Máster en Ingeniería de Sistemas y de Control

Trabajo de Fin de Máster

Predicción y estrategia de inversión de criptomonedas mediante técnicas de Machine Learning

Adrián Viéitez Mariño

Curso 2021/2022
Defensa: Septiembre 2022

Directores: Matilde Santos Peñas
Rodrigo Naranjo Mota

Máster en Ingeniería de Sistemas y de Control

Trabajo de Fin de Máster

Predicción y estrategia de inversión de criptomonedas mediante técnicas de Machine Learning

Adrián Viéitez Mariño

Proyecto específico propuesto por un profesor

Directores: Matilde Santos Peñas
Rodrigo Naranjo Mota

The realization of this work would not have been possible without the invaluable help of my directors, Matilde and Rodrigo, who guided me throughout the process and showed great availability in case of any questions. Thank you both.

I would also like to thank my family and friends for showing me support and understanding for so many cancelled plans and visits over the years.

And especially, I want to thank Tuğba, my wife, for the great support on a day-to-day basis, for the immense patience and infinite understanding shown along the whole way.

Thank you for being always there.

Autorización

Autorizo a la Universidad Complutense y a la UNED a difundir y utilizar con fines académicos, no comerciales y mencionando expresamente a sus autores, tanto la memoria de este Trabajo Fin de Máster, como el código, la documentación y/o el prototipo desarrollado.

Firmado: Adrián Viéitez Mariño

Firma del alumno

Resumen

Este trabajo examina en primera instancia cómo se puede construir un sistema de predicción de precios de criptomonedas utilizando técnicas de aprendizaje supervisado, sin tener en cuenta ningún indicador técnico extraído de la evolución del precio en sí, y basándose únicamente en los valores de otros índices bursátiles contextuales, indicadores de mercado y tendencias recogidas de servicios online o páginas web.

Bajo un enfoque de regresión, se desarrollan modelos basados en redes Long Short-Term Memory (LSTM) y Gated Recurrent Unit (GRU) y se comparan sus métricas para aclarar primero si alguno de ellos ofrece un mejor rendimiento. Los resultados obtenidos muestran que ambos son capaces de pronosticar con gran fidelidad la evolución real del precio de la criptomoneda Ethereum y que ninguno de ellos supera claramente al otro.

También se aplica el modelo de máquina de vectores de soporte (SVM) para resolver el pronóstico como un problema de clasificación, obteniendo en este caso resultados que apuntan a futuras investigaciones para mejorar su precisión.

Además, se analiza la viabilidad de la aplicación de los modelos presentados en el mercado real, creando una estrategia de inversión basada en las predicciones realizadas por los mismos. Tras aplicar la estrategia de inversión creada en dos períodos de tiempo diferentes, se confirma que aún con un enfoque simple como el presentado en este trabajo, es posible generar una ganancia sobresaliente con pocas operaciones.

Palabras clave: Deep learning, criptomonedas, Ethereum, selección de características, LSTM, GRU, SVM, mercado de valores, predicción, inversión

Abstract

This work examines how a cryptocurrency price forecast system can be built using supervised learning techniques, not taking into consideration any technical indicators extracted out of the price evolution itself, and solely based on the values of other contextual stock indices, market indicators and online trends.

Under a regression approach, Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks-based models are developed and their metrics compared in order to first clarify if any of them offers a better performance. The obtained results show that both are capable of forecasting very close to the actual evolution of the Ethereum cryptocurrency price and not having any of them clearly outperforming the other.

Also, support vector machine model (SVM) is applied to solve the forecast as a classification issue, obtaining in this case results that point towards further investigations in order to improve accuracy.

Moreover, the viability of its application on the real market is analyzed by creating an investment strategy based on the predictions done by the presented systems. After applying the created investment strategy in two different periods of time, it is confirmed that even with a simple approach as the one presented in this work, it is possible to generate an outstanding profit with few operations.

Key words: Deep learning, cryptocurrency, Ethereum, feature selection, LSTM, GRU, SVM, stock market, forecasting, investment

Index of Contents

CHAPTER 1: Introduction 1

1.1. Objective and specific goals 1

1.2. Structure of the Final Project Master report 2

CHAPTER 2: Technical Framework and State-of-Art 3

2.1. Cryptocurrencies 3

2.2. Ethereum 3

2.3. Cryptocurrency´s Market 4

2.3.1. Sessions 4

2.4. Artificial intelligence, Machine learning and Deep learning 6

2.4.1. Neural Networks 6

2.4.2. Support Vector Machines (SVM) 8

2.5. State of Art 9

CHAPTER 3: Materials. Data and processing 13

3.1. Databases and software 13

3.2. Data selected 13

3.3. Market indicators 15

3.4. Data pre-processing 16

3.5. Feature selection 18

3.5.1. Numerical input – Numerical output 18

3.5.2. Numerical input – Categorical output 20

3.6. Feature selection results 20

3.6.1. Regression approach 22

3.6.2. Classification approach 24

3.6.3. Summary 27

3.7. Sentiment analysis 27

3.7.1. Twitter 27

3.7.2. Reddit 28

3.7.3. Google news 28

CHAPTER 4: Methodology Regression approach 31

4.1. Evaluation: General information 31

4.2. Evaluation Regression approach 32

4.3. Investment strategy: General information 36

4.4. Investment strategy: Regression approach 37

CHAPTER 5: Methodology classification approach 47

5.1. Evaluation: Classification approach 47

5.2. Investment strategy: Classification approach 48
CHAPTER 6: Conclusions and future works 57
6.1. Conclusions 57
6.2. Future work 58
References 59
Abbreviations table 64
Appendix A 65
One day prediction 66
Seven days prediction 67
Fifteen days prediction 68
Thirty days prediction 69
Ninety days prediction 71
Appendix B 73

Index of Figures

Figure 1.- Overview of several contiguous sessions with candle sticks 5
Figure 2.- Representative values of a session with candlesticks [16] 5
Figure 3.- Diagram for LSTM cell [20] 7
Figure 4.- Diagram for GRU cell [20] 8
Figure 5.- Original data sample 14
Figure 6.- Ethereum “close” price evolution 14
Figure 7.- Original data with information gaps 17
Figure 8.- Data after filling the missing information gaps 17
Figure 9.- Data used after scaling it between 0 and 1 18
Figure 10.- Feature selection with RFE-RFR method 23
Figure 11.- Feature selection with RFE-RFC 25
Figure 12.- Feature selection with ANOVA 26
Figure 13.- Feature selection with ANOVA method removing 90 days forecast representation 26
Figure 14.- Comments extracted from Reddit, used for sentiment analysis 28
Figure 15.- News headers extracted from Google news, used for sentiment analysis 29
Figure 16.- Results after applying Vader sentiment to collected texts 30
Figure 17.- Ethereum data and training data selection 31
Figure 18.- Regression forecast comparation between LTSM, GRU and actual prices 32
Figure 20.- Graphical representation of regression investment strategy 38
Figure 20.- Graphical explanation of confusion matrix 47
Figure 21.- Graphical representation of classification investment strategy 49
Figure 22.- Feature selection results for 1 day predictions 66
Figure 23.- Feature selection results for 7 days predictions 67
Figure 24.- Feature selection results for 15 days predictions 68
Figure 25.- Feature selection results for 30 days predictions 69
Figure 26.- Feature selection results for 90 days predictions 71

Index of Tables

Table 1.- Features displayed from left to right in the 'x' axis of plots of this chapter 21
Table 2.- Feature selection with RFE-RFR 22
Table 3.- Feature selection with Pearson´s 23
Table 4.- Summary of Feature selection with RFE-RFC 24
Table 5.- Feature selection with ANOVA 25
Table 6.- Metrics of one day price prediction without sentiment 33
Table 7.- Metrics of one day price prediction with sentiment 33
Table 8.- Metrics of seven days price prediction without sentiment 34
Table 9.- Metrics of seven days price prediction with sentiment 34
Table 10.- Metrics of fifteen days price prediction without sentiment 35
Table 11.- Metrics of fifteen days price prediction with sentiment 35
Table 12.- Investment results for regression one day prediction with RFE features, without sentiment 38
Table 13.- Investment results for regression one day prediction with Pearson´s features, without sentiment 39
Table 14.- Investment results for regression one day prediction with RFE features, with sentiment 40
Table 15.- Investment results for regression one day prediction with Pearson´s features, with sentiment 40
Table 16.- Investment results for regression seven days prediction with RFE features, without sentiment 41
Table 17.- Investment results for regression seven days prediction with Pearson´s features, without sentiment. 41
Table 18.- Investment results for regression seven days prediction with RFE features, with sentiment 42
Table 19.- Investment results for regression seven days prediction with Pearson´s features, with sentiment 42
Table 20.- Investment results for regression fifteen days prediction with RFE features, without sentiment 43
Table 21.- Investment results for regression fifteen days prediction with Pearson´s features, without sentiment 43
Table 22.- Investment results for regression fifteen days prediction with RFE features, with sentiment 44
Table 23.- Investment results for regression fifteen days prediction with Pearson´s features, with sentiment 44
Table 24.- Metrics of trend predictions 48
Table 25.- Investment results for classification one day prediction with RFE features, without sentiment 50
Table 26.- Investment results for classification one day prediction with ANOVA features, without sentiment 50
Table 27.- Investment results for classification one day prediction with RFE features, with sentiment 51
Table 28.- Investment results for classification one day prediction with ANOVA features, with sentiment 51
Table 29.- Investment results for classification seven days prediction with RFE features, without sentiment 52
Table 30.- Investment results for classification seven days prediction with ANOVA features, without sentiment 52
Table 31.- Investment results for classification seven days prediction with RFE features, with sentiment 53
Table 32.- Investment results for classification seven days prediction with ANOVA features, with sentiment 53
Table 33.- Investment results for classification fifteen days prediction with RFE features, without sentiment 54
Table 34.- Investment results for classification fifteen days prediction with ANOVA features, without sentiment 54
Table 35.- Investment results for classification fifteen days prediction with RFE features, with sentiment 55
Table 36.- Investment results for classification fifteen days prediction with ANOVA features, with sentiment 55
Table 37.- Abbreviations table 64
Table 38.- Feature selection results for 1 day predictions 66
Table 39.- Table summarizing the feature selection results for 7 days predictions 67
Table 40.- Table summarizing the feature selection results for 15 days predictions 69
Table 41.- Table summarizing the feature selection results for 30 days predictions 70
Table 42.- Table summarizing the feature selection results for 90 days predictions 71
Table 43.- Iterations regression investment strategy 74

CHAPTER 1: Introduction

Thirteen years have passed since the mysterious Satoshi Nakamoto launched his blockchain-based Bitcoin cryptocurrency back in early 2009 [1].

Since then, many events have taken place until reaching the current situation where several thousand of different cryptocurrencies absorb a volume of more than 2.5 Trillion dollars[2], cryptocurrencies are accepted as payment method by more than 160000 business worldwide[3], and at least 103 countries have been identified, whose governments directed their financial regulatory agencies to develop regulations and priorities for financial institutions regarding cryptocurrencies and their use in AML/CFT, apart from many other countries that allow cryptocurrencies to be used [4]. But much recent public discussions of cryptocurrencies have been triggered by the substantial changes in their prices, and claims that the market for cryptocurrencies is a bubble without any fundamental value, and also concerns about evasion of regulatory and legal oversight. While these concerns have led to calls for increased regulation or even a total ban, ffurther debates concern inter alia: the classification of cryptocurrencies as commodities, money or something else; the potential development of cryptocurrency derivatives and of credit contracts in cryptocurrency; the use of initial coin offerings (ICO) employing cryptocurrency technology to finance start-up initiatives; and the issue of digital currencies by central banks employing cryptocurrency technologies. These discussions often shed more heat than light. There is as yet little clearly established scientific knowledge about the markets for cryptocurrencies and their impact on economies, businesses and people [5]. All these facts and noise around them has drawn a lot of attention from entrepreneurs, regulators, general public and of course, investors.

The attempts to forecast accurately the price evolution of stocks and other financial assets has always been an ambition for those who make out of the investments their working environment, a path to earn some additional income or just want to find long term solutions to save their money in a more productive way than leaving it in the savings account.

Since, as already mentioned, the cryptocurrency market has become another investment field for every time more people and companies, and in the framework of the development of my Master Thesis, I have found of great interest the study of computational models based on artificial intelligence and machine learning techniques that can help us to forecast the behaviour of the cryptocurrency´s prices. This is not only an excellent way to apply some of the knowledges acquired during the Master study, but also to extend some of the learnt concepts and gain new skills. Besides, it is a great challenge due to the lack of large amounts of historical data, the volatility of the market, the unknown factors that might influence its evolution and its development in the online platforms that offers conditions never seen before, as 24/7 trading windows, easy access to millions of people and the information and disinformation campaigns that all this audience is exposed. Hence, this Master Thesis is not only a duty, but a very motivating exercise.

The purpose of this work is, in one hand, understand how the last developments on the field of machine learning and artificial intelligence can help to forecast the price evolution of a selected cryptocurrency, in this case Ethereum; in the other hand, study the application of the forecasting results using investment strategies that would potentially help to obtain financial profits out of a complex and volatile environment, full of opportunities but also high risks, as it is the cryptocurrency market.

1.1. Objective and specific goals

The specific objectives that have been developed along this Master Thesis are the following:

- Study of the state of the art of forecasting models in the cryptocurrency market.
- Focus on ML and AI techniques used to develop models applicable to this market.
- Definition of steps to go from the selection of data to the final application of an investment strategy.
- Selection of input data for our models, deciding granularity, time frame, features and data pre-process needed.
- Considering which ML or AI techniques shall be used and implementation of models plus evaluation metrics.
- Development of additional functions that can extract sentiment analysis data out of different online platforms, in this case Reddit and Google news.
- Application of investment strategies that show how powerful can be the combination of the software tools exposed during this work, when it comes to investing on the cryptocurrency market.
- Comparation of the results that are being obtained during the different chapters of the project to end up offering a global summary and conclusions, suggesting as well lines of improvement and future work.

1.2. Structure of the Final Project Master report

The structure of this document follows the next line of thinking.

Chapter 1 presents the motivation of this work, together with the main objectives that have guided us in the development of it.

In **Chapter 2,** some technical background, clarifying concepts that are important to understand the base of the work done, is introduced. For instance, the cryptocurrencies, specially Ethereum and some ML and AI techniques, such as the deep learning networks. This section also includes an overview of some of the research done in this field, giving to the reader a hint on what is the status of the developments done in the same direction as this work.

In **Chapter 3**, all steps related to the data used in this work are explained, from the selection and pre-processing of it until the fine selection through FSM and the way to obtain the sentiment analysis data that will be added later as extra inputs of our models to study its impact on the predictions and investment results.

In **Chapter 4** the focus is placed on the forecasting of the Ethereum price, introducing the architecture of the LSTM and GRU networks used with this purpose, as well introducing, applying and commenting the results of the metrics that will evaluate their performance and the suggested investment strategy that will end up showing how profitable this regression approach to the problem can be. Also gives some general information that will be applying in both cases, when forecasting prices and when forecasting trends.

Chapter 5 is homologous to chapter 4, but setting the focus on the classification approach that tries to forecast the price trend of Ethereum. Different metrics are used and a new investment strategy needs to be applied due to the nature of the prediction itself.

The Master Thesis ends with **Chapter 6**, where conclusions and future work are presented, before moving to the list of references and Appendixes.

CHAPTER 2: Technical Framework and State-of-Art

2.1. Cryptocurrencies

Cryptocurrencies are a form of digital exchange that ensures that transactions are made through a robust encryption process, using the block chain as its back bone technology.

Blockchain consists of blocks with a growing list of records that are linked using cryptography to resist the alteration of the data. Each block in the blockchain possesses a cryptographic hash value of the previous block, a fresh timestamp, and transaction data. Because all blocks in the blockchain are decentralised and distributed in nature, any involved block in the blockchain cannot be modified retroactively without affecting all subsequent blocks. Blockchain technology contains its own key features such as security, transparency, decentralization, immutability, and programmability. Moreover, it doesn't depend on any centralized trusted authorities to process data transactions. In addition, there is no need for any intermediate third party to verify and validate the data transactions. This phenomenon allows users to easily verify and review the transactions independently and inexpensively. Once the data transactions are completed in a successful manner, the valid transactions are hashed and encoded into a Merkle tree in the block. The new blocks are added in the block chain onto the old blocks in such a way that they are extended as new blocks rather than being overwritten onto old blocks to significantly reduce the probability of an entry becoming outdated [6].

Electronic currencies are created through mining, an incentive process in which transactions are verified and new units are created and added to the core of existing ones. The miners are responsible for collecting the latest transactions into blocks and finding a solution to the algorithm of each currency. As a reward, a fixed amount of that currency is acquired by these miners [7].

From a policy maker approach, can be defined as a digital representation of value that is intended to constitute a peer-to-peer ("P2P") alternative to government-issued legal tender, is used as a general-purpose medium of exchange (independent of any central bank), is secured by a mechanism known as cryptography and can be converted into legal tender and vice versa [8].

2.2. Ethereum

Ethereum, announced in 2014 in Vitalik Buterin´s [9] and launched in 2015, aims at creating a universal blockchain-based application platform. It addresses several limitations of the Bitcoin's scripting language, incorporating a Turing complete language, meaning that Ethereum supports all types of computations, including loops.

That is, Ethereum supports the state of the transaction, as well as several other improvements over the blockchain structure. It provides an abstract layer enabling anyone to create their own rules for ownership, formats of transactions, and state transition functions. This is done by involving smart contracts, a set of cryptographic rules that are executed only if certain conditions are met [9]. The consensus in the Ethereum network is based on modified Ghost protocol (Greedy Heaviest Observed Subtree) [10]. It is created to tackle the issue of stale blocks in the network. The stale blocks can occur if one group of miners combined in a mining pool has more computing power than the others, meaning that the blocks from the first pool will contribute more to the network, thus creating the centralization issue. Ghost protocol includes those stale blocks into calculations of the longest chain. The centralization problem is removed through providing block rewards to stales, where the stale block receives 87.5% of the reward, and the nephew of that stale block receives the remaining 12.5% of the reward. In this way, the miners are still rewarded even if their block didn't become the part of the main blockchain (those blocks are called uncles). Ethereum uses the modification of the Ghost protocol which includes uncles up to seven generations [11][12].

Ethereum is the cryptocurrency chosen to be used for the study developed in this work. Most of works use Bitcoin, but it has been considered useful to have Bitcoin as input feature rather than object of study itself, since this might open the door to better predictions due to the influence that the Bitcoin price behaviour has over the other cryptocurrencies, also known as altcoins, and more specifically over Ethereum [13].

2.3. Cryptocurrency´s Market

A financial market is an environment where investors trade assets to obtain profit. The prices of assets change over time. Hence, by predicting these changes, investors can purchase assets which may increase in value and generate profit from those assets. In all markets, there is at least one asset which is considered to keep its value constant over time. Prices of all assets are measured with respect to this special asset, and it can be used to purchase any asset in the market. Hence, it is referred to as cash. In stock and fiat currency markets, the most common cash asset is the U.S. Dollar (USD). In cryptocurrency markets, on the other hand, the most popular is the Tether (USDT). This is because the USDT was designed to keep its exchange rate constant with respect to the USD; hence, cryptocurrency investors who want to close their positions or halt their trading sessions can rapidly exchange their assets for this risk-free asset.

2.3.1. Sessions

It is known as trading sessions in stock markets the specific hours during working days in which investors are allowed to exchange stocks. However, cryptocurrency markets do not have this limitation, since they are open 24 h. Nonetheless, we use the term trading session for the time in which the implemented algorithms are allowed to perform transactions in a simulated environment [14].

With that said, it is widely stated through all cryptocurrency's investment platforms that the opening time of this virtual trading sessions for cryptocurrency is the 00:00 UTC and the closing time the 23:59 UTC.

In both, cryptocurrency markets and stock markets, it is common to see the sessions represented as candlesticks, and indeed it is a good way to offer an overview of a session and its evolution. The candlesticks do not necessarily represent sessions, since usually they can be adjusted to represent different time frames, going from minutes, to hours, days, weeks, etc. In the Figure 1, a screenshot taken from Binance [15], can be observed the daily evolution of the Ethereum price displayed with this system, and it will help introducing the concepts of open, close, high and low prices during a session. These terms will be found also later in this work when collecting data for the different cryptocurrencies and financial indicators.

Figure 1.- Overview of several contiguous sessions with candle sticks

The Figure 2, taken from [16] offers a good visual explanation of the two candlesticks, the increasing or "bullish" candle in the left and the decreasing or "bearish" candle in the right. Some of the terms related to seasons in cryptocurrency market that might be of interest are:

Open price: will be the first price at which a cryptocurrency is traded after the opening time of the session.

Close price: price at which has been ultimately traded before the closing time deadline.

High price: highest price at which the cryptocurrency has been traded during the season.

Low price: lowest price at which the cryptocurrency has been traded during the season.

Adjusted close: in stock market, it amends the close price considering other factors like dividends, stock splits, and new stock offerings. In the Cryptocurrency market does not reflect any changes with respect to the close price.

Having these definitions, it is easy to understand that the increasing candles correspond to sessions where the price at the end of it is higher than at the beginning, and decreasing candles are those where the price at the end of the session is lower than at the beginning.

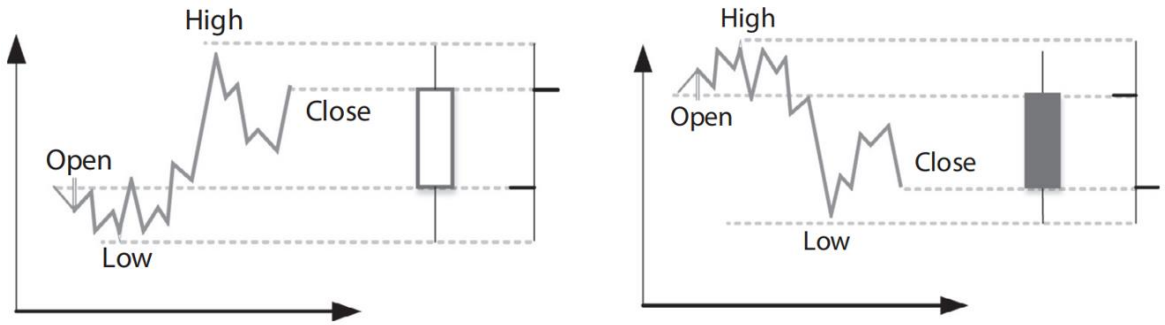

Figure 2.- Representative values of a session with candlesticks [16]

2.4. Artificial intelligence, Machine learning and Deep learning

Often, the terms Artificial Intelligence, Machine Learning and Deep Learning get mixed up and confused. Clearly the terms are closely related, but there are differences between them. Artificial Intelligence can be defined, according to Marvin Minsky, one of the pioneers in this field, as the science of building machines to do things that, if they would be done by humans, it would require intelligence. Said in other words, is a set of techniques that allowed machines to imitate human behaviours. Secondly, Machine Learning is a subset of techniques and algorithms of the Artificial intelligence that using data, improves automatically with experience. Last but not least, Deep Learning is a field of the machine learning that develops algorithms inspired on the structure and function of the huma brain, known as neural networks.

Focusing on the Deep Learning, it aims to understand concepts in a more precise way. In a neural network, successive layers of data are combined to learn the concepts. The simplest networks have only two layers: an input and an output, knowing that each one can have several hundreds, thousands, even millions of neurons. Among the most used deep learning algorithms, we have:

- Artificial neural networks (ANN): these are the simplest and are often used in addition because they sort information well.
- Convolutional neural networks (CNN): Applies apply filters to the information collected in order have new data (for example, bringing out the contours in an image can help to find where is the face).
- Recurrent neural networks (RNN): the best known are Long Short-term Memory (LSTM), which have the ability to retain information and reuse it soon after. They are used for text analysis (NLP), since each word depends on the previous few words (so that the grammar is correct).

As well as more advanced versions, such as auto-encoders, Boltzmann machines, self-organizing maps (SOM), etc [17].

2.4.1. Neural Networks

A standard neural network (NN) consists of many simple, connected processors called neurons, each producing a sequence of real-valued activations. Input neurons get activated through sensors perceiving the environment, other neurons get activated through weighted connections from previously active neurons. Some neurons may influence the environment by triggering actions. Learning or credit assignment is about finding weights that make the NN exhibit desired behaviour. Depending on the problem and how the neurons are connected, such behaviour may require long causal chains of computational stages, where each stage transforms the aggregate activation of the network. Deep Learning is about accurately assigning credit across many such stages [18].

With the ever-increasing availability of data, ANNs have become a dominant and popular technique for machine learning tasks. A feed-forward neural network (FFNN) is the most basic type of ANN. It has only forward connections in between the neurons, unlike RNNs, which have feedback loops [19].

