - \mu)^2 p(r_c) dr_c)^{\frac{1}{2}})] \tag{4}$$

Cryptocurrency returns are denoted with $r_c$ and benchmark asset returns are $r_b$. The P ($r_c$) is the probability density function of the cryptocurrency returns. Higher $S_h$ ratio indicates a more proportional risk-to-reward investment profile.

The $S_h$ ratio can be applied for deciding on preferred crypto assets based on their level of volatility and the expected level of returns. Table 5 shows the 90, 180 and 365-Day $S_h$ ratio of the three focus cryptocurrencies.

*Table 5: Cryptocurrency Sharpe ratio as of 1st of July 2018*

| | 90-Day $S_h$ | 180-Day $S_h$ | 365-Day $S_h$ |
|---|---|---|---|
| **BTC** | -0,23 | 0,43 | 1,87 |
| **LTC** | -0,06 | 1,39 | 1,56 |
| **ETH** | -1,89 | 1,35 | 1,69 |

*Adapted from Sifirdata (2018).*

Due to the last year, above average returns (see, Figure 7), we can notice higher, above 1, 360-Day $S_h$ ratios for all three focus currencies. Looking at the 365-day $S_h$ratio, we can observe high $S_h$ ratio for Bitcoin, Litecoin and Ether, which indicates positive long-term investment perspectives for all mentioned crypto assets.

The 180-day $S_h$ ratio results reveal similar positions for both Ether and Litecoin, whereas the Bitcoin ratio is rather low. The lower $S_h$ ratio of Bitcoin indicates that low probability that the Bitcoin returns will outperform the risk-free investments on a half-year time perspective.

On the other hand, the 90-day $S_h$ ratio is negative for all three crypto assets. This indicates that in the last three-month period there was a higher probability that the risk-free assets will outperform BTC, LTC and ETH return. Interesting is the case of Ether, which spoors the greatest discrepancy in the $S_h$ ratio periodic results.

Hence, we can observe a high level of ratio value inconsistency between different time periods. The ratio inconsistency goes in line with the observed speculative and volatile nature of cryptocurrencies. That, further on, challenges the applicability of Markowitz's modern portfolio theory and the $S_h$ ratio as an indicator. When deciding whether to invest in certain cryptocurrency, the investor can use $S_h$ ratio so to gain a certain risk-return overview for the particular crypto asset. However, the $S_h$ ratio has limited use as single asset allocation tool.

### 3.3 Network Value to Transaction (NVT) Ratio

Besides financial assets fundamentals, traditional financial analysts commonly apply relative valuation methodologies for defining stock values, relative to their peers. The relative valuation approach can be based on diverse financial ratios. Some of the most commonly applied ratios are Price-to-Earnings (P/E) ratio Price-to-Book (P/BV) ratio Enterprise-Value-to-EBITDA (EV/EBITDA) ratio etc. However, crypto assets do not have direct revenues, operating earnings, book or enterprise value. Following, the direct application of relative valuation for the matter of crypto assets is not possible.

C. Burniske and W. Woo (2017) have introduced the 4.3 Network Value to Transaction (hereinafter: NVT) ratio. The NVT ratio measures the cryptocurrency dollar value of transactions, relative to its overall network value. The NVT ratio is usually referred to as the "Crypto P/E ratio". As discussed, crypto assets, unlike securities, do not generate any earnings. Therefore, traditional P/E value analysis cannot be applied for comparing crypto assets relative to each other.

Nevertheless, Burniske and Woo (2017) decided to use the daily value of transactions flowing through the network, as a proxy of crypto asset utility. They use the daily transaction volume as a substitute to corporate earnings. They explain this approach by comparing cryptocurrencies with corporate stocks. Like the equity investors who look for the strong earning stocks, crypto investors are interested in cryptocurrencies which exhibit a satisfying amount of transaction activity on their network. Hence, they conclude that the fundamental value of cryptocurrencies depends on their ability for transacting value. NVT ratio for cryptocurrencies is calculated by relating their market capitalization, or the total network value, with the daily transaction volume:

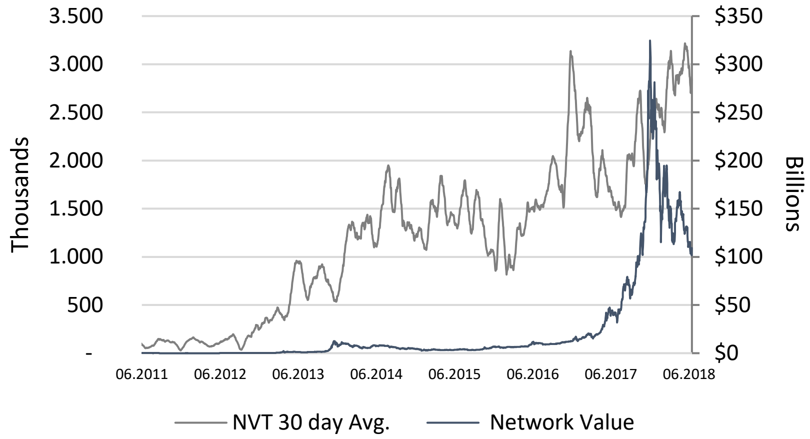
$$NVT\ ratio = \frac{\text{Network Value}}{\text{Daily Transaction Volume}} \tag{5}$$

where:

$$Network\ Value = Cryptocurrency\ Mkt.Price\ \times Total\ Supply \tag{6}$$

Relative to its historic range, a low NVT ratio implies that the cryptocurrency is currently undervalued for every unit of the asset's transaction volume. This can result from an expected decline in transaction volume (currently above historic average amount of transactions) or when the market value is below its perceived fundamental value. On the opposite, a historically higher NVT signals potential rise of transaction volume or that the asset is overvalued.

*Figure 8: Bitcoin (BTC) NVT ratio and Network Value*




*Adapted from Bitinfocharts (2018).*

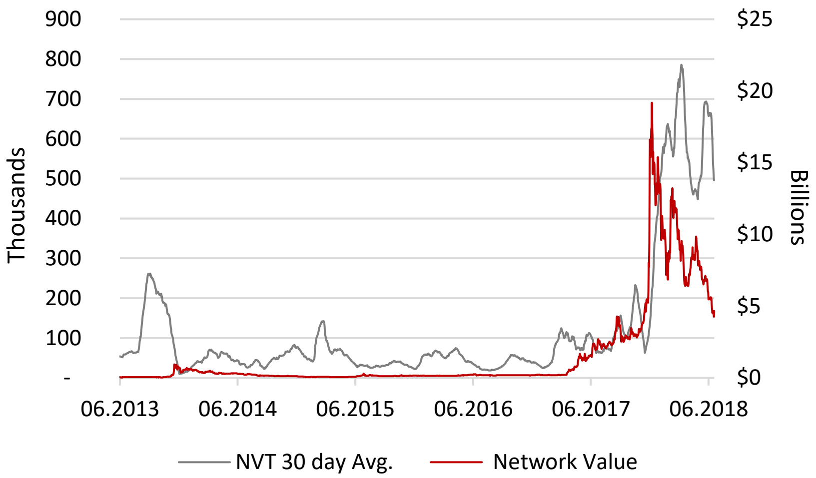
Looking from a historical perspective, we can observe continuous variations among the upward trend line in the case of Bitcoin. During the end of 2017, the NVT ratio falls below the level of 1.500 Th. and then continues to increase, up until the present. The decline in NVT ratio followed by the increase of Bitcoin value. Hence it can be concluded that the low NVT ratio indicated that Bitcoin price was underestimated by the market.

In 2013, Litecoin NVT ratio notes high values and then follows a decline with the start of 2014. The stable low NVT values continue with small fluctuations, up until the end of 2017, when the Litecoin NVT ratio reaches all-time record highs. Further on, after the first quarter of 2018, the LTC NVT ratio starts to decline, falling below 200 Th. and this trend continues in the second quarter of 2018.

In the case of Ether, since its introduction, the ETH NVT ratio has a continuous stable trend line. However, the second half of 2017 brought positive growth in NVT ratio for all three focus crypto assets, including Ether. At the end of 2017 Ether's NVT ratio reaches up above 44.000Th, following a sharp decline in the first quarter of 2018. The start of the second quarter marks an increase and again a decline in the ETH NVT ratio.

However, the applicability of the NVT ratio is rather questionable. The own work results show that transaction volumes have a tendency of following price changes, "post festum".

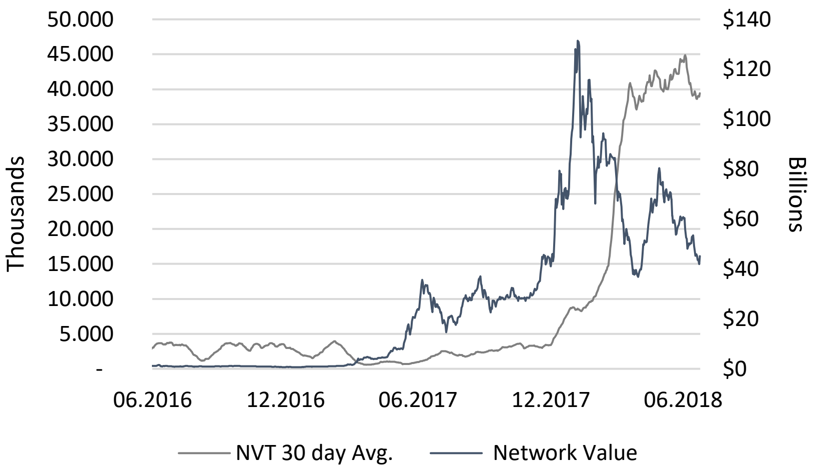
*Figure 9: Litecoin (LTC) NVT ratio and Network Value*




*Adapted from Bitinfocharts (2018).*

In addition, there is rather an unclearness regarding the correct estimation of cryptocurrency transaction volume (as the denominator in NVT ratio). What transaction value should we account for? For instance, Bitcoin uses an unspent transaction output model (UTXO), recording "spent" and "unspent" outputs, rather than user balances. Due to this fact, we face the double-counting of change outputs.

*Figure 10: Ether (ETH) NVT ratio and Network Value*




*Adapted from Bitinfocharts (2018).*

With off-the network value metrics, such as NVT ratio, we can capture only the on-chain transactions. To perceive the actual transactional potential of an asset, an off-chain scaling method needs to be developed. Hence, the NVT ratio is commonly prone to subjectivity, endogeneity and reflexive interdependency. This decreases the predictive power of the NVT ratio and its applicability as the key portfolio selection criteria.

# 4 NEURAL NETWORK METHOD

Financial forecasting is a single processing problem challenged by the low data availability, combined with the high noise, non-stationarity, and non-linearity of the data. This research study focuses on the prediction of cryptocurrency future value range, with the help of neural network algorithms. Neural networks are applied in forecasting financial outcome as a single processing application (Colianni, Rosales & Signorotti, 2015).

The following section presents an introduction towards neural networks, the Feedforward and the Recurrent neural networks. A detailed, comprehensive overview of neural networks is beyond the scope of this research study. Following is an overview of the neural network model applicability and the set-up introduction towards the predictive cryptocurrency model.

## 4.1 Introduction towards the Neural Networks

Neural networks present an artificial intelligence method applied for modelling complex functions (Medsker & Jain, 1999). Neural networks were firstly introduced in 1943 by Warren McCulloch and Walter Pitts. In their research study, McCulloch and Pitts (1990) develop a computational model for neural networks based on a complex mathematic algorithm, known as the threshold logic. Consequently, neural networks techniques have developed very fast and were recognized to have credible predictive proficiencies. Their structure complexity enables approximation of non-linear functions with a substantial level of accuracy.

Neural network models are classified as generalized, non-linear and non-parametric algorithms (Medsker & Jain, 1999). The inspiration for these models originates from the human brain anatomy and aims to replicate the human brain learning processes. Through the network learning process, neural networks generate proficient memory acquisition and auto-adaptability. The neural network models consist of many interconnected information processing units, known as the network nodes or neurons. The neurons perform summing functions which generate information stored as weights in the network connections (Hecht-Nielsen, 1992).

Two main categories of neural networks can be defined based on the temporal relationship among the network inputs and maintained internal state or the "network memory". Primarily, in the Feedforward neural network, inputs are fed to the network and transformed into an output. Network predictions do not depend on previous sequences. Feedforward networks do not hold inside to any of the network's intertemporal relationships or the so-called "network memory" (Medsker & Jain, 1999).

On the other hand, the Recurrent Neural Network models (RNNs) generate outcomes which are dependent on the previous network decisions or the network memory. What is specific to the RNNs is their possibility to selectively "memorize" past patterns. Hence, the current output of RNN depends on both, the previous output and the current input feedback connections and represent computational structures in a parsimonious manner (Cybenko, 1989).

## 4.2 Feedforward Neural Networks

The FF is a deep learning model with no feedback connections among the outputs of the network. Information flows only forward, in one direction – from the input nodes to the output layers. FF consists of a set of neurons, as being multiple information processing units and a set of interconnections among the neurons (Cybenko, 1989). Each neuron represents an activation function α, calculated as a weighted sum of all incoming signals to the neuron:

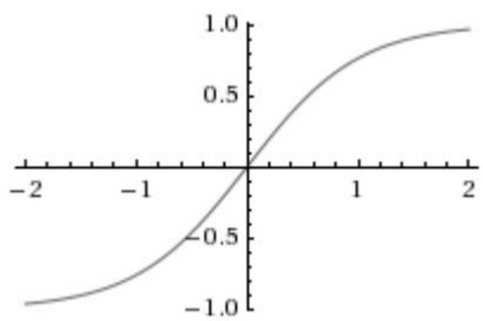
$$\alpha = \sum \omega_i \, x_i \tag{7}$$

The value of $f(\alpha)$ can be represented as an outgoing signal of the neuron, with $x_i$ as signal value and $\omega_i$ as the weight of incoming connection i. By increasing the weight of the connection, we enhance the influence per neuron. Most commonly applied activation function is the sigmoid function f(a):

$$f(\alpha) = \tanh(k \times a) \tag{8}$$

where k is a scaling factor which determines the steepness of the curve. The function value is bound to the range of ±1. Sigmoid function (Figure 11) is used for capturing non-linearity in the time data series.

*Figure 11: Symmetric Sigmoid Activation Function (k=1)*

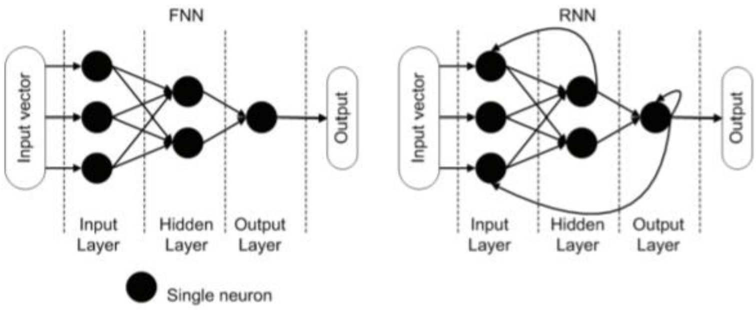



*Source: Cybenko (1989).*

The exact functionality of neuron connections is explained with the Hebb rule, established in 1949 by the Canadian psychologist Donald Hebb. According to the Hebbian principle, when a starting, source neuron is involved in the activation of a receiving neuron, the second becomes more sensitive to the signals, compared to the source neuron. On the opposite, when two neurons are connected, and the source is rarely involved in activation, the connection weakens over time, as the destination neuron “ignores” the source (Hebb, 1949). In correspondence with the Hebb rule, we can simulate network adaptations by adjusting the connection weight values. The adaptation method is called backpropagation.

A group of interconnected neurons forms the neural network layers. In general, FF NNs can be classified as single and multi-layer networks. Single-layer networks consist of a single layer of output nodes. The inputs are directly translated to the network outputs. On the other hand, a multi-layer deep learning network consists of three layers: input, hidden and an output layer. The input layer is a set of parameters without any incoming connection. These layers are known as the "sensors" to the NN. The value of the input layer is solely external. Hence, neurons from this layer send their values to the neurons of the next layer in the hierarchy. In between the input and output layer present the "hidden" segments. The hidden layers are invisible to external processes. The neurons in these layers have both incoming connections from preceding and outgoing connections for the succeeding layers. The hidden layers can be presented as the "cognitive brain" of the network. The output layer is the end result of the NN computations. Neurons in the output layer do not hold any outgoing connections since their f-values are translated directly as the "problem-solution" (Krenker, Bešter & Kos, 2011).

*Figure 12: Feedforward (FF) and Recurrent (RNN) Neural Network Topology*




*Source: Krenker, Bešter & Kos (2011).*

The resulting values in the output layer depend on the training process applied to the network. Adaptation of the network can be simulated by varying the weights of the included connections with the process of backpropagation of the error. The network is run based on the example of problem information from the training data set. Then, the calculation of the error is performed between the resulting output of the network and the actual correct solution. In addition, iterative backward alterations are performed for slight weight adjustments of existing connections, so to minimize the error of the output. This process is repeated over a set of training data. Weight alterations are performed with the Delta rule – based on the gradient descent principle (Krenker, Bešter & Kos, 2011). A detailed explanation of the Delta rule background mathematical calculations is beyond the scope of this master study.

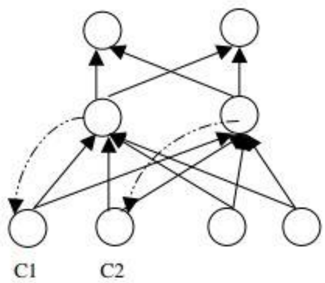
## 4.3 Recurrent Neural Networks

RNNs present complex algorithms designed for learning and predicting sequential or time-varying patterns. Traditional neural networks have a drawback in that they assume input independence. However, for many areas, such as price predictions, inputs are interdependent among each other. This anomaly is resolved with RNNs which share information between steps i.e. the output of a step is added to the input of the next one. This gives RNN a "memory feature".

RNNs have been an important focus of research during the 1990s. They can be applied to a wide variety of problems. For instance, RNNs find use in: predictive head tracking for virtual reality systems (Saad, Caudell & Wunsch, 1999), financial prediction using RNN (Giles, Lawrence & Tsoi, 1997), wind turbine power estimation (Li, Wunsch, O'Hair & Giesselmann, 1999) etc.

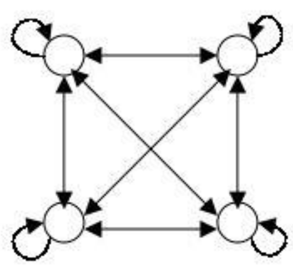
Neurons within the layer are not necessarily connected to all other neurons from the next layer. Hence, RNN architectures range from fully interconnected (Figure 14) to partially connected (Figure 13), including multilayer FF networks with distinct input and output layers.

*Figure 13: Partially Connected Network*




*Source: Medsker & Jain (1999).*

*Figure 14: Fully Connected Network*



*Source: Medsker & Jain (1999).*

Simple partially connected networks are applied for learning strings of data, where some nodes are part of the Feedforward structure, and some nodes provide the sequential context and receive feedback from other nodes. Weights from context units (marked with C1 and C2) are processed as input units with the beforementioned backpropagation process. The context units receive time-delayed feedback from the second layer units. The network is trained based on inputs and their desired successor outputs. Fully connected networks do not have distinct input layers (Medsker & Jain, 1999).

### 4.3.1 Long Short-Term Memory (LSTM) Algorithm

RNNs use their internal state (memory) to process sequences of inputs. The storage can be replaced by another network in case of time delays or feedback loops. The controlled states are referred to as gated memory. Theoretically, RNNs can track relationships over arbitrarily long sequences (Medsker & Jain, 1999). However, in practice, the influence of predictions more than a few steps in the past becomes insignificant to further impact subsequent predictions. The distorting nature of commonly applied activation functions (i.e. Sigmoid functions), one gradient can sequentially diminish over time, becoming condensed by layer after layer. On the other hand, a gradient can "explode" in case of being intensively magnified by the recurrent layer. This gradient behavior challenges long-term dependency performance. LSTM networks have been developed in order to address these shortcomings (Schuster & Paliwal, 1997)

LSTM is an extension of recurrent NNs used for solving the vanishing and exploding gradient problem faced by FF NNs (Schuster & Paliwal, 1997). It has a greater capacity of representing sequential information and it can incorporate attention mechanism for external memory acquisition.

A common LSTM algorithm consists of a cell, input, output and forgets "gates" on neural network layers. The cell is responsible for "remembering" values over certain time intervals. The three gates represent "conventional" artificial neurons that compute an activation of a weighted sum. LSTM is well-suited for classifying, processing and predicting time series, given unknown duration of time lags. Gates in LSTM are recurrently connected. The weights of these connections, which need to be learned during the training period, are later on used to direct operation of the gates. Each of the gates has its own parameter, weight, and biases. LSTM allows information to be stored across arbitrary time lags, and the error signal can be carried far back in time. This strength of the algorithm is the core memory block which is a recurrently self-connected linear unit (Schuster & Paliwal, 1997). By design, LSTM can discern connections over a long sequence of data. This makes LSTM applicable for financial predictions since it can recognize, and separate diverse market movements, causes and financial effects.

## 4.4 Literature Overview – Application of Neural Networks for Financial Series Predictions

The neural network approach differs significantly from the traditional commonly used technical and fundamental methodologies. A financial neural network model can be trained for the prediction of future asset prices with enhanced accuracy. Neural networks are considered to contribute with a state-of-the-art solution for prediction of noisy time series such as the financial series (Medsker & Jain, 1999).