2.4.1.1. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) networks are a distinct type of Recurrent Neural Networks. They are able to learn long-term dependencies and are very popular for working with sequential data such as time-series data. LSTM's address the vanishing gradient problem. When the time step is significant, the gradient gets too small or large, which leads to a vanishing gradient problem. This problem occurs while the optimizer back propagates, and makes the algorithm run,

while the weights almost do not change at all. The long-short-term memory cell uses an input gate; a forget gate and an output gate (a simple multilayer perceptron). These gates define whether the data can pass through or not depending on the data's priority. The gates also enable the network to learn what to save, what to forget, what to remember, what to pay attention, and what to output. The cell state and hidden state are used in gathering data for processing in the next state. The vanishing gradient can, therefore, be protected. *Figure 3* shows the structure of the nodes of the LSTM cell. The gates have the following equations:

Input Gate	$i_t = \sigma(W_i h_{t-1} + W_i h_t)$	(1)
Forget Gate	$f_t = \sigma(W_f h_{t-1} + W_f h_t)$	(2)
Output Gate	$o_t = \sigma(W_o h_{t-1} + W_o h_t)$	(3)
Intermediate Cell State	$\tilde{C} = tanh(W_c h_{t-1} + W_c h_t)$	(4)
Cell State (next memory input)	$c_t = (i_t * \tilde{C}) + (f_t * c_{t-1})$	(5)
New State	$h_t = o_t * tanh(c_t)$	(6)
Input Vector	X_t	(7)
Output Vector	h_t	(8)
Parameter matrices and vector	W, U, and f	(9)

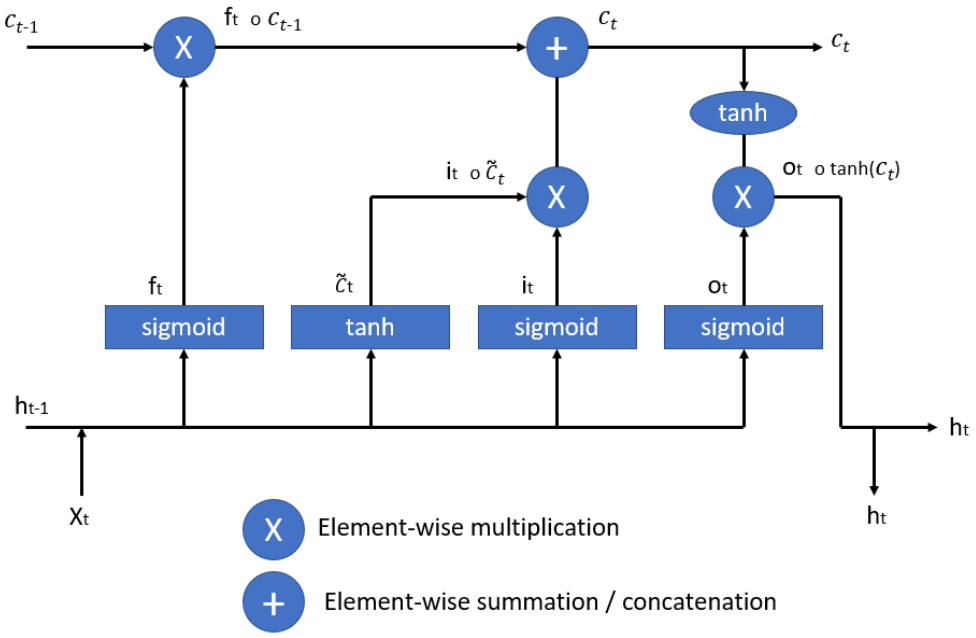

Figure 3.- Diagram for LSTM cell [20]

2.4.1.2. Gated Recurrent Unit (GRU)

The Gated Recurrent Unit, just like the LSTM, is a Recurrent Neural Network. It, however, has a less complicated structure compared to LSTM. It lacks an output gate but has an update z and a reset gate r. These gates are vectors which decide what information should be passed to the output. *Figure 4* shows the GRU cell. The Reset gate defines how to combine the new input with the previous memory. The GRU has the following equations:

Update gate	$z = \sigma(W_z h_{t-1} + U_z x_t)$	(10)
Reset gate	$r = \sigma(W_r h_{t-1} + U_r x_t)$	(11)
Cell state	$c = tanh(W_c\,(h_{t-1} * r) + U_c\, x_t)$	(12)
New state	$h_t = (z * c) + ((1- z) * h_{t-1})$	(13)

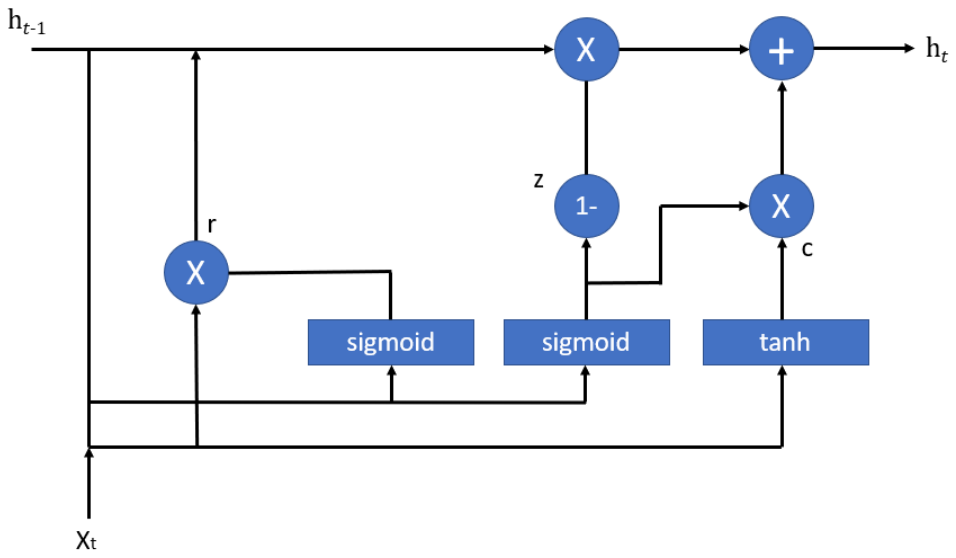

Figure 4.- Diagram for GRU cell [20]

2.4.2. Support Vector Machines (SVM)

Support Vector Machine (SVM) method or classifier was introduced as an induction principle that can avoid over-fitting the data at the assimilation of the training dataset and is known as the most flexible technique to construct the explicit and accurate boundaries. SVM works very well in various applications, provide fast training result and easy to use. Eventually, SVM has been invented to answer pattern recognition problems to fault diagnosis problems. It gives nonlinear and solid solution by applying kernel functions to map the input space into a higher dimensional feature. There are many benefits of the SVM including outperforms in generalization model and perform well with small datasets. SVM creates a lot of benefits in many fields including pattern classification problem. Besides, SVM is to produce a classification hyper-plane that differentiate two classes of data with maximum margin.

They map the input (x) into a high-dimensional feature space (z = ϕ(x)) and construct an optimal hyperplane defined by w · z − b = 0 to separate examples from the two classes. For SVMs with L1 soft-margin formulation, this is done by solving the primal problem

$$\min \frac{1}{2}\|w\|^2 + C \sum \xi_i$$
$$\text{s.t. } y_i(\mathrm{w} \cdot \mathrm{z} - \mathrm{b}) \geq 1 - \xi_i, \xi_i \geq 0 \; \forall \, i \tag{14}$$

where x_i is the ith example and y_i is the class label value which is either +1 or −1. (Throughout the paper, ι will denote the number of examples). This problem is computationally solved using the solution of its dual form

$$\max \sum \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j k(x_i, x_j)$$
$$\text{s.t. } 0 \leq \alpha_i \leq C \; \forall \, i, \sum_i y_i \alpha_i = 0 \tag{15}$$

where $k(x_i; x_j) = \phi(x_i) \cdot \phi(x_j)$ is the kernel function that performs the non-linear mapping. Some of the most popular kernel functions are the Gaussian kernel and the Polynomial kernel. [21]

2.5. State of Art

There are a number of works where ANNs are used for forecasting. According to them, ANNs possess various appealing attributes that make them good candidates for forecasting, compared to the statistical techniques. First, ANNs can model any form of unknown relationship in the data with few *a priori* assumptions. Second, ANNs can generalise and transfer the learned relationships to unseen data. Third, ANNs are universal approximators, meaning that they are capable of modelling any form of relationship in the data, especially non-linear relationships.

The range of functions modelled by an ANN is much higher than the range covered by the statistical techniques. [22] compared ANNs to the Box–Jenkins methodology for forecasting and showed that ANNs are comparatively better for forecasting problems with long forecasting horizons. [23] derived the same observations with respect to a tourism-demand forecasting problem. However, determining the best network structure and training procedure for a given problem is crucial to tune ANNs for maximal accuracy. An equally critical decision is the selection of the input variables for modelling [24].

ANNs have been applied to forecasting in a multitude of domains over the years. [25] used an ANN for electric load forecasting, along with a Euclidean norm applied to weather information in both the training data as well as the forecasting duration to derive the similarity between data points in the training data and the forecasts [19].

In this line, the application of ANN has been object of study for many years now on the field of stock market prediction, proving that they are more effective than regression methods [26]. Also, it has been proven that the error in the prediction of the price change magnitude using MLP neural network is smaller than using linear regression method. In addition to this, when the feed forward MLP neural network predicts the direction of the changes correctly, the change magnitude is very close to the real one in comparison with linear regression method [27]. This conclusion not only shows the convenience of using neural networks over regression methods, but also points towards recurrent neural networks. [28] compares the accuracy of ARIMA and LSTM, as representative techniques when forecasting time series data, and comes to the conclusion that LSTM-based algorithm improved the prediction by 85% on average compared to ARIMA.

And moving to the Cryptocurrency market and price prediction, it is easy to find many researches the last years, where this topic has gained more focus and interest. Again, studies are found where the performance of different Machine Learning strategies, as ARIMA, SVR, RF-Regressor, XGB-Regressor, MLP, LSTM, GRU, CNN, LSTM+CNN, GRU+CNN are compared, concluding that the worst performance attending to RMSE and R-squared measurements is the ARIMA method and the best

are the Recurrent Neural Networks [29] when the forecast is stated as a regression problem, while support vector machines have demonstrated to be the strongest classifier [30][31].

Recent studies compare the performance of different RNNs and come to the conclusion that the performances of LSTM and GRU are very similar when forecasting the prices of Bitcoin, Ethereum, Cardano, Tether and Binance using daily historical data [32]. [33] makes a great summary and offers a very complete overview of the works done in this field until the point in time when it was written, concluding that the LSTM networks are the ones offering the best performance for regression of Bitcoin price. This work has been of great inspiration for the present Master Thesis, since it gives good and well-structured information about the different methodologies followed by other works, including a large set of variables to be consider as features for our models.

And speaking about features, some works offer also results mentioning about their impact. [34] Shows once again how RNN as GRU and LSTM outperform traditional machine learning models, but also explores several financial features and brings up the importance of introducing dropout in the layers to reduce overfitting and improve performance. [35] makes use of Soft Gradient Boosting Machine and Rain Forest Tree to predict the prices of Bitcoin, Ethereum and Ripple, and they could confirm that the inclusion of additional features into the model led to an increased accuracy both on in- and out of samples on average from 1 to 3%. Additionally, this work tested that training datasets with larger lag depth of past values, does help to have a better trained model. In this work the largest lag depth used was 28 days.

[36] demonstrates that LSTM networks can do a good job using 30 lags -days- depth of past values for one day in the future prediction, but the performance increases using less lags, since due to the design of the model, it tends to rely more on short-term dynamics. Nevertheless, this might be different applying a large number of input features, what has not been done in the mentioned work.

Regarding the forecast horizon, it has been observed that, in the short term between 15 and 60 minutes in the future, trying to predict changes in the price of Bitcoin as a binary variable, the accuracy obtained was around 50-55%, getting better as the forecast horizon increases [37], which is not a great result, and seems to lead towards middle term predictions as one day rather than short term predictions. Towards the same forecast horizon but from a different approach, point other works [38], which despite not being able to find a good solution for the price prediction of Bitcoin -at least compared with other works-, they do throw some interesting conclusions showing that the RNN performance decreases, by increasing the error measurements, when the forecast horizon increases from one day onwards.

Going one step further, some studies try to understand the connection of world social and economic events in the price evolution of the Cryptocurrency´s. [39] shows that the cryptocurrency market actively reacts to economic policy uncertainty and geopolitical risks. Following these principles, many works tried to study in the last years the possibility of extracting Sentiment data from online websites, forums and social networks and use them as a feature to obtain better prediction results. [40] makes a comparison not using at first Sentiment data and making use of it later, on the forecast of Bitcoin, Ethereum and Ripple prices, using three different Machine learning techniques: MLP, SVM and RF. Their results don´t have a clear lesson to pick, since in some cases the Twitter Sentiment data helps to slightly improve the prediction, in some cases the prediction can be done by using just the Sentiment data itself and in other scenarios it might cause reduction in the model performance. Nevertheless, the sentiment collected has been only from Twitter and not well filtered, so they suggest that a better data pre-processing might help to improve the results. Some other works, as [41] go one step forward and, through more complex research, implementing a robust cryptocurrency-specific lexicon-based sentiment analysis approach in combination with bivariate Granger-causality tests, found that Twitter sentiment can be used to

predict the price returns of some of the top ten cryptocurrencies. But in general, is not easy task to find works that show a clear benefit on using sentiment analysis, even though most of them consider that is the right direction to investigate and bring the forecasting to the next level of accuracy.

CHAPTER 3: Materials. Data and processing

3.1. Databases and software

The data collection process has been automated using Python. Therefore, the focus was placed on using services that provide an API to implement requests using this programming language.

For the retrieval of stock market indices as well as cryptocurrencies historical data the library "*yfinance"* [42] has been used, which is an open-source tool that offers a threaded and pythonic way to download market data from Yahoo! Finance. This portal, part of the Yahoo! Network, offers news and information of financial nature, online tools for personal finance management and data -also historical- of more than 10000 cryptocurrencies and a myriad of stocks and indices.

In the same way, the goal was to automate the collection of data related to other indicators used as features by the neural networks that will be presented later in this work. Thus:

- To gather data regarding the number of visits to the Wikipedia page of the term "Ethereum", the library for Python "*requests"* [43] was used, which allows us to send HTTP requests in a very easy way directly to the Wikipedia API and get back the information with a convenient JSON format that is easy to parse and save as *.csv in our local disk.

- To be able to fetch historical data for the search volume on Google of the term "Ethereum", the Python library "*pytrends*" [44] has been used. This library provides a simple interface to Google backend and enables us to get daily and monthly data that can be used to compose our historical data record.

- In order to perform sentiment analysis on daily news headers, the python libraries "*urllib*"[45], "*requests*" [43] and "*bs4"* [46] were combined to first generate a local database, extracting strings that correspond to the mentioned news headers out of the iteration with daily granularity through google news searches URLs.

- Also, to perform sentiment analysis, it was decided to extract comments from Reddit which contain the term "Ethereum". To accomplish this task, the use of the Python library "*psaw*" [47] has been implemented, which consist of a minimalist wrapper for searching public Reddit comments/submissions via the pushshift.io API.

- EtherScan: Ethereum Gas price and Ethereum transactions volume are variables not available in all platforms. Their historical data has been retrieved manually from the website EtherScan [48].

3.2. Data selected

For the realization of this work, a total of twenty-one features has been considered, of which two derive from the sentiment analysis that will be presented later. But not all features have been ultimately used, being left out by the techniques of feature selection applied, as it will be shown in further sections. The two features related to sentiment analysis have not been included in the feature selection process, since the goal was to compare the performance of our systems with and without the sentiment analysis results, applying them as an extra on the top of an already working system. The period of time from which the data has been extracted goes from 9th November, 2017 until 13th Mai, 2022. The reason to choose those dates have been, difficulties to find data of some of the features from an earlier date in one hand, and the current date at the time that this work is being developed on the other hand. All data has been collected with daily granularity, having accumulated a total of 1647 datapoints per feature.

In the first place, as it is the main subject of study of this work, the historical data of Ethereum prices has been collected. Out of all retrieved values from the database for Ethereum, the one selected to be used has been the closing price for each day. In the *Figure 5* can be seen an example of the dataset received for Ethereum after request, where it is observed that contains not only the closing price but also the High, Low, Open and Adjusted Close. In general, the datasets received for all market indicators look like this one.

1	Date,High,Low,Open,Close,AdjClose
2	2017-11-09,329.4519958496094,307.0559997558594,308.6449890136719,320.8840026855469,320.8840026855469
3	2017-11-10,324.7179870605469,294.5419921875,320.6709899902344,299.25299072265625,299.25299072265625
4	2017-11-11,319.4530029296875,298.1919860839844,298.58599853515625,314.6809997558594,314.6809997558594
5	2017-11-12,319.15301513671875,298.51300048828125,314.69000244140625,307.9079895019531,307.9079895019531
6	2017-11-13,328.4150085449219,307.0249938964844,307.0249938964844,316.71600341796875,316.71600341796875
7	2017-11-14,340.177001953125,316.76300048828125,316.76300048828125,337.6310119628906,337.6310119628906
8	2017-11-15,340.9119873046875,329.81298828125,337.9639892578125,333.35699462890625,333.35699462890625
9	2017-11-16,336.15899658203125,323.6059875488281,333.4429931640625,330.92401123046875,330.92401123046875
10	2017-11-17,334.9639892578125,327.52301025390625,330.1669921875,332.3940124511719,332.3940124511719
11	2017-11-18,349.6159973144531,327.68701171875,331.9800109863281,347.61199951171875,347.61199951171875
12	2017-11-19,371.2909851074219,344.739990234375,347.4010009765625,354.385986328125,354.385986328125
13	2017-11-20,372.1369934082031,353.2890014648437,354.093994140625,366.7300109863281,366.7300109863281
14	2017-11-21,372.4700012207031,350.6929931640625,367.4429931640625,360.4010009765625,360.4010009765625

Figure 5.- Original data sample

Now, to get an overview of the evolution of the Ethereum price during the mentioned period of time, the *Figure 6* shows the plot of the closing prices of Ethereum that have been used to train and test our systems.

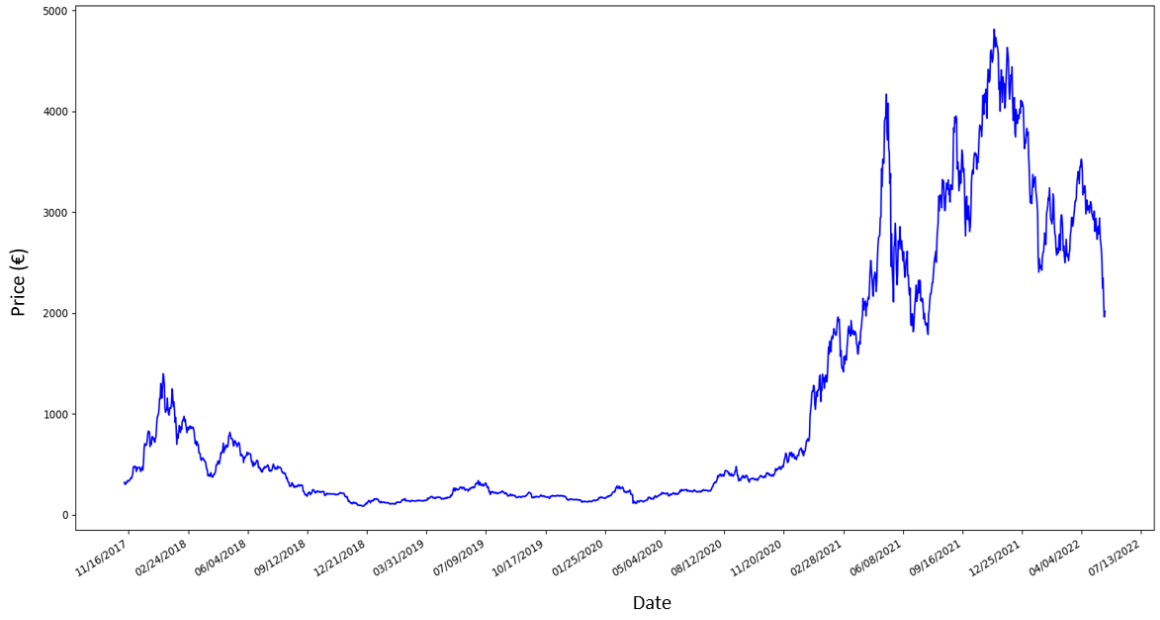

Figure 6.- Ethereum "close" price evolution

It can be observed that the plot can be divided in three different periods. A first period with quite some movement where the price sets, what it was back then, its historical maximum value. This period starts with a clear uptrend, multiplying its price by five. After reaching the maximum price at about 1500 Euros, close the end of the year 2017, it begins a downtrend with some ups and downs that continues until approximately December 2018. It is then, when a second period of quasi-stability begins, where the price of Ethereum shows slight changes until April-Mai 2020. At

this point the evolution of the price initiates the third period, and with it a rally up that will become much more pronounced at the beginning of the year 2021. This is the time of highest volatility and biggest changes on the price, setting what it is until nowadays the absolute historical maximum price of this cryptocurrency with more than 4600 Euros.

3.3. Market indicators

The selection of the market indicators, indices and other data inputs have been highly influenced by other papers [49], adding on the top personal analysis and self-thinking. As a result, the next collection has been selected as possible features:

- Bitcoin price (Btc). Historical data of Bitcoin price with daily granularity.
- Relation Bitcoin-Ethereum (BtcEth). Historical data of the relation between Bitcoin and Ethereum prices with daily granularity.
- S&P 500 index (SP500). One of the Standard & Poor's indices. It is a capitalization-weighted index, also known as value-weighted or market cap weighted instead of capitalization weighted. The simplest capitalization weighted index can be thought of as a portfolio consisting of all available shares of the stocks in the index. While one might track this portfolio's value in dollar terms, it would probably be an unwieldy number – for example, the S&P 500 float-adjusted market value is roughly US$ 11.0 trillion. Rather than deal with ten or more digits, the figure is scaled to a more easily handled number, currently around 1200. Dividing the portfolio market value by a factor, usually called the divisor, does the scaling. Its methodology can be consulted in the referred bibliography [50].
- Volatility index (VIX). This index measures the level of expected volatility of the S&P 500 Index over the next 30 days that is implied in the bid/ask quotations of SPX options. Thus, the VIX Index is a forward-looking measure, in contrast to realized (or actual) volatility, which measures the variability of historical (or known) prices. Its details can be found in [51].
- Bloomberg commodity index (BCOM). It is designed to be a highly liquid and diversified benchmark for commodities investments. The principal potential benefits of including commodities in a diversified financial portfolio include positive returns over time and low correlation with equities and fixed income. BCOM provides broad-based exposure to commodities as an asset class, since no single commodity or commodity sector dominates the Index. Rather than being driven by micro-economic events affecting one commodity market or sector, the diversified commodity exposure of BCOM potentially reduces volatility in comparison to non-diversified commodity baskets. For further information, see [52].
- Down Jones Industrial Average (DJI). 30-stock, price-weighted index that measures the performance of some of the largest U.S. companies. Except for the Global Industry Classification Standard (GICS) Transportation Industry Group from the Industrials Sector and the Utilities Sector, which are covered by the Dow Jones Transportation Average and the Dow Jones Utility Average, respectively, the index includes constituents from a variety of sectors. It is explained in [53].
- Nasdaq Composite (IXIC) The Nasdaq Composite is a stock market index that includes almost all stocks listed on the Nasdaq stock exchange. Along with the Dow Jones Industrial Average and S&P 500, it is one of the three most-followed stock market indices in the United States. The composition of the NASDAQ Composite is heavily weighted towards companies in the information technology sector [54]. Its methodology can be consulted in [55].
- Morgan Stanley Capital International – MSCI World index (MSCI) The MSCI World Index is a broad global equity index that represents large and mid-cap equity performance across all 23

developed markets countries. It covers approximately 85% of the free float-adjusted market capitalization in each country [56]. This index is described in [57].

- Hang Seng Index (HIS). The most widely quoted gauge of the Hong Kong stock market, includes the largest and most liquid stocks listed on the Main Board of the Stock Exchange of Hong Kong [58]. Its methodology can be read in [59].
- Shanghai stock exchange composite index (SSE). It is composed of all SSE listed stocks, including A Share and B Share stocks. The Index is weighted by total market capitalization to reflect price performance of listed stocks in Shanghai Stock Exchange [60].
- Euronext 100 index (N100). The Euronext 100 Index is the blue chip index of the pan-European exchange, Euronext NV. It comprises the largest and most liquid stocks traded on Euronext. Each stock must trade more than 20 percent of its issued shares over the course of the rolling one-year analysis period. The index is reviewed quarterly through a size and liquidity analysis of the investment universe [61]. More information is given in [62].
- Relation Euro-US Dollar (EurUsd). Historical data of the relation between Euro and USD prices with daily granularity.
- Crude Oil price (CrOil). This is one of the world´s top trading commodities. It trades on the major commodities exchanges in the form of contracts. More details can be found in [63].
- Gold price (Gold). Commodity primarily traded in the New York Mercantile Exchange (NYMEX). It has the peculiarity that is not affected by consumption, meaning that most of the mined gold is still accessible. During financial crisis investors trend to purchase it, since it has proven to be more stable, even though the difficult times. It is explained in [64].
- Daily visits to the English version of the Wikipedia entry “Ethereum” (WkViews). Data based on the records offered by Wikipedia.org [65].
- Google search daily volume for the term “Ethereum” (GTV). Data based on the records offered by Google, reporting the volume of searches done in the U.S. of the term “Ethereum”. Data taken with daily granularity from [66].
- Ethereum transactions fees price (EthGas). Gas refers to the unit that measures the amount of computational effort required to execute specific operations on the Ethereum network. Since each Ethereum transaction requires computational resources to execute, each transaction requires a fee. Gas refers to the fee required to conduct a transaction on Ethereum successfully. Further details about its methodology can be consulted in the referred link [67].
- Ethereum daily transactions volume (ETV). Record of Ethereum transactions registered per day.

A table summarizing all the data presented in this chapter can be found in *chapter “3.6 Feature selection results”,* where it is reminded the data available together with its abbreviations and clarified how it is displayed from left to right in the plots contained in that chapter.

3.4. Data pre-processing

Two tasks have been carried out to pre-processing the data:

a) Fill in missing data: for all market indicators or indices, the data pertaining to weekends and non-working days in general were not available. For those dates, the last known value has been used and the date inserted in our data structure.

1	date,price,open,high,low,vol,change,last
2	2022-03-07,36.45,35.88,36.55,32.59,-,13.98
3	2022-03-04,31.98,31.90,34.65,31.47,-,4.92
4	2022-03-03,30.48,30.52,32.01,29.31,-,-0.85
5	2022-03-02,30.74,34.20,34.41,30.12,-,-7.74
6	2022-03-01,33.32,29.45,35.19,29.44,-,10.51
7	2022-02-28,30.15,32.44,33.51,28.43,-,9.28
8	2022-02-25,27.59,31.68,32.00,26.93,-,-9.00
9	2022-02-24,30.32,37.50,37.79,29.45,-,-2.26
10	2022-02-23,31.02,28.04,31.07,27.20,-,7.67
11	2022-02-22,28.81,31.80,32.04,28.40,-,3.82
12	2022-02-21,27.75,26.66,29.71,26.38,-,-1.28
13	2022-02-18,27.75,26.66,29.71,26.38,-,-1.28
14	2022-02-17,28.11,24.83,28.37,24.76,-,15.73
15	2022-02-16,24.29,25.19,27.09,23.88,-,-5.49

Figure 7.- Original data with information gaps

First, in Figure 7, can be seen gaps in the dates, for instance, 2022-02-19 and 2022-03-05 are missing. Then, in Figure 8 it is possible to see how the missing information has been filled in, filling the values with the last known data. See as example how for the mentioned 2022-02-19, the values have been filled with the data obtained for 2022-02-18.

1	date,price,open,high,low,vol,change,last
2	2022-03-07,36.45,35.88,36.55,32.59,-,13.98
3	2022-03-06,31.98,31.90,34.65,31.47,-,4.92
4	2022-03-05,31.98,31.90,34.65,31.47,-,4.92
5	2022-03-04,31.98,31.90,34.65,31.47,-,4.92
6	2022-03-03,30.48,30.52,32.01,29.31,-,-0.85
7	2022-03-02,30.74,34.20,34.41,30.12,-,-7.74
8	2022-03-01,33.32,29.45,35.19,29.44,-,10.51
9	2022-02-28,30.15,32.44,33.51,28.43,-,9.28
10	2022-02-27,27.59,31.68,32.00,26.93,-,-9.00
11	2022-02-26,27.59,31.68,32.00,26.93,-,-9.00
12	2022-02-25,27.59,31.68,32.00,26.93,-,-9.00
13	2022-02-24,30.32,37.50,37.79,29.45,-,-2.26
14	2022-02-23,31.02,28.04,31.07,27.20,-,7.67
15	2022-02-22,28.81,31.80,32.04,28.40,-,3.82
16	2022-02-21,27.75,26.66,29.71,26.38,-,-1.28
17	2022-02-20,27.75,26.66,29.71,26.38,-,-1.28
18	2022-02-19,27.75,26.66,29.71,26.38,-,-1.28
19	2022-02-18,27.75,26.66,29.71,26.38,-,-1.28
20	2022-02-17,28.11,24.83,28.37,24.76,-,15.73
21	2022-02-16,24.29,25.19,27.09,23.88,-,-5.49

Figure 8.- Data after filling the missing information gaps

b) Scale each data source between 0 and 1, using the *MinMaxScaler* function of the *sklearn.preprocessing* library [68].

Earlier, Figure 6 showed the original data of the closing price of Ethereum, representing in its Y-axis the price in Euros and having therefore values up to almost 5000. In Figure 9 it is possible to observe the difference on the Y-axis after applying the Python MinMaxScaler function between 0 and 1.

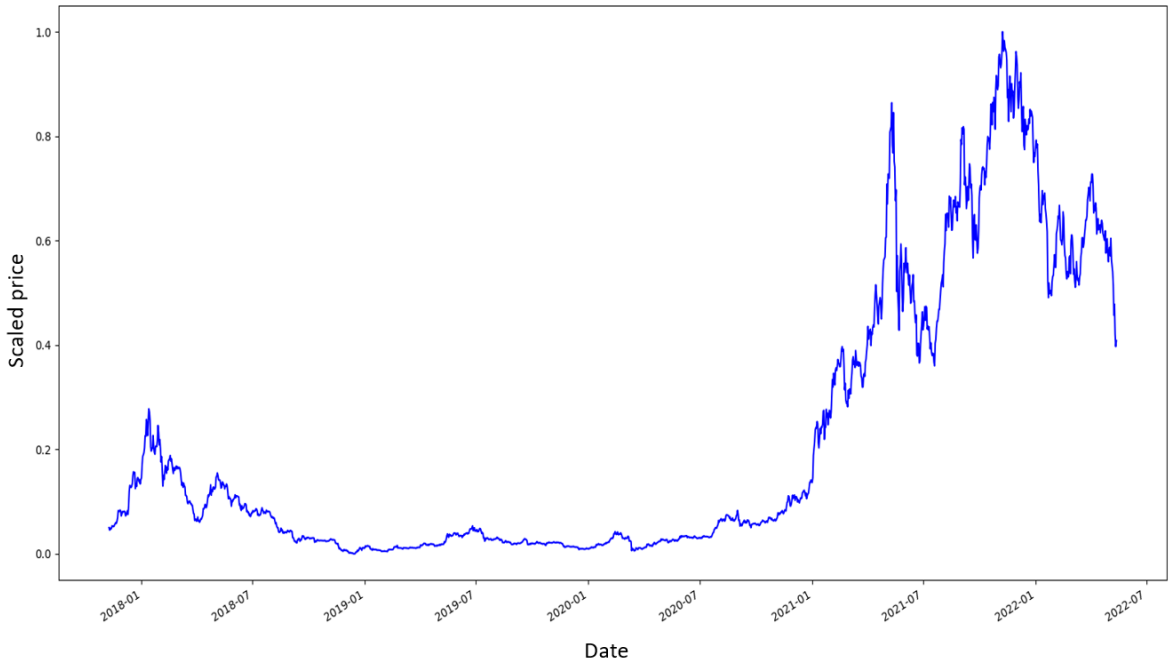

Figure 9.- Data used after scaling it between 0 and 1

Due to the daily granularity of all data and the nature of them, data do not change drastically from one day to another, so elimination of outliers or some filtering have not been necessary.

3.5. Feature selection

As it can be seen in section *"3.2. Data Selected"*, a large number of variables that can be used as inputs of our models has been collected. However, this is not always the most convenient condition to obtain a good performance out of a neural network.

Feature selection methods must be applied in this case, in order to reduce dimensionality of the feature space, gain accuracy in the forecasting results and reduce computational cost.

This work is based on supervised learning models, and it covers two cases:

- "Numerical input - Numerical output" when trying to predict future prices.
- "Numerical input - Categorical output" when trying to predict price trend.

In general, it does not result a simple task to select the best feature selection method, and some works [69] [70] suggest that wrapped methods outperform other techniques for very high dimensional feature spaces. In this case, two methods will be used and compared for each of the cases.

3.5.1. Numerical input – Numerical output

As Wrapped method it will be used the Recursive Feature Elimination implementing Random Forest Regressor. As alternative will be used the Pearson's Correlation Coefficient method.

a. RFE with Random Forest Regressor

Recursive Feature Elimination (RFE) is commonly used together with different classification algorithms (e.g., support vector machine, random forest, etc.) to build more efficient classifiers.

A ranking of features as well as candidate subsets is produced through RFE. A list of accuracy values corresponding to each subset is also generated through this procedure.

The RF comprises multiple decision trees, behaving as an ensemble classifier. For the training process, it uses the boot-strap resampling technique to randomly select sample subsets in each decision tree. The final classification result of the RF is determined by the scores derived from all the decision trees. The classification error depends on every tree's ability of classification, as well as the correlation between the trees.

The RF is a powerful classifier that can perform effectively and efficiently. As compared with other machine learning algorithms the RF has several advantages, such as unbiased estimator, easy to parallel, etc. One key advantage of RF is the importance measure, which reveals the impact of each variable of the predictor. Features with large importance values are ranked as more important than features with small values. The importance provides a method to evaluate the contribution of each feature. It is measured as follows. Firstly, the out-of-bag error is calculated for each decision tree. Then, the values of one feature are permutated across all the test samples, and the out-of-bag error is calculated again. The difference between the two out-of-bag errors measures the importance of that feature. If the error exhibits a large increase, that feature is important [71].

b. Pearson's Correlation Coefficient

The Pearson coefficient was first proposed by statistician Carl Pearson and was originally used as a statistical indicator to analyse and study the degree of linear correlation between variables [72]. There are several different correlation coefficients for different research subjects, and the most commonly used correlation coefficient is the Pearson correlation coefficient used in the paper. The correlation coefficients used in this paper are calculated by the product-difference method, i.e., first calculating the means of two variables separately, and then calculating the deviations between the two variables and their respective means separately, and multiplying the two deviations to obtain the correlation coefficients, which reflect the degree of correlation between the variables, and the Pearson coefficient is the most common one among many kinds of correlation coefficients. According to the characteristics of different correlated variables, different statistical indicators are selected for analysis. For example, if the statistical indicator of linear correlation is selected as the correlation coefficient and its square is used as the determination coefficient, the Pearson coefficient of this two-dimensional vector can be obtained by the formula.

$$\mathrm{r_{xy}} = \frac{\sum_{i-1}^{m}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i-1}^{m}(x_i-\bar{x})^2}\sqrt{\sum_{i-1}^{m}(y_i-\bar{y})^2}} \tag{16}$$

In this formula:

- 'x' is the mean of the elements in vector 1.

- 'y' is the mean of the elements in vector 2.

- 'rx' y is the degree of correlation between different variables.

- 'm' is the number of datapoints in the sequence.

3.5.2. Numerical input – Categorical output

In this case, the wrapped method used will be the Recursive Feature Elimination with Random Forest Classifier, and as alternative method ANOVA will be used.

a. RFE-Random Forest Classifier

The difference between Random Forest Classifier and the RFE-Random Forest Regressor already explained earlier in this section, is that the first uses the Majority Vote for the aggregation step whereas in the second case, it will actually find the mean of the particular output of all its decision trees.

b. ANOVA

Analysis of Variance (ANOVA) is a well-known statistical approach for comparing several independent means [73]. The ANOVA approach ranks features by calculating the ratio of variances between and within groups.

The ratio indicates how strongly the λ feature is linked to the group variables. The following equation is used to calculate the ratio F value of λ th in two benchmark datasets:

$$F(\lambda) = \frac{s_B^2(\lambda)}{s_W^2(\lambda)} \tag{17}$$

Where $s_B^2(\lambda)$ and $s_W^2(\lambda)$ are the sample variance between groups (also known as Mean Square Between, MSB) and within groups (also known as Mean Square Within, MSW), respectively, and can be calculated as,

$$s_B^2(\lambda) = \sum_{i=1}^{K} n_i \frac{\left(\left(\sum_{j=1}^{n_i} \frac{f_{ij}(\lambda)}{n_i}\right) - \left(\sum_{i=1}^{K}\sum_{j=1}^{n_i} \frac{f_{ij}(\lambda)}{\sum_{i=1}^{K} n_i}\right)\right)^2}{df_B} \tag{18}$$

$$s_W^2(\lambda) = \sum_{i=1}^{K}\sum_{j=1}^{n_i} \frac{\left(f_{ij}(\lambda) - \left(\sum_{i=1}^{K}\sum_{j=1}^{n_i} f_{ij}(\lambda) / \sum_{i=1}^{K} n_i\right)\right)^2}{df_W} \tag{19}$$

The degrees of freedom for MSB and MSW are $df_B = K - 1$ and $df_W = N - K$, respectively. The number of groups and the total number of samples are represented by K and N, respectively. The frequency of the λ th feature in the th sample in the j th group is denoted by $f_{ij}(\lambda)$. The number of samples in the j th group is denoted by n_i.

To implement all these techniques of feature selection, the Python library *Scikit* [74] has been used.

3.6. Feature selection results

The feature selection results will be presented for five different prediction cases: one day, seven days, fifteen days, thirty days and ninety days. The study object of this work will keep the focus on the prediction to one day, seven days and fifteen days. The feature selection for predictions to thirty days and ninety days have been added to offer a better overview of how much the day in the future that is intended to be predicted might affect the resulting selected features, and see if there is any method that offers more stable or volatile selection depending on this

parameter. At the same time, it allows us to observe if there are any features that stay always within the selection with independence of the method used and the future day that is intended to be forecasted.

The goal is to have a study of middle-term and long-term forecasting periods, assuming that the short-term is less than a day and it has been discarded for this work based on the state of art. The long-term might seem short, but, since the cryptocurrency prices in general, and the price of Ethereum in particular have proven to be sensitive to market changes and events occurred in the day to day of global economics and geopolitics, it would make no sense to consider trustful predictions for periods of time that are commonly classified as "long-term" (e.g., months or even years in the future).

To understand the plots displayed next, it is important to know that the RFE methods create a system of ranks where the features ranked with 1 are the best (Y-axis of the plots below). Pearson´s and ANOVA make use of a different approach, giving scores to the features. In the case of Pearson´s the selected features are the ones with highest scores, while ANOVA follows the opposite reasoning, and the selected features are the ones with lowest scores. The number of features selected has been given automatically by the Python functions for an optimal result.

The features displayed in the plots, from left to right, are:

Table 1.- Features displayed from left to right in the 'x' axis of plots of this chapter

Name	Meaning
Eth	Ethereum price
Btc	Bitcoin price
BtcEth	Relation of prices Bitcoin-Ethereum
SP500	S&P 500 index
EurUsd	Relation of prices Euro-USD
WkViews	Visits to Wikipedia entry "Ethereum"
VIX	Volatility index
GTV	Google Trends Volume
DJP	Bloomberg commodity index
DJI	Down Jones Industrial Average
IXIC	Nasdaq Composite
MSCI	Morgan Stanley Capital International
HSI	Hang Seng Index
SSE	Shanghai stock exchange composite index
N100	Euronext 100 index
CrOil	Crude Oil price
Gold	Gold price
EthGas	Price Ethereum Gas
ETV	Volume of Ethereum transactions

Once the different FS methods have been applied, we have grouped the features selected by each method in a table. In this way, it is easier to get the overview of how the selected features change with each method depending on the day ahead that will be forecasted.

Additionally, a plot is included for some methods to show how important the non-selected features are, to fully understand the difference between the subsets. It could be that a feature selected for seven days prediction, for example, appears as the most important of the non-selected features

for the fifteen days prediction, which could suggest that even though the selected features are not exactly the same, only one subset could be used for both cases.

3.6.1. Regression approach

Recursive Feature Elimination with Random Forest Regressor

The recursive feature elimination implemented with random forest regressor method shows some difference in the results for the proposed scenarios as can be seen in the Table 2. Some features are clearly not helpful when considering this dataset, as it can be seen that 'DJP' or 'HSI' have not been selected for any prediction. Others as N100 or VIX are clear candidates to be left out if the intention would be to build a group of features that can work with all predictions. And some features as 'Eth' or 'Btc' seem to be inputs that will help with most of predictions, being only not selected for 90 days predictions.

Still, it seems feasible to extract a pack of six or seven features that would work fine for all prediction cases. In Figure 10 can be seen that, for instance, 'Btc' has not been selected for 90 days predictions but it was not far from actually being selected, so 'Btc' could be part of that potential group of features that could work well for all cases. Also, it is interesting to observe that the N100 feature could be left out of a that generic group of features because it has been selected only for one day prediction and its importance is reduced for all other predictions.

Table 2.- Feature selection with RFE-RFR

	RFE-RFR				
	1d	7d	15d	30d	90d
Eth	X	X	X	X	
Btc	X	X	X	X	
BtcEth	X		X		X
SP500	X			X	X
EurUsd	X	X		X	X
WkViews		X	X		
VIX					X
GTV	X		X		
DJP					
DJI		X			
IXIC	X	X	X	X	X
MSCI		X	X	X	X
HIS					
SSE				X	X
N100	X				
CrOil				X	
Gold				X	X
EthGas	X	X	X		X
ETV		X	X		

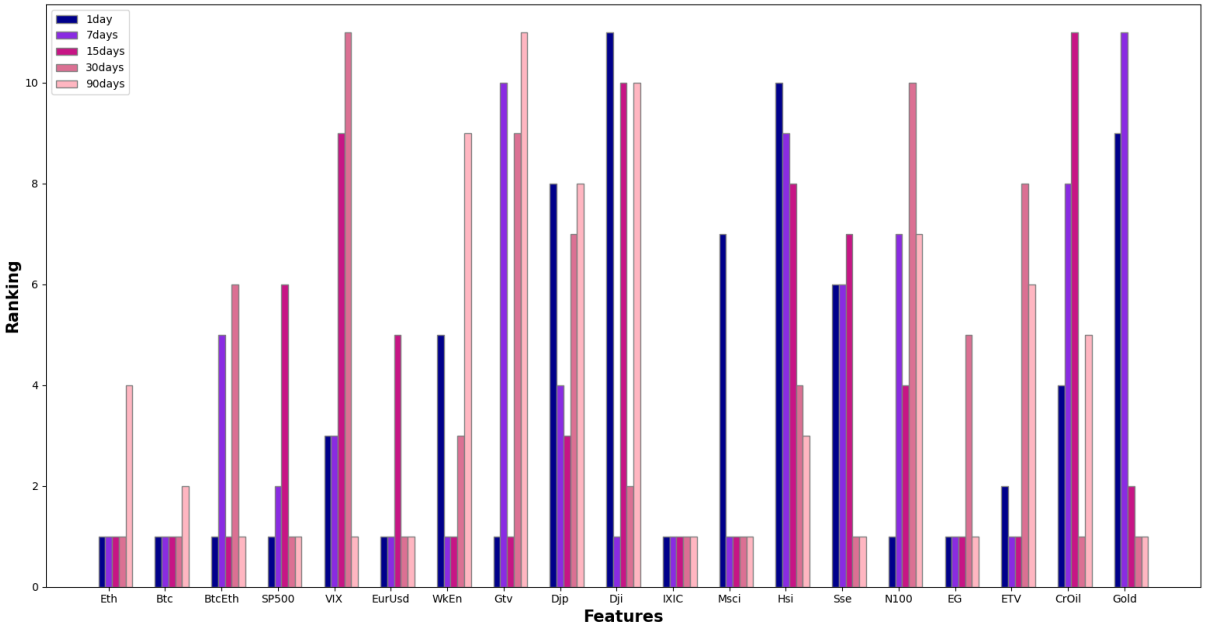

Figure 10.- Feature selection with RFE-RFR method

Pearson's Correlation Coefficient

The Pearson's Correlation Coefficient method suggests a much more stable selection of features for all cases. The selected features according to this method will be the next ten: Btc, Eth, BtcEth, EurUsd, WkViews, GTV, HSI, SSE and ETV. For this method, is not worth to display any plot, since the features are already quite generic, applicable for all prediction cases, although it can be observed that the 90 days prediction would require small adjustments if it would be needed to perform an optimal selection.

Table 3.- Feature selection with Pearson´s

	Pearson´s				
	1d	7d	15d	30d	90d
Eth	X	X	X	X	X
Btc	X	X	X	X	X
BtcEth	X	X	X	X	
SP500					X
EurUsd	X	X	X	X	X
WkViews	X	X	X	X	X
VIX					
GTV	X	X	X	X	
DJP					
DJI					
IXIC	X	X	X	X	X
MSCI					X
HIS	X	X	X	X	
SSE	X	X	X	X	X
N100					
CrOil					
Gold					X
EthGas					

ETV	X	X	X	X	X

3.6.2. Classification approach

Recursive Feature Elimination with Random Forest Classifier

Moving to the FSM for classification approach prediction, the results of the RFE-RFC method are shown first. In the Table 4 is observed that, as its regressive counterpart, this method shows significant differences between the raised forecast cases. It is surprising the big difference between the data selected for one day and seven days predictions, having in common only four features. Using this method, the case of one day predictions received up to four features that are not selected for any other case, which was not seen before. Until now, it was the 90 days prediction case the one with higher deviation from the others. Again, it is good to have a look in the plot of the final rankings created by this method, to have a better understanding of how different the sets of features really need to be.

Table 4.- Summary of Feature selection with RFE-RFC

	RFE-RFC				
	1d	7d	15d	30d	90d
Eth	X	X	X	X	X
Btc	X	X	X	X	X
BtcEth	X	X	X	X	
SP500		X			X
EurUsd		X			
WkViews	X		X		
VIX	X				
GTV	X				
DJP					X
DJI					X
IXIC			X	X	X
MSCI		X	X	X	X
HIS		X	X		
SSE		X		X	
N100				X	X
CrOil			X	X	X
Gold	X	X	X	X	
EthGas	X				
ETV	X				

Analyzing the Figure 11, it doesn´t seem to be possible to organize a group of features that could work well for all predictions. Most of the features selected for one or two prediction cases, are ranked in a bad position for the others. In this case is clear that the best would be to always use the results obtained for each individual case.

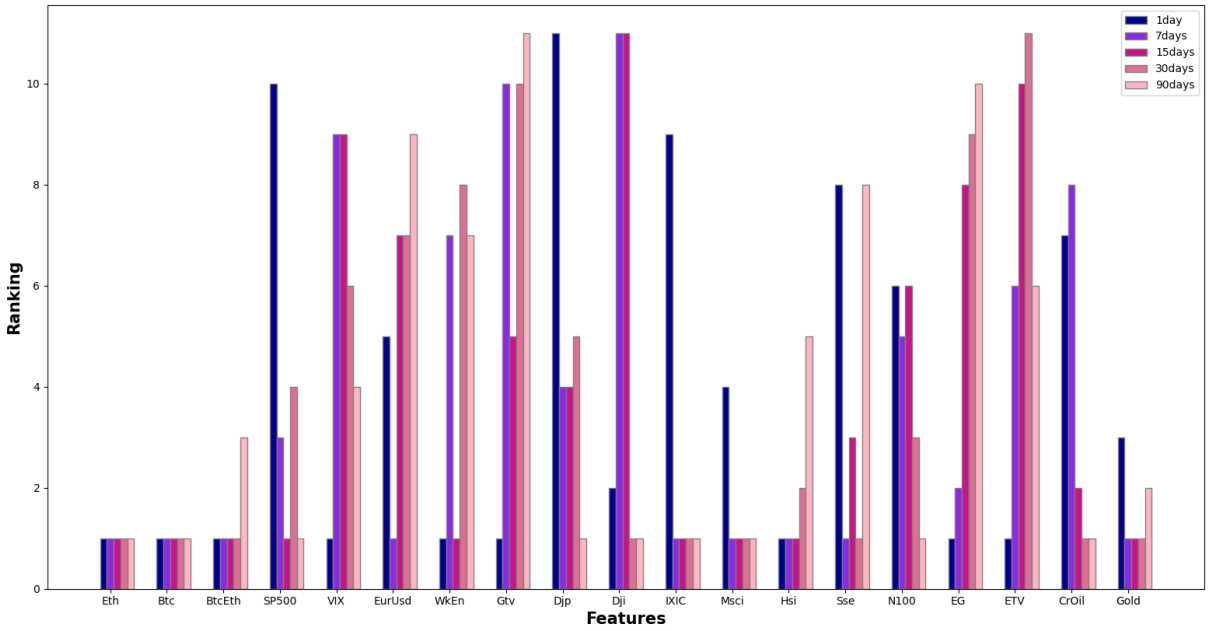

Figure 11.- Feature selection with RFE-RFC

ANOVA

Last but not least, the ANOVA method, also used to select features for prediction models following the classification approach, throws more homogeneous results between the prediction cases studied. There are around 10 features that seem to be almost common for all scenarios and therefore the study of the plot could help to understand how many features could be part of that potential generic group, applicable for all predictions.

Table 5.- Feature selection with ANOVA

	ANOVA				
	1d	7d	15d	30d	90d
Eth	X	X	X	X	
Btc	X	X		X	X
BtcEth	X		X		X
SP500	X		X	X	X
EurUsd	X	X	X	X	X
WkViews	X	X	X	X	X
VIX	X				
GTV	X	X	X	X	X
DJP					
DJI					
IXIC					X
MSCI					
HSI	X	X	X		
SSE		X		X	X
N100	X	X	X	X	
CrOil					
Gold					
EthGas		X	X	X	X
ETV		X	X	X	X

The full scope of the results can be seen in the Figure 12, but the scores for 90 days predictions are much higher than the rest and does not allow to do a proper analysis.

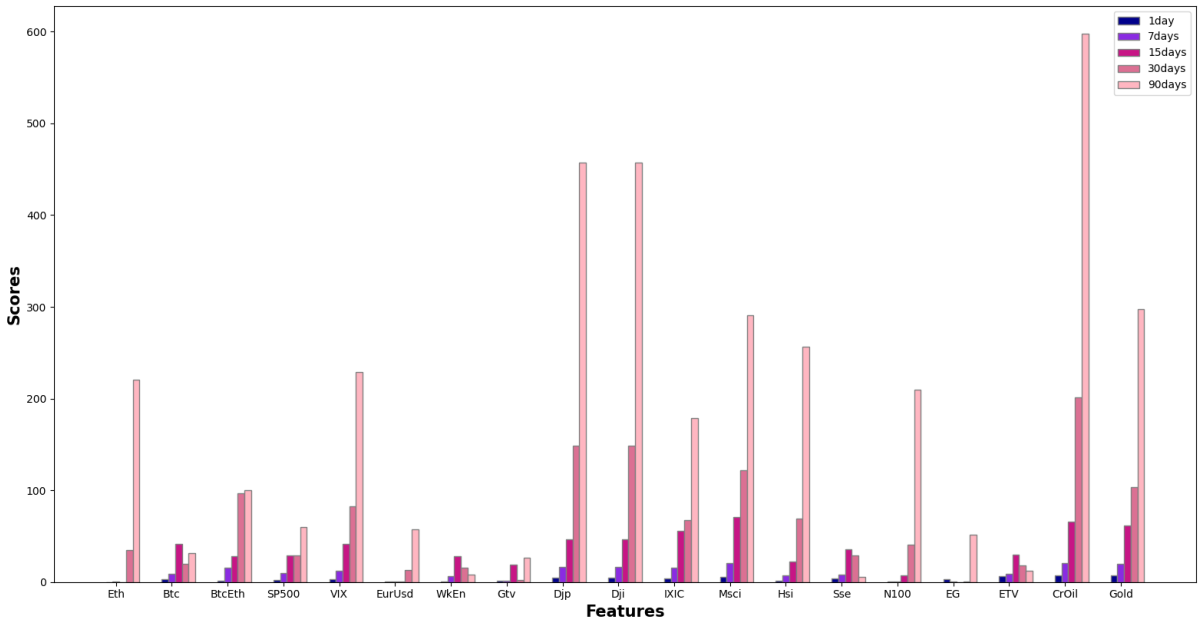

Figure 12.- Feature selection with ANOVA

Removing the feature selection for 90 days predictions to have a better view of the other cases, the Figure 13 is obtained:

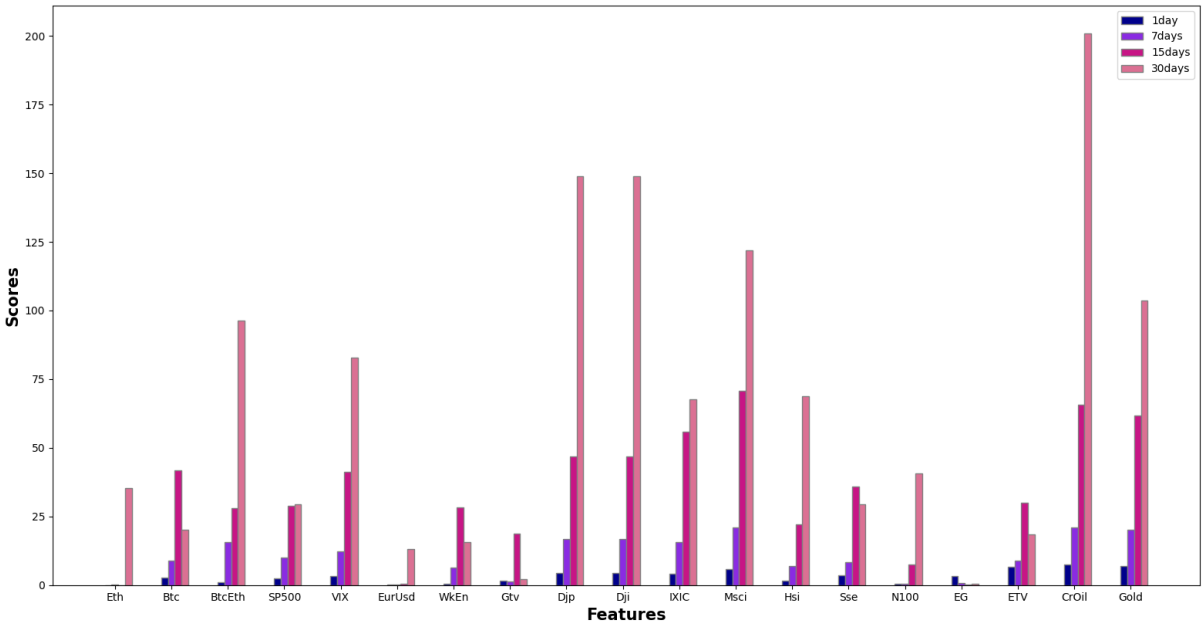

Figure 13.- Feature selection with ANOVA method removing 90 days forecast representation

Still, it is not easy to perform a full analysis of all data, but as an example, it could be studied the HSI feature, that has been selected for 1 day, 7 days and 15 days prediction cases and left out for 30 days and 90 days predictions. In the Figures, it can be seen that in one hand, its score is not far from the score of the last feature that has been selected for each case, and in the other hand, it is much lower than the worse scores present for each case. This means that this feature would be a good candidate to be part of the already mentioned, potential group of features applicable in all cases.