Numerous studies on the application of neural network technique in financial investment forecasting are available within the academic literature. For instance, Bernal, Fok, and Pidaparthi (2012) apply Echo State RNN (ESN) for predicting prices in the S&P 500. Their echo state network is a black box method which does not require any inside into the underlying dynamics of the time series. This method is, later on, opposed to the Klaman filter – a linear dynamic model based on a mean squared error reduction. ESN method proves to have a lower error rate, compared to the Kalman filter model. However, in periods of higher trading volume and higher volatility, the error rate is increased.

Hsieh and Yeh (2011) test the model of stock market prediction based on wavelet transforms and bee colony (ABC-RRN) algorithm. Their approach combines Haar wavelet for decomposition and noise elimination of stock price time series, with the fundamental and technical indicators. These indicators are used as inputs into the ABC-RNN. The model is tested on several stock markets: Dow Jones Industrial Average Index (DJIA), London FTSE-100 Index (FTSE), Tokyo Nikkei-225 Index (Nikkei) and Taiwan Stock Exchange Index (TAIEX). The results of the system application are promising for further prediction possibilities.

Oliveira, Nobre, and Zarate (2013) introduce a neural network model for the purpose of predicting stock prices on the Brazilian stock exchange for short term trading. In their approach, they combine both technical and fundamental data as model data inputs.

Therefore, it can be concluded that the neural network method has been by far, developed and utilized for extracting predictive information into the financial world of analysis. However, due to their novelty, very few studies can be found for the matter of crypt asset value prediction.

Among the few, Madan, Saluja and Zhao (2014) present application of machine learning methods: binominal GLM, SVM algorithm and random forest algorithm for prediction of the Bitcoin price. Their data set is based on 25 features related to Bitcoin network, achieving 98,7 % model accuracy.

Nakano, Takahashi, and Takahashi (2018) apply several layered neural network structures on technical indicator data, calculated on a 15-minute time frame. Afterward, they confront their three developed investment model strategies with the standard Buy-and-Hold strategy.

Results show significant outperformance of the developed model compared to the simple Buy-and-Hold scenario.

In this study, the author focuses on explaining cryptocurrency value returns based on the Feedforward and Recurrent (LSTM) NN model. The predictive models will be tested initially on technical and network-specific features, and later enhanced with the sentiment analysis feature. The neural network model is applied for developing an investment strategy for three focus crypto assets: Bitcoin, Litecoin and Ether.

## 4.5 Quantitative and Crypto-specific Input Variables

Forecasting movements of future stock prices is a widely studied topic in many fields, including finance, statistics, and computer sciences. However, the plausibility of predicting prices and subsequently "beating the market" still presents an open matter of discussion. The challenge is to construct a solid model that will detect the factors that drive prices of particular financial asset. Efficient Market Hypothesis (EMH) challenges the likelihood of market price prediction, arguing that market prices are random, hence they cannot be forecasted (Markowitz, 1952). However, on the spot, market investors base their beliefs on their predictive price models built on fundamental technical analysis tools (see, Appendix 6) and general market sentiment understanding. Applying the same approach, we introduce quantitative and crypto-specific variables as inputs for the developed predictive model.

The predictive model grounds on Feedforward and Recurrent (LSTM) neural network algorithm, applied for forecasting the future value range of cryptocurrencies. Basic Feedforward neural network and the LSTM RNN model are selected to be tested upon underlying cryptocurrency variables.

*Table 6: NN Model Input variables – Quantitative and Crypto specific variables*

| Input Variable | Explanation |
|---|---|
| Cryptocurrency Market Price | Daily closing market price |
| Network Value (Market Capitalization) | Market price x Circulating Supply |
| Number of Transactions | The volume of Transactions on the Blockchain |
| Sent from Address | Number of Transactions sent from one address |
| Difficulty Rate | Computational power needed for block creation |
| Hash Rate | Number of Hashes per second |
| Mining profitability | Daily return from mining (in U.S. dollars) |
| Block size | Data contained per transaction block (KB/MB) |
| Sent in U.S. dollars | Value of daily sent coins |
| Transaction Fees | Fee paid per unit of blockchain transaction |
| Confirmation (Block) Time | The time required for block creation (in minutes) |
| Median Transaction Value | Median U.S. dollar value of daily transactions |
| Number of Active Addresses | Number of unique addresses included in daily transactions |
| CRIX Index | Cryptocurrency Market Index |

*Adapted from Bitinfocharts (2018).*

The algorithmic models are based on quantitative and crypto-specific value indicators. Different types of quantitative and crypto-specific metrics are introduced as inputs to the neural network models: technical data presented with the cryptocurrencies closing market prices, market capitalization, volume of transactions and crypto-specific type of data: block size, sent transactions from address, difficulty, hash rate, mining profitability, sent in U.S. dollars, transaction fees, confirmation time, transaction value, number of active addresses and CRIX index.

The data was obtained from bitinfocharts.com – publicly available, online data library, excluding the CRIX index. CRIX data was gathered from the Humboldt University of Berlin online database (Chen, Chen, Härdle & Lee, 2016). Following is a brief description of the quantitative metrics employed in the LSTM model.

Cryptocurrency Market Price – daily closing market prices (exchange rates) of Bitcoin, Litecoin and Ether, obtained from Bitinfocharts.com, as of 1st of July 2018.

Network Value or the Market Capitalization is a quantitative metric indicating the relative size of a cryptocurrency market. It is calculated as:

$$Network\ Value_{it} = Price_{it} \times Circulating\ Supply_{it} \tag{9}$$

where circulating supply is the representation of the number of coins circulating on the market and it can, de facto, influence cryptocurrency price movements.

Number of Transaction – a quantitative indicator of the daily volume of transactions on the Blockchain. It is commonly applied for the purpose of technical analysis. Transaction volume can relate to the market sentiment. Quantitative traders look at the volume of transactions as an indicator of future trend shifts (Murphy,1986).

Sent from Addresses is a measure of the number of transactions (MB/KB) sent from given cryptocurrency address at a moment of time (Bitinfocharts, 2018).

Difficulty metrics indicates the average difficulty per day for a given Blockchain system. The difficulty as variable measures the computational power required for finding a new block with a fixed level of hash power. Due to the Difficulty Retarget mechanism, the difficulty is adjusted resulting from the aggregate mining effort employed or removed from the network (Bitinfocharts, 2018).

Hash rate (Hash/S) – the average hash rate (Hash/S) per day represents the estimated number of hashes per second that the Blockchain network is performing. A “hash” is a function which converts data input into an encrypted output with a fixed length.

Mining profitability is a measure of U.S. dollar daily returns of miners, measured in Terra Hash per second – THash/S (Bitinfocharts, 2018).

Block size indicates the amount of data contained in one block per transaction. Block size on the Blockchain is measured in Kilobytes (KB) and Megabytes (MB). Certain cryptocurrencies have a limited block size. For instance, Bitcoin transactions are limited to 1 MB (since 2010) for security reasons – preventing against Denial of Service attacks (DOS) to which large block is prone to. On the opposite, higher blocks size can increase transaction processing, achieving a larger scale to foster mass adoption competing with existing rival payment systems (i.e. VISA or PayPal)

Sent in U.S. dollars presents as well, quantitative crypto-specific metrics. It reveals the number of coins sent in U.S. dollars, on a daily level, within the Blockchain (Bitinfocharts, 2018).

Transaction Fees – the U.S. dollar amount of fees paid for executing cryptocurrency transactions. Diverse cryptocurrency exchanges impose different fee rates for performing cryptocurrency transactions. For the purpose of the analysis, we account for the average transaction fee amount for given cryptocurrency.

Confirmation or block time metrics is expressed in minutes of time. A block is founded by mining over a certain time interval, on average, regardless of the magnitude of the mining effort. This time is captured with a block time indicator. For instance, BTC blocks will be found, on average, once every 10 minutes (Cermak, 2017).

Median Transaction Value is a crypto-specific indicator which measures the U.S. dollar value of transactions for a certain crypto asset, at a certain moment of time. Data on median transaction value is used as input to the neural network models (Bitinfocharts, 2018).

Number of Active Addresses – the cryptocurrency network is built out of different addresses representing a single person's account. Nevertheless, one person has the possibility of owning several addresses. Hence, this feature provides inside into the number of unique addresses which have been included in daily transactional activity.

CRIX is a cryptocurrency market index, calculated based on the weighted averages on market capitalization of a variating number of different cryptocurrencies. The CRIX index is constructed and calculated by the Humboldt University of Berlin.

CRIX is computed by calculating the differences in the log returns of the market against a selection of possible benchmarks. The aim of any index is to weight the prices of the assets by it quantitates. CRIX, grounds on Laspeyres index which takes the value of a basket of k assets and compares it against a base period:

$$P_{Qt}^{L}(k) = \frac{\sum_{i=1}^{k} PitQiq}{\sum_{i=1}^{k} PitQiq} \qquad (10)$$

CRIX index:

$$CRIXt(k, \beta) = \frac{\sum_{i=1}^{k} \beta it - PitQiq}{Divisor\ (k)ti} \tag{11}$$

Where P and Q stand for price and quantity, respectively, and βit is the adjustment factor of a set i at given time period t. In addition, i=1 indicates that this is the i-th adjustment factor, and i-1 is the last time point when Qi, t-l, Divisor(k)i, t-l and βi, t-l were updated βi, t-l is defined to be βi, t-l=1 for all i and l (Trimborn & Härdle, 2016).

CRIX consists of a selection of cryptocurrencies that are representative of the cryptocurrency market. The frequent change experienced in the cryptocurrency market requires a dynamic index structure, in order to capture and understand these changes. The number of index constituents is rescheduled quarterly so to ensure an up-to-date fit to the current market structure. As of 1st of July 2018, the CRIX index consists of 19 different crypto assets (Humboldt University of Berlin, 2018).

CRIX Index data since 31st July 2014 is used as an input variable to the neural network predictive model.

### 4.6 Neural Network Predictive model

The neural network predictive model is based on FF and the RNN algorithm and it is trained and tested based on diverse quantitative and crypto-specific input variables, as elaborated in section 5.5. The training phase of the networks includes the overall available data as of 28th of February 2018. On the other hand, the performance of the train model is tested on 120-day data records, starting from 28th of February, up until the 26th of June 2018. Graphical representation of the FF and RNN (LSTM) network model is represented in Appendix 2 and 3.

The network algorithm makes predictions on the cryptocurrency price range for three days ahead, and the price change threshold was set to ± 3 %. The model prediction results are presented with values 0,1 and 2, where each value represents one trading position forecast (investment recommendation):

- Short: Predictive value = 0, predicted price decline by more than 3 %
- Hold: Predictive value = 1, predicted price within the range of ± 3 %
- Long: Predictive value = 2 predicted price increase by more than 3 %

The trading strategy setup is predefined before testing the neural network. In the beginning, the U.S. dollar worth of one cryptocurrency (BTC, LTC or ETH) is invested with the current market price. During the test period, two accounts are maintained: Account_USD (return on investment earn in U.S dollars) and Account_Crypto (value of the crypto asset on disposal, valued by the spot crypto market price). The compounded account combines both Account_USD and Account_Crypto.

Following, if the model predicts a price increase by more than 3 %, then the algorithm takes the long position in the cryptocurrency – invests total amount available on the Account_USD. If the balance on this account is zero, then no action is taken. On the other hand, if the model predicts a price decrease by more than 3 %, the algorithm takes the short position on the crypto asset – sells all cryptocurrencies accumulated on the Account_Crypto. If the balance on this account is zero, no action should be taken. In the situation where the network predicts price movement within the range of ± 3 %, the algorithm takes hold position – no further changes are undertaken.

*Table 7: Trading strategy summarized overview*

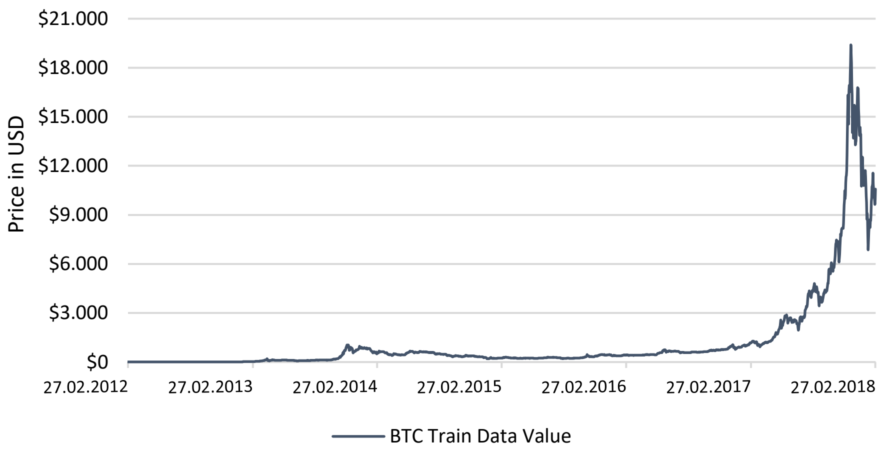
| Predicted value | Prediction | Condition 1. | Action | Condition 2. | Strategy |
|---|---|---|---|---|---|
| 2 | Price increase by > +3 % | If balance on Account_USD > 0 Invest total amount | Buy | Otherwise, hold | Long |
| 1 | Price moves in the range of ± 3 % | No change | Hold | No change | Hold |
| 0 | Price decrease by > -3 % | If balance on Account_Crypto > 0 Sell all | Sell | Otherwise, no change | Short |

*Source: Own work.*

### 4.6.1 FF and RNN (LSTM) Model Results – Bitcoin

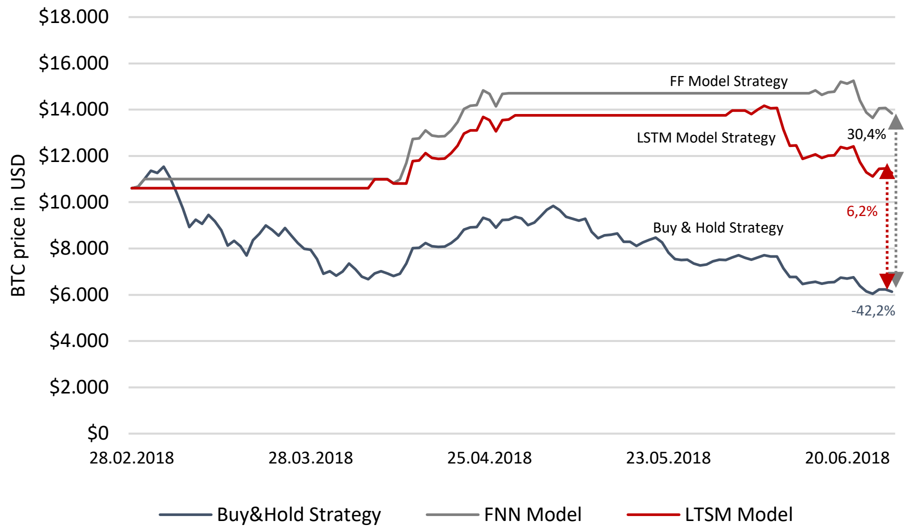
The Feedforward and the Recurrent neural network model were trained based on Bitcoin data available for the period of 17.10.2010 – 27.02.2018 (2.783 calendar days or 95,9 % of the data used for training of the network). The models are trained to predict Bitcoin price ranges based on quantitative and crypto specific input variables, as elaborated above. Figure 15 shows the Bitcoin price development during the period of network training.

*Figure 15: BTC Train Value*




*Adapted from Bitinfocharts (2018).*

*Figure 16: BTC Predictive Model Results*




*Source: Own work.*

Figure 16 presents the strategy performance on Bitcoin price prediction for a period of 120 days. Both the FF (grey trendline) and LSTM (red trendline) strategies are compared with the simple Buy-and-Hold strategy threshold.

*Table 8: Bitcoin Strategy Results Overview*

| | | Buy & Hold Strategy | FF Model | LSTM Model |
|---|---|---|---|---|
| BTC | Return on Investment (ROI %) | -42,2 | 30,4 | 6,2 |

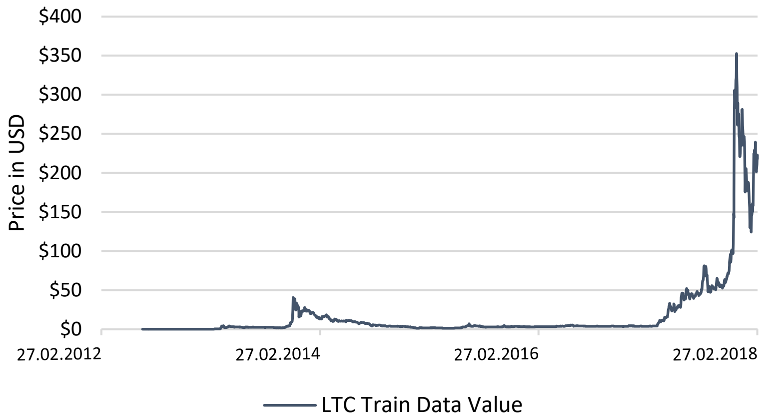
*Source: Own work.*

If the investor took the long position in one Bitcoin as of 28th of February 2018 and held the asset up until the end of the testing period – 26th of June 2018, his return on investment will be negative, -42,2 % overall. However, algorithmic trading brings investors positive returns. The results show the best performance of the Feedforward NN model strategy, which generates 30,4 % returns on investment. Following is the LSTM model strategy with significantly lower returns of 6,2 %.

### 4.6.2 Feedforward and LSTM Model Results – Litecoin

The FF and the LSTM networks have been trained with Litecoin specific data available from the period of 13th July 2012 up until 27th of February 2018 (2.056 calendar days or 94,5 % of the data used for training of the network). Figure 17 presents the Litecoin price development during the period of network training.

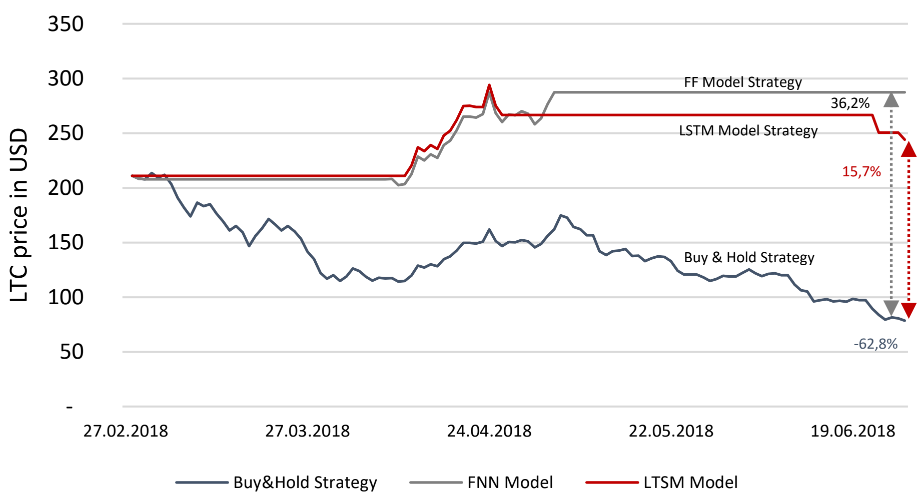
*Figure 17: LTC Train Value*




*Adapted from Bitinfocharts (2018).*

Figure 18 presents the strategy performance on FF (grey trendline) and LSTM (red trendline) model strategies are opposite to the Buy-and-Hold strategy. In the case of Litecoin price prediction and algorithmic trading, the Feedforward network model shows again superior performance, with ROI of 36,2 % while the LSTM model exhibits 15,7 % ROI. During the test period, the Litecoin value has decreased by 62,8 % in total (Table 9).

*Figure 18: LTC Predictive Model Results*




*Source: Own work.*

*Table 9: Litecoin Strategy Results Overview*

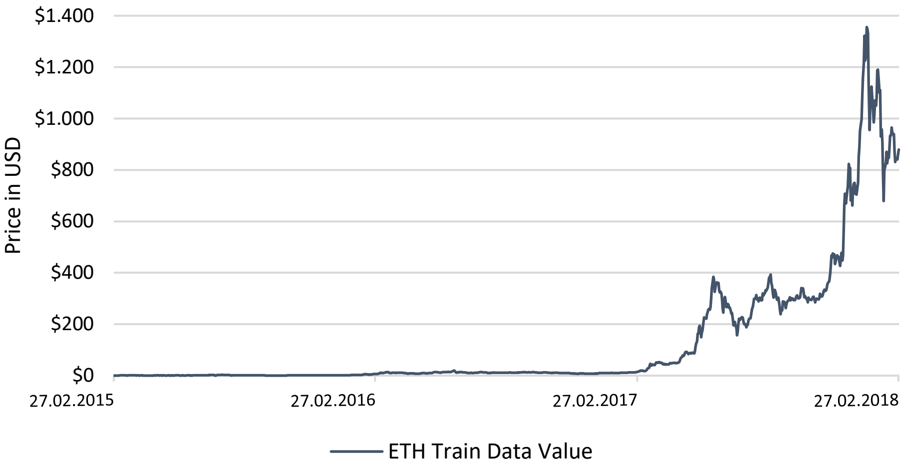
| | | Buy & Hold Strategy | FF Model | LSTM Model |
|---|---|---|---|---|
| LTC | Return on Investment (ROI %) | -62,8 | 36,2 | 15,7 |

*Source: Own work.*

### 4.6.3 Feedforward and LSTM Model Results – Ether

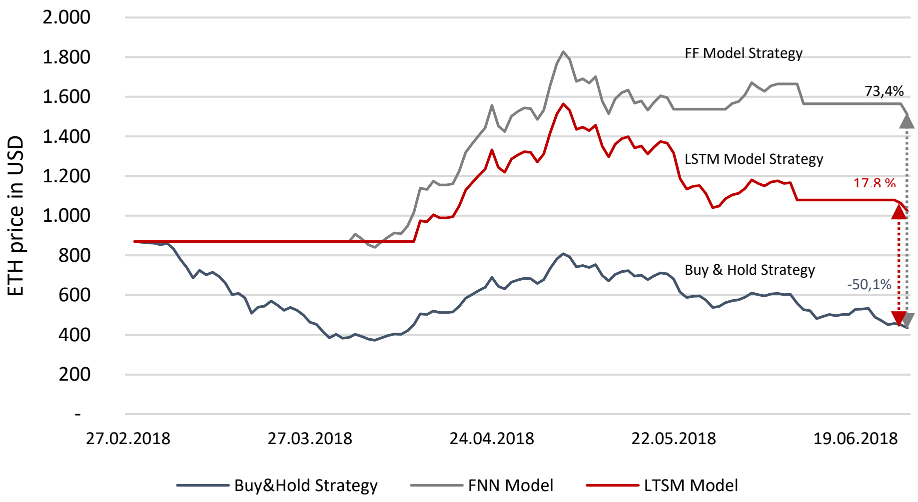
Figure 19 presents the Ether price movements during the period of the network training, from the period of 30th September 2014 until 27th of February 2018 (1.247 calendar days or 92,2 % of the data used for training of the network).