3.6.3. Summary

In this section has been shown that using FSM, the number of inputs can be reduced by approximately 50%. Also, depending of the FSM and the day that wants to be forecasted, the subset of features selected changes. For both approaches, classification and selection, the RFE method applied finds subsets with bigger difference among themselves for the given prediction cases. In three of the four shown FSM, it seems feasible to compose a subset of features that could be applied for all different days ahead predictions.

Now, how good these subsets are, will be studied in the section “4.2 Evaluation”, where this data will be used as input of the ML models to perform the forecast and it will be done a comparation of the obtained error metrics. Each forecast will be done two times per model, using the features selected by each FSM, always with the specific subset that was obtained by method and prediction day. In addition to that, another two forecast cases per model will be done, adding as input features the two data arrays that will be obtained from the sentiment analysis. In total, four forecasts with different features will be performed per model.

Also, in Appendix B can be found a different overview of the results shown in this section, grouping them by prediction day instead of by method. Refer to the Appendix B to see the tables, figures and analysis corresponding to this different approach of the results.

3.7. Sentiment analysis

The initial idea to perform sentiment analysis was to extract data from Twitter, Reddit and Google news. This chapter could be just by itself a whole study object, but the approach used for this work has been to analyze if with the current available tools in Python would be possible to develop a simple solution that improves the results of our forecast. As a result of the effort invested on this part of the work, we have ended with:

3.7.1. Twitter

Twitter is a microblogging and social networking service on which users post and interact with messages known as "tweets", owned by American company Twitter, Inc. Registered users can post, like, and retweet tweets, however, unregistered users have the ability to only read tweets that are publicly available. Users interact with Twitter through browser or mobile frontend software, or programmatically via its APIs. As of June 2022, Twitter ranks as the 6th-most-visited website in the world with 4.9B visits according to Semrush [75].

The idea was to use the API that Twitter offers for developers to interface with the application and its database [76]. The API V2 of Twitter offers three levels of access: Essential, Elevated and Academic Research Access. To search tweets from every day and for the previously described period of time that is used in this work, it would be needed to use the Full-archive search endpoint, which is limited to project with Academic Research Access. This endpoint allows to programmatically access public Tweets from the complete archive dating back to the first Tweet in March 2006, based on a search query.

It was not possible to get this access, since all documentation provided to Twitter was, according to them, not enough to prove any of their requirements:

- A link to my profile in my institution’s faculty or student directory
- A link to my Google Scholar profile
- A link to my research group, lab or departmental website where I am listed

Therefore, for the realization of this work has not been used any Twitter data.

3.7.2. Reddit

It is an American social news aggregation, content rating, and discussion website. Registered users submit content to the site such as links, text posts, images, and videos, which are then voted up or down by other members. As of June 2022, Reddit ranks as the 8th-most-visited website in the world and 7th most-visited website in the U.S., with 3.7B visits according to Semrush [75].

Using the Python Pushift.io API Wrapper [77], described as a minimalist wrapper for searching public reddit comments/submissions via the pushshift.io API, we did extract with daily granularity comments containing the word Ethereum.

As a result, it was obtained 1646 files with the comments of every day for the time period of this work. After reviewing the comments, many of them have been filtered out by using keywords as ".co.uk", "snapchat" or "@gmail" that in most of cases were spam and would only add noise to our sentiment analysis.

In the *Figure 14* is shown a small sample of the obtained file for one day, where each row is a comment. Each comment has been stored with its UTC time and author, but this information is not visible in this screenshot, to be able to show a larger comment length with bit enough size that makes it readable. Some rows don´t seem to contain the work Ethereum, but actually this is because they have been cut by the screenshot itself.

Observe that some comments seem to be very positive, as for examples one mentioning that "you can build the future of finance on Ethereum", while other comments as for example "wasn´t that vitalik´s talk on Ethereum 2.0 and going forward? I don´t remember a unicorn" seem to show some disappointment about Ethereum technology. Other cases as "If you replaced the word 'bitocoin' with the word 'ethereum' in my reply above, it would still hold true…" seem to be mostly neutral, not providing any feeling/sentiment information.

)"They'd probably port it over to Ethereum so they could take advantage of the superior network effects.How exactly are you classifying a ""successful"" DApp?"
i"Because this is the Internet of money, for real. Bitcoin can't do it. You can basically build the entire upcoming global financial system on Ethereum.
"Forget ""decentralized apps"". Think ""programmable money"". As in, you can build the future of finance on Ethereum. And that's what people are doing already."
'If you replaced the word 'bitcoin' with the word 'ethereum' in my reply above, it would still hold true. There are differences though, in what sense are you asking?
"the reason ""they"" want to keep the energy grid intact is more for law enforcement. power bills offer a valuable source of information for cops looking to bust grower
know, it's the only ethereum token being used by hundreds of thousands of people in a popular production application.Any idea how it works on the back end? Are they
h"Yes and no. I work in consulting and crypto is going to be like 2000, most will be gone but big players will emerge stronger than ever. Bitcoin and Ethereum arenâ€™t go
"As to your first question I'm not entirely sure, I believe it has something to do with the stability of the tech though. If blocks are faster than 12 seconds it will be an inher
"Invest their money into ethereum. They'll have a million each by the time they grow up
i"wasn't that vitalik's talk on ethereum 2.0 and going forward? i don't remember a unicorn
("https://blog.decentraland.org/making-ethereum-more-usable-ethalarm-233660030d30 could be useful for etheroll notifications? for the devs in here
"Fuck it, it's good we're not joined to the hip with Bitcoin. Once Bitcoin comes crashing everybody will be buying Ethereum
4"This fork is similiar to what Ethereum did earlier in October. Not a split of a coin but merely fixing the network.
;""Yes, just like when you send any Bitcoin transaction, there is a small transaction fee. In Ethereum you also pay a transaction fee when interacting with contracts or sen
"Yes, just like when you send any Bitcoin transaction, there is a small transaction fee. In Ethereum you also pay a transaction fee when interacting with contracts or sen
"well i wouldn't recommend picking a ssd for a pc that's 1100$. just a 2th 7200 rpm hdd is fine i think. also the rx580 is hard to get at a reasonnable price at the moment. th

Figure 14.- Comments extracted from Reddit, used for sentiment analysis

3.7.3. Google news

It is an aggregator service developed by Google. It presents a continuous flow of links to articles organized from thousands of publishers and magazines.

The interesting part of this service, unlike the two previous described, is the great collection of news headlines that can be collected out of it, that it could be seen as a more objective view of the context that surrounds the daily evolution of the cryptocurrency's prices.

To get the headlines of the news published by this service, it has been used a technique known as web scrapping, which consists on, through the use of certain python functions, open websites links and extract all their content to then, parse it to a structured language that can be easier to access, in our case XML. Once this is done, it is possible to read the XML text and find the root to the different headlines, so they can be extracted and written to a *.csv file. In this case, each headline has been saved together with its publication date and link. To perform these actions, the python packages have been used:

- Library Requests, that allows to send HTTP/1.1 requests [43].
- Module request of the library Urllib, to open and read URL's [45].
- Module BeautifulSoup of the library bs4, to pull data out of HTML and XML files [46].

In the *Figure 15* is shown a small sample, part of one of the obtained files with headlines belonging to one day. As it can be observed, although the queries have been done for English language, there are other languages present in the files, as Spanish, German, French, Turkish, etc. The files contain strange characters as well, as it can be seen in the screenshot. The screenshot has been edited to display only the headlines, for an easier readability. During the development of this work, it has been attempted to program the use of a VPN that would simulate requests done from America, trying to get results only in English language, but this task could not be finished with successful results. Also, to filter the headlines and improve the quality of the data would mean a much higher effort in this case than it was for the comments extracted from Reddit.

Besides the difficulties, it is possible to observe in the Figure some sentences that, by reading them, clearly suggest positive feeling/sentiment, as "Ethereum Addresses Holding > 1 ETH Reach All-Time High" and "This Ethereum price chart pattern suggest ETH can reach $6.5K in Q4", and some with more neutral feeling as "Ethereum Merge Estimated in May-June 2022 - Developers". In the screenshot can not be found a clear negative headline against the interests of Ethereum, but it is easy to imagine that through the 1646 files obtained, many will turn negative.

Figure 15.- News headers extracted from Google news, used for sentiment analysis

Now, regarding the way to analyze the sentences and get a negative, positive or neutral sentiment result out of them, it has been used VADER Sentiment Analysis. VADER (Valence Aware Dictionary and sentiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media, and works well on texts from other domains [78].

After processing one sentence, the VADER Setiment Analysis returns four values: pos, neu, neg and compound. According to their own documentation [79], the compound score is computed by summing the valence scores of each word in the lexicon, adjusted according to the rules, and then normalized to be between -1 (most extreme negative) and +1 (most extreme positive). This is the

most useful metric if you want a single unidimensional measure of sentiment for a given sentence. Calling it a 'normalized, weighted composite score' is accurate. The pos, neu, and neg scores are *ratios for proportions of text that fall in each category* (so these should all add up to be 1 or close to it with float operation). These are the most useful metrics if you want to analyze the context & presentation of how sentiment is conveyed or embedded in rhetoric for a given sentence.

In the case of this work, the compound was the value used. To get a compound for each day, a loop was programmed to run through the headlines every single day in part and each one of them was processed by the VADER Sentiment Analysis algorithm. At the end, all results were added up and the average calculated, saving these average values all together one file. Even though all values were saved, only the compound value has been used as feature input of our neural networks. These compound values can be seen in the *Figure 16* framed in blue colour.

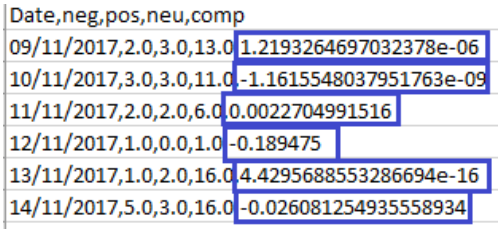

Figure 16.- Results after applying Vader sentiment to collected texts

CHAPTER 4: Methodology Regression approach

4.1. Evaluation: General information

As already mentioned, the focus will be on one day, seven days and fifteen days predictions, so the features selected for 30 days and 90 days are not evaluated.

To evaluate the features selected by each method, each set of features obtained in the "*Chapter 3.6. Feature selection results"* has been used as inputs to train each network.

The selection of the right architecture and tuning of parameters means by itself a topic worthy of further study. In this work, some tests have been done in order to find a fair solution that balances good performance with a small investment in terms of time and effort, therefore no hyperparameter optimization has been carried out.

Usually, between 70 to 80% of the data should be used for training the network and the remaining 20 to 30 % left for testing. In this case, since the data is so limited, having the 20% of data for testing would mean to have less than a year time, and using a 30% would only consider the long period of more stability in the price of Ethereum leaving the time where the most significant changes have taken place, out. This has an obvious negative impact on the capability of the network to generalize. Therefore, the goal has been to use a shorter period of time that could give us as much information as possible and represent a good variety of behaviours on the Ethereum price evolution. Only 650 samples out of the 1647 have been used for training, what represents a 39 % of the total data. The training data is the one represented with shadowed background again on the Figure 17, and it covers the period that goes from 08/10/2019 until 19/07/2021.

This selection leaves two very different periods for testing. The first with almost two years of data with mostly small changes and the second with around ten months of big jumps in the price of Ethereum. These two periods will be used and analyzed independent from each other when the investment strategy is applied.

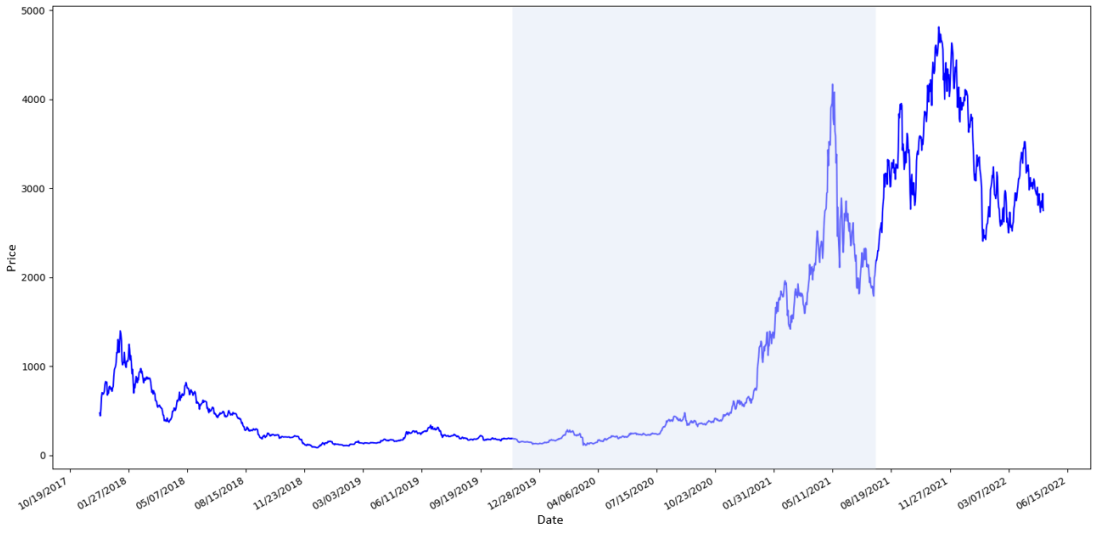

Figure 17.- Ethereum data and training data selection

4.2. Evaluation Regression approach

The network architecture used in this work consists of one input layer and one hidden layer, both with 128 nodes and Dropout regularization with rate 0.2 that helps to prevent overfitting, plus a Dense layer to reduce the output to only one value. The input and the hidden layers will be LSTM or GRU depending on the implemented model.

The LSTM and GRU models have been fit using the efficient Adam optimization and the mean squared error loss function, using a batch size of 32 and 60 epocs.

The selected metrics to evaluate the forecast results obtained from the networks are: Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). MSE calculates the average square value between ground truth and the predicted values. MAE demonstrates the percentage difference between predicted variables, while RMSE computes percentage difference between actual and predicted value.

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|e_i| \tag{20}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(e_i)^2 \tag{21}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(e_i)^2} \tag{22}$$

$$e = y - \bar{y} \tag{23}$$

The metrics resulting of stating the forecast as a regression problem can be found sorted by combination of different conditions in tables. But first, a graphical representation of those results is shown in the Figure 18, where it can be seen how the networks LSTM (green) and GRU (red) have predicted the evolution of the price of Ethereum. Again, consider that the time frame that goes from 08/10/2019 until 19/07/2021 has been used for training.

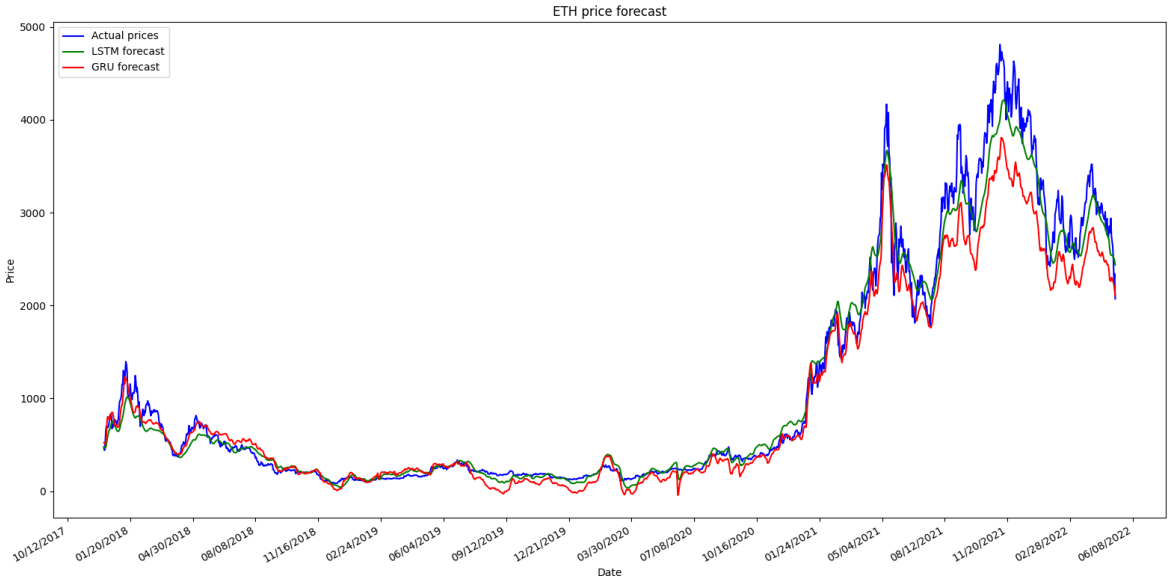

Figure 18.- Regression forecast comparation between LTSM, GRU and actual prices

One day prediction without sentiment

Table 6.- Metrics of one day price prediction without sentiment

	LSTM			GRU		
	MSE	RMSE	MAE	MSE	RMSE	MAE
			1st period			
RFE-RFC	0.045	0.207	0.015	0.057	0.228	0.019
Pearson´s	0.095	0.298	0.024	0.157	0.379	0.033
			2nd period			
RFE-RFC	1.382	1.152	0.101	1.152	1.050	0.092
Pearson´s	1.688	1.247	0.111	1.236	1.081	0.097

One day prediction with sentiment

Table 7.- Metrics of one day price prediction with sentiment

	LSTM			GRU		
	MSE	RMSE	MAE	MSE	RMSE	MAE
			1st period			
RFE-RFC	0.055	0.228	0.017	0.083	0.273	0.023
Pearson´s	0.094	0.299	0.024	0.161	0.379	0.033
			2nd period			
RFE-RFC	1.541	1.200	0.106	1.304	1.120	0.098
Pearson´s	1.844	1.312	0.118	1.251	1.092	0.098

When forecasting one day lag under the regression approach with LSTM and GRU networks, it is to observe that:

- LSTM outperforms GRU in the first period, but GRU performs better in the second period.
- The features selected by the RFE method have an average overall better performance. Just in the case of using GRU network with sentiment analysis, the features selected by Pearson´s method result in a better performance, but only in the second period.
- The predictions made using sentiment analysis are worse than the predictions made without sentiment analysis.
- With or without sentiment analysis, both networks have smaller errors in the first period, which can be understandable looking at the evolution of the prices in this time frame.

Seven days prediction without sentiment

Table 8.- Metrics of seven days price prediction without sentiment

	LSTM			GRU		
	MSE	RMSE	MAE	MSE	RMSE	MAE
			1st period			
RFE-RFC	0.231	0.468	0.037	0.200	0.430	0.035
Pearson´s	0.120	0.340	0.026	0.231	0.466	0.040
			2nd period			
RFE-RFC	1.884	1.327	0.113	1.347	1.136	0.096
Pearson´s	1.905	1.326	0.112	2.499	1.546	0.138

Seven days prediction with sentiment

Table 9.- Metrics of seven days price prediction with sentiment

	LSTM			GRU		
	MSE	RMSE	MAE	MSE	RMSE	MAE
			1st period			
RFE-RFC	0.376	0.602	0.049	0.282	0.512	0.042
Pearson´s	0.123	0.344	0.025	0.200	0.430	0.036
			2nd period			
RFE-RFC	1.719	1.257	0.107	1.479	1.192	0.101
Pearson´s	2.049	1.388	0.118	2.612	1.591	0.142

Predictions to seven days show that:

- GRU improves the performance of LSTM when both are using features selected by RFE method, while LSTM offers better performance than GRU when using the features selected by Pearson´s. This behaviour is present for both periods and with independence of using or not using sentiment analysis.
- Again, the predictions using sentiment analysis offer, in general, worse results.
- The results show a worsening compared to the predictions made to one day in the future.
- Following the same pattern as the previous results for one day prediction, the second period obtains worse results than the first period, with and without sentiment analysis.

Fifteen days prediction without sentiment

Table 10.- Metrics of fifteen days price prediction without sentiment

	LSTM			GRU		
	MSE	RMSE	MAE	MSE	RMSE	MAE
			1st period			
RFE-RFC	0.180	0.420	0.031	0.175	0.409	0.031
Pearson´s	0.192	0.427	0.030	0.235	0.475	0.035
			2nd period			
RFE-RFC	2.081	1.404	0.116	1.679	1.271	0.104
Pearson´s	1.891	1.355	0.110	2.366	1.512	0.126

Fifteen days prediction with sentiment

Table 11.- Metrics of fifteen days price prediction with sentiment

	LSTM			GRU		
	MSE	RMSE	MAE	MSE	RMSE	MAE
			1st period			
RFE-RFC	0.393	0.615	0.049	0.382	0.601	0.050
Pearson´s	0.198	0.431	0.030	0.273	0.505	0.037
			2nd period			
RFE-RFC	1.783	1.312	0.105	1.625	1.249	0.102
Pearson´s	2.522	1.525	0.124	2.483	1.544	0.129

Predictions to fifteen days show that:

- GRU in combination with the features selected by RFE obtain the best performance for both, the first period and the second period when no using sentiment analysis data.
- Using sentiment analysis, LSTM with features selected by Pearson´s are the best option for the first period, but again GRU with RFE are the best on the second period.
- One more time, the predictions made using sentiment analysis obtain in general slightly worse results than the forecast without sentiment analysis data.
- In line with the previous results, the second period shows higher deviations between the predicted prices and the actual prices.

In general, it is observed that:

- The features selected by RFE method result in smaller prediction errors.
- Using these features, the GRU network is the one with best performance for all three prediction lags.
- It has not been possible to obtain any advantage using sentiment analysis, and the errors have been actually bigger when adding it.

- Regarding the prediction lag, the results for one day prediction is the best, while predictions to seven days and fifteen days throw similar errors.
- The forecasts obtained for the first period have smaller errors than the ones obtained for the second period.

4.3. Investment strategy: General information

To have a good investment strategy that suits well the forecast data, is a key factor. It could be that even not having good metrics measurements, the investment strategy throws profitable results. As the main purpose of this work is to study if the use of deep learning can somehow help to have a reliable investment system, the focus has been to find a strategy that using the obtained forecast data, would be profitable.
Again, in this step it is possible to go as far as desired and create really complex strategies, but the approach of this study has been to keep it simple. Below, it is presented the investment strategy developed for both approaches: the regression and the classification cases, but first it is introduced the metrics used to study the performance of the investment strategies, since they are common for both approaches.

Earnings

It is the difference between the final capital and the initial capital. It can be positive in case of obtaining profit or negative is there is a loss.

$$\text{Earnings} = \text{Initial fiat} - \text{Final fiat} \tag{24}$$

% Earnings

It is the earnings expressed in %, this means the percentage over the initial fiat that has been obtained as earnings.

$$\%\ \text{Earnings} = (\text{Earnings}*100) / \text{Initial fiat} \tag{25}$$

Max drawdown

Maximum difference between a maximum amount of capital owned and the minimum amount of capital reached after that maximum and before the capital surpasses the mentioned maximum.

$$\text{Max drawdown} = \text{Max}[(P_i - L_i)/P_i] \tag{26}$$

P = Local maximum value.
L = Local minimum value reached before the previous P is surpassed.
i = iteration of all local maximum values in the serie

Profit factor

Difference between the accumulated profit and the accumulated loss. This performance metric relates the amount of profit per unit of risk, with values greater than one indicating a profitable system.

$$\text{Pf}^* = \frac{\sum_{i=1}^{n} P_i}{\sum_{k=1}^{m} L_k} \tag{27}$$

*Pf: profit factor; P: profit when an operation has been profitable; L: loss when an operation has ended with loss.

Average profit

Profit average per operation, considering as an operation the process of buying Ethereum and selling it back.

$$\text{AvgP} = \frac{\sum_{i=1}^{n} P_i}{nOperations} \tag{28}$$

Average loss

Loss average per operation, considering as an operation the process of buying Ethereum and selling it back.

$$\text{AvgL} = \frac{\sum_{k=1}^{m} L_k}{nOperations} \tag{29}$$

Furthermore, also the number of buying operations, selling operations, stoploss triggered and final fiat will be shown for each case.

4.4. Investment strategy: Regression approach

Under this approach the work is done with price prediction and as it can be seen in *Figure 18*, the behaviour predicted by the networks has been very similar to the actual evolution of the Ethereum price. The price predicted by itself is not so important, since at the end is common to work with percentage values, rather than with absolute values.

The solution implemented recognizes local maximums or minimums in the forecast data, understanding those peaks as the price that on a specific day is higher or lower than the two previous and the two following days in the case of maximum and minimum respectively. The search will be done for each iteration (t) in the third day before the current iteration day (t-3). Following this logic, the two days after the maximum or minimum found, will be precisely the two days before our current iteration day. To end up accepting the maximum or minimum of our forecast data, the current iteration day must follow as well the trend of the previous two days, in order to confirm that the peak found is indeed stable.

To do this, the system uses the forecast data received from the previous step, and it will iterate through it looking only to the past forecast data and to the forecast data of "today", never making use of the forecast data of days in the future, since in a real scenario that would not be available. The forecast data of today is understood to be the value that the Ethereum price will have in the future day that has been defined (might be one day, seven days or fifteen days), and therefore the peak found three days before the current iteration day corresponds also to the future. Of course, this does not apply for the forecast to one day in the future. In that specific case, the system places the order for as soon as possible after the maximum or minimum has been recognized, which is the day after the current iteration day.

This search for maximums and minimums starts on the fifth day of the available data, so the needed previous lags have been already forecasted.

Additionally, a stoploss strategy has been implemented and, after some initial testing, set to the seven percent of the value of the Ethereum owned at the time. To select this value, the balance

between two factors was the main focus: first, the number of stoploss triggered when running the investment strategy, trying to reduce them as much as possible to let the investment strategy be the one selecting the points when to sell Ethereum, and second, to avoid big loses when the investment strategy could not predict correctly a drop in the price. The stoploss is raised every time that the value of the owned Ethereum increases. The value of the owned Ethereum is known in this strategy as virtual fiat, and it is recalculated every day according to its new price. If the new virtual fiat is less than the previously calculated stoploss, a selling order is triggered on that very same day and the stoploss becomes the actual fiat, while virtual fiat goes to zero, since all Ethereum is being sold.

In this strategy, there will always be either only fiat or only Ethereum (virtual fiat), not having any progressive buying system based on trust or selling system to guarantee a minimum profit.