*Figure 19: ETH Train Value*




*Adapted from Bitinfocharts (2018).*

*Figure 20: ETH Predictive Model Results*




*Source: Own work.*

The results show superior performance of the FF NN model. The FF model brings returns of 73,4 %, while the LSTM model brings 17,8 % returns on investment. If the investor has bought one Ether as of 28th of February 2018 and hold it until the end of the testing period – 26th of June 2018, his return on investment will be negative, -50,1 % (Table 10).

*Table 10: Ether Strategy Results Overview*

| | | Buy & Hold Strategy | FF Model | LSTM Model |
|---|---|---|---|---|
| ETH | Return on Investment (ROI %) | -50,1 | 73,4 | 17,8 |

*Source: Own work.*

## 4.7 Predictive Model Accuracy

The following section presents an overview of the predictive accuracy achieved with the application of FF and RNN (LSTM) models. Table 11 summarizes the accuracy outcome for all three cryptocurrencies.

For the case of the three in-focus cryptocurrencies: Bitcoin, Litecoin and Ether, the Feedforward neural network model shows better predictive accuracy (45,8 %, 52,5 % and 49,2 % correspondingly). The RNN model makes lower accurate predictions, falling within the range of 33-47 %.

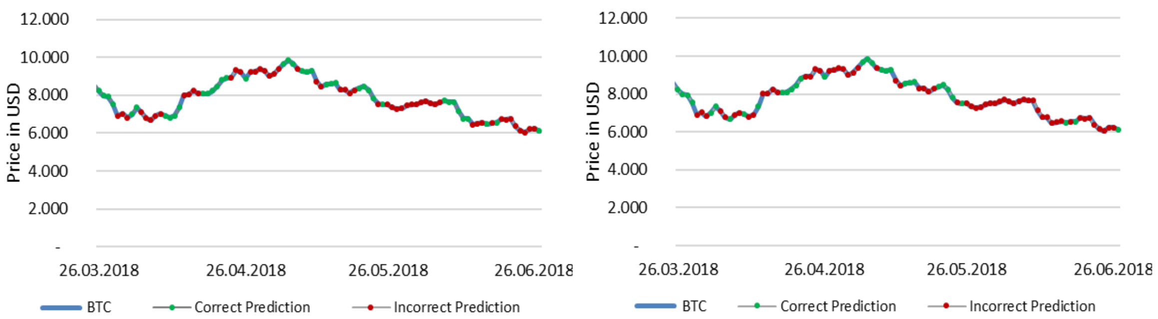
*Table 11: Summarized Result Accuracy Overview – BTC, LTC & ETH price prediction*

| **Bitcoin** | **FF Model** | **LSTM Model** |
|---|---|---|
| Correct predictions | | |
| Actual Price = Predicted Price | 55 | 45 |
| Incorrect predictions | | |
| Actual Price > Predicted Price | 46 | 44 |
| Actual Price < Predicted Price | 19 | 31 |
| **Predictive accuracy** | **45,8 %** | **37,5 %** |
| **Litecoin** | | |
| Correct predictions | | |
| Actual Price = Predicted Price | 63 | 57 |
| Incorrect predictions | | |
| Actual Price > Predicted Price | 44 | 56 |
| Actual Price < Predicted Price | 13 | 7 |
| **Predictive accuracy** | **52,5 %** | **47,5 %** |
| **Ether** | | |
| Correct predictions | | |
| Actual Price = Predicted Price | 59 | 40 |
| Incorrect predictions | | |
| Actual Price > Predicted Price | 15 | 22 |
| Actual Price < Predicted Price | 46 | 58 |
| **Predictive accuracy** | **49,2 %** | **33,3 %** |

*Source: Own work.*

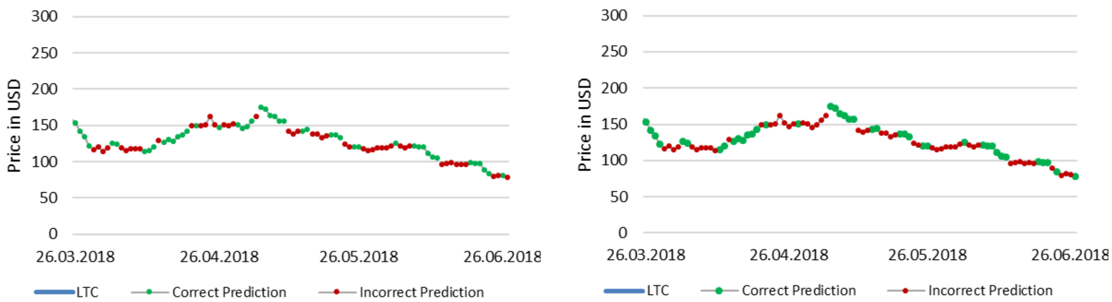
Figures 21, 22 and 23 display the neural network model accuracy for making BTC, LTC and ETH price predictions. Green dots present the accurate price prediction – the model correctly indicated price increase or decrease. Red dots present the wrong model prediction – the model failed to predict price increase/decrease according to the pre-defined criteria.

*Figure 21: Bitcoin – FF and LSTM Predictive Model Accuracy*



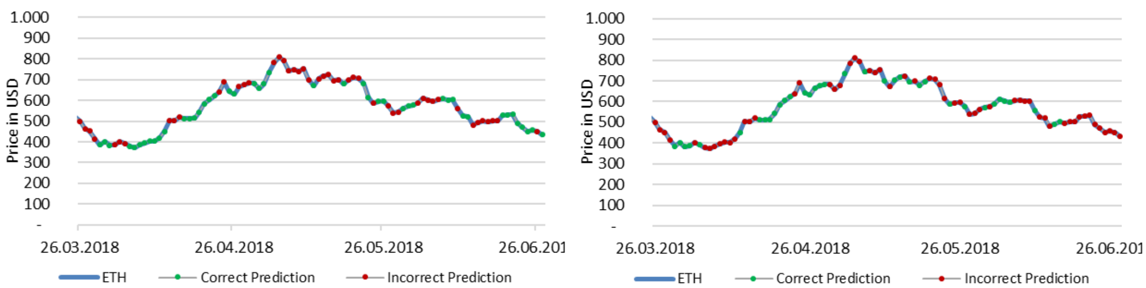

*Source: Own work.*

*Figure 22: Litecoin – FF and LSTM Predictive Model Accuracy*




*Source: Own work.*

*Figure 23: Ether – FF and LSTM Predictive Model Accuracy*




*Source: Own work.*

When predicting the price of BTC and LTC, the two algorithmic models make over-conservative price predictions (see, Table 11). The highest level of misclassification is present in the case when Actual price > Predicted price, or the model predicts lower prices than the observed. For the case of Ether, it can be noted that both models make opposite predictions: the FF model makes over-optimistic predictions (Actual price < Predicted price misclassification), whereas the RNN model's predictions are over-pessimistic (Actual price > Predicted price misclassification).

# 5 SENTIMENT ANALYSIS

Sentiment Analysis (hereinafter: SA) is a text classification problem of analyzing people sentiment contained in certain text data. SA has its origins in natural language processing and linguistics and it was established as a separate field of research in the early 2000s (Pang & Lee, 2004). SA algorithms present an analysis tool for disclosing sentiment, emotions, opinion and attitudes from a written text towards a certain subject matter. In its essence, SA is applied for identifying and examining sentiment contained in a given text, so to comprehend key features that influence human behavior. By far, SA has been applied on social media data gathered from online social communities (blogs, microblogging platforms, and other collaborative social media). The majority of SA applications are based on spotting polarity (positive/negative) and intensity of the sentiment contained in the given text (Chaturvedi, Cambria, Welschb & Herrera, 2017).

The task of SA is a classification problem considering feature extraction from certain text. The features that are mainly analyzed are: terms presence and frequency (words and word n-grams frequency) parts-of-speech (POS) – focus on adjectives as important opinion indicators opinion words and phrases and negations (Pang & Lee, 2004).

Feature selection techniques treat text either as a group of words (Bag of Words – BOWs), or as a string retaining the sequence of words in the document. Several statistical methods can be applied for the feature selection process. Point-wise Mutual Information (PMI), Chi-square, Latent Semantic Indexing (LSI) are just some of these methods (Pang & Lee, 2008).

Supervised learning depends on the existence of labeled training documents. Several classifiers have been developed for this matter. Probabilistic or generative classifiers assume that each class is a component of the mixture, and each mixture component provides a probability of sampling terms of the component. A linear classifier is a supervised learning method where: $X = \{x_1, x_2 ... x_n\}$ is the normalized document word frequency, vector $A = \{a_1, a_2 \ldots\ldots a_n)$ is a vector of linear coefficients with the same dimensionality as the feature space, and b is a scalar. Hence, the output of the linear predictor is defined as $p = AX + b$ where P stands for the separating hyperplane between classes. Support Vector Machines (SVM) and NNs are most commonly used linear classifiers (Vukotic, Claveau & Raymond, 2015).

The decision tree classifier provides a hierarchical decomposition of training data with the condition on the attribute value for data division. The condition is the presence or absence of value used to divide the data. The process occurs recursively until the leaf nodes contain minimum records (Melville, Gryc & Lawrence, 2009).

The rule-based classifier groups data according to some set of rules. The condition is set on the term presence. Hence, the left-hand side is a condition on the feature set expressed in disjunctive normal form and the right-hand side is the class label. The training phase is the process of rule construction. Rules are constructed based on criteria of support and confidence. Support refers as the absolute number of instances in training data relevant to the rule Confidence refers to the conditional probability that the right-hand side of one rule is satisfied if the left-hand side is satisfied. The main difference between decision trees and rule-based classifiers is that decision trees use strictly hierarchical portioning of the data space. Rule-based classifiers, however, allow for overlapping of data classes (Kotsiantis, 2007).

## 5.1 Sentiment Analysis – Application in Financial Trading

News flow and sentiment are important sources of signals in quantitative investment and systematic trading. However, a vast amount of social media data is unstructured, making it difficult to understand if the market perceives information as positive, negative or irrelevant. The advancement of machine learning tools and digital data records has reduced costs, hence making sentiment analysis a growing area of experimental research.

As discussed, the task of SA is text sentiment classification. It is a collection of effective computing research aiming for text classification into either the positive, negative or neutral domain (Chaturvedi, Cambria, Welschb & Herrera, 2017). It requires tackling many natural language processing (NLP) tasks: word polarity disambiguation, personality recognition, name entity recognition, sarcasm detection, and aspect extraction. At present, we witness various researches in the field of applying SA tools for trading purposes.

Sul (2016) in the paper Trading on Twitter: Using Social Media Sentiment to Predict Stock Return uses 2.5 million tweets as a data source in order to predict stock movements of S&P 500 firms. He applies his own sentiment classifier and connects it to the stock returns. Results show that sentiment tracked through Twitter posts is expected to be reflected in stock returns on the same trading day. Basing a trading strategy on the sentiment study, the author exhibits 11–15 % predicted annual gains (Sul, 2016).

In the paper Bayesian Regression and Bitcoin, Shah and Zhang (2014), achieved 89 % ROC over fifty days of buying and selling Bitcoins by applying Bayesian regression results. Madan, Saluja and Zhao (2014) applied a random forest model to predict the price change of Bitcoin. In this manner, they achieved 98,7 % predictive accuracy, while still failing to account for people's emotions, not managing to harness their potential in the applied learning algorithms.

The study Algorithmic Trading of Cryptocurrency Based on Twitter Sentiment Analysis performed by Colianni, Rosales and Signorotti (2015) examined how tweet sentiment can be used to impact investment decisions on Bitcoin. The researchers applied supervised machine learning techniques gaining accuracy of above 90 % hour-by-hour and day-by-day. This accuracy was achieved through robust error analysis of the input data, reducing error rates after cleaning the dataset noise.

Most recently, the SA algorithm has found its applicability in cryptocurrency trading tools. For instance, Thomson Reuters Corp. launched a new BTC market data feed in order to help traders when making their decisions (Thomson Reuters, 2018). In cooperation with the company Market Psych, they constructed an algorithm that uses metrics like "greed" and "fear" to identify hedging opportunities or create buy-sell orders at certain periods of time. The core of this product is a program called Bayesian filter used for turning unstructured data into actionable insight for maximizing gains through the ML processes. In this way, Thomson Reuters Market Psyche Indices use feelings as an indicator for providing alpha, while at the same time providing content filtering, detecting data only relevant to cryptocurrencies.

Another example is Sentdex algorithm. The Senddex platform uses Reddit as the source of content for tracking sentiment around Bitcoin. The algorithm defines sentiment based on Reddit articles linked to posts and comments, as well. Sentdex reads articles and it pulls out "named entities" with the "named entity recognition algorithm" based on "chunking" group bits of text into noun-phrases and broad text information. What is common to the above mentioned, real-world applications, is the application of sentiment for the purpose of providing insight into cryptocurrency-related public sentiment and opinion.

### 5.2 VADER – Simple Rule-based Model for Sentiment Analysis

VADER or Valence Aware Dictionary for Sentiment Reasoning is an extensive lexicon and a rule-based sentiment analysis model. The VADER model presents a tool specifically designed for the requirements of sentiment expressed in social media. VADER text sentiment analysis was introduced in 2014 by C. J. Hutto from the Georgia Institute of Technology. The model incorporates lexical features with specific rules of grammatical and syntactical conventions used in expressing and emphasizing sentiment intensity. It incorporates both lexicon and rule-based model characteristics.

The VADER sentiment lexicon is sensitive to both the polarity and the intensity of sentiments expressed in a social media context. VADER combines qualitative and quantitative methods in the application of the gold-standard (human validated) list of lexical features connected to the microblogging space sentiment. The "Gold standard" approach includes features from the traditional sentiment lexicons (established sentiment word-banks – LIWC, ANEW, General Inquirer). VADER uses a human-centric approach, incorporating qualitative analysis with empirical validation of human raters and the wisdom of the crowd (collective opinion is more trustworthy than individual opinion). Human raters are retrieved from Amazon Mechanical Turk (Hutto & Gilbert, 2014).

By applying the wisdom-of-the-crowd approach (Surowiecki, 2004), sentiment valence is estimated. Each feature is granted with a sentiment polarity score (positive/negative) and a sentiment intensity score in the ± 4 range. The -4 score being the most negative and +4 the most positive sentiment score. The sentiment score is obtained by summing up the intensity of each word in the text. The sentiment score of a sentence is the normalization to the total, so it is mapping in the range between ± 1. The ± 1 normalization is the following:

$$x = \frac{\mathrm{x}}{\sqrt{x^2+a}}, \textit{ where } \alpha = 15 \qquad (12)$$

Denoted with $x$ is the sum of the sentiment score of the constituent words in the sentence and α is a normalization parameter set to 15.

It can be noted that as $x$ increases, it gets closer to ± 1 values. A similar effect can be observed when VADER is applied to a large amount of word data. Hence, VADER SA works best on short documents, such as tweets and sentences, rather than on large documents.

This gold standard list of features with associated valence for each feature comprises VADER's sentiment lexicon. In addition, numerous lexical features or heuristics attuned to microblogging space were included (emoticons, sentiment related acronyms and initialisms, sentiment slang etc.).

The model distinguishes itself from traditional sentiment models in that it is more sensitive to sentiment expressions frequently applied in a social media context. In addition, five general rules are introduced to enhance the performance of the model: punctuation (ex. exclamation point), capital letters, degree modifiers (degree adverbs: "extremely", "marginally" etc.), contrastive conjunctions (ex. "but") and negation flips of text polarity. All these rules present the heuristics that according to Huto & Gilbert (2014), humans use to assess sentiment intensity in the text.

Compared to seven SA lexicons: Linguistic Inquiry Word Count (LIWC), General Inquirer (GI), Affective Norms for English Words (ANEW), SentiWordNet (SWN), SenticNet (SCN), Word-Sense Disambiguation (WSD) using WordNet, and the Hu-Liu opinion lexicon; the VADER model performs exceptionally well in the social media domain. It records a significant Person Product Moment Correlation Coefficient ($r = 0{,}881$) and improved classification accuracy ($F_1 = 0{,}96$) compared to the individual human rates (Huto et al, 2014).

The author uses VADER's simple rule-based approach due to its good performance, computational simplicity, and accessibility. Furthermore, the model is both self-contained and domain agnostic, hence it does not require an extensive set of training data and has superior performance (Hutto & Gilbert, 2014). It is an open-source code, available as a SA package in Python.

### 5.3 Sentiment Analysis – Twitter Data Set

Social media and microblogging networks serve as a platform for expressing an opinion and tracking public sentiment. Social networks attract millions of users and are characterized with high-level efficiency in terms of reach, frequency, usability, immediacy, and permanence.

At present, most of the existing SA researches were based on data gathered from social media platforms. Twitter, in particular, is widely adopted as a social media source for tracking sentiment features in the investment analytics field. An enlarging number of investors and the general public use Twitter for sharing their opinion trough out special text format known as the "tweet". "Tweet" stands for a text-based post containing only 140 characters, analogical to headlines of the traditional newspaper. Short messages are simple and convenient for both the sender and the reader to share a matter of interest and in real time – to communicate their opinion. In November 2017, Twitter expanded the character limit up to 280. The purpose was to improve user convenience and engagement. In the second quarter of 2018, there were about 335 million active Twitter users with above 400 "tweets" sent per minute (Statista, 2018).

The limitation on the length of tweets motivates users to convey emotion in one direction, enabling categorization based on its valence. Individual trading behavior is not always based on rational assessment of relevant information, but rather it is impacted by emotions. As noted, Twitter can introduce a solid environment for fostering emotions. Emotions can generate an intense effect on individual trading decisions. SA is the process of automatically detecting if one text unit contains emotional or opinionated content and if yes, determining its polarity.

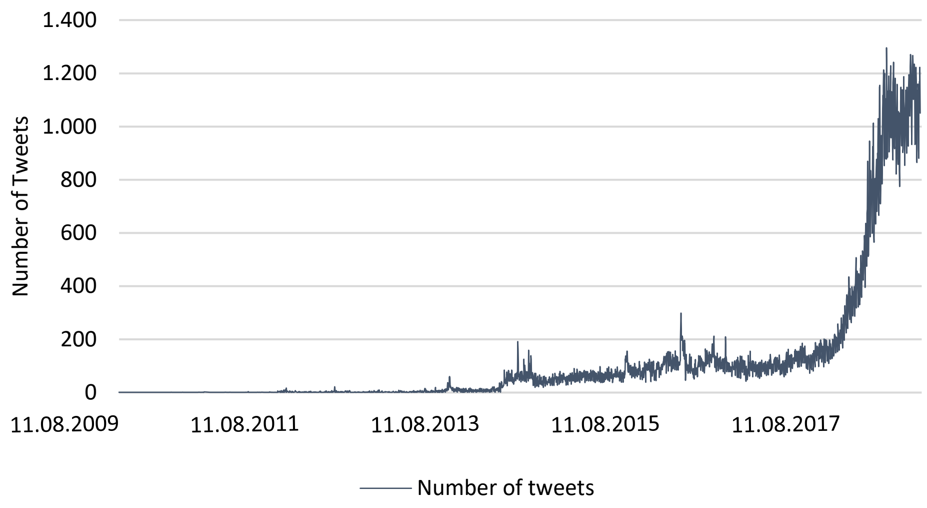
*Table 12: Twitter Hashtags used for Twitter data download.*

| Hashtags: | Used since: | Explanation: |
|---|---|---|
| #Crypto | 23.02.2008 | Cryptocurrency abbreviation |
| #LTC | 15.05.2008 | Litecoin abbreviation |
| #ICO | 02.10.2008 | Initial Coin Offering |
| #Token | 20.10.2008 | Digital currency |
| #ETH | 30.10.2008 | Ether abbreviation |
| #Ether | 22.03.2009 | Ether |
| #Bitcoin | 20.03.2010 | Bitcoin |
| #Crypto#News | 03.04.2010 | News on cryptocurrencies |
| #Trading | 08.05.2010 | Cryptocurrency context |
| #Cryptocurrency | 11.07.2010 | Encrypted, digital currency |
| #Ethereum | 10.10.2010 | Ethereum network |
| #Litecoin | 13.10.2011 | Litecoin |
| #Altcoin | 19.12.2011 | Alternative coin |

*Source: Advanced Twitter Search Tool (2018).*

In general, the trading community uses the convention of specifically tagging crypto-related tweets. This makes Twitter an ideal source for spotting the information about social interests and general public sentiment toward cryptocurrencies. For our purpose, tweets have been diligently selected with the help of the Advanced Twitter Search Tool. Tweets were downloaded with the help of TweetScraper tool. The tool is developed on the Scrapy platform without using Twitter's APIs, thus, it eliminates the Twitter API's rate limitations and restrictions.

*Figure 24: Number of Tweets*




*Source: Advanced Twitter Search Tool (2018).*

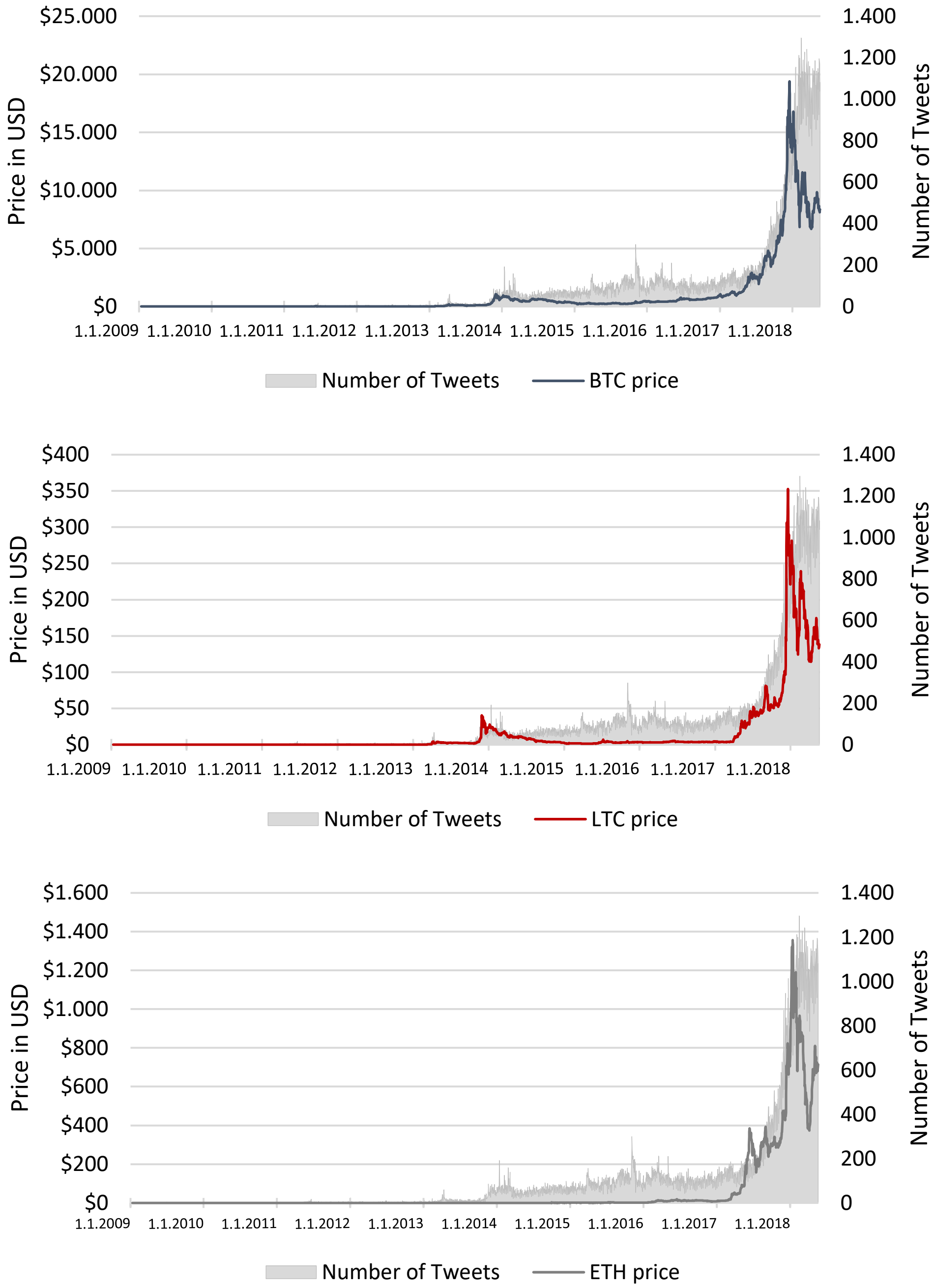
Tweets applied for sentiment tracking were selected based on the containing hashtags. Selection of cryptocurrency-related hashtags was performed based on the own work findings. The focus was especially put on the three most well-known cryptocurrencies with the longest data histories: Bitcoin, Litecoin and Ethereum. Hashtags used for tagging Twitter posts serve as an indicative filter for content data selection.

TweetsScraper tool has been used for dowloading tweet data. The tool is built on Scrapy.org platform which is independent from the Twitter's API limitations and restrictions. The downloaded tweet data series began on the 11th of January 2009 and ended on the 1st of July 2018. Following, the VADER model leverages on the base of data of 2.608 calendar days or 386.111 observations in total.

## 5.4 Sentiment Analysis Results

The sentiment analysis dataset consists of tweets posted in the period of 11.01.2009 – 01.07.2018 (2.608 calendar days). For this period, 386.111 tweets have been analyzed, and out of them, daily compound sentiment score has been derived.

*Figure 25: BTC/ LTC/ ETH Price and Number of Tweets*




*Source: Advanced Twitter Search (2018); Bitinfocharts (2018).*

The relation of cryptocurrency prices (BTC, LTC, ETH) and the number of top cryptocurrency tweets per day are presented on Figure 25. In the case of all three cryptocurrencies, it can be noted that in the earlier period of the cryptocurrency market development, the number of tweets is low to insignificant (2009–2014). At the end of 2014 and the beginning of 2015, we note a graduated increase in the number of cryptocurrency related tweets. Value wise, prices of Bitcoin, Litecoin and Ether for that period were significantly lower, compared to the nowadays benchmark. At the end of 2017, the number of crypto-related tweets exponentially increased, following the significant rise in cryptocurrency prices. What can be observed from these graphs (from December 2017 onwards) is that a higher delta of cryptocurrency price changes is followed by a higher number of tweets in relation to the coin.

The VADER Sentiment model was obtained from GitHub and run in Python. The outcome is a compounded score generated based on the Twitter dataset. This compound score is computed by summing the valence scores of each word in the lexicon, and later adjusted according to the Golden rules, explained in section 6.2. Following, the score is normalized to be between –1 (extreme negative) and +1 (most extreme positive). The result is a single unidimensional measure of sentiment for a given sentence. The score presents "a normalized, weighted composite score". The defined sentiment compound score is the following:

- Positive sentiment – the compound score is higher or equal to 0.05
- Neutral sentiment – the compound score is between -0.05 and 0.05
- Negative sentiment – the compound score is lower or equal to -0.05

Resulting from the VADER model analysis, during the period of 2.608 days, in 1.122 days the compounded mean score was revealed to the positive sentiment towards these cryptocurrencies, in 346 days the compounded tweet score was neutral, whereas in 1.140 days the mean compounded tweet score showed negative sentiment.

*Table 13: VADER Result: Positive Sentiment – Correct Compound Score*

| Tweet | VADER Score | Sentiment |
|---|---|---|
| The more governments fight against #Bitcoin the stronger it becomes, the more valuable it becomes and the more clear the reason for its creation becomes... #BuyAndHODL<br>8:19 AM - 6 Feb 2018<br>10 Retweets 33 Likes | Compound Score: **0.797**<br>*Negative score: 0.07,*<br>*Neutral score: 0.64,*<br>*Positive score: 0.29.* | Positive |

*Source: Own work.*

*Table 14: VADER Result: Negative Sentiment – Correct Compound Score*

| Tweet | VADER Score | Sentiment |
|---|---|---|
| "Bitcoin and other crypto-currencies are a way for bad dictators or criminals to bypass sanctions.." Bill Browder at the #WEF18 in #Davos cnbc.com/2018/01/23/cri ... He said that #Bitcoin's usefulness for criminal activity will be the death of it and other crypto-currencies.<br>5:32 AM - 28 Jan 2018<br>7 Retweets 12 Likes | Compound Score: -**0.923**<br>*Negative score: 0.26,*<br>*Neutral score: 0.69,*<br>*Positive score: 0.04.* | Negative |

*Source: Own work.*

As previously mentioned, the VADER tool classification was used in order to define sentiment contained in the Twitter database. According to the model developers, VADER's

FI classification accuracy is up to 0.96 (Hutto & Gilbert, 2014). Nevertheless, sentiment classification is a challenging task, hence, the majority of the models, including VADER, are prone to misclassification errors. Due to this awareness for the model imperfection, an overview of the sentiment classification results follows along with samples from correct and incorrect tweet valence estimation

Table 13 and 14 present a sample of correct sentiment assessed tweets: positive content tweets graded as positive and negative tweets, assessed as negative. The sentiment is correctly assigned based on both valence and polarity of the sentiment contained in the tweet.

On the opposite, Table 15 shows an example of sentiment misclassification. The tweet is assessed as containing a strong negative sentiment. The words: "fail", "sell", "loss", "fear" – have influenced that this tweet is graded as an expression of a strong negative sentiment toward Bitcoin. However, this tweet contains positive sentiment towards the Litecoin, Bitcoin and the cryptocurrency community, in general.

*Table 15: VADER Result: Incorrect Sentiment Classification/Compound Score*

| Tweet | VADER Score | Sentiment |
|---|---|---|
| 90% of people are destined to fail with crypto, because they are too emotional, and will sell at a loss anyway. I have seen this ALOT recently! Fear Is the Biggest Killer As An Investor! #LitecoinFam #Ltc #Litecoin $LTC #Crypto #Bitcoin #cryptocurrency 8:51 PM - 2 Apr 2018 10 Retweets 47 Likes | Compound Score: **-0.929**<br>*Negative score: 0.27, Neutral score: 0.69, Positive score: 0.03* | Negative |

*Source: Own work.*

## 5.5 Neural Network Predictive model and Sentiment Analysis

### 5.5.1 Feedforward and LSTM Model with Sentiment Analysis – Bitcoin

The sentiment feature is introduced to the Feedforward and the Recurrent NN model trained to predict the range of Bitcoin price. Figure 26 shows the neural model performance on Bitcoin price prediction for a period of 120 days. Both the FF and LSTM model predictions incorporate the SA feature. The neural network models are compared to the simple Buy-and-Hold strategy.

Superior performance is noted again, by the FF model, with 24,2 % return on investment. The LSTM trading model strategy earns negative -2,4 % ROI. In both cases, algorithmic trading shows better result compared to the Buy-and-Hold scenario (-42,2 % ROI).

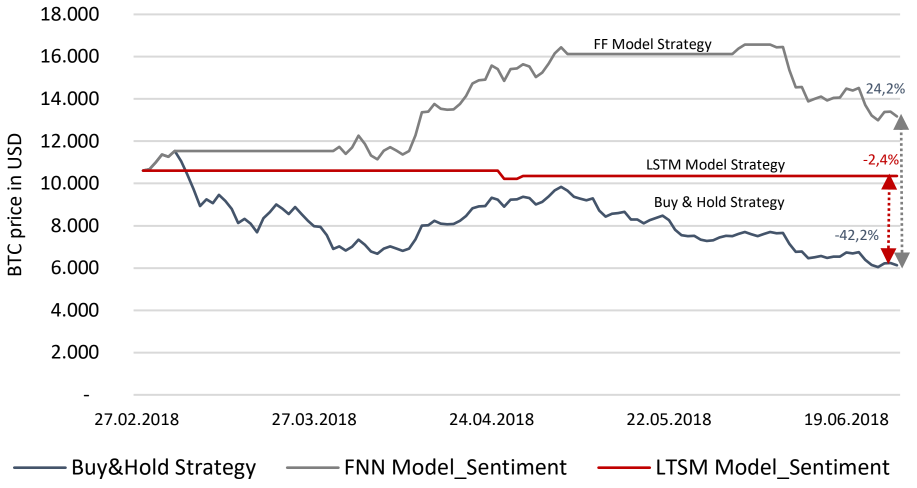
However, by adding the sentiment feature to the predictive model we have decreased the achieved return on investments, compared to the neural network models based only on quantitative and crypto-specific data.

*Table 16: Bitcoin Strategy Results Overview*

| | Buy & Hold Strategy | FF Model | FF Sentiment Model | LSTM Model | LSTM Sentiment Model |
|---|---|---|---|---|---|
| BTC Return on Investment (%) | -42,2 | 30,4 | 24,2 | 6,2 | -2,4 |

*Source: Own work.*

*Figure 26: Bitcoin Predictive Model Results with Sentiment Feature*



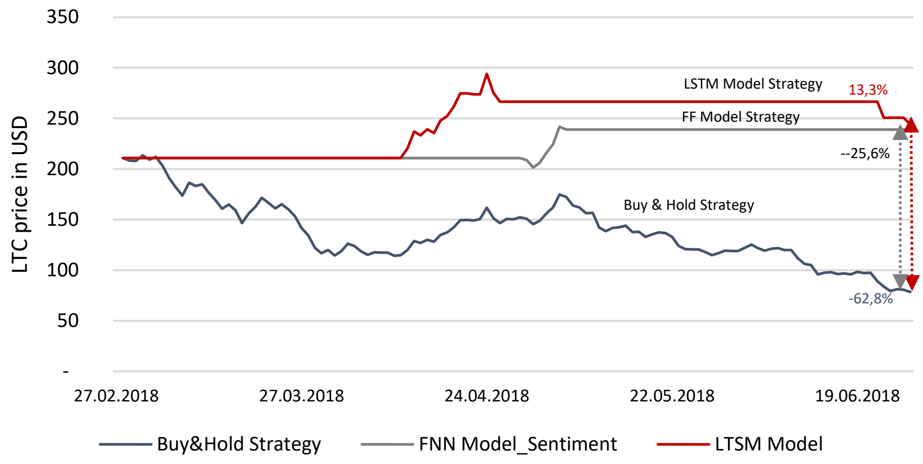

*Source: Own work.*

### 5.5.2 Feedforward and LSTM Model with Sentiment Analysis – Litecoin

Figure 27 shows the neural model performance for Litecoin price prediction with the FF (grey trendline) and LSTM (red trendline) models including the sentiment analysis feature. Again, the models are compared with the Buy-and-Hold scenario.

The Feedforward model performs better compared to the LSTM model, 13,3 % vs. -25,6 % returns on average. Similarly, algorithmic trading performs better compared to the Buy-and-Hold strategy (-62,8 %). However, by adding the sentiment feature to the predictive model we have again decreased the achieved return on investment, compared to the neural network models based only on quantitative and crypto specific data.

*Figure 27: Litecoin Predictive Model Results with Sentiment Feature*




*Source: Own work.*

*Table 17: Litecoin Strategy Results Overview*

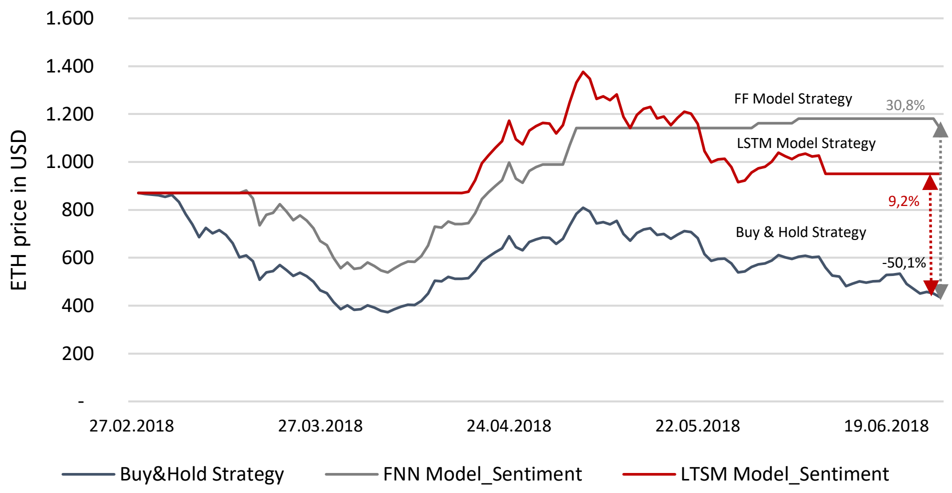
| | | Buy & Hold Strategy | FF Model | FF Sentiment Model | LSTM Model | LSTM Sentiment Model |
|---|---|---|---|---|---|---|
| LTC | Return on Investment (%) | -62,8 | 36,2 | 13,3 | 15,7 | -25,6 |

*Source: Own work.*

### 5.5.3 Feedforward and LSTM Model with Sentiment Analysis – Ether

Figure 28 presents the strategy performance on FF, LSTM and the Buy-and-Hold strategy.

*Figure 28: Ether Predictive Model Results with Sentiment Feature*




*Source: Own work.*

Again, it can be observed the superior performance of the Feedforward neural network model. The Feedforward model brings returns of 30,8 % on average. The LSTM model brings 9,2 % average returns. However, by adding the sentiment feature to the predictive model we have achieved lower returns on investment, compared to the neural network models based only on quantitative and crypto specific data (Table 18).

*Table 18: Ether Strategy Results Overview*

| | | Buy & Hold Strategy | FF Model | FF Sentiment Model | LSTM Model | LSTM Sentiment Model |
|---|---|---|---|---|---|---|
| ETH | Return on Investment (%) | -50,1 | 73,4 | 30,8 | 17,8 | 9,2 |

*Source: Own work.*

### 5.6 Predictive Model Accuracy

The resulting model accuracy of the sentiment neural network models is presented in Table 19. Predictive accuracy is presented for all three crypto assets: Bitcoin, Litecoin and Ether.

*Table 19: Summarized Result Accuracy Overview – BTC, LTC & ETH price prediction.*

| **Bitcoin** | **FF Sentiment Model** | **LSTM Sentiment Model** |
|---|---|---|
| Correct predictions | | |
| Actual Price = Predicted Price | 51 | 43 |
| Incorrect predictions | | |
| Actual Price > Predicted Price | 27 | 75 |
| Actual Price < Predicted Price | 42 | 2 |
| **Predictive accuracy** | **42,5 %** | **35,8 %** |
| | | |
| **Litecoin** | | |
| Correct predictions | | |
| Actual Price = Predicted Price | 64 | 60 |
| Incorrect predictions | | |
| Actual Price > Predicted Price | 34 | 33 |
| Actual Price < Predicted Price | 22 | 27 |
| **Predictive accuracy** | **53,3 %** | **50,0 %** |
| | | |
| **Ether** | | |
| Correct predictions | | |
| Actual Price = Predicted Price | 54 | 40 |
| Incorrect predictions | | |
| Actual Price > Predicted Price | 46 | 27 |
| Actual Price < Predicted Price | 20 | 53 |
| **Predictive accuracy** | **45,0 %** | **33,3 %** |

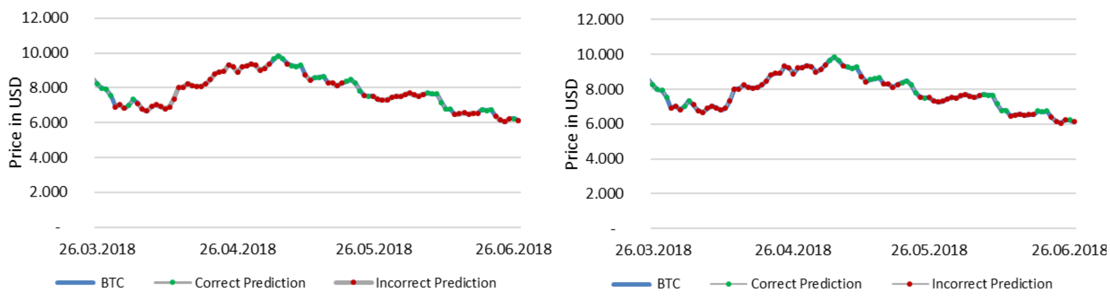
*Source: Own work.*

Simple Feedforward network enhanced with the sentiment shows superior performance over the Recurrent LSTM network model. The weak performance of the LSTM model is resulting from the short data history which is not sufficient for proper training of the LSTM network.

The short data history is a common characteristic for the novelty of all three analyzed cryptocurrencies.

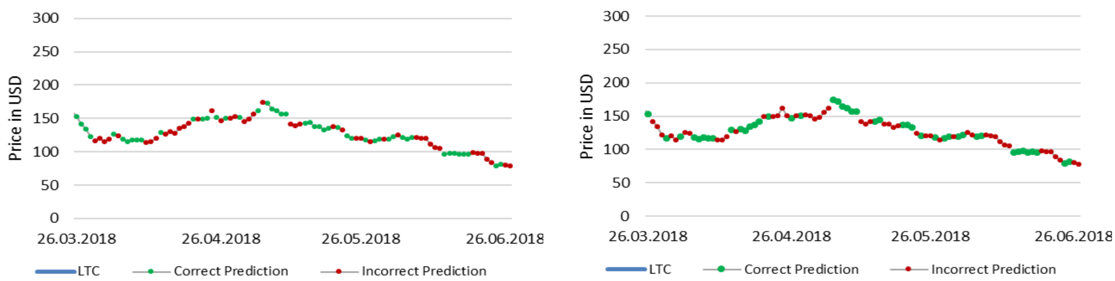
The Feedforward neural network model shows accuracy within the range of 40–53 %. The RNN model has lower accurate predictions, falling within the range of 33–50 %. Figures 29, 30 and 31 display the neural network model accuracy for making Bitcoin, Litecoin and Ether price predictions.