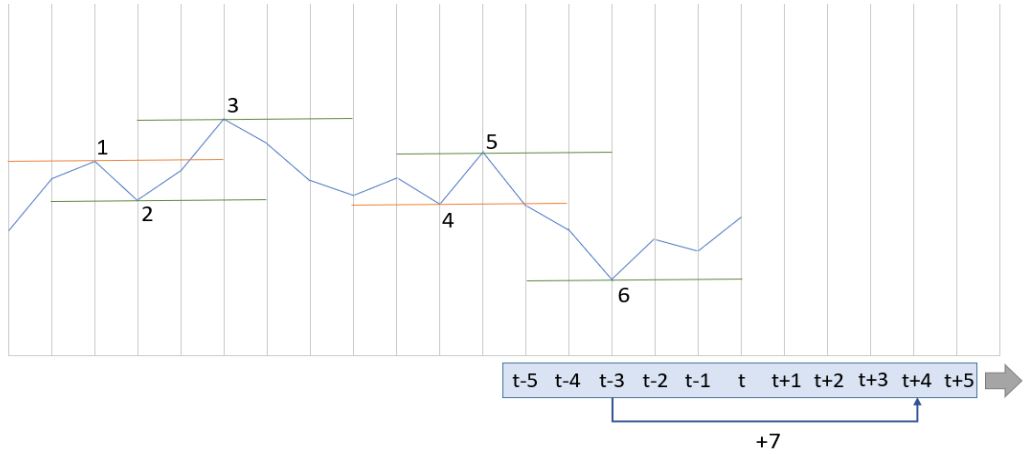

Figure 19.- Graphical representation of regression investment strategy

In the *Figure 20* can be seen a representation of what has been explained in previous paragraphs. The points 1 and 4 would be a maximum and a minimum, but they would not apply for this strategy, since in the third lag after they take place, the price of Ethereum does not follow the trend as it would be needed. Points 3 and 5 are maximums and 2 and 6 minimums, all of them meeting the stated conditions required to place selling and buying orders respectively. Under the plot, it has been represented the shifting cluster of lags that will iterate through time finding the peaks and placing the orders. In this case, it is being found the minimum number 6, corresponding to t-3, and having a prediction to 7 days later, the buying order will be placed in t+4. In this way, four days later the strategy will invest all owned fiat on buying Ethereum.

Now, having run each network with the different features and with both scenarios, with and without sentiment analysis one hundred times, each iteration has also run the resulting forecast values through the presented investment strategy. At the end, all data has been collected and the average calculated and summarized in the following tables:

One day prediction without sentiment

Table 12.- Investment results for regression one day prediction with RFE features, without sentiment

RFE-RFR	1st period		2nd period	
Network	LSTM	GRU	LSTM	GRU
Final Fiat	28191.95	45431.01	12147.00	7974.71
Earnings	18191.95	35431.01	2147.00	-2025.28

% Earnings	181.91	354.31	21.47	-20.25
Max. Drawdown (%)	18.00	29.72	22.98	34.08
Profit factor	2.32	2.48	2.10	0.70
Avg. Profit	1185,72	1418,78	453,12	289,4
Avg. Loss	519,14	571,84	228,64	410,28
Buy Operations	27.72	42.23	10.08	16.31
Sell Operations	26.85	41.58	10.08	16.31
StopLoss	18.15	15.93	4.71	8.61

Table 13.- Investment results for regression one day prediction with Pearson´s features, without sentiment

Pearson´s	1st period		2nd period	
Network	LSTM	GRU	LSTM	GRU
Final Fiat	31653.23	81466.81	14249.73	11088.11
Earnings	21653.23	71466.81	4249.73	1088.11
% Earnings	216.53	714.66	42.49	10.88
Max. Drawdown (%)	24.24	20.35	18.95	30.61
Profit factor	2.27	3.26	3.73	1.20
Avg. Profit	1189,72	2135,78	781,26	428.7
Avg. Loss	533.8	673,24	239.92	357,32
Buy Operations	33.56	48.93	8.0	15.58
Sell Operations	33.38	48.58	8.0	15.58
StopLoss	17.37	16.2	4.4	8.35

Can be observed that in most of cases the strategy has been highly profitable. It results surprising and very remarkable the profit obtained when using the GRU network combined with the features selected by the Pearson´s method during the first period. In the *Appendix B* can be seen a log of all iterations of the investment strategy for this specific case, so it can be seen how it ended up reaching these earnings.

It is also curious to see how, although in general the results are very positive, there is a case that returns a loss (GRU with RFE in the second period). Also, it is worth to mention that the GRU network has generated a higher volume of operations in compared with LSTM, and that the LSTM network seems to be more adequate for the second period while the GRU performs better on the first period.

Looking for the best overall performance, the best option to be safe would be to use the features selected by Pearson´s with a LSTM network, since it has a very solid performance in terms of risk, as it can be seen through the calculated profit factor. In the other hand, it could also be an option to pursue better earnings under higher risk, and then the GRU network together again with the Person´s features could be selected.

One day prediction with sentiment

Table 14.- Investment results for regression one day prediction with RFE features, with sentiment

RFE-RFR	1st period		2nd period	
Network	LSTM	GRU	LSTM	GRU
Final Fiat	30580.78	33030.72	13153.05	8378.12
Earnings	20580.78	23030.72	3153.05	-1621.87
% Earnings	205.80	230.30	31.53	-16.21
Max. Drawdown (%)	23.97	45.12	17.29	32.76
Profit factor	2.34	1.82	2,97	0.76
Avg. Profit	1048,38	965,6	271.59	156.69
Avg. Loss	450,44	526,54	91,26	205.33
Buy Operations	35.52	53.85	9.38	16.63
Sell Operations	35.36	53.67	9.38	16.63
StopLoss	18.42	20.02	4.40	7.68

Table 15.- Investment results for regression one day prediction with Pearson´s features, with sentiment

Pearson´s	1st period		2nd period	
Network	LSTM	GRU	LSTM	GRU
Final Fiat	28134.64	47297.83	14813.86	10696.88
Earnings	18134.64	37297.83	4813.86	696.88
% Earnings	181.34	372.97	48.13	6.96
Max. Drawdown (%)	31.94	38.27	16.98	28.45
Profit factor	2.10	2.15	5.16	1.14
Avg. Profit	948,6	1290,2	809,56	389,7
Avg. Loss	460,62	594,36	188,94	344,98
Buy Operations	38.83	54.81	7.83	16.68
Sell Operations	38.82	54.65	7.83	16.68
StopLoss	18.73	18.85	4.16	7.27

In line with the metrics shown in the *4.2 Evaluation Regression approach,* the results obtained when using sentiment analysis don´t improve the results when no sentiment analysis is used.

In any case, as it happened before, the strategy had mostly positive earnings and again, the LSTM network has performed better under conditions of higher volatility while GRU has offer better performance in the first periods, when the Ethereum price has a more stable evolution.

It is observed that, as in the previous case, the best option being a bit more conservative, would be to use the features selected by Pearson´s in combination with LSTM architecture, which in this case is even more clear if we look again to the profit factor. If the decision would be to search the best earnings assuming more risks, again the Pearson´s features would be the way to go, but this time in combination with a GRU network.

Seven days prediction without sentiment

Table 16.- Investment results for regression seven days prediction with RFE features, without sentiment

RFE-RFR	1st period		2nd period	
Network	LSTM	GRU	LSTM	GRU
Final Fiat	19120.35	11568.01	12273.81	13153.78
Earnings	9120.35	1568.01	2273.81	3153.78
% Earnings	91.20	15.68	22.73	31.53
Max. Drawdown (%)	31.92	56.85	13.73	22.20
Profit factor	1.76	1.06	2.95	1.81
Avg. Profit	667.50	429.94	522,08	461,50
Avg. Loss	376,22	394,68	202.04	265.50
Buy Operations	32.18	55.30	7.33	15.55
Sell Operations	31.20	54.30	7.33	15.55
StopLoss	17.82	28.66	5.22	7.76

Table 17.- Investment results for regression seven days prediction with Pearson´s features, without sentiment

Pearson's	1st period		2nd period	
Network	LSTM	GRU	LSTM	GRU
Final Fiat	20910.54	21191.29	12610.26	16267.86
Earnings	10910.54	11191.29	2610.26	6267.86
% Earnings	109.10	111.91	26.10	62.67
Max. Drawdown (%)	35.51	40.43	17.76	15.57
Profit factor	1.93	1.67	3.51	3.17
Avg. Profit	787,28	635.38	779.62	645.80
Avg. Loss	409,94	377.72	254.66	210.64
Buy Operations	30.25	44.95	5.18	15.18
Sell Operations	29.51	43.95	5.18	14.80
StopLoss	16.512	19.15	3.76	6.612

Again, the results are very positive, but this time the earnings in the first period are lower than the ones derived of the forecast to one day lag, while for the second period are in average higher, but actually due to the improvement given in the GRU network, since the LSTM stays with very similar results are before.

The profit factor suffers a worsening in the first period for both networks, while is clearly better in the second period when using GRU, not having any case of loss.

GRU combined with Pearson´s throws the best earnings, again with -this time slightly- higher risk than LSTM. For this prediction case, it could be done again the differentiation between the conservative and more aggressive strategy by using LSTM or GRU respectively, always with Pearson´s features.

Seven days prediction with sentiment

Table 18.- Investment results for regression seven days prediction with RFE features, with sentiment

RFE-RFR	1st period		2nd period	
Network	LSTM	GRU	LSTM	GRU
Final Fiat	16041.14	9414.00	13599.36	12922.87
Earnings	6041.14	-585.99	3599.36	2922.87
% Earnings	60.41	-5.85	35.99	29.22
Max. Drawdown (%)	36.81	58.11	11.64	28.82
Profit factor	1.38	0.96	3.86	1.81
Avg. Profit	515.32	351.36	662.66	440.78
Avg. Loss	371.34	359.20	194.72	252.78
Buy Operations	42.32	59.82	7.92	15.66
Sell Operations	42.3	59.0	7.92	15.63
StopLoss	22.17	24.68	4.57	6.85

Table 19.- Investment results for regression seven days prediction with Pearson´s features, with sentiment

Pearson's	1st period		2nd period	
Network	LSTM	GRU	LSTM	GRU
Final Fiat	21947.88	14342.31	14200.68	14900.93
Earnings	11947.88	4342.31	4200.68	4900.93
% Earnings	119.47	43.42	42.00	49.00
Max. Drawdown (%)	34.57	52.28	11.41	20.73
Profit factor	1.69	1.22	10.77	2.87
Avg. Profit	789.98	472.32	1025.30	553.56
Avg. Loss	459.16	384.84	95.12	204.18
Buy Operations	37.07	51.40	4.77	15.68
Sell Operations	37.06	51.27	4.70	14.95
StopLoss	19.78	21.13	3.11	5.36

Making use of sentiment analysis does not clearly bring improvement over not using sentiment analysis. Nevertheless, it does improve the earnings obtained by LSTM with Pearson´s features and gets to be the best option for any forecast done to seven days lags. The most remarkable improvement is the outstanding profit factor gotten for the second period when using LSTM with Pearson´s, which means an extremely robust and trustfully performance of the strategy, even though for the first period would be classified almost as weak.

Fifteen days prediction without sentiment

Table 20.- Investment results for regression fifteen days prediction with RFE features, without sentiment

RFE-RFR	1st period		2nd period	
Network	LSTM	GRU	LSTM	GRU
Final Fiat	12944.95	8659.39	12112.30	16229.93
Earnings	2944.95	-1340.60	2112.30	6229.93
% Earnings	29.44	-13.40	21.12	62.29
Max. Drawdown (%)	29.18	39.84	11.30	15.13
Profit factor	1.37	0.90	3.57	3.13
Avg. Profit	359.84	242.16	519.64	869.30
Avg. Loss	260.68	266.56	145.48	297.20
Buy Operations	31.83	53.42	5.56	10.97
Sell Operations	31.80	52.82	5.56	10.97
StopLoss	15.80	21.10	3.01	5.02

Table 21.- Investment results for regression fifteen days prediction with Pearson´s features, without sentiment

Pearson's	1st period		2nd period	
Network	LSTM	GRU	LSTM	GRU
Final Fiat	15284.78	14432.43	12129.69	15050.13
Earnings	5284.78	4432.43	2129.69	5050.13
% Earnings	52.84	44.32	21.29	50.50
Max. Drawdown (%)	31.75	33.64	10.11	16.63
Profit factor	1.54	1.26	4.72	2.69
Avg. Profit	481.70	435.22	567.10	806.24
Avg. Loss	309.52	338.42	119.90	306.44
Buy Operations	30.22	46.11	4.77	10.06
Sell Operations	29.65	45.15	4.77	10.06
StopLoss	14.67	21.57	3.00	4.25

For fifteen days lag predictions, the results given for the first period are as it could be expected, collecting the worst earnings of the three prediction lags. However, for the second period the earnings are actually as good as before or even better. It seems like the earnings get higher when predicting further lags in times of higher volatility, while for periods of time when the price remains more stable happens the opposite.

Once again, combining profit factor and earnings, the best option is LSTM with Pearson´s if the intention is to be more conservative and if the goal is to obtain more profit under some extra of risk, it is worth to use GRU architecture with Pearson´s features.

Fifteen days prediction with sentiment

Table 22.- Investment results for regression fifteen days prediction with RFE features, with sentiment

RFE-RFR	1st period		2nd period	
Network	LSTM	GRU	LSTM	GRU
Final Fiat	14120.84	13877.54	12414.01	12884.22
Earnings	4120.84	3877.54	2424.01	2884.22
% Earnings	41.20	38.77	24.24	28.84
Max. Drawdown (%)	31.71	39.65	11.46	16.80
Profit factor	1.29	1.19	3.67	1.84
Avg. Profit	451.22	375.02	545.54	521.8
Avg. Loss	345.46	309.6	148.6	283.34
Buy Operations	39.98	61.20	6.67	12.82
Sell Operations	38.98	60.55	6.67	12.82
StopLoss	18.80	22.10	3.81	5.75

Table 23.- Investment results for regression fifteen days prediction with Pearson´s features, with sentiment

Pearson's	1st period		2nd period	
Network	LSTM	GRU	LSTM	GRU
Final Fiat	13300.69	15701.08	12002.09	13882.92
Earnings	3300.69	5701.08	2002.09	3882.92
% Earnings	33.00	57.01	20.02	38.82
Max. Drawdown (%)	36.40	44.21	11.16	15.12
Profit factor	1.22	1.25	3.73	2.49
Avg. Profit	445.58	484.94	565.08	604.18
Avg. Loss	361.78	387.18	151.22	242.46
Buy Operations	38.27	57.03	5.43	10.90
Sell Operations	37.36	56.03	5.43	10.90
StopLoss	18.13	20.48	3.60	5.43

Although some specific cases show an improvement, the application of sentiment analysis gives in general worse results. Again, it could be accepted that for a conservative strategy the best would be to work with the forecast given by LSTM with Pearson´s features, and for a more risky but also profitable scenario it would be advisable to use the forecast given by GRU with Pearson´s features, which has been a constant during all studied cases.

In general, it has been seen that both, LSTM and GRU can offer very good performance, but with some clear differences. Speaking about the Max. Drawdown, this metric reveals that both strategies could be improved for the first period, where the LSTM ends with an average of ~32% for 15 days and 7 days predictions, and 26% for 1 day predictions, and the GRU with an average of 39% for 15 days, 51% for 7 days predictions and 33% for 1 day predictions. All these results are pretty bad and it should be studied how it could be avoided to lose so much capital once that it has been earned. Perhaps, modifying the stop-loss percentage could be a good starting point. In this case and being consequent with other results observed in this section, GRU offers Max. Drawdown

numbers and the use of sentiment analysis impoverishes the performance of both models. As for the second period, the results are much better, and it could be seen as acceptable, getting an average below 15% for LSTM and under 20% for GRU. It is also interesting to mention that the average profit per operation is better for short term predictions in periods of small changes and as the prediction horizon increases, the best performance shifts towards periods of higher volatility (2^{nd} period in this case). And again, it can not be concluded with the methodology used in this work, that the sentiment analysis improves the forecast results and therefore the outcome of the investment strategy.

It can be concluded that the optimal combination would be LSTM with features selected by Pearson´s method without sentiment analysis data.

CHAPTER 5: Methodology classification approach

5.1. Evaluation: Classification approach

Before reading this chapter, please refer to chapter "4.1. Evaluation: General information".

The classification approach is based on working with the trend of the prices rather than with the price itself. To do this, to the price of Ethereum at every iteration, it has been subtracted the price of Ethereum "X" number of days later, being "X" the number of days in the future that it is being forecasted (one, seven or fifteen). If the result of that subtraction is positive, means that the price will be higher, and if it is negative, it means that the price will be lower. This has been coded in Boolean variables as True and False respectively and given as the expected output for training, as well as the reference output expected when testing.

The metric used to evaluate the performance of the Support Vector Machine used is the F-measure, which is defined as a harmonic mean of precision (P) and recall (R) [80]:

$$F = \frac{2P\,R}{P + R} \tag{30}$$

Also, to give an overview of the performance, for each case is shown its confusion matrix. By definition a confusion matrix C is such that $C_{i,j}$ is equal to the number of observations known to be in group i and predicted to be in group j. Thus in binary classification, the count of true negatives is $C_{0,0}$, false negatives is $C_{1,0}$, true positives is $C_{1,1}$ and false positives is $C_{0,1}$. A confusion matrix whose i-th row and j-th column entry indicates the number of samples with true label being i-th class and predicted label being j-th class [81].

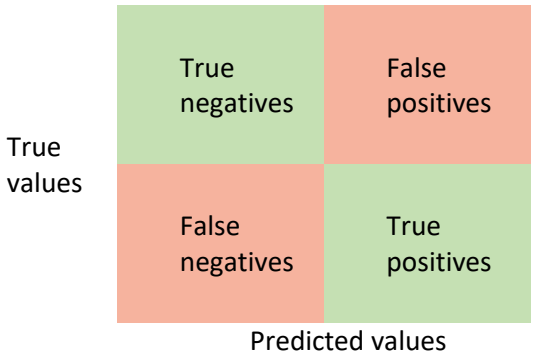

Figure 20.- Graphical explanation of confusion matrix

As done for the regression approach, each SVM model has been used to forecast one day, seven days and fifteen days lags, with and without sentiment analysis and using for all cases the features selected by the RFE method and the ones selected by the ANOVA method. The results are shown in the *Table 12.*

In this case, there is no display of the prediction results, since the most accurate way to show them would be drawing the actual prices and give to each day a colour depending on the sign of the prediction (e.g. green for ascending uptrend and red for downtrend), having to visually connect that colour with what actually happens as many days later as it corresponds to the prediction lag.

Table 24.- Metrics of trend predictions

	RFE-RFC				ANOVA			
1 day								
SVM	without sentiment		with sentiment		without sentiment		with sentiment	
	f1-score: 0.5		f1-score: 0.51		f1-score: 0.48		f1-score: 0.51	
	Confusion matrix		Confusion matrix		Confusion matrix		Confusion matrix	
	604	230	563	271	512	322	581	253
	583	198	516	265	516	256	543	238
7 days								
SVM	without sentiment		with sentiment		without sentiment		with sentiment	
	f1-score: 0.49		f1-score: 0.48		f1-score: 0.51		f1-score: 0.51	
	Confusion matrix		Confusion matrix		Confusion matrix		Confusion matrix	
	478	368	544	302	290	556	413	433
	451	312	540	223	235	528	351	412
15 days								
SVM	without sentiment		with sentiment		without sentiment		with sentiment	
	f1-score: 0.5		f1-score: 0.47		f1-score: 0.52		f1-score: 0.49	
	Confusion matrix		Confusion matrix		Confusion matrix		Confusion matrix	
	485	354	509	330	318	521	375	464
	440	322	517	245	251	511	359	403

It can be observed that in general the features selected by ANOVA method help to obtain slightly better accuracy. Also, the use of sentiment analysis data does not provide any improvement, neither a clear worsening of the predictions. Besides, there is no clear trend on the evolution of the accuracy depending on the lag that is being predicted. Therefore, the results are in general very similar and around the 50% accuracy. In general, it cannot be said that this is a good result for any of the predictions.

5.2. Investment strategy: Classification approach

Again, please refer first to chapter “4.3 Investment strategy: General information” before continue to read this chapter.

In this case, the goal was not to calculate the future price of Ethereum, but its trend. Therefore, the investment strategy had to be readjusted, but the idea has been to follow a similar concept as the regression approach, keeping it simple. The process has consisted on set a buying order when the trend is, for the price, to go higher. Then, when a purchase of Ethereum is done, the strategy blocks new orders for the forecasted time, this means, if the prediction was done for seven days and there was indeed a purchase, during seven days no buying or selling orders will be processed. Actually, there will not be any selling orders, what will happen is that at the end of every blocked period of time, the future trend will be checked, and if the estimation if once again for the price to go higher, the Ethereum is not sold back to fiat, but if the trend says that the price will be lower, the Ethereum is converted back to fiat, blocking again potential orders for buying Ethereum during the forecasted period of time, and so on.

Also, here a stoploss technique has been implemented, so even during the blocked period of time if the price drops a certain percentage, the Ethereum will be converted back to fiat. Again, if the price of Ethereum increases, the stoploss is updated with a new higher value.

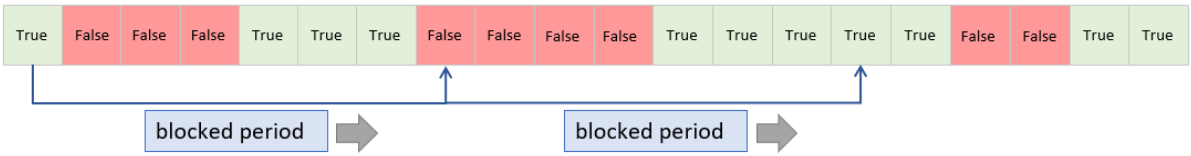

Figure 21.- Graphical representation of classification investment strategy

In the *Figure 21* can be observed how for the first iteration, having a "true" value meaning that the prediction estimates that the price will be higher in seven days, the order would be to invest in Ethereum. Then, the next check would be seven days later, where the forecast results states with a "false" value that the Ethereum price will be seven days later lower, therefore the Ethereum would be sold back to fiat on this iteration before the second blocked period begins. If during the first blocked period, where due to the first "true" value all our assets are Ethereum, the value of the virtual fiat drops below the stoploss, the owned Ethereum would be immediately sold back to fiat, not waiting for the end of the blocked period.

As done for the regression approach, the forecast data obtained in the *4.4. Evaluation* have been used to study how good this investment strategy could be and see if it would be possible to obtain profits with such a simple approach in the two period of time already presented before. The results are shown in the tables below:

One day prediction without sentiment

Table 25.- Investment results for classification one day prediction with RFE features, without sentiment

RFE-RFC	1st period	2nd period
Network	SVM	SVM
Final Fiat	8501.71	11662.68
Earnings	-1498.28	1662.68
% Earnings	-14.98	16.62
Max. Drawdown (%)	30.45	16.58
Profit factor	0.72	1.80
Avg. Profit	199.8	532.98
Avg. Loss	276.06	295.44
Buy Operations	18	7
Sell Operations	17	7
StopLoss	0	7

Table 26.- Investment results for classification one day prediction with ANOVA features, without sentiment

ANOVA	1st period	2nd period
Network	SVM	SVM
Final Fiat	9917.69	14972.85
Earnings	-82.30	4972.85
% Earnings	-0.82	49.72
Max. Drawdown (%)	2.43	33.96
Profit factor	0	1.39
Avg. Profit	0	761
Avg. Loss	82.3	544.78
Buy Operations	1	23
Sell Operations	1	23
StopLoss	0	23

The behaviour shown by the developed strategy reminds to the behaviours obtained in the regression approach, where the first period would have more operations than the second period. This happens when using the features selected by the RFE method. Then, when the features are derived from using ANOVA, the strategy shows a big difference and has only one operation. In any case, this second scenario is the most convenient looking at the earnings obtained for the first and the second periods and calculating the final fiat that would be resulting from the combination of both. In reality, this is not a good approach and it would not be a good way to invest.

One day prediction with sentiment

Table 27.- Investment results for classification one day prediction with RFE features, with sentiment

RFE-RFC	1st period	2nd period
Network	SVM	SVM
Final Fiat	13106.46	12351.85
Earnings	3106.46	2351.85
% Earnings	31.06	23.51
Max. Drawdown (%)	57.08	24.70
Profit factor	1.09	1.62
Avg. Profit	344.9	612.96
Avg. Loss	314.52	377.78
Buy Operations	111	10
Sell Operations	110	10
StopLoss	0	10

Table 28.- Investment results for classification one day prediction with ANOVA features, with sentiment

ANOVA	1st period	2nd period
Network	SVM	SVM
Final Fiat	7805.64	16346.61
Earnings	-2194.35	6346.61
% Earnings	-21.94	63.46
Max. Drawdown (%)	46.03	22.21
Profit factor	0.83	1.97
Avg. Profit	185.08	754.44
Avg. Loss	220.76	381.12
Buy Operations	62	17
Sell Operations	61	17
StopLoss	0	17

Now, using the forecast done with sentiment analysis data, it can be seen how the strategy results profitable using the features derived of RFE method. Still, seems to be quite risky looking at the profit factor, but the earnings are for both periods over 20%, which is a good result.

For the case of using ANOVA to select the features of our system, the prediction would not work well with this strategy, and even though the second period throws a very good profit, the loss of over 20% of the initial investment in the first period would discard this as a good option.

Seven days prediction without sentiment

Table 29.- Investment results for classification seven days prediction with RFE features, without sentiment

RFE-RFC	1st period	2nd period
Network	SVM	SVM
Final Fiat	13078.94	7601.03
Earnings	3078.94	-2398.96
% Earnings	30.78	-23.98
Max. Drawdown (%)	4.23	45.63
Profit factor	inf	0.84
Avg. Profit	1332.24	127.84
Avg. Loss	0	150.9
Buy Operations	3	104
Sell Operations	2	104
StopLoss	0	0

Table 30.- Investment results for classification seven days prediction with ANOVA features, without sentiment

ANOVA	1st period	2nd period
Network	SVM	SVM
Final Fiat	5031.07	7760.42
Earnings	-4968.92	-2239.57
% Earnings	-49.68	-22.39
Max. Drawdown (%)	94.98	29.54
Profit factor	0.65	0.48
Avg. Profit	719.2	101.9
Avg. Loss	1093.42	208.54
Buy Operations	15	21
Sell Operations	14	21
StopLoss	2	0

Working with the SVM forecast for seven days lag without sentiment analysis data, this strategy has a pretty bad performance, giving losses between 20 to 50 % approximately. Only one case ends up being profitable, but this is not enough to conclude that this would be a useful strategy, neither to say that the use of SVM to perform forecasts to seven days lag is appropriate.

Seven days prediction with sentiment

Table 31.- Investment results for classification seven days prediction with RFE features, with sentiment

RFE-RFC	1st period	2nd period
Network	SVM	SVM
Final Fiat	12752.24	8867.35
Earnings	2752.24	-1132.64
% Earnings	27.52	-11.32
Max. Drawdown (%)	12.57	23.80
Profit factor	3.14	0.86
Avg. Profit	807.6	118.54
Avg. Loss	257.16	136.82
Buy Operations	5	62
Sell Operations	5	62
StopLoss	0	0

Table 32.- Investment results for classification seven days prediction with ANOVA features, with sentiment

ANOVA	1st period	2nd period
Network	SVM	SVM
Final Fiat	25672.81	8259.56
Earnings	15672.81	-1740.43
% Earnings	156.72	-17.40
Max. Drawdown (%)	82.49	28.49
Profit factor	1.46	0.86
Avg. Profit	1042.72	136.94
Avg. Loss	710.42	157.4
Buy Operations	47	85
Sell Operations	46	85
StopLoss	16	0

Using sentiment analysis data for this prediction lag makes a big difference on the behaviour of the strategy. It triggers many more operations in the case of using ANOVA to select the features and this derives on excellent earnings with help of the stop-losses. Nonetheless, the weakness of this combination is exposed when the price evolution shows higher volatility, as it does during the second period, and depending on the time frame when the strategy would be applied, there would be an important risk of losing part of the investment. Therefore, even though the sentiment analysis helps once again to improve the performance of the strategy, it would not be advisable to make use of it.