*Figure 29: Bitcoin – FF and LSTM Predictive Model Accuracy*



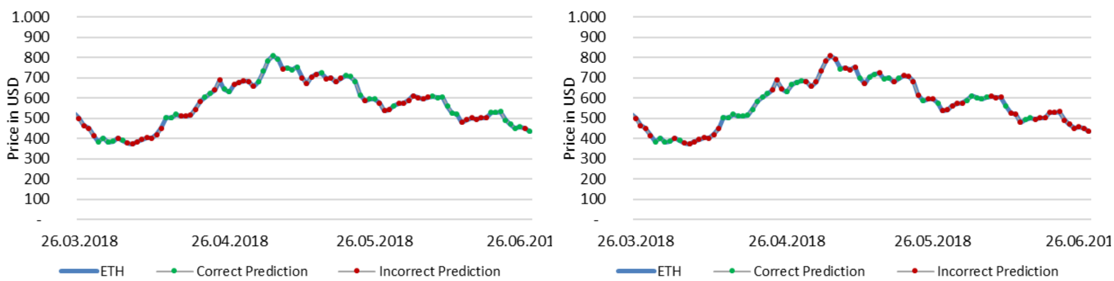

*Source: Own work.*

*Figure 30: Litecoin – FF and LSTM Predictive Model Accuracy*




*Source: Own work.*

*Figure 31: Ether – FF and LSTM Predictive Model Accuracy*




*Source: Own work.*

The Feedforward sentiment model records over-pessimistic predictions for the case of Litecoin and Ether (Actual price > Predicted price). The FF model predicts lower LTC and ETH prices than the ones actually observed. Conversely, the Feedforward Sentiment model forecast of the Bitcoin price proves to be over-optimistic, predicting higher prices than the actual.

The RNN (LSTM) Sentiment model makes over-pessimistic price predictions for Bitcoin and Litecoin, whereas, in the case of Ether, it can be noted predominantly over-optimistic misclassifications (Actual price < Predicted price misclassification).

Cryptocurrency prices are highly related to the analyzed sentiment results. However, sentiment shifts happen with a certain time lag. The results show that price changes are causing changes in investor's sentiment, post-festum. Therefore, the sentiment feature derived from Twitter data set, cannot be used as a predictive variable in foreseeing future cryptocurrency price developments.

# 6 RESULT DISCUSSION

The first research question has been tested with the application of two trading models based on fourteen crypto-specific metrics. The results show a moderate level of predictive accuracy for both neural network models. However, the use of algorithmic trading models improves the cryptocurrency risk-adjusted returns, opposite to the basic Buy-and-Hold strategy.

During the four-month period of the model testing, the general downside trend in the prices of all three analyzed cryptocurrencies can be observed. The highest drop can be noted in the case of Litecoin (-62,8 %), followed by Ether (-50,1 %) and Bitcoin (-42,2 %). Therefore, if the investor took a long position in BTC, LTC or ETH, as of 28$^{th}$ of February, and hold the asset up until 26$^{th}$ of June, their portfolio will have suffered negative returns.

The neural network modeled strategies report positive returns on investment. The highest return is achieved with the FF model strategy, based only on quantitative and crypto specific metrics. Following is the performance of the LSTM model based solely on quantitative and crypto-specific inputs. The superior performance of the Feedforward network has been anticipated, since the outcome of the LSTM model is substantially more limited by the insufficiency of historic data records on cryptocurrencies. The introduction of the sentiment feature reduces the predicted profitability from the algorithmic trading (Table 20).

Similar results can be observed in terms of the predictive model accuracy. The FF model before the Sentiment introduction has the highest predictive accuracy, followed by the LSTM model. In the same way, the introduction of the sentiment feature reduces the accuracy of our model predictions. The comparable outcome can be observed for the case of all three crypto assets.

*Table 20: Neural Network Model Results – Summary*

| | | Buy & Hold Strategy | FNN Model | FNN Sentiment Model | LTSM Model | LTSM Sentiment Model |
|---|---|---|---|---|---|---|
| BTC | Return on Investment (%) | -42,2 | 30,4 | 24,2 | 6,2 | -2,4 |
| LTC | Return on Investment (%) | -62,8 | 36,2 | 13,3 | 15,7 | -25,6 |
| ETH | Return on Investment (%) | -50,1 | 73,4 | 30,8 | 17,8 | 9,2 |
| BTC | Model Accuracy (%) | | 45,8 | 42,5 | 37,5 | 35,8 |
| LTC | Model Accuracy (%) | | 52,5 | 53,3 | 47,5 | 50,0 |
| ETH | Model Accuracy (%) | | 49,2 | 45,0 | 33,3 | 33,3 |

*Source: Own work.*

The second research question tested whether the predictive performance of the FF and LSTM model can be enhanced by introducing the cryptocurrency derived public sentiment feature. On the contrary to the initial expectations, the sentiment feature reduces the predictive power of the Feedforward and LSTM model.

There are several plausible reasons for this outcome. The analyzed sentiment results show that the changes in investment sentiment happen with a certain time lag. The price changes are inducing changes in the public sentiment, post-festum. Hence, once prices of crypto assets increase or decrease by some significant amount, the social media activity is triggered, and strong investor sentiment is in place. Therefore, the sentiment derived from Twitter posts cannot be used as a valuable source for predicting future cryptocurrency prices. Furthermore, the limited access to the overall Twitter data set narrows down the profound sentiment research study based on "tweet" data. The disclosure of the data provider algorithm is rather limited, which raises the question of the selection criteria bias. All things considered, it can be concluded that the introduction of the sentiment analysis study does not contribute towards the explanation of cryptocurrency value.

## 6.1 Feedforward Neural Network Model Strategy – Iterative Analysis

Due to the superior outcome of the Feedforward algorithmic trading model, iterative analysis has been carried out in order to address the performance of the model within different time periods. Cryptocurrency markets are excessively volatile and various external factors can cause instantiations shifts in cryptocurrency prices. Hence, the model assessment has been conducted by adjusting the network training and model testing periods. With the backward adjustment of the neural network training and testing period, we decrease the train time interval and the data available for training of the network, while the testing period remains the same, 120 days (Appendix 5).

The results of the iterative analysis indicate once again, the superior performance of the FF algorithm over the basic Buy-and-Hold strategy. A similar outcome can be noted for all three cryptocurrencies: Bitcoin, Litecoin and Ether. However, by adjusting the period of training of the FF neural network, changes in the predictive accuracy of the model can be observed. The FF model predictive accuracy has a range of 23 % and 53 % for the referred crypto assets. It can be observed that the algorithmic FF model makes much more accurate predictions during the presence of a steady upward or downward price trend, while it is significantly less efficient during the time of big regulation and social media changes.

It can be noted that, by going back in time, there is a considerable decline in the predictive model accuracy. In specific, the period t-2 (18.03.2018 – 15.07.2018) shows the lowest model accuracy and return on investments. The lowest performance of the model during this period can be explained with several cryptocurrency market events and regulation changes at that moment of time. For instance, in March 2018 social media ban of cryptocurrencies ads has been introduced (Facebook, Google, and Twitter). This step had significant consequences on the crypto markets since the majority of the crypto assets seek and gain publicity within the social media platforms. In addition, in July 2018, South Korea has issued a regulation which categorizes and regulates cryptocurrency exchanges in the same manner as the other financial institutions and banks. In the same period, the Financial Stability Board announced its framework for monitoring cryptocurrency markets. All these sudden changes and unforeseen market events have challenged the predictive power of the FF model. Following, the developed FF model has not been able to correctly predict and account for such unforeseen cryptocurrency market events.

*Table 21: Feedforward Neural Network Model – Iterative Analysis train vs. test data*

| | | FF model Tested period | FF model Period t | FF model Period t-1 | FF model Period t-2 | FF model Period t-3 | FF model Period t-4 | FF model Period t-5 |
|---|---|---|---|---|---|---|---|---|
| BTC | Train data (%) | 95,7 | 96,1 | 95,9 | 95,7 | 95,5 | 95,3 | 95,1 |
| BTC | Test data (%) | 4,3 | 3,9 | 4,1 | 4,3 | 4,5 | 4,7 | 4,9 |
| LTC | Train data (%) | 94,4 | 95,0 | 94,8 | 94,5 | 94,2 | 93,8 | 93,4 |
| LTC | Test data (%) | 5,6 | 5,0 | 5,2 | 5,5 | 5,8 | 6,2 | 6,6 |
| ETH | Train data (%) | 91,2 | 92,5 | 92,0 | 91,3 | 90,4 | 89,4 | 88,2 |
| ETH | Train data (%) | 8,8 | 7,5 | 8,0 | 8,7 | 9,6 | 10,6 | 11,8 |

*Source: Own work.*

The rationality for this outcome is that the data applied for the training of the FF network is insufficient for proper price forecasting. In fact, there is not enough data of comparable historic events in the employed cryptocurrency price database, so that the algorithm can become proficient in forecasting such reactional price changes. Due to this shortcoming of the cryptocurrency dataset, the network cannot be trained properly for predicting market movements of this kind.

In addition, the embodied volatility of the cryptocurrency market and its sensitivity towards changes in regulation and media influences can be plausibly better forecasted with a proper, adjusted SA tool. As a room for further research, cryptocurrency-specific SA tool can be developed in order to track shifts in the market sentiment and the media on an overall and larger media scale. Further on, by adding special focus on the publications and announcements of the governments, regulators and various relevant legal institutions around the globe, such price shifts in the cryptocurrency prices can be plausibly predicted with a higher degree of model accuracy.

*Table 22: Feedforward Neural Network Model – Iterative Analysis results*

| | | FF model Tested period | FF model Period *t* | FF model Period *t-1* | FF model Period *t-2* | FF model Period *t-3* | FF model Period *t-4* | FF model Period *t-5* |
|---|---|---|---|---|---|---|---|---|
| BTC | ROI (%) | 30,4 | 2,5 | 0,0 | -30,3 | -17,8 | 72,5 | 77,3 |
| LTC | ROI (%) | 36,2 | 33,0 | 0,0 | -34,6 | 60,6 | 26,9 | 182,8 |
| ETH | ROI (%) | 73,4 | 1,2 | -37,6 | 1,7 | 158,6 | 32,6 | 149,4 |
| BTC | Model Accuracy (%) | 45,8 | 53,3 | 24,2 | 22,5 | 34,2 | 32,5 | 38,3 |
| LTC | Model Accuracy (%) | 52,5 | 39,2 | 39,2 | 22,5 | 39,2 | 30,8 | 33,3 |
| ETH | Model Accuracy (%) | 49,2 | 36,7 | 42,5 | 39,2 | 46,7 | 35,0 | 31,7 |

*Source: Own work.*

## CONCLUSION

With this research study, the author aimed to develop an adequate predictive cryptocurrency model, accounting for volatility as a notorious feature of the crypto assets. The models developed within this research ground on the Feedforward and Recurrent (LSTM) neural network algorithm, which combine data on crypto-specific financial and technical metrics. In addition to these quantitative metrics, the sentiment study is performed by using Twitter as the predominant social media data source. The sentiment study was performed with the application of Valence Aware Dictionary for Sentiment Reasoning (VADER) model which evaluates positivity/negativity of word units contained. “Tweets” served as an input data source for the VADER sentiment model. This comparative approach was applied in order to investigate the predictive power of historical prices, crypto specific metrics and public sentiment for forecasting future movements on the cryptocurrency market. Conclusively, the algorithmically-defined trading strategy performance is compared with the basic Buy-and-Hold strategy.

The research study results revealed moderate model predictive accuracy and enhanced investment returns. As discussed above, the moderate level of predictive accuracy of the algorithmic models can result from several reasons. The small input database is the initial model burden. Due to the short cryptocurrency history, the size of the obtained data is not sufficient for proper training of the neural network. The short history of crypto assets reduces the predictive power of the neural network models. In addition, algorithmic models are challenged by the outstanding noise and complex dimensionality of the financial data. The algorithm does not take into account market corrections, human sentiment and unforeseen events which can cause shifts in market prices.

Moreover, the sentiment analysis study results show that, in the case of cryptocurrencies, price changes are causing shifts in investor's sentiment, and not the vice versa. This challenges the application of the sentiment analysis study in foreseeing future prices of the cryptocurrencies. In addition, the sentiment study presented in this research is based on VADER sentiment analysis tool which presents a pre-defined lexicon-based algorithm. For the purpose of achieving a better understanding of the investor's sentiment, future work should consider the analysis of the sentiment based on a tailored, cryptocurrency-adjusted lexicon tool. Rather than applying a pre-defined general lexicon tool, a new cryptocurrency tailored tool can plausibly improve the predictive power of the investor's sentiment for forecasting future movements on the cryptocurrency market.

In conclusion, the neural network algorithm, trained on crypto specific and sentiment data is a moderate solution for providing accurate predictions of the future cryptocurrency price developments. However, the algorithmic trading strategy succeeds in generating significantly increased returns on investment, as opposite to the simple Buy-and-Hold strategy. The moderate predictive power of the model can be explained with the highly volatile and inconsistent nature of crypto assets. Cryptocurrency prices are highly reliant on speculation, media hype and other external influencers. The algorithm fails to account for the relevant market corrections, investment speculation, and further unforeseen events. It can be expected that once the cryptocurrency playfield is cleared from the involved media hype and market speculation, a room for application of neural network valuation algorithms is going to be created, resulting with more accurate and reliable cryptocurrency value predictions.

## REFERENCE LIST


1. Agaian, S. & Kolm, P. (2017, August 23). Financial Sentiment Analysis Using Machine Learning Techniques. *International Journal of Investment Management and Financial Innovations, 3*(1), 1–9.
2. Bell, S. (2001, March 1). The Role of the State and the Hierarchy of Money. *Cambridge Journal of Economics, 25*(2), 149–163.
3. Bernal, A., Fok, S. & Pidaparthi, R. (2012, December 14). *Financial Market Time Series Prediction with Recurrent Neural Networks*. Retrieved on March 22, 2018, from https://ieeexplore.ieee.org/document/836215/references
4. Bitinfocharts. (2018). *BitInfoCharts*. Retrieved Feb-July 2018, from https://bitinfocharts.com/
5. Bouoiyour, J., Tiwari, A.K., Selmi, R., & Olayeni, A.K. (2016, May 16). What drives Bitcoin prices? *Economics Bulletin*, *36*(2), 843–850.

6. Burniske, C. & White, A. (2017). Bitcoin: Ringing the Bell for a New Asset Class. *ARK Invest Research, 1*(1), 3–24.
7. Buterin, V. (2014). *Ethereum White Paper: A Next Generation Smart Contract & Decentralized Application Platform.* Retrieved on February 5, 2018, from https://cryptorating.eu/whitepapers/Ethereum/Ethereum_white_paper.pdf
8. Button, S. (2017, December 1). Cryptocurrency and Blockchains in Emerging Economies. *Software Quality Professional, 20*(3), 39–46.
9. Caginlap, C. & Caginlap, G. (2018, February 6). Valuation, Liquidity Price and Stability of Cryptocurrencies. *Proceeding of the National Academy of Sciences of the USA 115*(6), 1131–1134.
10. Caro, A. (2017). Blockchain: *The Beginners Guide to Understanding the Technology Behind Bitcoin & Cryptocurrency* (1st ed.). New York: CreateSpace Independent Publishing Platform, 10–45.
11. Cermak, V. (2017, May 13). Can Bitcoin Become a Viable Alternative to Fiat Currencies? An Empirical Analysis of Bitcoin's Volatility Based on a GARCH Model. *Journal of Empirical Financial Management, 10*(17), 16–42.
12. Chaturvedia, I., Cambria, E., Welschb, R.E. & Herrera, F. (2018). Distinguishing Between Facts and Opinions for Sentiment Analysis: Survey and Challenges. *Information Fusion, 44,* 65–77.
13. Cheah, T. & Fry, J. (2015, February 18). Speculative Bubbles in Bitcoin Markets? An Empirical Investigation into the Fundamental Value of Bitcoin. *Economics Letters*, *130,* 32–36.
14. Chen, S., Chen, C.Y., Härdle, W. K. & Lee, T. (2016, August 30). A First Econometric Analysis of the CRIX Family. *SFB 649 Discussion Papers, 1*(31), 1–7.
15. Chiou, W. J. P., Lee, A. C. & Lee, C. F. (2010). Stock Return, Risk, and Legal Environment Around the World. *International Review of Economics & Finance, 19*(1), 95–105.
16. Chishti, S. & Barberis, J. (2016, March 10). *The FINTECH Book: The Financial Technology Handbook for Investors, Entrepreneurs and Visionaries*. London: John Wiley & Sons Inc., 10–28.
17. Chu, J., Chan, S., Nadarajah, S. & Osterrieder, J. (2017, May 31). A Statistical Analysis of Cryptocurrencies. *Journal of Risk Financial Management, 10*(2), 12.
18. Chuen, D. K. L., Guo, L. & Wang, Y. (2017, June 30). Cryptocurrency: A New Investment Opportunity? *The Journal of Alternative Investments, 20*(3) 16–40.

19. CoinMarketCap. (2018). Cryptocurrency Market Capitalizations. Retrieved Jan-Aug 2018, from https://CoinMarketCap.com/
20. Colianni, S., Rosales, S. & Signorotti, M. (2015, December 12). *Algorithmic Trading of Cryptocurrency Based on Twitter Sentiment.* Retrieved on April 13, 2018, from https://www.semanticscholar.org/paper/Algorithmic-Trading-of-Cryptocurrency-Based-on-Colianni-Rosales/9b838a3177523b8179511283b9489caa0f69910d?tab=citations
21. CRIX. (2018). *CRIX – Crypto Index*. Retrieved on June 27, 2018, from http://crix.hu-berlin.de/data/crix.json
22. Crosby, M., Nachiappan, Pattanyak, P., Verma, S. & Kalyanaraman, V. (2016). BlockChain Technology: Beyond Bitcoin. *Applied Innovation Review, 2,* 8–18.
23. Culpan, T. (2017, June 14). *Getting High on Cryptocurrencies.* Retrieved on May 15, 2018, from https://www.bloomberg.com/gadfly/articles/2017-06-14/this-cryptocurrency-boom-is-getting-a-little-absurd
24. Cybenko, G. (1989). Approximation by Superpositions of a Sigmoidal Function. *Springer: Mathematics of Control, Signals and Systems, 2*(4), 303–314.
25. D'Alfonso, A., Langer, P. & Vandelis, Z. (2016, October 17). *The Future of Cryptocurrencies.* Toronto: The Economist, Ryerson University, 4–25.
26. Dornbusch, R (1985). Purchasing Power Parity. *National Bureau of Economic Research, 1591,* 5–35.
27. Filippi, D. P. (2014). *Bitcoin: A Regulatory Nightmare to Libertarian Dream.* Retrieved on February 12, 2018, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2468695
28. Gantori, S, Danovan, P., Ganesh, K., DeMichiel, M., Dannean, K., Trussardi, F. & Klien, M. (2017, October 12). Cryptocurrencies – Beneath the bubble. *UBS, Chief Investment Office Americas, Wealth Management,* 2–18.
29. Giles, C. L., Lawrence, S. & Tsoi, A. C. (1997). Rule Inference for Financial Prediction using Recurrent Neural Networks. *Proceedings of IEEE/IAFE Conference on Computational Intelligence for Financial Engineering (CIFEr),* 253–259.
30. GitHub. (2018). *GitHub Software Development Platform*. Retrieved on July 1, 2018, from https://github.com/cjhutto/vaderSentiment
31. Gomber, P. & Haferkorn, M. (2015). High Frequency Trading. *Encyclopaedia of Information Science and Technology,* (3rd ed.), 1–9.

32. Gourieroux, C. & Hencic, A. (2015, May 15*)*. Non-causal Autoregressive Model in Application to Bitcoin/USD Exchange Rates. *Economics of Risk, 583(1),* 17–40.
33. Hayes, S. A. (2017). Cryptocurrency Value Formation: An Empirical Study Leading to a Cost of Production Model for Valuing Bitcoin. *Telematics and Informatics, 34*(7), 1308–1321.
34. Hebb, D. (1949). *Emotional Disturbances in The Organization of Behaviour.* New York: Wiley & Sons Inc., 235–274.
35. Hecht-Nielsen, R. (1992). Neural Networks for Perception. *Computation, Learning and Architectures, 2*(1), 65–93.
36. Houben, R. & Snyers, A. (2018). *Cryptocurrencies and Blockchain. Legal Context and Implications for Financial Crime, Money Laundering and Tax Evasion*. Brussels: Policy Department for Economic, Scientific and Quality of Life Policies, PE 619.024, 15–85.
37. Howell, T. S., Niessner, M. & Yermack, D. (2018, August 1). Initial Coin Offerings: Financing Growth with Cryptocurrency Token Sales. *The National Bureau of Economic Research, 1*(1), 1–33.
38. Hsieh, T. J. & Yeh, W. C. (2011). Forecasting Stock Markets Using Wavelet Transforms and Recurrent Neural Networks: An Integrated System Based on Artificial Bee Colony Algorithm. *Applied Soft Computing, 11*(2), 2510–2525.
39. Hubrich, S (2017, October 28). *Know When to Hodl Them, Know When to Fodl Them: An Investigation of Factor Based Investing in the Cryptocurrency Space*. Retrieved on April 1, 2018, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3055498
40. Hutto, C. J. & Gilbert, E. E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Proceedings of the 8th International Conference on Weblogs and Social Media (ICWSM–14),* 216–224.
41. Inardi, P., Saa, O., & Popov, S. (2017, November 12). *STEX's valuation analysis, version 0.0. Smart Token Exchange.* Retrieved on April 15, 2018, from https://stex.exchange/media/pdf/STeX_Valuation_by_Paulo_Finardi__Olivia_Saa__Serguei_Popov.pdf
42. International Monetary Fund. (2016). Virtual Currencies and Beyond: Initial Considerations. *International Monetary Fund (IMF), 3*(16), 6–35.
43. Katsiampa, P. (2017, June 20). Volatility Estimation for Bitcoin: A Comparison of GARCH models. *Economics Letters, 158*(3–6), 1-2.
44. Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques. *Springer Nature, 3(*26), 3-23.

45. Krafft, P. M., Penna, D. N. & Pentland, A. (2018, April 12). An Experimental Study of Cryptocurrency Market Dynamics. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18, ACM Press,* 1–11.
46. Krenker, A., Bešter, J. & Kos, A (2011). *Artificial Neural Networks – Methodological Advances and Biomedical Applications.* Retrieved on May 30, 2018, from http://www.intechopen.com/books/artificial-neural-networksmethodological-advances-and-biomedical-applications/introduction-to-the-artificial-neural-networks
47. Li, S., Wunsch, D., O'Hair, E. & Giesselmann, M. (1999). Neural Network for Wind Power with Compressing Function. *Proceedings of IEEE International Conference on Neural Network*, 115–120.
48. Litecoin Project. (2017). *Litecoin - Open source P2P digital currency*. Retrieved March 28, 2018, from https://litecoin.org/
49. Luria, G. & Turner, A. (2013). *Bitcoin: Intrinsic Value as Conduit for Disruptive Payment Network Technology.* Retrieved on January 1, 2018, from https://bravenewcoin.com/insights/bitcoin--intrinsic-value-as-conduit-for-disruptive-payment-network-technology
50. Madan, I. Saluja, S. & Zhao, A. (2014). *Automated Bitcoin Trading via Machine Learning Algorithms.* California: Department of Computer Science, Stanford University, 1–5.
51. Mandeng, J. O. (2018, February 12). *Cryptocurrencies, Monetary Stability and Regulation: Germany's Nineteenth Century Private Banks of Issue*. London: The London School of Economics and Political Science, 2–18.
52. Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance. 7*(1), 77–91.
53. McCulloch, W.S. & Pitts, W. (1990). A Logical Calculus of the Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biology, 52*(1–2), 99–115.
54. Medsker, L. R., Jain, L. C. (1999). *Recurrent Neural Networks Design and Applications (1st ed.).* Texas: The CRC Press, Inc. Boca Raton, 11–20.
55. Melville, P., Gryc, W. & Lawrence, R.D. (2009). Sentiment Analysis of Blogs by Combining Lexical Knowledge with Text Classification. *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '09,* 1275–1284.
56. Murphy, J. J. (1986). *Technical Analysis of the Futures Markets: A Comprehensive Guide to Trading Methods and Applications*. New York: Institute of Finance, 2–85.

57. Naik, V., Devarajan, M., Nowobilski, A., Page, S. & Pedersen, N. (2016). *Factor Investing and Asset Allocation – A Business Cycle Perspective*. Retrieved on May 1, 2018, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2930857
58. Nakamoto, S. (2008). *Bitcoin: A Peer-to-Peer Electronic Cash System.* Retrieved on February 1, 2018, from https://bitcoin.org/bitcoin.pdf
59. Nakano, M., Takahashi, A. & S. (March 2018). *Bitcoin Technical Trading with Artificial Neural Network.* Retrieved on March 1, 2018, from http://www.cirje.e.u-tokyo.ac.jp/research/03research02dp.html
60. OECD. (2012). *Purchasing Power Parities (PPP)*. Retrieved on September 23, 2018, from: https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm
61. Oliveira, F. A., Nobre, C. N., & Zárate, L. E. (2013). Applying Artificial Neural Networks to Prediction of Stock Price and Improvement of the Directional Prediction Index. *Expert Systems Applications*, *40*(18), 7596–7606.
62. Pang, B. & Lee, L. (2004). *A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts*. New York: Department of Computer Science, Cornell University Ithaca, 1–7.
63. Pang, B. & Lee, L. (2008). Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval. *Now Publishers, 2*(1–2), 1–13.
64. Pástor, L. & Stambaugh, R. F. (2003). Liquidity Risk and Expected Stock Returns. *The Journal of Political Economy*, *111*(3), 642–685.
65. Peters, W. G., Panayl, E. & Chapelle, A (2015). Trends in Cryptocurrencies and Blockchain Technologies: A Monetary Theory and Regulation Perspective. *The Journal of Financial Perspectives, 3*(3), 4–41.
66. Saad, E. W., Caudell, T. P. & Wunsch, D. C. (1999). *Predictive Head Tracking for Virtual Reality*. Retrieved on May 30, 2018, from https://www.researchgate.net/profile/Thomas_Caudell/publication/3839113_Predictive_head_tracking_for_virtual_reality/links/00b7d52f26056306fd000000/Predictive-head-tracking-for-virtual-reality.pdf
67. Saidov, U. (2018, April 3). *Cryptocurrencies: The Rise of Decentralized Money.* Retrieved on May 1, 2018, from https://blogs.cfainstitute.org/investor/2018/04/03/cryptocurrencies-the-rise-of-decentralized-money/
68. Schaller, R.R. (1997). Moore's Law: past, present and future. *IEEE Spectrum, 34*(6), 53–59.

69. Schuster, M. & Paliwal, K. K. (1997). Bidirectional Recurrent Neural Networks. IEEE Transactions on Signal Processing. *Institute of Electrical and Electronics Engineers (IEEE), 45*(11), 2673–2681.
70. Quantopian. (2018). *Sentdex Sentiment Analysis Data*. Retrieved on June 14, 2018, from https://www.quantopian.com/data/sentdex/sentiment
71. Shah, D. & Zhang, K. (2014). *Bayesian regression and Bitcoin*. Retrieved on March 23, 2018, from https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=7008053&filter=issueId%20EQ%20%227028426%22&pageNumber=3
72. Stambaugh, R. F., Yu, J., and Yuan, Y (2011). The Short of it: Investor Sentiment and Anomalies. *Journal of Financial Economics, 104*(2), 288–302.
73. Sul, H. K. (2016, June 23). Trading on Twitter: Using Social Media Sentiment to Predict Stock Returns. *Journal of the Decision Sciences Institute, 48*(3), 3–34.
74. Surowiecki, J. (2004). *The Wisdom of The Crowds: Why the Many are Smarter than the Few and how Collective Wisdom Shapes Business, Economies, Societies and Nations.* New York: Doubleday & Co., 193–224.
75. Swamy, K., Pashley, M. & Gilbert, E. (1997). Neural Network Applications in Finance. *Proceedings from the 4th Annual Meeting of The American Society of Business and Behavioral Sciences Conference*, 1–10.
76. Tapscott D., & Tapscott A. (2016). *Blockchain Revolution: How the Technology Behind Bitcoin is Changing Money, Business, and the World*. London: Informa UK Limited, 28–51.
77. Tetlock, P. C. (2007). Giving content To Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance, 62*(3), 1139–1168.
78. Thomson Reuters. (2018). *News Analytics and MarketPsych Indices*. Retrieved on June 14, 2018, from https://financial.thomsonreuters.com/en/products/data-analytics/financial-news-feed/world-news-analysis.html
79. Trimborn, S. & Härdle, W. K (2016). CRIX or Evaluating Blockchain Based Currencies. *SFB 649 Discussion Paper 2015-048*, 2–18.
80. TweetScraper (2018). *TweetScraper.* Retrieved April-May 2018, from https://github.com/jonbakerfish/TweetScraper
81. Twitter (2018). *Twitter Advanced Search.* Retrieved May-July 2018, from https://twitter.com/search-advanced?lang=en

82. Vukotic, V., Claveau, V. & Raymond, C. (2015). IRISA at DeFT 2015: Supervised and Unsupervised Methods in Sentiment Analysis. *Proceedings from 22ème Traitement Automatique des Langues Naturelles,* 2–12.
83. Wilder, J. W. (1978). *New Concepts in Technical Trading Systems*. North California: Hunter Publishin Company, 7–22.
*84.* Wilson, D. & Ateniese, G. (2015). From Pretty Good to Great: Enhancing PGP Using Bitcoin and the Blockchain. Network and Systyem Security. *Lecture Notes in Computer Science, 9408,* 368–375.
85. Zetzsche, D. A., Buckley, R. P., Armer D. W. & Föhr, L. (2018). *The ICO Gold Rush: It's a Scam, It's a Bubble, It's a Super Challenge for Regulators.* Retrieved on March 1, 2018, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3116193

# APPENDICES

## Appendix 1: Povzetek (Summary in Slovene language)

Namen te raziskovalne naloge je bil razvoj primernega modela za napovedovanje vrednosti kriptovalut. Modeli, razviti v okviru te raziskave temeljijo na dveh nevronskih mrežah: Feedforward in Recurrent (LSTM) neural network, ki združujeta podatke o posebnih finančnih in tehničnih meritvah. Poleg teh kvantitativnih meritev, je bila s pomočjo Twitterja izvedena mnenjska raziskava. Za pozitivno oziroma negativno oceno izbranih besednih zvez je bil uporabljen VADER model. Podatkovni vir za VADER so bili tviti.

Namen tega pristopa je preizkusiti napovedno moč zgodovinskih cen, specifičnih podatkih kriptovalut in javnega mnenja za napovedovanje prihodnjih gibanj na trgu kriptovalut. Na koncu se algoritemsko definirana strategija trgovanja primerja z osnovno Buy-and-Hold strategijo.

Rezultati raziskave so pokazali zmerno natančnost napovedi in večjo donosnost naložb. Zmerna stopnja napovedne natančnosti algoritmičnega modela je posledica več razlogov. Nezadostna velikost podatkov je začetna omejitev modela. Zaradi kratke zgodovine kriptovalut, količina pridobljenih podatkov ne zadošča za ustrezno učenje nevronske mreže. Kratka zgodovina kriptovalut zmanjšuje napovedno moč modela nevronse mreže. Poleg tega je zmogljivost algoritmičnega modela omejena s kompleksno dimenzionalnostjo finančnih podatkov. Algoritem trgovanja kriptovalut prav tako ne more upoštevati popravkov na trgu in učinka nepredvidenih dogodkov.

Poleg tega rezultati mnenjske raziskave kažejo, da v primeru kriptovalut spremembe cen spreminjajo razpoloženje vlagateljev, in ne obratno. To omenjuje uporabo mnenjske raziskave pri predvidevanju cen kriptovalut v prihodnsti. Da bi dosegli boljše razumevanje razpoloženja investitorja, bi bilo treba pri prihodnjem delu upoštevati mnenjsko raziskavo, ki temelji na prilagojenem leksikonu za kriptovalute. Takšen pristop lahko izboljša napovedno moč modela pri napovedovanju prihodnjih gibanj na trgu kriptovalute.

Algoritmi nevronskih omrežij, ki temeljijo na specifičnih podatkih kriptovalut in podatkih o razpoloženju, s sedanjim modelom in omejitvami podatkov, niso optimalna rešitev za zagotavljanje natančnih napovedi prihodnjih gibanj cen kriptovalut. Vendar pa strategija algoritemskega trgovanja uspe ustvariti precejšnje pozitivne naložbene donose, za razliko od strategije Buy-and-Hold. Zmerno napovedno moč modela lahko razložimo z zelo nestanovitno in nedosledno naravo kriptovalut. Cene kriptovalut so zelo odvisne od špekulacij, medijskim poročanjem in drugih zunanjih vplivov. Algoritem trgovanja ne upošteva ustreznih popravkov trga, investicijskih špekulacij in drugih nepredvidenih dogodkov. Pričakovati je, da se bo priložnost za uporabo nevronskih mrež pokazala, ko se bosta medijska pozornost in špekulativnost zmanjšala, kar bo imelo za posledico natančnejše in zanesljivejše napovedi vrednosti kriptovalut.

## Appendix 2: Feedforward Neural Network Model

*Figure 1: Feedforward Neural Network Model Structure*

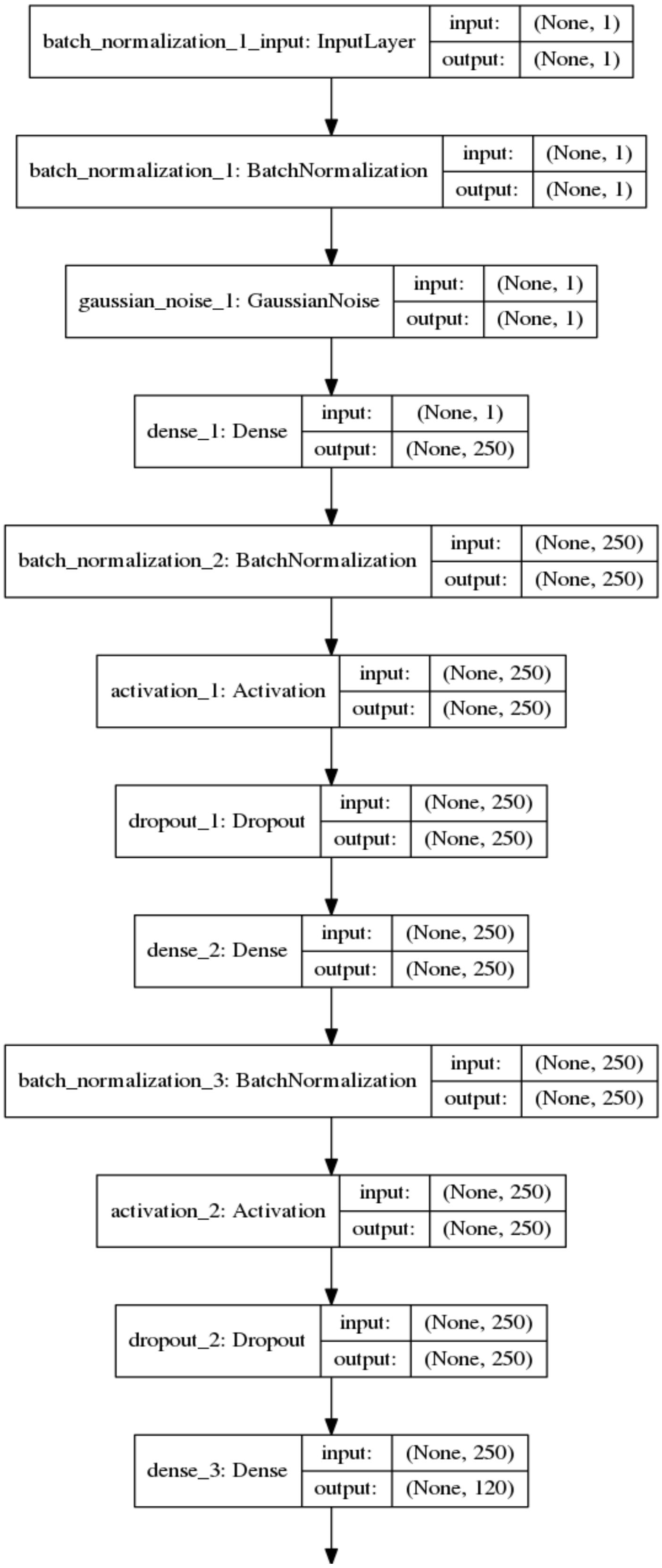
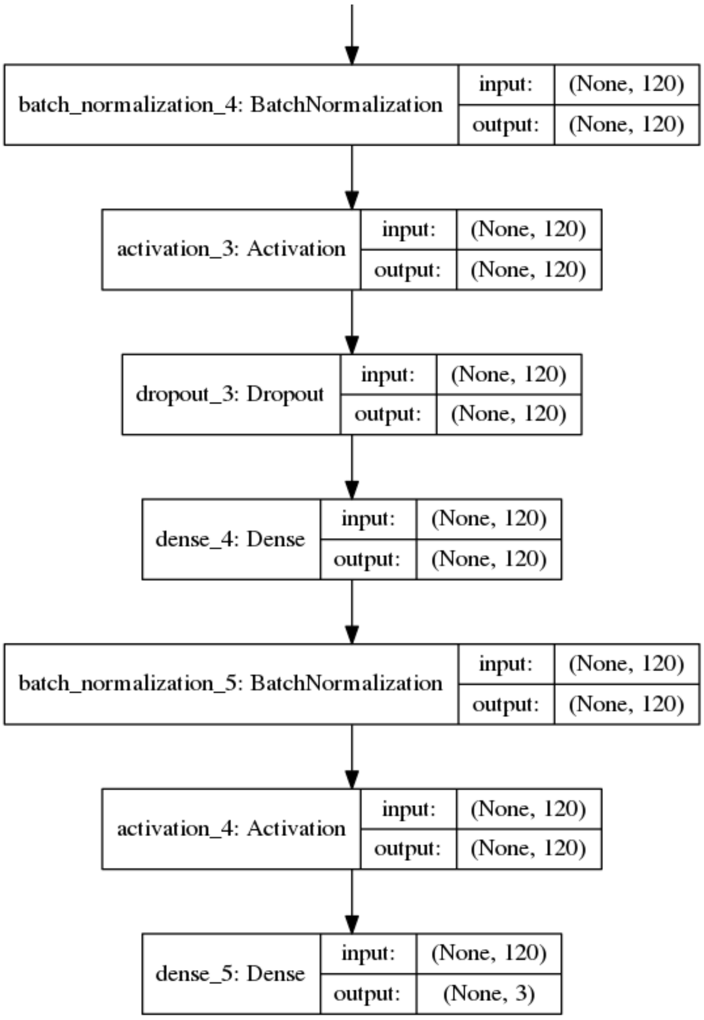



*Figure continues*

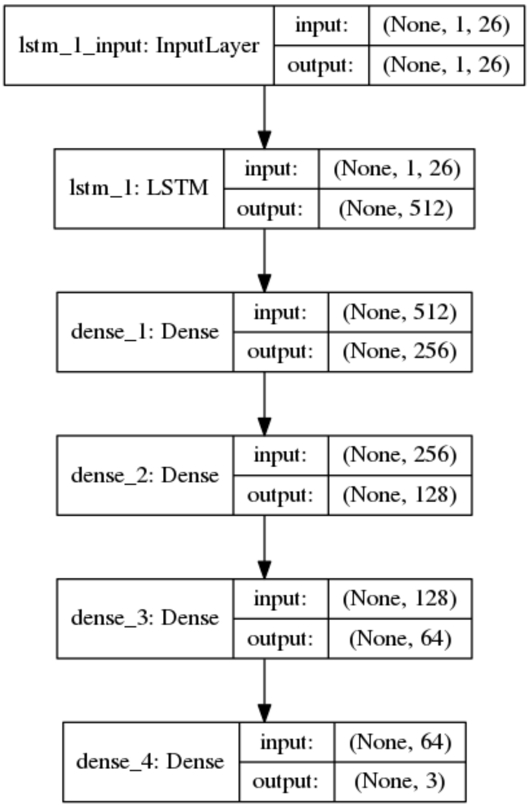
*Figure 1: Feedforward Neural Network Model Structure (continues)*




*Source: Own work.*

## Appendix 3: Recurrent Neural Network Model (LSTM)

*Figure 2: Recurrent Neural Network (LSTM) Model Structure*




*Source: Own work.*

## Appendix 4: Sentiment Classification – VADER Analysis results

*Table 1: Positive Tweet Sentiment Classification – Correct Sentiment Classification*

| | Tweet | VADER Score | Sentiment |
|---|---|---|---|
| 1 | In God we Trust → In the Government we Trust → In Math we Trust. #Crypto #Cyptocurrencies #Blockchain<br>3:45 AM - 18 Oct 2017<br>71 Retweets 82 Likes | Compound Score: **0.900**<br>*{neg: 0.0, neu: 0.52, pos: 0.48}* | Positive |
| 2 | The more governments fight against #Bitcoin the stronger it becomes, the more valuable it becomes and the more clear the reason for its creation becomes... #BuyAndHODL<br>8:19 AM - 6 Feb 2018<br>10 Retweets 33 Likes | Compound Score: **0.797**<br>*{neg: 0.076, neu: 0.634, pos: 0.29}* | Positive |
| 3 | Your business can benefit from #bitcoin! Here are 5 reasons to accept bitcoin at your small business - bit.ly/1qXuTRC #SMB #payments<br>12:41 AM - 22 Jul 2014<br>3 Retweets 3 Likes | Compound Score: **0.708**<br>*{neg: 0.0, neu: 0.753, pos: 0.247}* | Positive |
| 4 | Litecoin is one of the most popular and credible crypto currencies with huge usability potential #ltc #LTC #PayWithLitecoin #crypto<br>4:20 AM - 8 Apr 2018<br>6 Retweets 20 Likes | Compound Score: **0.659**<br>*{neg: 0.0, neu: 0.769, pos: 0.231}* | Positive |
| 5 | "Asking a CEO of a bank what they think of #Bitcoin is like asking the CEO of a taxi company what they think of Uber."<br>Still like this tweet a lot.<br>1:58 AM - 23 Apr 2018<br>56 Retweets 164 Likes | Compound Score: **0.612**<br>*{neg: 0.0, neu: 0.833, pos: 0.167}* | Positive |