Fifteen days prediction without sentiment

Table 33.- Investment results for classification fifteen days prediction with RFE features, without sentiment

RFE-RFC	1st period	2nd period
Network	SVM	SVM
Final Fiat	6113.27	8596.67
Earnings	-3886.72	-1403.32
% Earnings	-38.86	-14.03
Max. Drawdown (%)	68.30	39.05
Profit factor	0.59	0.87
Avg. Profit	827.56	136.72
Avg. Loss	1382.82	156.76
Buy Operations	7	70
Sell Operations	7	70
StopLoss	2	0

Table 34.- Investment results for classification fifteen days prediction with ANOVA features, without sentiment

ANOVA	1st period	2nd period
Network	SVM	SVM
Final Fiat	7661.96	8596.67
Earnings	-2338.03	-1403.32
% Earnings	-23.38	-14.03
Max. Drawdown (%)	91.54	19.11
Profit factor	0.82	0.87
Avg. Profit	901.98	136.72
Avg. Loss	1096.82	156.76
Buy Operations	12	70
Sell Operations	12	70
StopLoss	5	0

It becomes clear by looking at the results that this approach is not a good idea when trying to invest in Ethereum based on fifteen days lag forecasts. There is no case where the system makes profit and so, there is nothing positive to be extracted out of this case.

Fifteen days prediction with sentiment

Table 35.- Investment results for classification fifteen days prediction with RFE features, with sentiment

RFE-RFC	1st period	2nd period
Network	SVM	SVM
Final Fiat	13517.67	6804.52
Earnings	3517.67	-3195.47
% Earnings	35.17	-31.95
Max. Drawdown (%)	32.36	35.42
Profit factor	1.87	0.63
Avg. Profit	1219.44	97.98
Avg. Loss	652.08	155.04
Buy Operations	7	56
Sell Operations	6	56
StopLoss	3	0

Table 36.- Investment results for classification fifteen days prediction with ANOVA features, with sentiment

ANOVA	1st period	2nd period
Network	SVM	SVM
Final Fiat	19555.35	8340.44
Earnings	9555.35	-1659.55
% Earnings	95.55	-16.59
Max. Drawdown (%)	71.88	22.80
Profit factor	1.20	0.81
Avg. Profit	1674.92	114.96
Avg. Loss	1385.36	141.30
Buy Operations	33	63
Sell Operations	33	63
StopLoss	21	0

As it happened in the previous predictions to one day and seven days, and in opposition of what was observed in the regression approach, the sentiment analysis helps again to improve the performance of the strategy. Also, once again the combination with the features selected by the ANOVA method is the one that benefits the most of making use of the sentiment data. In any case, one more time it is clear that using this strategy could end up being profitable, but sure it would be very risky due to the poor performance that it offers under circumstances as the ones applying in the second period, where actually there is a negative earnings of more than 15%.

It can be observed that the development followed in this work for the trend prediction has not been successful. As the metrics were announcing with around 50% accuracy, the investment strategy implemented has followed that pattern offering close to 50% of cases with profit and the other half with loss. There is no clear winner between the two FSM, since both offer results with too high risk. Again, the addition of sentiment analysis data has not added value to the final results.

CHAPTER 6: Conclusions and future works

6.1. Conclusions

As the cryptocurrency market has been growing and accumulating value, the study and understanding of its behaviour and the variables that might affect it, has gained importance. Along with its growth, an increasingly strong link has been stablished with the traditional stock markets, and this facilitates the use of new forms of prediction that can be as accurate or even more than technical analysis or time series models as ARMA, ARIMA or MA. Although this market is still a young environment, its exponential growth and call for attention have dragged already many authors to study and write works about it, how it evolves and how to forecast it, which has been of great help to elaborate this Master Thesis.

In this context, the supervised learning techniques have taken a great importance and through this work it has been proven that even a simple approach, choosing the right features and training period, can lead to a prediction that does not deviate much from the current price evolution. Indeed, the use of RNN as the presented LSTM and GRU models offers a very good performance and requires low architecture design and hyperparameter tuning effort. Then again, following the same principle of simplifying the study as much as possible for the implementation of the investment strategy, metrics as the Maximum Favourable Excursion and Maximum Adverse Excursion or indicators as the Take Profit have been left out of the scope. Despite all this, the results have shown an extraordinary outcome with earnings that have exceeded expectations at the beginning of this work.

Nevertheless, when it comes to trend prediction, the results have not been that successful, obtaining an average accuracy of around 50% using support vector machine, which is not a good result. The investment strategy used in this work for the regression approach has also revealed some weakness in the obtained forecast data, as it can be seen that numerous times the stoploss was what actually triggered a selling operation, instead of the plan generated by the algorithm after the prediction. This is, however, understandable taking into consideration that both, the neural networks and the investment strategy, have big room for improvement. It is also expected that as time goes on and more historical data is accumulated, in the future it will be possible to have better trained models, and together with the growing connection between cryptocurrency markets and traditional stock markets, the cryptocurrency prices should reduce their volatility and therefore become easier to predict with higher level of accuracy, at least when it is faced as a regression problem. In the case of the classification approach, it has not been possible to generate a solid and trustful strategy that could be applied for the three different intended predictions. Still, it has been shown that it is possible to get profits with the very simple concept developed, but the risks would be high. Nevertheless, this does not mean that it is not possible to design a better system. Quite the opposite. The results suggest that putting some more effort on the steps followed for this work, there would be a good chance to end up building a profitable combination between the prediction system and the investment strategy.

Furthermore, it has been exposed the lack of improvements when using sentiment analysis, mainly in the regression approach. This can be due to the very simple techniques followed to extract data from the online platforms and the missing analysis of important sources as Twitter. The potential of this input seems to be out of doubt, but due to the large amount of data that can be retrieved, it deserves some further study to be able to filter the most significant for each study case.

Clearly, the use of deep learning techniques to predict price rather than trends is a trustful and helpful tool that combined with the right algorithms can reduce risks and increase profits when investing in cryptocurrencies. Besides, the market evolution in the last years, together with the

continuous improvements on the sentiment analysis developments, points towards conditions where would be easier to perform accurate predictions and will make the Deep Learning an essential technology for all those who wish to transact safely in this field.

6.2. Future work

With all this said, it would be a good interesting to perform actions through the whole process to improve each step and tune the results towards better predictions and more adequate investment strategies. Some of the futures lines of work would be defined by:

- Introduction of different inputs that could come from the traditional stock market, global financial indicators, cryptocurrency market, etc. Other features could be found that would provide better information for the training of the neural networks.
- Use not only of neural networks as the RNN but also applying combinations as CNN with RNN, as some studies have been doing recently.
- Hyperparameter optimization. Doing adjustments in the networks hyperparameters could help to learn more accurately the relation between the different inputs with the expected output, increasing the performance of the networks.
- Add sources to the sentiment analysis data. No only Twitter, but also other forums or news feed websites related with the economy field in general and with the cryptocurrencies more specifically.
- Staying in the sentiment analysis topic, another improvement could come from doing a better selection of the terms to be searched, related to the cryptocurrency which price wants to be forecasted.
- Once that the sources have been added and the terms selected, it would be good to implement a proper filter to avoid spam messages and focus the study in only one language. Even extracting metadata as the rating of a comment or the support given to a header can help to filter the most significant texts.
- An extension of the Vader Sentiment tool would be also needed. This tool gives the chance to define dictionaries with words that get weights or scores assigned by the user and then applies these pre-sets to the words if they are found in the analyzed text.
- Implementation of a more complex investment strategy would be another step to take if it is desired to have a more reliable system. As already mentioned, the implementation of metrics as the Maximum Favorable Excursion and Maximum Adverse Excursion or indicators as the Take Profit, plus the tuning of all parameters and percentages, as the ones involved in buying or selling Ethereum, the stoploss, etc. would with no doubt help to determine when to start or end an operation as well as to protect the accumulated profit.

It is indeed a fascinating topic, with multitude of subjects to study and challenges to be overcome, but also very rewarding for those who enjoy exploring the edges of the deep learning. Probably there was never before a better moment to dive and explore the forecasting of cryptocurrencies due to the increasing amount of historical data, the availability of big data coming from online websites, social networks and forums, and the unstoppable number of researches and advances done in the field of the neural networks.

References

[1] Satoshi Nakamoto (2008). Bitcoin: A Peer-to-Peer Electronic Cash System

[2] Follak, K.P.(2022): Crypto Assets: Evolution and Revolution in International Financial Markets, 1(1), 2-4.

[3] Holtfort, T., Horsch, A. Schwarz, J. (2022): Economic, technological and social drivers of cryptocurrency market evolution and its managerial impact, 5-8.

[4] Bajpai, P. (2021): Countries where Bitcoin is legal and illegal, article in Investopedia: www.investopedia.com, last accessed 2022/07/12.

[5] Giudici, G., Milne, A., Vinogradov, D., Cryptocurrencies: market analysis and perspectives, (2020), doi: 10.1007/s40812-019-00138-6.

[6] Rajasekaran, A. S., Azees, M., Al-Turjman, F., (2022), Sustainable Energy Technologies and Assessments, 52.

[7] García-Corral, F.J., Cordero-García, J.A., de Pablo-Valenciano, J., Uribe-Toril, J., (2022),A bibliometric review of cryptocurrencies: how have they grown?, doi: 10.1186/s40854-021-00306-5.

[8] Dr. Houben, R., Snyers, A. (2018), Cryptocurrencies and blockchain, legal context and implications for financial crime, money laundering and tax evasion. Published by Policy Department for Economic, Scientific and Quality of Life Policies, European Parlament.

[9] V. Buterin, "Ethereum white paper: a next generation smart contract & decentralized application platform," 2013.

[10] Y. Sompolinsky and A. Zohar, "Secure high-rate transaction processing in Bitcoin," Financial Cryptography, pp. 507-527, 2015

[11] Ethereum Community, "A next-generation smart contract and decentralized application platform", White paper.

[12] Vujičić, D., (2018), Blockchain Technology, Bitcoin, and Ethereum: A Brief Overview, 17th International Symposium INFOTEH-JAHORINA.

[13] Sifat, I.M., Mohamad, A., Shariff, M.S., (2019), Lead-Lag relationship between Bitcoin and Ethereum: Evidence from hourly and daily data, 50, 306-321

[14] Betancourt, C., Chen, W-H., (2021) Deep reinforcement learning for portfolio management of markets with a dynamic number of assets, doi: 10.3390/app11167377

[15] Price evolution of Ethereum under the market view of Binance, available under: https://www.binance.com/es/trade/ETH_BUSD?_from=markets&theme=dark&type=spot, last accessed on 20.08.2022

[16] R. Naranjo, M. Santos, Expert systems with applications, (2019), 133, 34-48

[17] Abdellaoui, M., *et al,* (2020), Artificial Intelligence approach for Target Classification: A State of the Art., 5(4), 445-456.

[18] Schmidhuber, J., Neural Networks (2015), 61, 85-117.

[19] Hewamalage, H., Bergmeir, C., Bandara, K., Recurrent Neural Networks for Time Series Forecasting: Current status and future directions, (2021), International Journal of Forecasting, 37, 388-427.

[20] Yamak, P., Yujian, L., Gadosey, P., (2019), A Comparation between ARIMA, LSTM and GRU for Time Series Forecasting, doi: 10.1145/3377713.3377722

[21] Duan, K., Keerthi, S., Poo, A., Evaluation of simple performance measures for tuning SVM hyperparameters, (2003), 51, 41-59

[22] Tang et al., Time series forecasting using neural networks vs. Box–Jenkins methodology, 1991, 57, pp. 303-310

[23] Claveria and Torra, Forecasting tourism demand to Catalonia: Neural networks vs. time series models, (2014), 36, 220-228.

[24] Zhang and Kline, Quarterly time-series forecasting with neural networks, 2007, 18, 1800-1814

[25] Mandal et al., A neural network based several-hour-ahead electric load forecasting using similar days approach, 2006, 28, 367-373.

[26] Kimoto T. *et al*, Stock Market Prediction System with Modular Neural Networks, (1990), doi: 10.1109/IJCNN.1990.137535.

[27] Naeini, M., Taremian, H., Hashemi, H., Stock Market Value Prediction Using Neural Networks(2010), doi: 10.1109/CISIM.2010.5643675.

[28] NAMIN, S. , NAMIN A., Forecasting Economic and Financial Time Series: ARIMA vs LSTM, (2018), Texas Tech. University, available under: https:// arxiv.org/ftp/arxiv/papers/1803/1803.06386.pdf, last accessed on 17.07.2022.

[29] Nasirtafreshi, I., Forecasting cryptocurrency prices using Recurrent Neural Network and Long-Short Term Memory, (2022), 139, doi: https://doi.org/10.1016/j.datak.2022.102009.

[30] Scikit-learn, Confusion matrix function website, available under: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html, last accessed on 26.07.2022.

[31] Akyildirim, E., Goncu, A., Sensoy, A., Prediction of cryptocurrency returns using machine learning, (2020), doi: 10.1007/s10479-020-03575-y.

[32] Hansun, S., Wicaksana, A., (2022) Multivariate cryptocurrency prediction: comparative analysis of three recurrent neural networks approaches, doi: 10.1186/s40537-022-00601-7

[33] Sabry, F., *et al*, Cryptocurrencies and Artificial Intelligence: Challenges and Opportunities, (2020), doi: 10.1109/ACCESS.2020.3025211.

[34] Dutta, A., A Gated Recurrent Unit Approach to Bitcoin Price Prediction, (2020), doi: 10.3390/jrfm13020023.

[35] Derbentsev, V., *et al*, (2021), Comparative Performance of Machine Learning Ensemble Algorithms for Forecasting Cryptocurrency Prices, 31(01), 140-148.

[36] Yiying, W., Yeze, Z., Cryptocurrency Price Analysis With Artificial Intelligence, (2019), doi: 10.1109/INFOMAN.2019.8714700.

[37] Jaquart, P., Dann, D., Weinhardt, C., Short-term bitcoin market prediction via machine learning, (2021), 7, 45-66.

[38] Felizardo, L., *et al*, (2019), Comparative study of Bitcoin price prediction using WaveNets, Recurrent Neural Networks and other Machine Learning Methods, doi: 10.1109/BESC48373.2019.8963009

[39] Colon, F., *et al,* The effect of political and economic uncertainty on the cryptocurrency market, (2021), 39, doi: 10.1016/j.frl.2020.101621.

[40] Valencia, F., *et al*, (2019), Price movement prediction of cryptocurrencies using sentiment analysis and machine learning, doi:10.3390/e21060589.

[41] Kraaijeveld, O., The predictive power of public Twitter sentiment for forecasting cryptocurrency prices, (2020), doi: https://doi.org/10.1016/j.intfin.2020.101188.

[42] Yfinance project website, available under https://pypi.org/project/yfinance/, last accessed on 24.07.2022

[43] Python Request library documentation website, available under: https://requests.readthedocs.io/en/latest/, last accessed on 24.07.2022.

[44] Pytrends project website, available under https://pypi.org/project/pytrends/, last accessed on 24.07.2022

[45] Python Urllib library documentation website, available under: https://docs.python.org/3/library/urllib.html, last accessed on 24.07.2022.

[46] Python Beautifulsoup library documentation, website available under: https://beautiful-soup-4.readthedocs.io/en/latest/, last accessed on 24.07.2022.

[47] Python PSAW library documentation website, available under: https://psaw.readthedocs.io/en/latest/#, last accessed on 24.07.2022.

[48] Etherscan home site, available under https://etherscan.io/, last accessed on 24.07.22.

[49] Sabri, F,. *et al, Cryptocurrencies and Artificial Intelligence: Challenges and opportunities, (2020), doi: 10.1109/ACCESS.2020.3025211*

[50] The McGraw-Hill Companies (2011), S&P Index Mathematics Methodology, (2011), available under: https://web.archive.org/web/20111106115820/https://www.sp-indexdata.com/idpfiles/shariah/prc/active/whitepapers/Methodology_Index_Math_Web.pdf, last accessed on 16.07.2022

[51] Cboe website, available under: https://www.cboe.com/tradable_products/vix/faqs/, last accessed on 16.07.2022

[52] Index Methodology, The Bloomberg Commodity Index Family, (2014), available under: https://web.archive.org/web/20150319064733/http://www.bloombergindexes.com/content/uploads/sites/3/content/uploads/sites/3/2014/06/Bloomberg-Commodity-Index-Methodology.pdf , last accessed on 16.07.2022.

[53] Dow Jones Averages Methodology (2022), available under: https://www.spglobal.com/spdji/en/documents/methodologies/methodology-dj-averages.pdf , last accessed on 16.07.2022

[54] English version of the Wikipedia entry "Nasdaq Composite", available under: https://en.wikipedia.org/wiki/Nasdaq_Composite, last accessed on 16.07.2022.

[55] Nasdaq index Methodology Guide, (2021), available under: https://indexes.nasdaqomx.com/docs/Nasdaq_Index_Methodology_Guide%20(1).pdf , last accessed on 16.07.2022.

[56] MSCI World Index (USD) Factsheet, available under: www.msci.com/documents/10199/149ed7bc-316e-4b4c-8ea4-43fcb5bd6523, last accessed on 16.07.2022.

[57] MSCI Index Calculation Methodology, Index Calculation Methodology for the MSCI Equity Indexes, (2020), available under: https://www.msci.com/eqb/methodology/meth_docs/MSCI_IndexCalcMethodology_Jan2020.pdf , last accessed on 16.07.2022.

[58] HSI Factsheet, available under: www.hsi.com.hk/static/uploads/contents/en/dl_centre/factsheets/hsie.pdf, last accessed on 16.07.2022.

[59] Index Methodology, General Guide. For Managing the Hang Seng Family Indexes. (2022), available under: https://www.hsi.com.hk/static/uploads/contents/en/dl_centre/methodologies/index_methodology_guide_e.pdf , last accessed on 16.07.2022.

[60] SSE Indices Calculation & Maintenance, available under: http://www.sse.com.cn/sseportal/en_us/ps/sczn/sse_indices_cal_and_main_en.pdf , last accessed on 16.07.2022.

[61] English version of the Wikipedia entry "Euronext 100", available under: https://en.wikipedia.org/wiki/Euronext_100, last accessed on 16.07.2022.

[62] Index Rule Book, Euronext 100 Index, Next 150 Index, (2015), available under: https://live.euronext.com/sites/default/files/documentation/index-rules/euronext_100_next_150_index_rules_version_15-01_oct_2015.pdf , last accessed on 16.07.2022.

[63] CME Group website, available under: https://www.cmegroup.com/markets/energy/crude-oil/light-sweet-crude.contractSpecs.html, last accessed on 16.07.2022.

[64] ICE Benchmark Administration Statement of Compliance With the Benchmarks Regulation and Independent Assurance, (2020), Published by ICE Benchmark Administration, available under:https://www.theice.com/publicdocs/Statement_of_Compliance_with_the_EU_Benchamrks_Regulation.pdf , last accessed on 16.07.2022.

[65] English version of the Wikipedia API, available under: https://www.mediawiki.org/wiki/API:Main_page, last accessed on 16.07.2022.

[66] Google trends website, available under: https://trends.google.es/trends/?geo=US, last accessed on 16.07.2022.

[67] English version of the website of ethereum.org, available under: https://ethereum.org/en/, last accessed on 16.07.2022.

[68] Python library Scikit-leran website, available under: https://scikit-learn.org/stable/modules/classes.html, last accessed on 27.07.22

[69] Bagherzadeh, F,. Comparative study on total nitrogen prediction in wastewater treatment plant and effect of various feature selection methods on machine learning algorithms performance, (2021), 41, 102033.

[70] Pirbazari, A., Evaluating Feature Selection Methods for Short-Term Load Forecasting, (2019), doi: 10.1109/BIGCOMP.2019.8679188.

[71] Chen, Q. *et al* ,Decision Variants for the Automatic Determination of Optimal Feature Subset in RF-RFE, (2018), doi: 10.3390/genes9060301

[72] Chen, H., Chagn, X., Photovoltaic power prediction of LSTM model based on Pearson feature selection, (2021), 7, 1047-1054.

[73] Nsiri, H., Alavi, S., A Novel Framework Based on Deep Learning and ANOVA Feature Selection Method for Diagnosis of COVID-19 Cases from Chest X-Ray Images (2022), doi: 10.1155/2022/4694567.

[74] Scikit-learn API Reference website, available under: https://scikit-learn.org/stable/modules/classes.html, last accessed on 24.07.2022.

[75] Reditt.com statistics profile in Semrush website, available under: https://www.semrush.com/website/reddit.com/overview/, last accessed on 24.07.2022.

[76] Twitter developer platform website, available under: https://developer.twitter.com/en, last accessed on 24.07.2022.

[77] PSAW: Python Pushshift.io API Wrapper website, available under: https://psaw.readthedocs.io/en/latest/, last accessed on 24.07.2022.

[78] VaderSentiment python library website, available under: https://pypi.org/project/vader-Sentiment/, last accessed on 24.07.2022.

[79] VADER Sentiment Analysis Guithub repository website, available under: https://github.com/cjhutto/vaderSentiment, last accessed on 24.07.2022.

[80] Sasaki, Y., The truth of the F-measure, (2007), Research fellow School of Computer Science, University of Manchester, M1 7DN.

[81] Hitam, N., Ismail, A., Comparative Performance of Machine Learning Algorithms for Cryptocurrency Forecasting, p-ISSN: 2502-4752, e-ISSN: 2502-4760.

Abbreviations table

Table 37.- Abbreviations table

AI	Artificial Intelligence
AML	Anti-Money Laundering
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
API	Application Programming Interface
Btc	Bitcoin price
BtcEth	Relation of prices Bitcoin-Ethereum
CFT	Combating the Financing of Terrorism
CNN	Convolutional Neural Network
CrOil	Crude Oil price
DJI	Down Jones Industrial Average
DJP	Bloomberg commodity index
Eth	Ethereum price
EthGas	Price Ethereum Gas
ETV	Volume of Ethereum transactions
EurUsd	Relation of prices Euro-USD
FSM	Feature Selection Method
Gold	Gold price
GRU	Gated Recurrent Unit
GTV	Google Trends Volume
HSI	Hang Seng Index
ICO	Initial Coin Offering
IXIC	Nasdaq Composite
LSTM	Long-Short Term Memory
ML	Machine learning
MSCI	Morgan Stanley Capital International
NLP	Natural Language Processing
N100	Euronext 100 index
RFC	Rain Forest Classifier
RFE	Recursive Feature Elimination
RFR	Rain Forest Regressor
RNN	Recurrent Neural Network
SSE	Shanghai stock exchange composite index
SOM	Self-Organising Maps
SP500	S&P 500 index
SVM	Support Vector Machine
USD	United States Dollar
UTC	Universal Coordinated Time
VIX	Volatility index
WkViews	Visits to Wikipedia entry “Ethereum”

Appendix A

As extension of *"chapter 3.6. Feature selection results"*, and as already mentioned at the end of it, below are summarized by prediction day target the results of applying the FSM used in this work to the dataset shown in the Table 1.

One day prediction

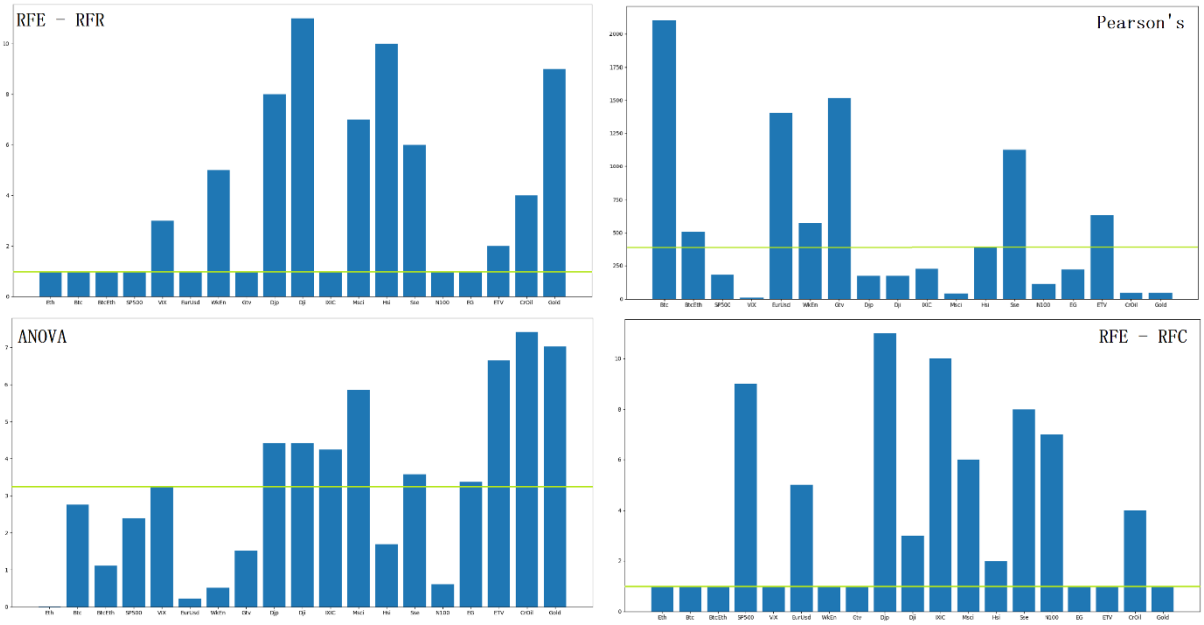

Figure 22.- Feature selection results for 1 day predictions

Each method gives different results, but it seems clear that there are some core features common to all subsets. Also, it is worth to note that both, the Wikipedia visits and the Google trends search volume, would be part of this core of features, being the two of them indicators of the interest that people show on the term Ethereum online.

Table 38.- Feature selection results for 1 day predictions

1 Day	**Regression**		**Classification**	
	RFE - RFR	*Pearson's*	*ANOVA*	*RFE – RFC*
Eth	X	X	X	X
Btc	X	X	X	X
BtcEth	X	X	X	X
SP500	X		X	
EurUsd	X	X	X	
WkViews		X	X	X
VIX			X	X
GTV	X	X	X	X
DJP				
DJI				
IXIC	X	X		
MSCI				
HSI		X	X	
SSE		X		
N100	X		X	
CrOil				

Gold			X
EthGas	X		X
ETV		X	X

See that for the Pearson´s results shown in *Figure 22*, the bar belonging to "Eth" has been eliminated to be able to offer a better view of the score obtained for the rest of features, since the score for Eth was 140000, leaving the others almost invisible next to it.