*Source: Own work.*

*Table 2: Negative Tweet Sentiment Classification – Correct Sentiment Classification*

| | Tweet | VADER Score | Sentiment |
|---|---|---|---|
| 1 | "Bitcoin and other crypto-currencies are a way for bad dictators or criminals to bypass sanctions.." Bill Browder at the #WEF18 in #Davos cnbc.com/2018/01/23/cri ... He said that #Bitcoin's usefulness for criminal activity will be the death of it and other crypto-currencies.<br>5:32 AM - 28 Jan 2018<br>7 Retweets 12 Likes | Compound Score: **-0.923**<br>*{neg: 0.265, neu: 0.695, pos: 0.04}* | Negative |
| 2 | The Dark Web's Devil- How #Bitcoin fuels Terrorism & Child trafficking sayingtruth.com/bitcoins-stay- ... #TruthMustBeSaid<br>2:13 AM - 11 Jun 2017<br>24 Retweets 13 Likes | Compound Score: **-0.875**<br>*{neg: 0.429, neu: 0.571, pos: 0.0}* | Negative |
| 3 | What a disaster: @coinbase processes #bitcoin transaction 5 days late at wrong price... before 30% drop! goo.gl/Gslh38<br>4:19 PM - 18 Dec 2013<br>9 Retweets 9 Likes | Compound Score: **-0.862**<br>*{neg: 0.425, neu: 0.575, pos: 0.0}* | Negative |
| 4 | At a basic level, the core developers do not understand what Bitcoin is or why it even exists. They simply don't get it.<br>This is why the core developers failed in 2017... and why they will continue to fail in 2018, 2019, 2020, etc.<br>#BitcoinCash is #Bitcoin<br>11:51 AM - 14 Jan 2018<br>44 Retweets 116 Likes | Compound Score: **-0.778**<br>*{neg: 0.131, neu: 0.869, pos: 0.0}* | Negative |
| 5 | 'You will lose everything': #Bitcoin worse than casino gambling – Russian economy minister on.rt.com/8q18<br>8:23 AM - 17 Oct 2017<br>44 Retweets 49 Likes | Compound Score: **-0.700**<br>*{neg: 0.309, neu: 0.691, pos: 0.0}* | Negative |

*Source: Own work.*

*Table 3: Incorrect Tweet Sentiment Classification – Example*

| | Tweet | VADER Score | Sentiment |
|---|---|---|---|
| 1 | 90% of people are destined to fail with crypto, because they are too emotional, and will sell at a loss anyway. I have seen this ALOT recently! Fear Is the Biggest Killer As An Investor! #LitecoinFam #Ltc #Litecoin $LTC #Crypto #Bitcoin #cryptocurrency<br>8:51 PM - 2 Apr 2018<br>10 Retweets 47 Likes | Compound Score: **-0.929**<br>*{neg: 0.27, neu: 0.699, pos: 0.031}* | Negative |
| 2 | They tried to copy it, fork it, ban it, clone it, ridicule it, launch at it, discredit it and kill it and yet #Bitcoin is still sound money.<br>9:44 AM - 16 Sep 2017<br>274 Retweets 592 Likes | Compound Score: **-0.906**<br>*{neg: 0.303, neu: 0.697, pos: 0.0}* | Negative |
| 3 | Nowadays the corrupt sends youth to die in war. In the #Bitcoin times the youth has the power to send the corrupt to die in oblivion.<br>8:54 PM - 17 Nov 2014<br>14 Retweets 6 Likes | Compound Score: -**0.913**<br>*{neg: 0.337, neu: 0.663, pos: 0.0}* | Negative |
| 4 | Don't be worried if people talk bad about #Bitcoin, #CloakCoin, #Ethereum and other #Altcoins.<br>Be worried the minute they start talking good about #cryptocurrencies, because the opportunity is over by then.<br>#BeSmart #DontMissOut #TakeYourChance<br>4:08 AM - 6 May 2018<br>8 Retweets 14 Likes | Compound Score: -**0.296**<br>*{neg: 0.239, neu: 0.761, pos: 0.0}* | Negative |
| 5 | "1 day it will no matter what kind of currency we use. If banks don't get on board,as #bitcoin has proved, we will find a way around" #a16z<br>9:36 AM - 20 May 2014<br>1 Retweet 4 Likes | Compound Score: -**0.273**<br>*{neg: 0.078, neu: 0.883, pos: 0.039}* | Negative |

*Source: Own work.*

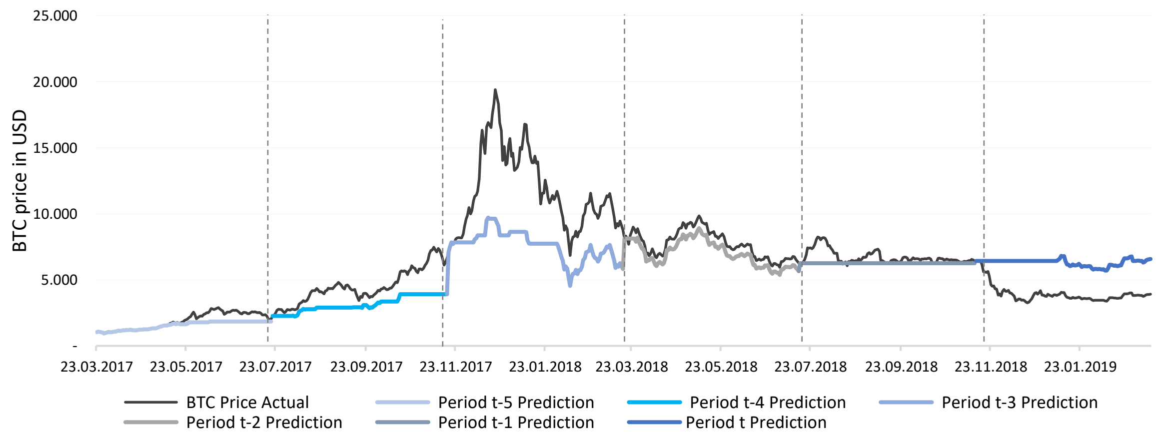
## Appendix 5: Feedforward Neural Network Model – Iterative Analysis

*Table 4: Periods for training and testing of the FF over different time intervals*

| | Train from: | Train until: | *Days of Model training* | Test from: | Test until: | *Overall Daily records* | % of Data for Training | % of Data for Testing |
|---|---|---|---|---|---|---|---|---|
| **BTC** | 17.10.2010 | 27.02.2018 | *2.690* | 28.02.2018 | 28.06.2018 | *2.811* | **95,7%** | **4,3%** |
| | 17.10.2010 | 11.11.2018 | *2.947* | 12.11.2018 | 12.03.2019 | *3.068* | **96,1%** | **3,9%** |
| | 17.10.2010 | 14.07.2018 | *2.827* | 15.07.2018 | 12.11.2018 | *2.948* | **95,9%** | **4,1%** |
| | 17.10.2010 | 17.03.2018 | *2.708* | 18.03.2018 | 16.07.2018 | *2.829* | **95,7%** | **4,2%** |
| | 17.10.2010 | 16.11.2017 | *2.587* | 17.11.2017 | 17.03.2018 | *2.708* | **95,5%** | **4,4%** |
| | 17.10.2010 | 19.07.2017 | *2.467* | 20.07.2017 | 17.11.2017 | *2.588* | **95,3%** | **4,6%** |
| | 17.10.2010 | 21.03.2017 | *2.347* | 22.03.2017 | 20.07.2017 | *2.468* | **95,1%** | **4,9%** |
| **LTC** | 13.07.2012 | 27.02.2018 | *2.055* | 28.02.2018 | 28.06.2018 | *2.176* | **94,4%** | **5,5%** |
| | 13.07.2012 | 11.11.2018 | *2.312* | 12.11.2018 | 12.03.2019 | *2.433* | **95,0%** | **4,9%** |
| | 13.07.2012 | 14.07.2018 | *2.192* | 15.07.2018 | 12.11.2018 | *2.313* | **94,8%** | **5,2%** |
| | 13.07.2012 | 17.03.2018 | *2.073* | 18.03.2018 | 16.07.2018 | *2.194* | **94,5%** | **5,5%** |
| | 13.07.2012 | 16.11.2017 | *1.952* | 17.11.2017 | 17.03.2018 | *2.073* | **94,2%** | **5,8%** |
| | 13.07.2012 | 19.07.2017 | *1.832* | 20.07.2017 | 17.11.2017 | *1.953* | **93,8%** | **6,1%** |
| | 13.07.2012 | 21.03.2017 | *1.712* | 22.03.2017 | 20.07.2017 | *1.833* | **93,4%** | **6,5%** |
| **ETH** | 30.09.2014 | 27.02.2018 | *1.246* | 28.02.2018 | 28.06.2018 | *1.367* | **91,1%** | **8,8%** |
| | 30.09.2014 | 11.11.2018 | *1.503* | 12.11.2018 | 12.03.2019 | *1.624* | **92,5%** | **7,4%** |
| | 30.09.2014 | 14.07.2018 | *1.383* | 15.07.2018 | 12.11.2018 | *1.504* | **92,0%** | **8,0%** |
| | 30.09.2014 | 17.03.2018 | *1.264* | 18.03.2018 | 16.07.2018 | *1.385* | **91,3%** | **8,7%** |
| | 30.09.2014 | 16.11.2017 | *1.143* | 17.11.2017 | 17.03.2018 | *1.264* | **90,4%** | **9,5%** |
| | 30.09.2014 | 19.07.2017 | *1.023* | 20.07.2017 | 17.11.2017 | *1.144* | **89,4%** | **10,5%** |
| | 30.09.2014 | 21.03.2017 | *903* | 22.03.2017 | 20.07.2017 | *1.024* | **88,2%** | **11,7%** |

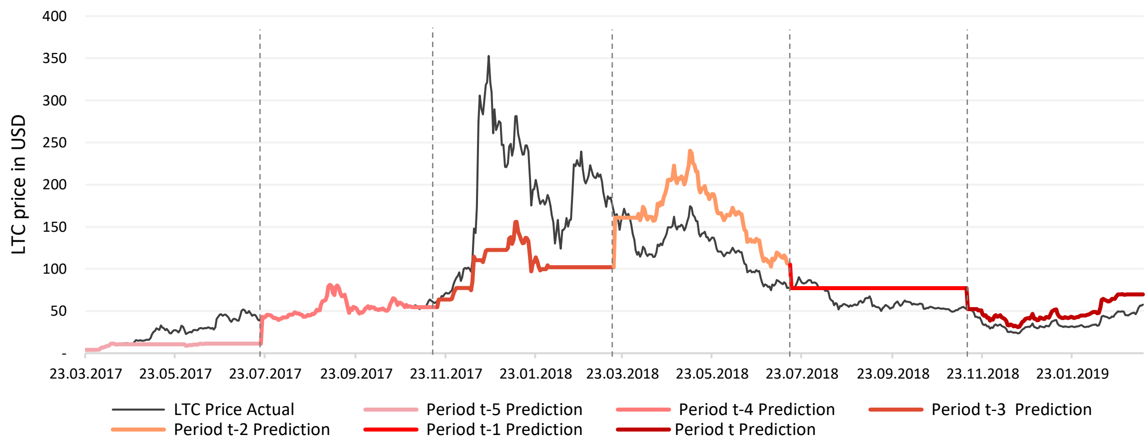
*Source: Own work.*

*Figure 3: FF NN Model Strategy Iteration results – Bitcoin*



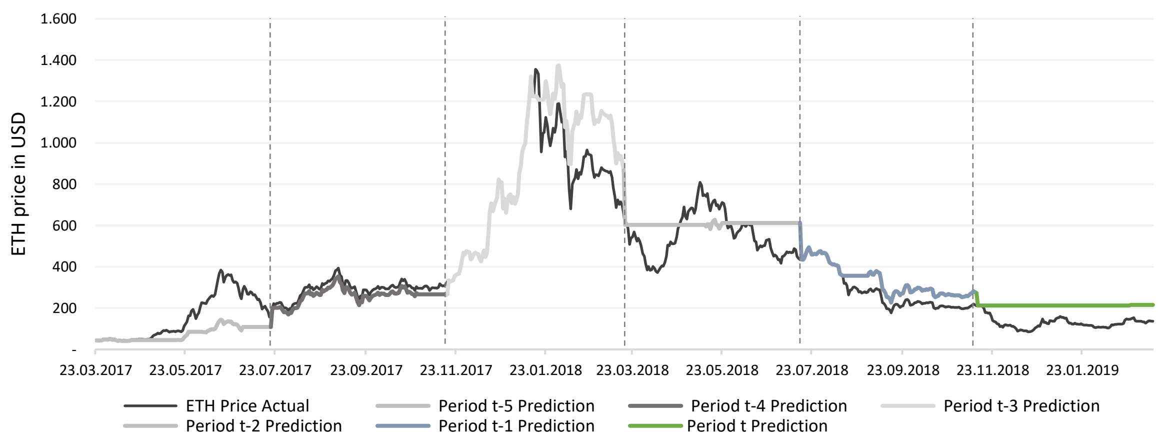

*Source: Own work.*

*Figure 4: FF NN Model Strategy Iteration results – Litecoin*




*Source: Own work.*

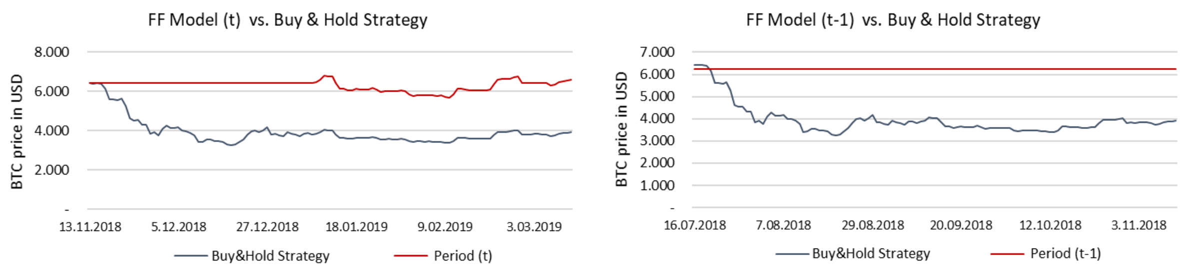
*Figure 5: FF NN Model Strategy Iteration results – Ether*

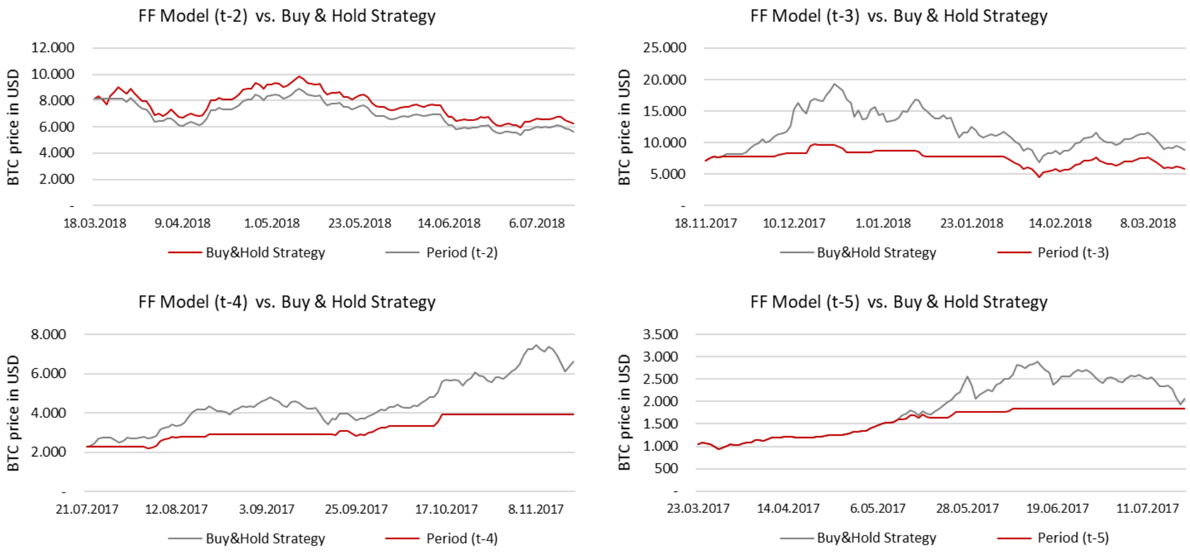



*Source: Own work.*

- FF NN model iteration results – BTC, LTC & ETH:

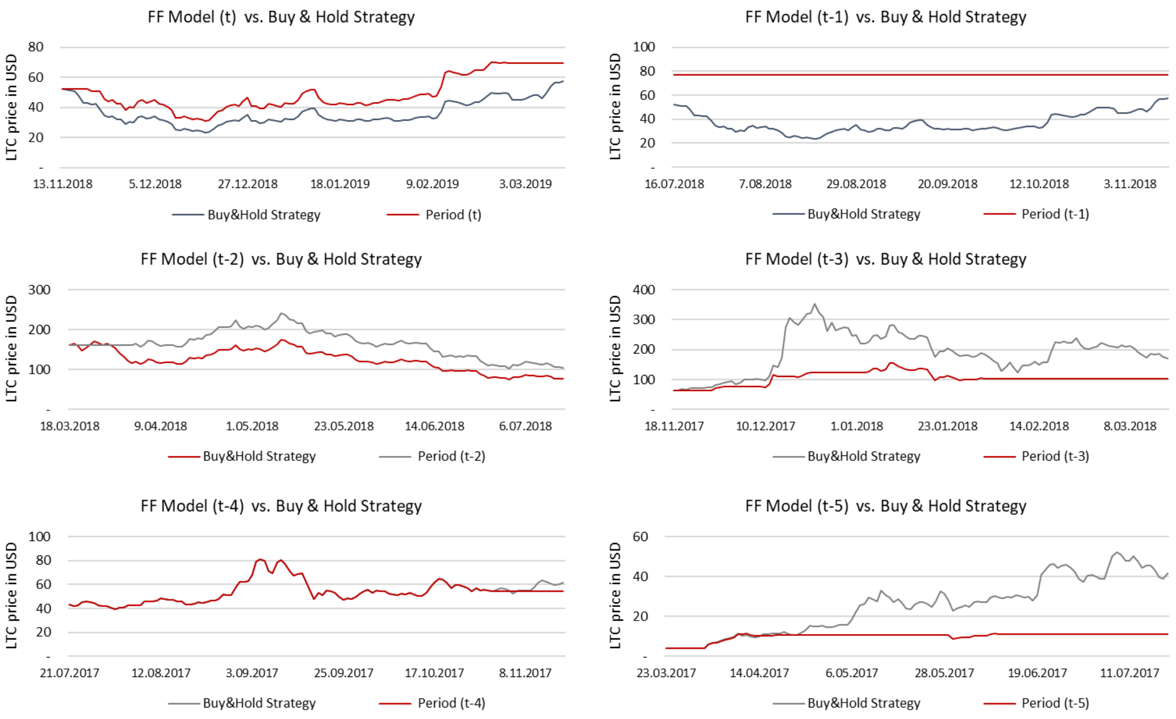
*Figure 6: BTC FF model iteration results*





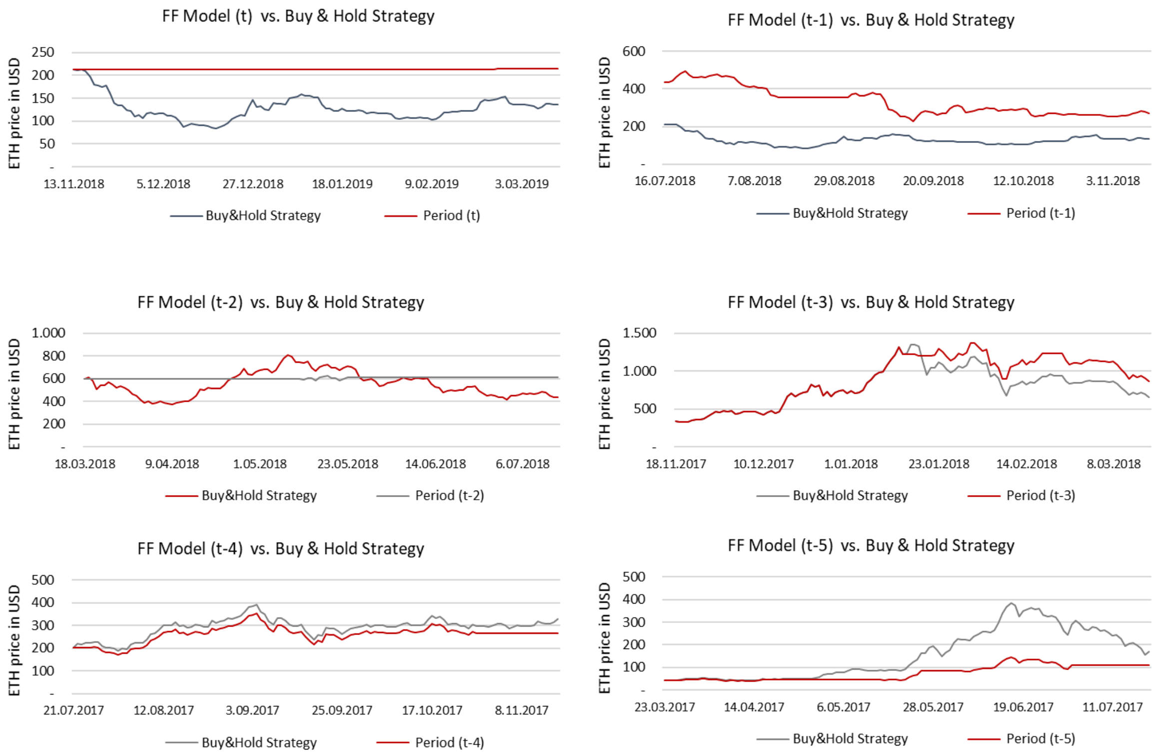

*Source: Own work.*

*Figure 7: LTC FF model iteration results*




*Source: Own work.*

*Figure 8: ETH FF model iteration results*




*Source: Own work.*

## Appendix 6: Technical Analysis of Cryptocurrencies – Technical Trading Strategies

Moving average (MA) presents a simple technical analysis tools which are used for defining the current price trend and smoothing out noisy data. MA does not predict the price direction, but rather help to identify the trend and plausibly define potential support and resistance levels.