Seven days prediction

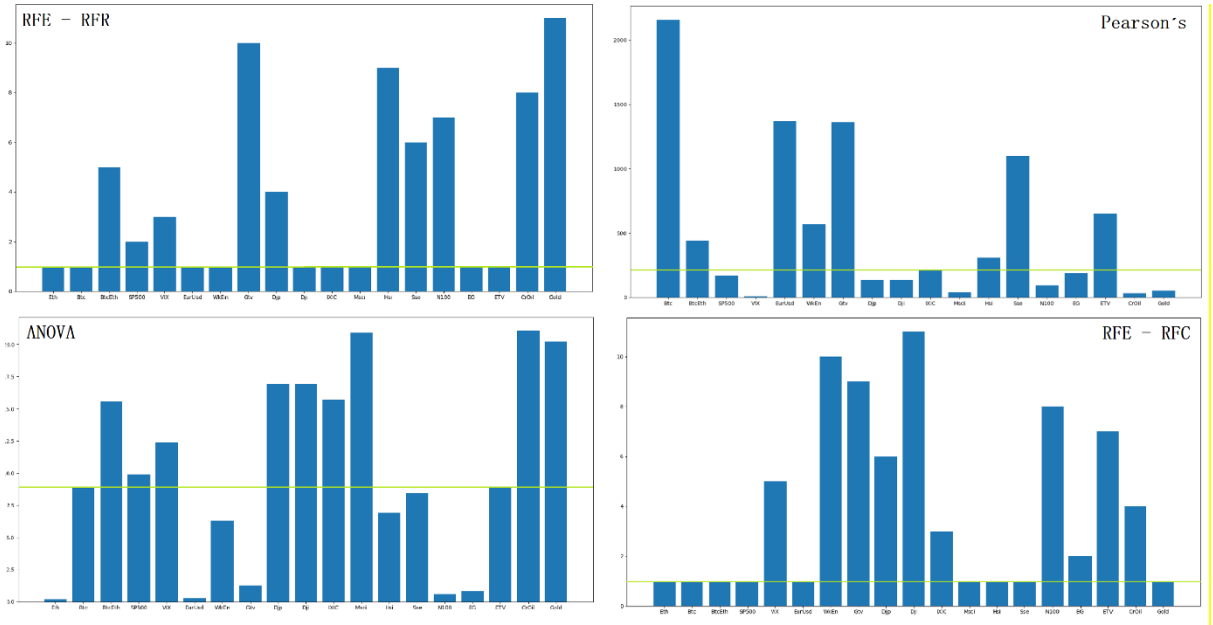

Figure 23.- Feature selection results for 7 days predictions

There is a significant difference between the subsets selected for one day forecast and seven days forecast, with exception of the Pearson´s method, that has given the same subset. It is interesting to see how the core of features obtained for the predictions one day lag loses importance and it shifts towards Chinese market indicators (HSI, SSE) when it comes to select the features for the predictions seven days lag.

Also features directly related to the use of Ethereum, as they are the price of Ethereum Gas and the Ethereum transaction volume, seem to be more helpful for the forecasting seven days lag.

It seems like the best option would be to use indeed different features when the forecast target changes from one day in the future to seven days in the future. These results will be taken into consideration in the next chapters of this work, applying different features for the different forecast scenarios.

Table 39.- Table summarizing the feature selection results for 7 days predictions

7 Days	Regression		Classification	
	RFE – RFR	Pearson's	ANOVA	RFE – RFC
Eth	X	X	X	X
Btc	X	X	X	X

BtcEth		X		X
SP500				X
EurUsd	X	X	X	X
WkViews	X	X	X	
VIX				
GTV		X	X	
DJP				
DJI	X			
IXIC	X	X		
MSCI	X			X
HSI		X	X	X
SSE		X	X	X
N100			X	
CrOil				
Gold				X
EthGas	X		X	
ETV	X	X	X	

Again, we must note that for the Pearson´s results shown in *Figure 23,* the bar belonging to "Eth" has been eliminated to be able offer a better view of the score obtained for the rest of features, since the score for Eth was much higher than the others.

Fifteen days prediction

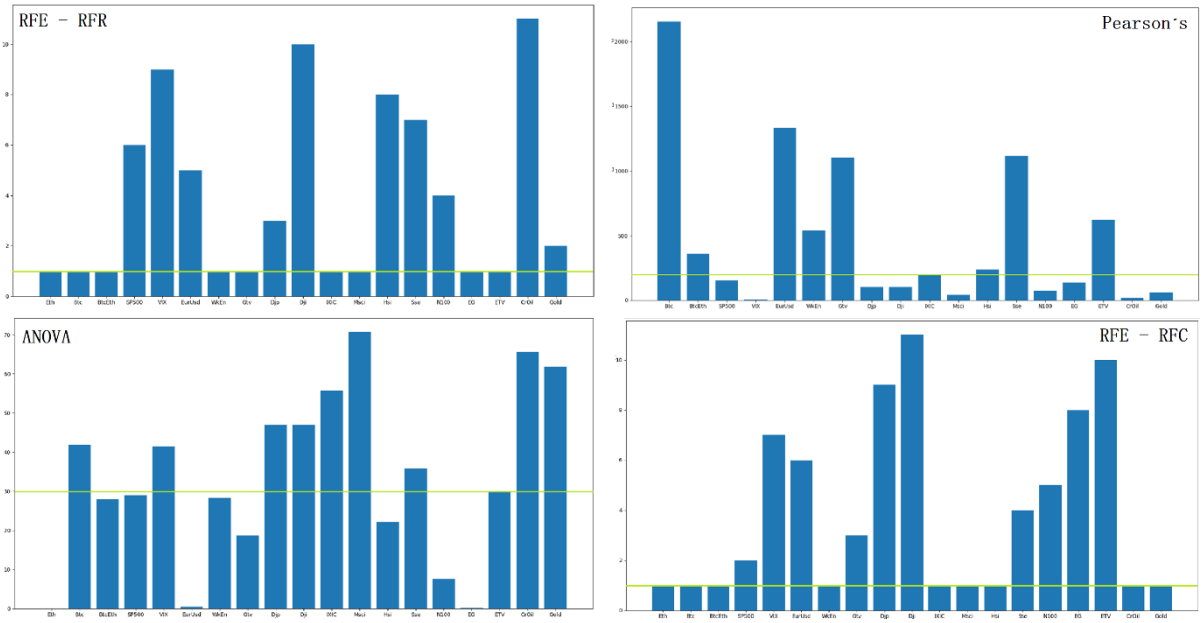

Figure 24.- Feature selection results for 15 days predictions

Pearsons´s method responds once again with the exact same subset as it gave in the two previous cases. The other three methods are also very stable compared to the seven days prediction, obtaining only a difference of two features each one of them. The results suggest that for these two cases it could be built a tool that makes use of only one subset of features for both cases, and most likely the impact on the prediction would be very small.

Table 40.- Table summarizing the feature selection results for 15 days predictions

15 Days	Regression		Classification	
	RFE - RFR	Pearson's	ANOVA	RFE – RFC
Eth	X	X	X	X
Btc	X	X		X
BtcEth	X	X	X	X
SP500			X	
EurUsd		X	X	
WkViews	X	X	X	X
VIX				
GTV	X	X	X	
DJP				
DJI				
IXIC	X	X		X
MSCI	X			X
HSI		X	X	X
SSE		X		
N100			X	
CrOil				X
Gold				X
EthGas	X		X	
ETV	X	X	X	

Once again, for the Pearson´s results shown *in Figure 24*, the bar belonging to "Eth" has been eliminated to be able offer a better view of the score obtained for the rest of features, since the score for "Eth" was much higher than the others.

Thirty days prediction

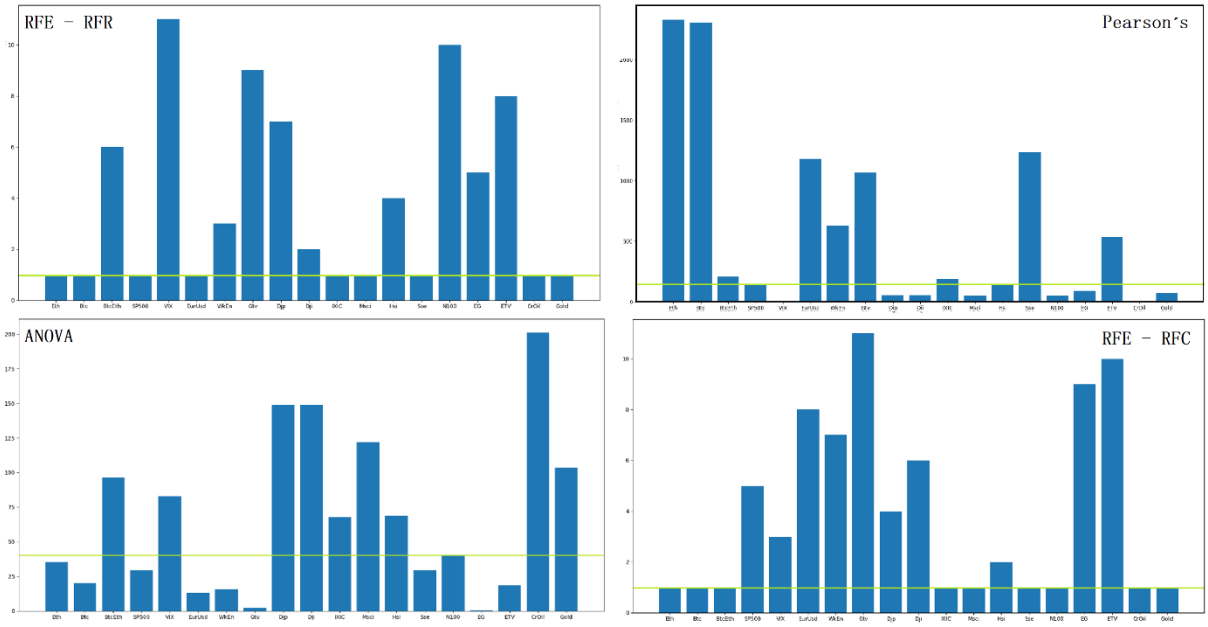

Figure 25.- Feature selection results for 30 days predictions

For the thirty days lag prediction there are again small differences for all methods, having the Pearson´s method as the most consistent, with the same subset of features as the obtained in all previous cases. Again, the biggest difference is with respect to the prediction one day lag. With seven days and fifteen days lag predictions, it remains in the same line, as far as the main feature are concerned.

Table 41.- Table summarizing the feature selection results for 30 days predictions

30 Days	Regression		Classification	
	RFE – RFR	Pearson's	ANOVA	RFE – RFC
Eth	X	X	X	X
Btc	X	X	X	X
BtcEth		X		X
SP500	X		X	
EurUsd	X	X	X	
WkViews		X	X	
VIX				
GTV		X	X	
DJP				
DJI				
IXIC	X	X		X
MSCI	X			X
HSI		X		
SSE	X	X	X	X
N100			X	X
CrOil	X			X
Gold	X			X
EthGas			X	
ETV		X	X	

Ninety days prediction

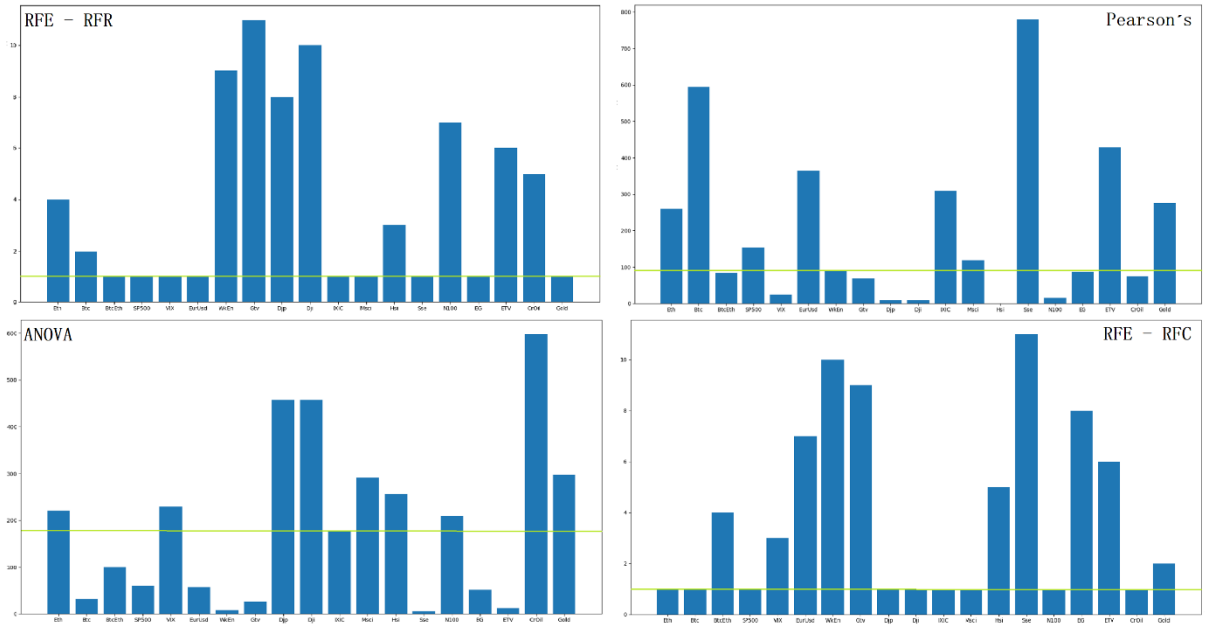

Figure 26.- Feature selection results for 90 days predictions

In this case, all methods offer very different results from the previous cases, including for first time changes in the Pearson´s method. This could be an indicator of the complexity of doing what for cryptocurrency prices might be considered as long-term forecasts, where perhaps both, the features selected as well as the forecasting results, would not be the most reliable, as already stated by the state of art

Table 42.- Table summarizing the feature selection results for 90 days predictions

90 Days	Regression		Classification	
	RFE – RFR	Pearson's	ANOVA	RFE – RFC
Eth		X		X
Btc		X	X	X
BtcEth	X		X	
SP500	X	X	X	X
EurUsd	X	X	X	
WkViews		X	X	
VIX	X			
GTV			X	
DJP				X
DJI				X
IXIC	X	X	X	X
MSCI	X	X		X
HIS				
SSE	X	X	X	
N100				X
CrOil				X
Gold	X	X		
EthGas	X		X	
ETV		X	X	

Appendix B

In this appendix can be seen a log of all iterations of the investment strategy for the prediction given by the GRU model, for one day prediction without sentiment analysis data, during the first period, so it can be seen how it ended up reaching the earnings shown in *4.4. Investment strategy: Regression approach.*

The columns displayed in the table are:

Day: Iteration day.

EthPrice: Price of Ethereum on that iteration day.

Plan: action planned by the investment strategy.

Done: What has actually happened. It could be that for example the investment strategy was planning to not sell Ethereum, but due to the fall on its price, a stop loss triggered the selling operation. Also can happen that a selling operation planned by the investment strategy doesn't take place because the owned Ethereum has been sold due to a stop loss and when the day to sell according to the plan comes, there is no Ethereum to sell.

nStopLoss: number of stop loss triggered until that iteration.

Fiat: The capital owned in "traditional" currency, in this case USD.

Virtual: Capital owned in Ethereum. It is calculated multiplying EthPrice*nEth.

nEth: number of Ethereum owned.

StopLoss: Price at which the stop loss is set.

Table 43.- Iterations regression investment strategy

Day	EthPrice	Plan	Done	nStopLos	Fiat	Virtual	nEth	StopLoss
0	473,5			0	10000	0	0	0
1	441,72			0	10000	0	0	0
2	515,13			0	10000	0	0	0
3	651,43			0	10000	0	0	0
4	702,76			0	10000	0	0	0
5	695,81			0	10000	0	0	0
6	684,44	buy	buy	0	0	10000	14,6103	9300
7	696,2			0	0	10171,8	14,6103	9459,80
8	719,97			0	0	10519,0	14,6103	9782,72
9	794,64			0	0	11610,0	14,6103	10797,3
10	826,82	sell	sell	0	12080,1	0	0	0
11	819,08			0	12080,1	0	0	0
12	821,06			0	12080,1	0	0	0
13	674,85	buy	buy	0	0	12080,1	17,9002	11234,5
14	719,38			0	0	12877,1	17,9002	11975,7
15	694,14			0	0	12425,4	17,9002	11975,7
16	765,83			0	0	13708,6	17,9002	12749
17	773,83	sell	sell	0	13851,8	0	0	0
18	762,84			0	13851,8	0	0	0
19	737,02			0	13851,8	0	0	0
20	753,59			0	13851,8	0	0	0
21	717,25	buy	buy	0	0	13851,8	19,3122	12882,2
22	756,73			0	0	14614,2	19,3122	13591,2
23	772,64			0	0	14921,4	19,3122	13876,9
24	884,44			0	0	17080,5	19,3122	15884,9
25	962,71	sell	sell	0	18592,2	0	0	0
26	980,92			0	18592,2	0	0	0
27	997,71			0	18592,2	0	0	0
28	1041,68	buy	buy	0	0	18592,2	17,8483	17290,8
29	1153,17			0	0	20582,1	17,8483	19141,4
30	1148,53			0	0	20499,3	17,8483	19141,4
31	1299,73			0	0	23198,2	17,8483	21574,3
32	1255,81			0	0	22414,3	17,8483	21574,3
33	1154,93		sell	1	21574,3	0	0	0
34	1273,19			1	21574,3	0	0	0
35	1396,42			1	21574,3	0	0	0
36	1366,77			1	21574,3	0	0	0
37	1291,92			1	21574,3	0	0	0
38	1053,68			1	21574,3	0	0	0
39	1014,25			1	21574,3	0	0	0
40	1036,28			1	21574,3	0	0	0
41	1039,09			1	21574,3	0	0	0
42	1155,15			1	21574,3	0	0	0
43	1049,58			1	21574,3	0	0	0
44	1003,26			1	21574,3	0	0	0
45	986,229			1	21574,3	0	0	0
46	1058,78			1	21574,3	0	0	0

Day	EthPrice	Plan	Done	nStopLos	Fiat	Virtual	nEth	StopLoss
47	1056,03			1	21574,3	0	0	0
48	1055,17			1	21574,3	0	0	0
49	1107,07			1	21574,3	0	0	0
50	1246,01			1	21574,3	0	0	0
51	1182,36			1	21574,3	0	0	0
52	1071,13	buy	buy	1	0	21574,3	20,1416	20064,1
53	1118,31			1	0	22524,6	20,1416	20947,9
54	1036,79		sell	2	20947,9	0	0	0
55	915,785			2	20947,9	0	0	0
56	964,019			2	20947,9	0	0	0
57	834,682			2	20947,9	0	0	0
58	697,951			2	20947,9	0	0	0
59	793,122			2	20947,9	0	0	0
60	757,068			2	20947,9	0	0	0
61	817,807			2	20947,9	0	0	0
62	883,865			2	20947,9	0	0	0
63	860,415			2	20947,9	0	0	0
64	814,66			2	20947,9	0	0	0
65	868,707			2	20947,9	0	0	0
66	845,258	buy	buy	2	0	20947,9	24,7828	19481,5
67	923,561			2	0	22888,4	24,7828	21286,2
68	936,976	sell	sell	2	23220,9	0	0	0
69	944,21	buy	buy	2	0	23220,9	24,5929	21595,4
70	974,115			2	0	23956,3	24,5929	22279,4
71	923,921			2	0	22721,9	24,5929	22279,4
72	943,865			2	0	23212,4	24,5929	22279,4
73	895,371		sell	3	22279,4	0	0	0
74	849,971			3	22279,4	0	0	0
75	812,845			3	22279,4	0	0	0
76	864,189			3	22279,4	0	0	0
77	840,515			3	22279,4	0	0	0
78	844,81			3	22279,4	0	0	0
79	869,315			3	22279,4	0	0	0
80	878,265			3	22279,4	0	0	0
81	855,199			3	22279,4	0	0	0
82	872,2			3	22279,4	0	0	0
83	856,853	buy	buy	3	0	22279,4	26,0014	20719,8
84	857,225			3	0	22289,1	26,0014	20728,8
85	866,678			3	0	22534,9	26,0014	20957,4
86	853,684	sell	sell	3	22197,0	0	0	0
87	816,951			3	22197,0	0	0	0
88	752,831			3	22197,0	0	0	0
89	704,596			3	22197,0	0	0	0
90	728,916			3	22197,0	0	0	0
91	686,89			3	22197,0	0	0	0
92	723,339			3	22197,0	0	0	0
93	699,831			3	22197,0	0	0	0
94	690,827			3	22197,0	0	0	0

Day	EthPrice	Plan	Done	nStopLos	Fiat	Virtual	nEth	StopLoss
95	614,291			3	22197,0	0	0	0
96	611,304			3	22197,0	0	0	0
97	601,666			3	22197,0	0	0	0
98	552,779			3	22197,0	0	0	0
99	538,644			3	22197,0	0	0	0
100	556,726			3	22197,0	0	0	0
101	557,175			3	22197,0	0	0	0
102	561,732			3	22197,0	0	0	0
103	539,702			3	22197,0	0	0	0
104	539,619	buy	buy	3	0	22197,0	41,1346	20643,2
105	526,439			3	0	21654,8	41,1346	20643,2
106	524,287			3	0	21566,3	41,1346	20643,2
107	489,951		sell	4	20643,2	0	0	0
108	450,116			4	20643,2	0	0	0
109	446,279			4	20643,2	0	0	0
110	385,968			4	20643,2	0	0	0
111	394,645			4	20643,2	0	0	0
112	396,457			4	20643,2	0	0	0
113	379,607			4	20643,2	0	0	0
114	386,425			4	20643,2	0	0	0
115	416,893			4	20643,2	0	0	0
116	380,543			4	20643,2	0	0	0
117	383,232			4	20643,2	0	0	0
118	370,285			4	20643,2	0	0	0
119	385,314	buy	buy	4	0	20643,2	53,5751	19198,2
120	400,508			4	0	21457,2	53,5751	19955,2
121	398,526	sell	sell	4	21351,0	0	0	0
122	414,243			4	21351,0	0	0	0
123	430,54			4	21351,0	0	0	0
124	492,941	buy	buy	4	0	21351,0	43,3136	19856,5
125	492,735			4	0	21342,1	43,3136	19856,5
126	501,478			4	0	21720,8	43,3136	20200,4
127	531,702			4	0	23029,9	43,3136	21417,8
128	511,147			4	0	22139,6	43,3136	21417,8
129	502,894			4	0	21782,1	43,3136	21417,8
130	524,789			4	0	22730,5	43,3136	21417,8
131	567,889			4	0	24597,3	43,3136	22875,5
132	615,718			4	0	26669,0	43,3136	24802,1
133	605,395			4	0	26221,8	43,3136	24802,1
134	621,86			4	0	26935,0	43,3136	25049,5
135	642,548			4	0	27831,1	43,3136	25882,9
136	708,158			4	0	30672,9	43,3136	28525,8
137	615,416		sell	5	28525,8	0	0	0
138	662,809			5	28525,8	0	0	0
139	647,032			5	28525,8	0	0	0
140	683,678			5	28525,8	0	0	0
141	688,881			5	28525,8	0	0	0
142	669,924			5	28525,8	0	0	0

Day	EthPrice	Plan	Done	nStopLos	Fiat	Virtual	nEth	StopLoss
143	673,613			5	28525,8	0	0	0
144	687,149			5	28525,8	0	0	0
145	779,543	buy	buy	5	0	28525,8	36,593	26529,0
146	785,624			5	0	28748,3	36,593	26735,9
147	816,12			5	0	29864,2	36,593	27773,7
148	792,311			5	0	28993,0	36,593	27773,7
149	753,725		sell	6	27773,7	0	0	0
150	752,857			6	27773,7	0	0	0
151	752,275			6	27773,7	0	0	0
152	727,277			6	27773,7	0	0	0
153	679,586			6	27773,7	0	0	0
154	686,048			6	27773,7	0	0	0
155	733,496			6	27773,7	0	0	0
156	730,549			6	27773,7	0	0	0
157	708,871			6	27773,7	0	0	0
158	707,05			6	27773,7	0	0	0
159	672,657			6	27773,7	0	0	0
160	694,367	buy	buy	6	0	27773,7	39,9987	25829,6
161	696,53			6	0	27860,3	39,9987	25910,0
162	715,369			6	0	28613,8	39,9987	26610,8
163	699,222	sell	sell	6	27967,9	0	0	0
164	647,741			6	27967,9	0	0	0
165	583,585			6	27967,9	0	0	0
166	601,755			6	27967,9	0	0	0
167	586,734			6	27967,9	0	0	0
168	587,28			6	27967,9	0	0	0
169	572,668			6	27967,9	0	0	0
170	516,036			6	27967,9	0	0	0
171	565,388			6	27967,9	0	0	0
172	559,59			6	27967,9	0	0	0
173	577,645			6	27967,9	0	0	0
174	580,043			6	27967,9	0	0	0
175	591,808			6	27967,9	0	0	0
176	618,329	buy	buy	6	0	27967,9	45,2315	26010,2
177	592,985			6	0	26821,6	45,2315	26010,2
178	609,303			6	0	27559,7	45,2315	26010,2
179	607,124			6	0	27461,1	45,2315	26010,2
180	605,187			6	0	27373,5	45,2315	26010,2
181	601,077			6	0	27187,6	45,2315	26010,2
182	597,562			6	0	27028,6	45,2315	26010,2
183	526,479		sell	7	26010,2	0	0	0
184	533,284			7	26010,2	0	0	0
185	496,843			7	26010,2	0	0	0
186	477,494			7	26010,2	0	0	0
187	519,742			7	26010,2	0	0	0
188	491,004			7	26010,2	0	0	0
189	499,642			7	26010,2	0	0	0
190	500,448			7	26010,2	0	0	0

Day	EthPrice	Plan	Done	nStopLos	Fiat	Virtual	nEth	StopLoss
191	518,891			7	26010,2	0	0	0
192	537,957			7	26010,2	0	0	0
193	536,268			7	26010,2	0	0	0
194	527,367			7	26010,2	0	0	0
195	465,817	buy	buy	7	0	26010,2	55,8378	24189,5
196	474,52			7	0	26496,1	55,8378	24641,4
197	457,672			7	0	25555,4	55,8378	24641,4
198	460,31			7	0	25702,7	55,8378	24641,4
199	432,772	sell	sell	7	24165,0	0	0	0
200	442,365			7	24165,0	0	0	0
201	422,365			7	24165,0	0	0	0
202	436,009			7	24165,0	0	0	0
203	455,18			7	24165,0	0	0	0
204	453,918			7	24165,0	0	0	0
205	475,347			7	24165,0	0	0	0
206	464,195			7	24165,0	0	0	0
207	467,318			7	24165,0	0	0	0
208	474,412			7	24165,0	0	0	0
209	474,012	buy	buy	7	0	24165,0	50,9798	22473,5
210	491,66			7	0	25064,7	50,9798	23310,2
211	489,119			7	0	24935,2	50,9798	23310,2
212	476,682			7	0	24301,1	50,9798	23310,2
213	434,424		sell	8	23310,2	0	0	0
214	446,518			8	23310,2	0	0	0
215	430,074			8	23310,2	0	0	0
216	434,027			8	23310,2	0	0	0
217	436,086			8	23310,2	0	0	0
218	449,85			8	23310,2	0	0	0
219	480,658			8	23310,2	0	0	0
220	501,002			8	23310,2	0	0	0
221	480,513			8	23310,2	0	0	0
222	469,619			8	23310,2	0	0	0
223	450,698	buy	buy	8	0	23310,2	51,7202	21678,5
224	462,436			8	0	23917,3	51,7202	22243,1
225	459,657			8	0	23773,5	51,7202	22243,1
226	450,853			8	0	23318,2	51,7202	22243,1
227	479,373	sell	sell	8	24793,3	0	0	0
228	472,493			8	24793,3	0	0	0
229	464,037			8	24793,3	0	0	0
230	469,666			8	24793,3	0	0	0
231	466,898	buy	buy	8	0	24793,3	53,1021	23057,7
232	466,665			8	0	24780,9	53,1021	23057,7
233	457,081	sell	sell	8	24271,9	0	0	0
234	433,867			8	24271,9	0	0	0
235	420,747			8	24271,9	0	0	0
236	412,621			8	24271,9	0	0	0
237	418,262			8	24271,9	0	0	0
238	407,252			8	24271,9	0	0	0