### Moving Average Convergence/Divergence (MACD)

Moving Average Convergence/Divergence (MACD) is a technical analysis tool used for detecting changes in the trend direction, duration, strength, and momentum in the stock prices. It is calculated as the difference between two exponential moving averages (EMAs) of closing prices. MACD is a time lagging indicator calculated based on historical price data. The exponential moving averages (EMA) highlight recent changes in asset prices, showing the trend in price development. Following, MACD can be a useful tool for highlighting the changes in the asset price trends (Murphy,1986).

MACD disadvantage is the low response rate to a very low or high volatility market conditions. The MACD measures the divergence between EMA, hence it can provide meaningful insight into the trend changes. However, the indicator is less useful when the market is facing sudden changes and countervailing price movements.

The period for the moving average calculation can differ, but the most commonly applied parameters involve a faster EMA of 12 days, a slower EMA of 26 days, and the signal line as 9-day EMA of the difference between the two. Following, basic components of MACD [12;26;9] strategy are:

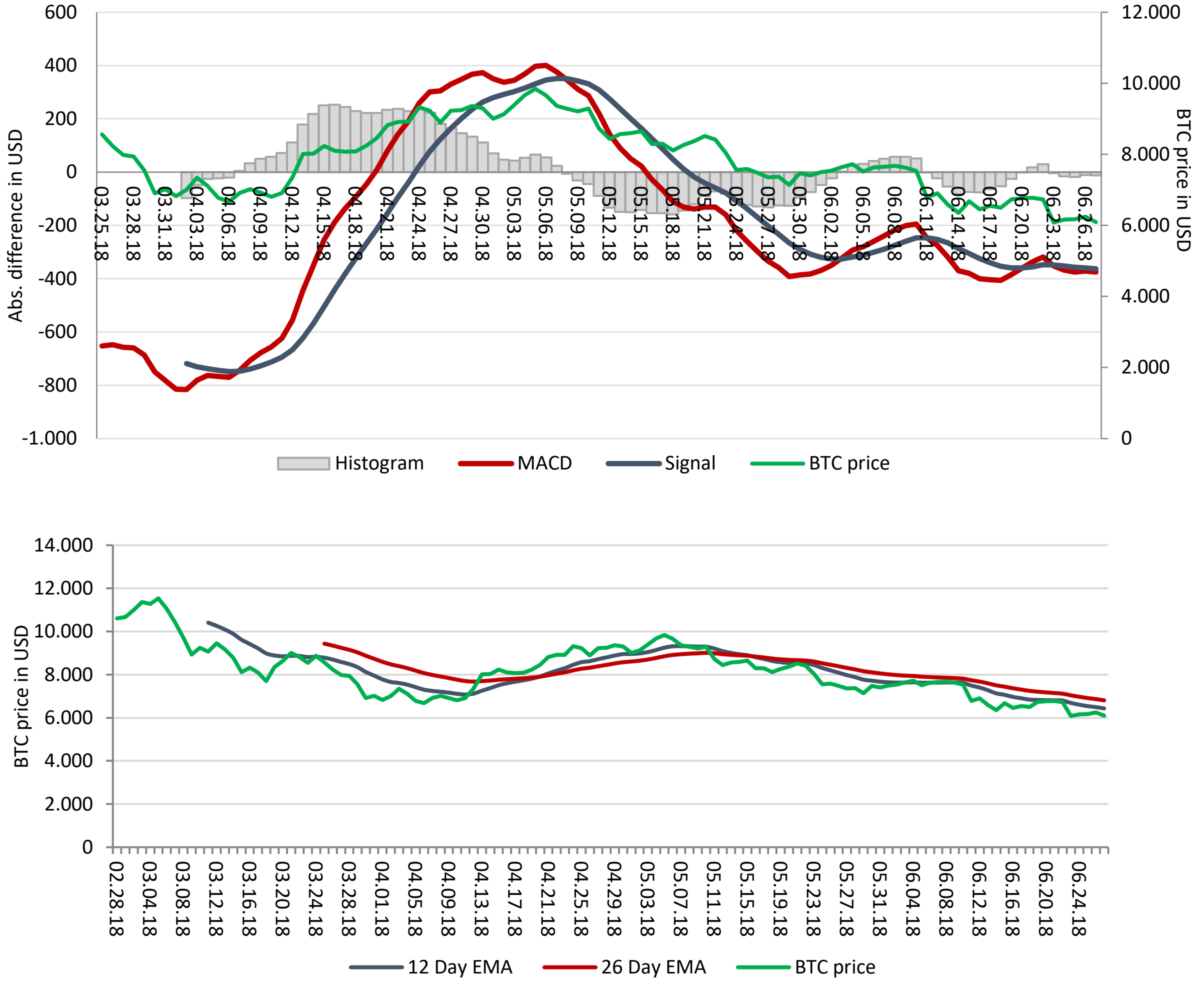
- MACD: the difference between the 12 and 26-day EMA
- Signal: 9-day EMA of the MACD line
- Histogram: the difference between MACD and the Signal

Based on the above stated, two trading strategies can be applied by using MACD indicator:

- Signal line crossover MACD Strategy – when the MACD line crosses up through the signal line we have bullish/buy scenario; when the MACD line crosses down through the signal line – bearish/sell scenario.
- Zero crossover MACD Strategy – when the EMA [fast, 12] is equal to the EMA [slow, 26]. A move from positive to negative is bearish and from negative to positive, bullish.

Moving Average Convergence/Divergence (MACD) Strategy – Bitcoin

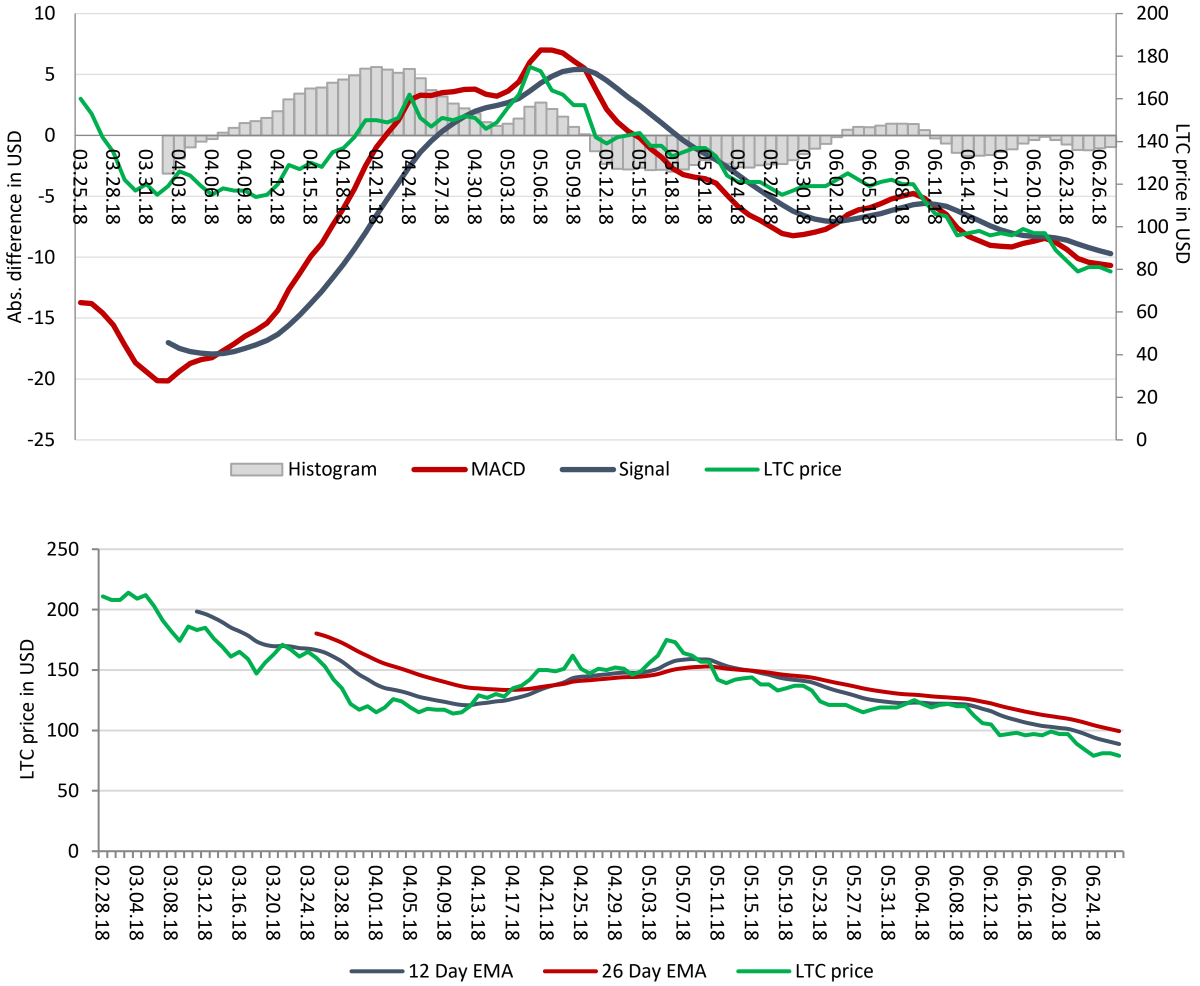
*Figure 9: BTC Moving Average Convergence/Divergence (MACD) strategy*




*Source: Own work.*

Moving Average Convergence/Divergence (MACD) Strategy – Litecoin

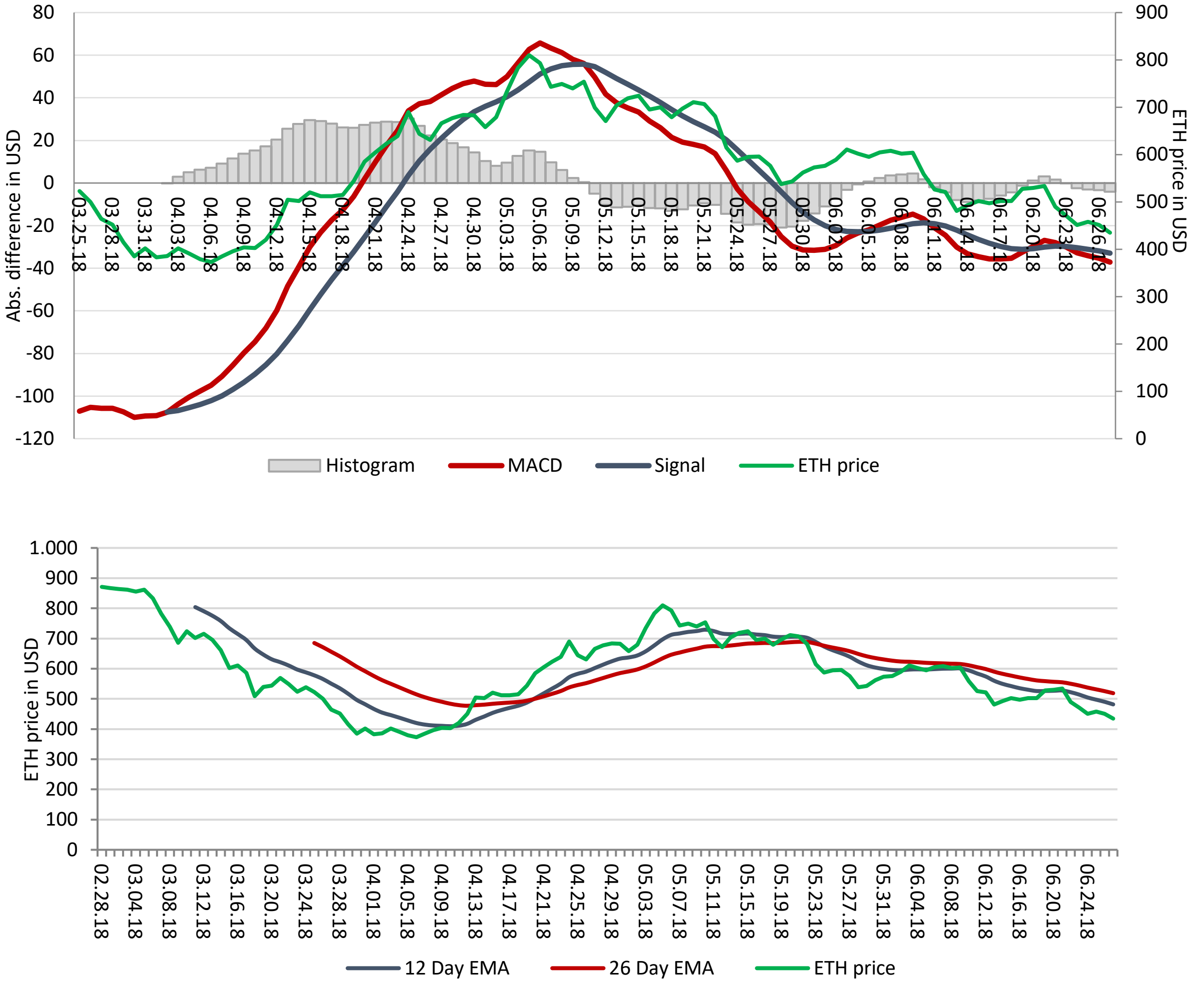
*Figure 10: LTC Moving Average Convergence/Divergence (MACD) strategy*




*Source: Own work.*

Moving Average Convergence/Divergence (MACD) Strategy – Ether

*Figure 11: ETH Moving Average Convergence/Divergence (MACD) strategy*




*Source: Own work.*

**Relative Strength Index (RSI)**

The Relative Strength Index is a technical indicator which shows the current and historical strength or weakness of an asset value based on the recent period closing prices. It measures the velocity and magnitude of directional price changes. It is classified as momentum oscillator calculated as the ratio of higher versus lower closed prices. High RSI means stronger positive changes, and lower RSI indicates stronger negative changes. It has values from 0 to 100 (Wilder, 1978).

RSI is calculated with upward and downward changes. Upward changes are characterized by the current period close price being higher than the previous:

$$U = close_t - close_{t-1} \quad (1)$$
$$D = 0$$

On the opposite, a downward period is characterized by the current period close price being lower than previous:

$$U = 0$$
$$D = close_{t-1} - close_t \quad (2)$$

The average U and D are calculated using n-period EMA and their ratio represents the price Relative Strength (RS):

$$RS = \frac{EMA\ (U,n)}{EMA\ (D,N)} \quad (3)$$

$$RSI = 100 - \frac{100}{1+RS} \quad (4)$$

The level of RSI reveals the asset's recent trading strength. When the price moves up rapidly, a certain point it is considered as overbought. Likewise, when the price falls rapidly, at some point becomes oversold. Following, if the RSI reaches a level above 70 this indicates a sell signal. When RSI falls below 30, it indicates that the asset has been oversold, hence a buy position is signaled. When RSI falls below 30, it indicates that the asset has been oversold, hence a buy position is signaled. RSI above 70 and below 30 indicates a market reversal.

In addition, the divergence between RSI and price action can be used as a strong indication for a market turning point. A bearish divergence occurs when the price reaches a new low, and RSI makes a lower high. A bullish divergence occurs when price makes a higher low (Wilder, 1978).

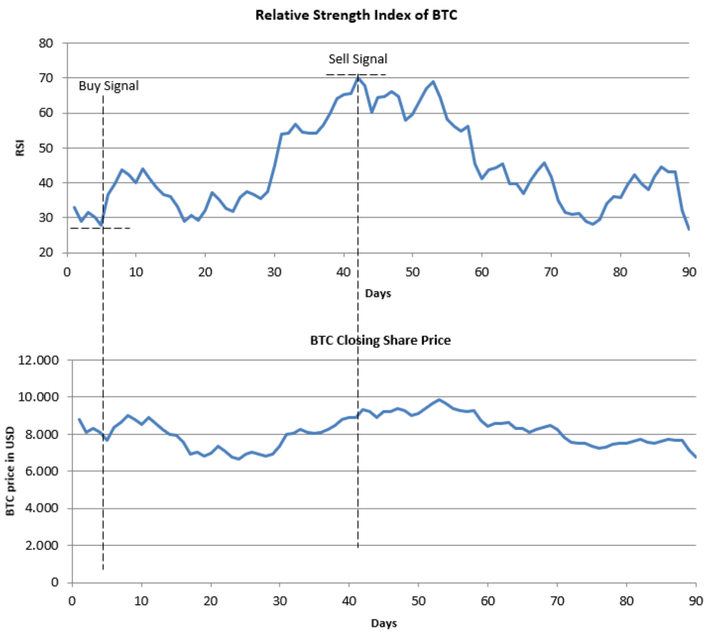
*Table 5: Trading Results – Summarized Overview*

| | | Buy & Hold Strategy | FNN Model | LSTM Model | MACD Signal line crossover | MACD Zero crossover | RSI Strategy |
|---|---|---|---|---|---|---|---|
| BTC | ROI (%) | -42,2 | 22,1 | 11,6 | 11,9 | 2,2 | 14,8 |
| LTC | ROI (%) | -62,8 | 19,8 | 19,8 | 6,8 | -4,0 | 15,6 |
| ETH | ROI (%) | -50,1 | 75,8 | 18,9 | 63,2 | 13,1 | 28,6 |

*Source: Own work.*

- Relative Strength Index (RSI) Strategy – Bitcoin

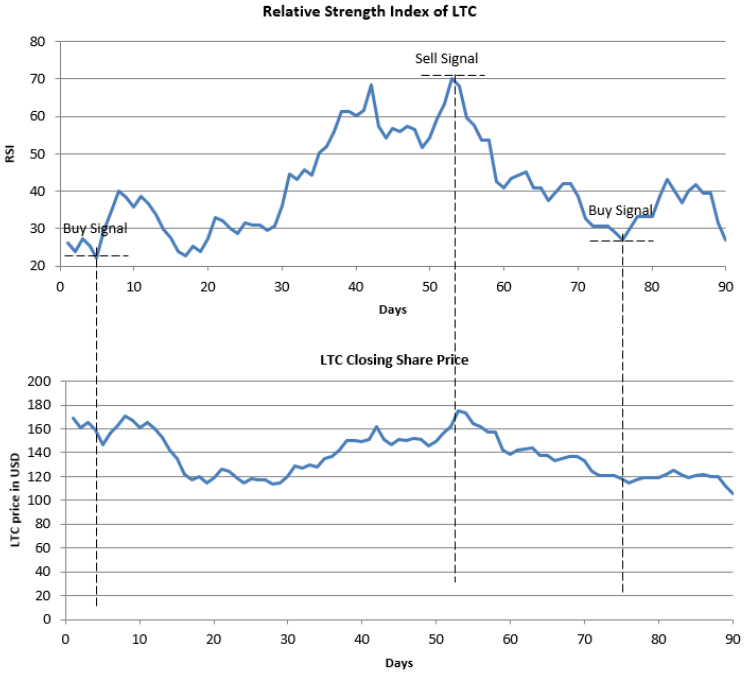
*Figure 12: BTC Relative Strength Index (RSI) strategy*




*Source: Own work.*

- Relative Strength Index (RSI) Strategy – Litecoin

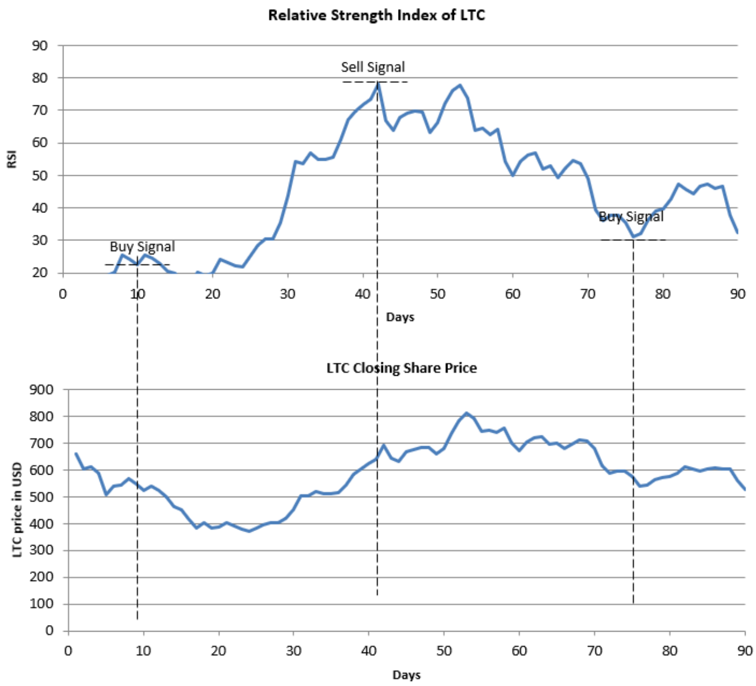
*Figure 13: LTC Relative Strength Index (RSI) strategy*




*Source: Own work.*

- Relative Strength Index (RSI) Strategy – Ether

*Figure 14: ETH Relative Strength Index (RSI) strategy*




*Source: Own work.*