Day	EthPrice	Plan	Done	nStopLos	Fiat	Virtual	nEth	StopLoss
239	410,516			8	24271,9	0	0	0
240	406,658			8	24271,9	0	0	0
241	380,215			8	24271,9	0	0	0
242	356,613			8	24271,9	0	0	0
243	365,588			8	24271,9	0	0	0
244	334,176			8	24271,9	0	0	0
245	322,112			8	24271,9	0	0	0
246	319,57			8	24271,9	0	0	0
247	286,495			8	24271,9	0	0	0
248	278,932			8	24271,9	0	0	0
249	282,364			8	24271,9	0	0	0
250	288,046			8	24271,9	0	0	0
251	315,729			8	24271,9	0	0	0
252	295,812			8	24271,9	0	0	0
253	300,834			8	24271,9	0	0	0
254	274,315			8	24271,9	0	0	0
255	281,944			8	24271,9	0	0	0
256	271,341			8	24271,9	0	0	0
257	277,104			8	24271,9	0	0	0
258	282,967			8	24271,9	0	0	0
259	279,646			8	24271,9	0	0	0
260	275,197			8	24271,9	0	0	0
261	285,603			8	24271,9	0	0	0
262	296,499			8	24271,9	0	0	0
263	289,312			8	24271,9	0	0	0
264	284,105			8	24271,9	0	0	0
265	283,004			8	24271,9	0	0	0
266	295,341	buy	buy	8	0	24271,9	82,1829	22572,9
267	294,371			8	0	24192,2	82,1829	22572,9
268	289,259			8	0	23772,1	82,1829	22572,9
269	285,723			8	0	23481,5	82,1829	22572,9
270	232,331		sell	9	22572,9	0	0	0
271	230,215			9	22572,9	0	0	0
272	217,203			9	22572,9	0	0	0
273	197,951			9	22572,9	0	0	0
274	196,924			9	22572,9	0	0	0
275	197,079			9	22572,9	0	0	0
276	185,066			9	22572,9	0	0	0
277	183,331			9	22572,9	0	0	0
278	211,354			9	22572,9	0	0	0
279	211,749			9	22572,9	0	0	0
280	223,074			9	22572,9	0	0	0
281	220,589			9	22572,9	0	0	0
282	197,875	buy	buy	9	0	22572,9	114,076	20992,8
283	209,975			9	0	23953,2	114,076	22276,5
284	209,969			9	0	23952,6	114,076	22276,5
285	224,591			9	0	25620,6	114,076	23827,1
286	246,584	sell	sell	9	28129,5	0	0	0

Day	EthPrice	Plan	Done	nStopLos	Fiat	Virtual	nEth	StopLoss
287	240,48			9	28129,5	0	0	0
288	244,33			9	28129,5	0	0	0
289	228,73	buy	buy	9	0	28129,5	122,981	26160,4
290	218,505			9	0	26872,0	122,981	26160,4
291	215,847			9	0	26545,1	122,981	26160,4
292	228,494			9	0	28100,5	122,981	26160,4
293	222,402	sell	sell	9	27351,2	0	0	0
294	231,635			9	27351,2	0	0	0
295	232,848			9	27351,2	0	0	0
296	230,768			9	27351,2	0	0	0
297	227,181			9	27351,2	0	0	0
298	220,489	buy	buy	9	0	27351,2	124,048	25436,7
299	222,218			9	0	27565,7	124,048	25636,1
300	227,601			9	0	28233,5	124,048	26257,1
301	225,12			9	0	27925,7	124,048	26257,1
302	226,119	sell	sell	9	28049,6	0	0	0
303	229,255			9	28049,6	0	0	0
304	227,982			9	28049,6	0	0	0
305	225,769			9	28049,6	0	0	0
306	189,499			9	28049,6	0	0	0
307	196,727	buy	buy	9	0	28049,6	142,581	26086,2
308	199,841			9	0	28493,6	142,581	26499,1
309	195,715			9	0	27905,3	142,581	26499,1
310	209,704	sell	sell	9	29899,9	0	0	0
311	210,12			9	29899,9	0	0	0
312	207,083			9	29899,9	0	0	0
313	203,352			9	29899,9	0	0	0
314	203,727	buy	buy	9	0	29899,9	146,764	27806,9
315	205,43			9	0	30149,9	146,764	28039,4
316	205,142			9	0	30107,6	146,764	28039,4
317	204,044			9	0	29946,4	146,764	28039,4
318	204,336			9	0	29989,3	146,764	28039,4
319	203,852	sell	sell	9	29918,3	0	0	0
320	202,719			9	29918,3	0	0	0
321	203,329			9	29918,3	0	0	0
322	204,21			9	29918,3	0	0	0
323	205,366			9	29918,3	0	0	0
324	197,247			9	29918,3	0	0	0
325	197,556			9	29918,3	0	0	0
326	197,381			9	29918,3	0	0	0
327	198,872			9	29918,3	0	0	0
328	200,635	buy	buy	9	0	29918,3	149,118	27824,0
329	200,186			9	0	29851,3	149,118	27824,0
330	207,486	sell	sell	9	30939,9	0	0	0
331	209,091	buy	buy	9	0	30939,9	147,973	28774,1
332	218,452			9	0	32325,1	147,973	30062,3
333	217,183			9	0	32137,3	147,973	30062,3
334	212,231			9	0	31404,5	147,973	30062,3

Day	EthPrice	Plan	Done	nStopLos	Fiat	Virtual	nEth	StopLoss
335	210,074			9	0	31085,3	147,973	30062,3
336	212,533			9	0	31449,2	147,973	30062,3
337	211,34			9	0	31272,7	147,973	30062,3
338	210,418			9	0	31136,2	147,973	30062,3
339	206,826			9	0	30604,7	147,973	30062,3
340	181,397	sell	sell	9	26841,9	0	0	0
341	180,806			9	26841,9	0	0	0
342	175,177			9	26841,9	0	0	0
343	174,001			9	26841,9	0	0	0
344	177,067			9	26841,9	0	0	0
345	149,175			9	26841,9	0	0	0
346	130,339			9	26841,9	0	0	0
347	136,701			9	26841,9	0	0	0
348	126,706			9	26841,9	0	0	0
349	123,296			9	26841,9	0	0	0
350	113,494			9	26841,9	0	0	0
351	116,45			9	26841,9	0	0	0
352	108,335			9	26841,9	0	0	0
353	110,01			9	26841,9	0	0	0
354	122,438			9	26841,9	0	0	0
355	117,5426			9	26841,9	0	0	0
356	113,1714	buy	buy	9	0	26841,9	237,179	24963,0
357	118,6356			9	0	28137,9	237,179	26168,2
358	116,3889			9	0	27605,0	237,179	26168,2
359	108,925	sell	sell	9	25834,7	0	0	0
360	110,2142			9	25834,7	0	0	0
361	102,4756			9	25834,7	0	0	0
362	91,76105			9	25834,7	0	0	0
363	93,29456			9	25834,7	0	0	0
364	92,15964			9	25834,7	0	0	0
365	95,14206			9	25834,7	0	0	0
366	91,68565			9	25834,7	0	0	0
367	88,9453			9	25834,7	0	0	0
368	90,5933			9	25834,7	0	0	0
369	86,53931			9	25834,7	0	0	0
370	84,3083			9	25834,7	0	0	0
371	84,44081			9	25834,7	0	0	0
372	85,2621			9	25834,7	0	0	0
373	95,13383			9	25834,7	0	0	0
374	101,1125			9	25834,7	0	0	0
375	101,2686			9	25834,7	0	0	0
376	116,2169			9	25834,7	0	0	0
377	109,4962	buy	buy	9	0	25834,7	235,942	24026,3
378	116,7595			9	0	27548,5	235,942	25620,1
379	130,7653			9	0	30853,0	235,942	28693,3
380	140,238			9	0	33088,0	235,942	30771,9
381	129,8443		sell	10	30771,9	0	0	0
382	131,8656			10	30771,9	0	0	0

Day	EthPrice	Plan	Done	nStopLos	Fiat	Virtual	nEth	StopLoss
383	116,5759			10	30771,9	0	0	0
384	137,647			10	30771,9	0	0	0
385	138,0183			10	30771,9	0	0	0
386	139,8595			10	30771,9	0	0	0
387	133,3683			10	30771,9	0	0	0
388	140,8194			10	30771,9	0	0	0
389	155,0477			10	30771,9	0	0	0
390	149,135			10	30771,9	0	0	0
391	154,5819			10	30771,9	0	0	0
392	155,6386			10	30771,9	0	0	0
393	157,7462			10	30771,9	0	0	0
394	151,6992			10	30771,9	0	0	0
395	150,3596			10	30771,9	0	0	0
396	150,8031			10	30771,9	0	0	0
397	128,6252			10	30771,9	0	0	0
398	127,5483			10	30771,9	0	0	0
399	125,9665			10	30771,9	0	0	0
400	116,8978			10	30771,9	0	0	0
401	129,0687			10	30771,9	0	0	0
402	122,0327			10	30771,9	0	0	0
403	123,5471			10	30771,9	0	0	0
404	123,7419			10	30771,9	0	0	0
405	121,0103	buy	buy	10	0	30771,9	254,291	28617,8
406	124,519			10	0	31664,1	254,291	29447,6
407	119,4748			10	0	30381,4	254,291	29447,6
408	117,1579			10	0	29792,2	254,291	29447,6
409	118,7476			10	0	30196,5	254,291	29447,6
410	117,4526			10	0	29867,2	254,291	29447,6
411	117,3629			10	0	29844,4	254,291	29447,6
412	116,3783			10	0	29594,0	254,291	29447,6
413	116,4887			10	0	29622,1	254,291	29447,6
414	113,4053		sell	11	29447,6	0	0	0
415	106,59			11	29447,6	0	0	0
416	105,5982			11	29447,6	0	0	0
417	108,9075			11	29447,6	0	0	0
418	107,061			11	29447,6	0	0	0
419	107,6098			11	29447,6	0	0	0
420	110,4303			11	29447,6	0	0	0
421	107,4927			11	29447,6	0	0	0
422	107,8216			11	29447,6	0	0	0
423	107,4435			11	29447,6	0	0	0
424	104,9193			11	29447,6	0	0	0
425	104,5353			11	29447,6	0	0	0
426	119,2677	buy	buy	11	0	29447,6	246,903	27386,3
427	119,4309			11	0	29487,9	246,903	27423,7
428	124,8065			11	0	30815,2	246,903	28658,1
429	121,2984			11	0	29949,0	246,903	28658,1
430	122,5723			11	0	30263,5	246,903	28658,1

Day	EthPrice	Plan	Done	nStopLos	Fiat	Virtual	nEth	StopLoss
431	122,5536			11	0	30258,9	246,903	28658,1
432	121,3902			11	0	29971,7	246,903	28658,1
433	122,1004			11	0	30147,0	246,903	28658,1
434	123,2602			11	0	30433,4	246,903	28658,1
435	133,5989			11	0	32986,0	246,903	30677,0
436	146,0984			11	0	36072,2	246,903	33547,1
437	145,3462			11	0	35886,5	246,903	33547,1
438	149,5543			11	0	36925,5	246,903	34340,7
439	146,1308			11	0	36080,2	246,903	34340,7
440	149,0925			11	0	36811,4	246,903	34340,7
441	158,7954			11	0	39207,2	246,903	36462,6
442	135,8527		sell	12	36462,6	0	0	0
443	139,8229			12	36462,6	0	0	0
444	138,0897			12	36462,6	0	0	0
445	136,126			12	36462,6	0	0	0
446	136,7462			12	36462,6	0	0	0
447	136,4436			12	36462,6	0	0	0
448	134,207			12	36462,6	0	0	0
449	132,2494			12	36462,6	0	0	0
450	127,7741			12	36462,6	0	0	0
451	137,8224			12	36462,6	0	0	0
452	138,7894			12	36462,6	0	0	0
453	138,0345			12	36462,6	0	0	0
454	135,1271			12	36462,6	0	0	0
455	137,9653			12	36462,6	0	0	0
456	136,7587			12	36462,6	0	0	0
457	133,834			12	36462,6	0	0	0
458	134,4429	buy	buy	12	0	36462,6	271,213	33910,3
459	133,2631			12	0	36142,7	271,213	33910,3
460	133,7116	sell	sell	12	36264,3	0	0	0
461	137,6918			12	36264,3	0	0	0
462	142,3982			12	36264,3	0	0	0
463	140,0005	buy	buy	12	0	36264,3	259,030	33725,8
464	139,3746			12	0	36102,2	259,030	33725,8
465	140,5192			12	0	36398,7	259,030	33850,7
466	140,6864			12	0	36442	259,030	33891,0
467	136,6361			12	0	35392,8	259,030	33891,0
468	137,1814			12	0	35534,1	259,030	33891,0
469	138,2399			12	0	35808,2	259,030	33891,0
470	136,9869			12	0	35483,7	259,030	33891,0
471	135,031	sell	sell	12	34977,1	0	0	0
472	135,4604			12	34977,1	0	0	0
473	140,9882			12	34977,1	0	0	0
474	139,4165			12	34977,1	0	0	0
475	142,4974	buy	buy	12	0	34977,1	245,457	32528,7
476	142,0927			12	0	34877,7	245,457	32528,7
477	141,5141			12	0	34735,7	245,457	32528,7
478	141,8303			12	0	34813,3	245,457	32528,7

Day	EthPrice	Plan	Done	nStopLos	Fiat	Virtual	nEth	StopLoss
479	163,9617			12	0	40245,6	245,457	37428,4
480	161,4588			12	0	39631,3	245,457	37428,4
481	158,0525			12	0	38795,2	245,457	37428,4
482	165,5148			12	0	40626,9	245,457	37783,0
483	165,9472			12	0	40733,0	245,457	37881,7
484	174,5316			12	0	42840,1	245,457	39841,3
485	180,2588	sell	sell	12	44245,9	0	0	0
486	176,1146			12	44245,9	0	0	0
487	177,3375			12	44245,9	0	0	0
488	165,497	buy	buy	12	0	44245,9	267,351	41148,7
489	164,7363			12	0	44042,5	267,351	41148,7
490	164,5134			12	0	43982,9	267,351	41148,7
491	167,8385			12	0	44871,9	267,351	41730,8
492	161,5742	sell	sell	12	43197,1	0	0	0
493	167,6234			12	43197,1	0	0	0
494	166,9508			12	43197,1	0	0	0
495	173,814			12	43197,1	0	0	0
496	173,7084			12	43197,1	0	0	0
497	173,7513			12	43197,1	0	0	0
498	170,0465			12	43197,1	0	0	0
499	171,8747	buy	buy	12	0	43197,1	251,329	40173,3
500	171,4478			12	0	43089,8	251,329	40173,3
501	165,8261	sell	sell	12	41676,9	0	0	0
502	154,4594			12	41676,9	0	0	0
503	156,3471	buy	buy	12	0	41676,9	266,566	38759,5
504	158,4349			12	0	42233,5	266,566	39277,1
505	157,2951	sell	sell	12	41929,6	0	0	0
506	155,202			12	41929,6	0	0	0
507	162,166			12	41929,6	0	0	0
508	160,8183			12	41929,6	0	0	0
509	162,1228	buy	buy	12	0	41929,6	258,629	38994,5
510	167,9524			12	0	43437,3	258,629	40396,7
511	164,0266			12	0	42422,0	258,629	40396,7
512	163,4507			12	0	42273,1	258,629	40396,7
513	172,6532	sell	sell	12	44653,1	0	0	0
514	169,7987			12	44653,1	0	0	0
515	170,9482			12	44653,1	0	0	0
516	170,2893			12	44653,1	0	0	0
517	173,1427	buy	buy	12	0	44653,1	257,897	41527,4
518	194,3043			12	0	50110,6	257,897	46602,9
519	187,3253			12	0	48310,7	257,897	46602,9
520	196,8463			12	0	50766,2	257,897	47212,5
521	217,1486	sell	sell	12	56002,1	0	0	0
522	247,1531			12	56002,1	0	0	0
523	264,01			12	56002,1	0	0	0
524	243,7646			12	56002,1	0	0	0
525	234,5952	buy	buy	12	0	56002,1	238,718	52081,9
526	261,2922			12	0	62375,2	238,718	58008,9

Day	EthPrice	Plan	Done	nStopLos	Fiat	Virtual	nEth	StopLoss
527	251,8644			12	0	60124,6	238,718	58008,9
528	255,2159			12	0	60924,6	238,718	58008,9
529	244,6989			12	0	58414,0	238,718	58008,9
530	245,9657			12	0	58716,4	238,718	58008,9
531	249,485			12	0	59556,6	238,718	58008,9
532	251,7599	sell	sell	12	60099,6	0	0	0
533	267,0696			12	60099,6	0	0	0
534	272,8626			12	60099,6	0	0	0
535	271,7687			12	60099,6	0	0	0
536	269,4557			12	60099,6	0	0	0
537	255,8589			12	60099,6	0	0	0
538	268,1136	buy	buy	12	0	60099,6	224,157	55892,6
539	265,3913			12	0	59489,4	224,157	55892,6
540	270,2255			12	0	60573,0	224,157	56332,9
541	252,6079	sell	sell	12	56623,9	0	0	0
542	241,3388			12	56623,9	0	0	0
543	247,051			12	56623,9	0	0	0
544	249,4743			12	56623,9	0	0	0
545	250,9304			12	56623,9	0	0	0
546	245,7381			12	56623,9	0	0	0
547	233,0867			12	56623,9	0	0	0
548	247,5201	buy	buy	12	0	56623,9	228,765	52660,2
549	245,7818			12	0	56226,2	228,765	52660,2
550	260,9033			12	0	59685,5	228,765	55507,5
551	256,1688	sell	sell	12	58602,4	0	0	0
552	264,0875	buy	buy	12	0	58602,4	221,905	54500,2
553	269,0193			12	0	59696,8	221,905	55518,0
554	269,2232			12	0	59742,1	221,905	55560,1
555	274,3511			12	0	60880,0	221,905	56618,4
556	265,0518			12	0	58816,4	221,905	56618,4
557	269,4318			12	0	59788,4	221,905	56618,4
558	271,695			12	0	60290,6	221,905	56618,4
559	294,9103			12	0	65442,2	221,905	60861,2
560	309,3803			12	0	68653,2	221,905	63847,4
561	307,8278			12	0	68308,6	221,905	63847,4
562	310,4247			12	0	68884,9	221,905	64063
563	318,1263			12	0	70593,9	221,905	65652,4
564	336,7532			12	0	74727,3	221,905	69496,4
565	294,2676		sell	13	69496,4	0	0	0
566	311,2261			13	69496,4	0	0	0
567	320,0589			13	69496,4	0	0	0
568	290,696			13	69496,4	0	0	0
569	293,6411			13	69496,4	0	0	0
570	291,5964			13	69496,4	0	0	0
571	303,1			13	69496,4	0	0	0
572	284,5232			13	69496,4	0	0	0
573	287,9975			13	69496,4	0	0	0
574	287,5471			13	69496,4	0	0	0

Day	EthPrice	Plan	Done	nStopLos	Fiat	Virtual	nEth	StopLoss
575	305,7006			13	69496,4	0	0	0
576	313,2514			13	69496,4	0	0	0
577	308,881			13	69496,4	0	0	0
578	290,0023			13	69496,4	0	0	0
579	268,704	buy	buy	13	0	69496,4	258,635	64631,7
580	276,2767			13	0	71455,0	258,635	66453,1
581	269,4588			13	0	69691,6	258,635	66453,1
582	227,5781		sell	14	66453,1	0	0	0
583	229,776			14	66453,1	0	0	0
584	199,1887			14	66453,1	0	0	0
585	211,485			14	66453,1	0	0	0
586	226,5662			14	66453,1	0	0	0
587	221,3334			14	66453,1	0	0	0
588	229,1192			14	66453,1	0	0	0
589	225,6308			14	66453,1	0	0	0
590	217,56			14	66453,1	0	0	0
591	212,7285			14	66453,1	0	0	0
592	217,0464			14	66453,1	0	0	0
593	219,618			14	66453,1	0	0	0
594	219,6296			14	66453,1	0	0	0
595	207,4083			14	66453,1	0	0	0
596	211,1866	buy	buy	14	0	66453,1	314,665	61801,4
597	211,2681			14	0	66478,8	314,665	61825,3
598	210,5226	sell	sell	14	66244,2	0	0	0
599	218,6541			14	66244,2	0	0	0
600	217,8084			14	66244,2	0	0	0
601	217,8716	buy	buy	14	0	66244,2	304,051	61607,1
602	222,4903			14	0	67648,6	304,051	62913,2
603	222,6697			14	0	67703,1	304,051	62963,9
604	234,215			14	0	71213,5	304,051	66228,5
605	226,0206	sell	sell	14	68722	0	0	0
606	226,391			14	68722	0	0	0
607	220,9419			14	68722	0	0	0
608	210,4889			14	68722	0	0	0
609	206,7335			14	68722	0	0	0
610	216,0929			14	68722	0	0	0
611	211,2881			14	68722	0	0	0
612	208,709			14	68722	0	0	0
613	186,6077			14	68722	0	0	0
614	188,5021			14	68722	0	0	0
615	185,4401			14	68722	0	0	0
616	185,6877			14	68722	0	0	0
617	194,4935			14	68722	0	0	0
618	203,0919			14	68722	0	0	0
619	196,5654			14	68722	0	0	0
620	186,8916			14	68722	0	0	0
621	191,3329	buy	buy	14	0	68722	359,175	63911,4
622	194,7062			14	0	69933,6	359,175	65038,2

Day	EthPrice	Plan	Done	nStopLos	Fiat	Virtual	nEth	StopLoss
623	191,2896			14	0	68706,4	359,175	65038,2
624	186,8424	sell	sell	14	67109,1	0	0	0
625	188,9294			14	67109,1	0	0	0
626	187,5167			14	67109,1	0	0	0
627	173,89			14	67109,1	0	0	0
628	169,5167			14	67109,1	0	0	0
629	168,8349			14	67109,1	0	0	0
630	172,4698			14	67109,1	0	0	0
631	171,6294			14	67109,1	0	0	0
632	178,3474			14	67109,1	0	0	0
633	179,4993			14	67109,1	0	0	0
634	175,9929			14	67109,1	0	0	0
635	174,2171			14	67109,1	0	0	0
636	169,9562			14	67109,1	0	0	0
637	178,2626			14	67109,1	0	0	0
638	181,3555			14	67109,1	0	0	0
639	181,1495			14	67109,1	0	0	0
640	179,7872			14	67109,1	0	0	0
641	178,7255			14	67109,1	0	0	0
642	181,0161	buy	buy	14	0	67109,1	370,735	62411,4
643	181,1097			14	0	67143,8	370,735	62443,7
644	188,1055			14	0	69737,4	370,735	64855,7
645	189,7891			14	0	70361,5	370,735	65436,2
646	197,1132			14	0	73076,8	370,735	67961,4
647	208,6087			14	0	77338,6	370,735	71924,9
648	211,393			14	0	78370,9	370,735	72884,9
649	221,2809			14	0	82036,6	370,735	76294,1
650	218,05			14	0	80838,9	370,735	76294,1
651	215,5162			14	0	79899,5	370,735	76294,1
652	211,5454			14	0	78427,4	370,735	76294,1
653	201,9213		sell	15	76294,1	0	0	0
654	168,1107			15	76294,1	0	0	0
655	170,8927			15	76294,1	0	0	0
656	166,7271			15	76294,1	0	0	0
657	174,711			15	76294,1	0	0	0
658	174,5729			15	76294,1	0	0	0
659	170,5035			15	76294,1	0	0	0
660	179,8722			15	76294,1	0	0	0
661	177,3404			15	76294,1	0	0	0
662	180,7105			15	76294,1	0	0	0
663	175,1993			15	76294,1	0	0	0
664	176,985	buy	buy	15	0	76294,1	431,076	70953,5
665	176,3515			15	0	76021,0	431,076	70953,5
666	173,0591			15	0	74601,7	431,076	70953,5
667	181,1863	sell	sell	15	78105,2	0	0	0
668	182,0216			15	78105,2	0	0	0
669	193,2933			15	78105,2	0	0	0
670	191,6597			15	78105,2	0	0	0

Day	EthPrice	Plan	Done	nStopLos	Fiat	Virtual	nEth	StopLoss
671	182,5697	buy	buy	15	0	78105,2	427,810	72637,8
672	180,8266			15	0	77359,5	427,810	72637,8
673	182,0751			15	0	77893,6	427,810	72637,8
674	186,9609			15	0	79983,8	427,810	74384,9
675	181,4061	sell	sell	15	77607,4	0	0	0
676	176,0135			15	77607,4	0	0	0
677	178,0284			15	77607,4	0	0	0
678	173,6213	buy	buy	15	0	77607,4	446,992	72174,9
679	172,913			15	0	77290,8	446,992	72174,9
680	175,5344			15	0	78462,5	446,992	72970,1
681	174,921			15	0	78188,3	446,992	72970,1
682	172,3009	sell	sell	15	77017,1	0	0	0
683	162,4028			15	77017,1	0	0	0
684	162,1685			15	77017,1	0	0	0
685	181,5232	buy	buy	15	0	77017,1	424,282	71625,9
686	179,8355			15	0	76301,1	424,282	71625,9
687	184,2422			15	0	78170,8	424,282	72698,8
688	182,6628			15	0	77500,6	424,282	72698,8
689	190,3426			15	0	80759,0	424,282	75105,9
690	184,6922	sell	sell	15	78361,7	0	0	0
691	183,9669			15	78361,7	0	0	0
692	183,9699			15	78361,7	0	0	0
693	183,9257	buy	buy	15	0	78361,7	426,050	72876,3
694	182,425			15	0	77722,3	426,050	72876,3
695	186,3552			15	0	79396,7	426,050	73839,0
696	189,3042	sell	sell	15	80653,2	0	0	0
697	191,5938			15	80653,2	0	0	0
698	187,9765			15	80653,2	0	0	0