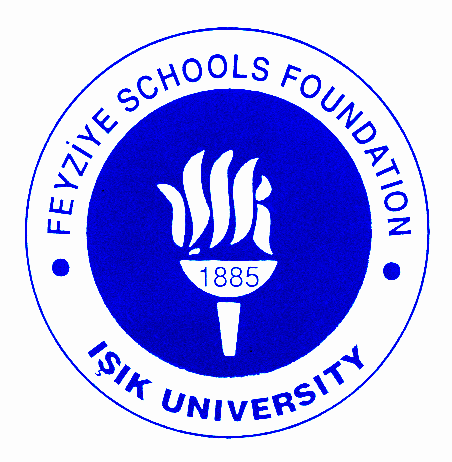
****

**IŞIK UNIVERSITY**

**Faculty of Arts and Sciences**

**Department of Information Technologies**

***B. S. Thesis***

**Cryptocurrency Information System and Future Price Analysis**

**by**

**Hayrican Yılmaz**

**217MI1191**

**Supervised by: *Elif Deniz Yiğitbaşı***

**December - 2021**

**Cryptocurrency Information System and Future Price Analysis**

A Project Presented

by

**Hayrican Yılmaz**

**217MI1191**

to

The Department of Information Technologies

in partial fulfillment of the requirements

for the degree of

**Bachelor of Science**

in the field of

Management Information Systems

Işık University

İstanbul, Turkey

December - 2021

© Copyright 2021 by Hayrican Yılmaz

All Rights Reserved

### Abstract

Cryptocurrencies are one of the biggest debates in recent years. Some people argue that cryptocurrencies specifically bitcoin are the currency of the future. Others think that cryptocurrencies are just a bubble and have no future. In addition to these discussions and dilemmas, the number of people who meet these markets and make losses in the hope of making money is increasing day by day. This paper is written to explain cryptocurrencies and technologies to people who do not yet have enough knowledge in this field. In addition, it is aimed to predict future prices with machine learning models and to base these estimations on rational foundations. Three different time series and machine learning methods were used in this project. Facebook Prophet is used for long-term forecasts, Neural Prophet is used for medium-term forecasts, and Vector Auto-Regressive Model is used for short-term forecasts.

### Acknowledgments

This project would not have been possible without the support of many people. I would like to thank my advisor, Elif Deniz Yelmenoğlu, who examined my project countless times, guided me, always supported me, and forced me to be better. It was a great pleasure working with her. I would also like to thank my friends and family who have always helped me in different stages of my thesis.

### Contents

[Abstract 2](#_Toc33256551)

[Acknowledgments 3](#_Toc33256552)

[Contents 4](#_Toc33256553)

[List of Figures 6](#_Toc33256554)

[Chapter 1 Introduction 7](#_Toc33256555)

[1.1 Overview 7](#_Toc33256556)

[1.2 Motivation 8](#_Toc33256557)

[1.3 Contributions 8](#_Toc33256558)

[1.4 Development Tools 8](#_Toc33256559)

[Chapter 2 Literature Review 9](#_Toc33256560)

[2.1 Measuring Information Systems Success 9](#_Toc33256562)

[2.2 Artificial Intelligence in Finance 9](#_Toc33256563)

2.3 Deep Learning in Finance…………………………………………………………………………………………………………..10

2.4 Algorithmic Trade………………………………………………………………………………………………………………………10

2.5 Bitcoin's Price Prediction……………………………………………………………………………………………………………11

2.6 Time Series Forecasting……………………………………………………………………………………………………………..11

2.7 Forecasting at Scale…………………………………………………………………………………………………………………..11

[Chapter 3 Methods 1](#_Toc33256564)3

[3.1 Overview 1](#_Toc33256565)3

[3.2 Blockchain and Cryptocurrency 1](#_Toc33256566)3

[3.3 Stationarity Tools 1](#_Toc33256567)6

[3.4 Vector Autoregressive Model(VAR) 1](#_Toc33256568)7

3.5 Facebook Prophet……………………………………………………………………………………………………………………..20

3.6 Neural Prophet………………………………………………………………………………………………………………………….23

[Chapter 4 Experimental Results](#_Toc33256569) 26

[4.1 Overview](#_Toc33256570) 26

[4.2 Data Preprocessing and Normalization](#_Toc33256571) 26

4.3 Vector Auto Regression……………………………………………………………………………………………………………..31

4.4 Prophet……………………………………………………………………………………………………………………………………..32

4.5 Neural Prophet………………………………………………………………………………………………………………………….35

[Chapter 5 Implementation of Project](#_Toc33256572) 38

[5.1 Overview](#_Toc33256573) 38

[5.2 Project’s Web Interface](#_Toc33256574) 38

[Chapter 6 Results and Discussion 4](#_Toc33256575)0

[6.1 Results 4](#_Toc33256576)0

[6.2 Discussion 4](#_Toc33256577)0

[References 4](#_Toc33256579)1

### List of Figures

Figure 1 – Flow of Blockchain……………………………………………………………………………….14

Figure 2 - Differences Between Cryptocurrencies..……………………………..………..…………….……15

Figure 3 – Human in the loop….……………………………………………………………………………..21

Figure 4 –Time Series Data………………………………………………………………………………..…27

Figure 5 – Bitcoin Closing Price Before Normalization …………………………………………………….27

Figure 6 – Bitcoin Volume Before Normalization …………………………………………………………..28

Figure 7 – US Dollar Price After Normalization……………………………………………………..………28

Figure 8 – Normalization Process..…………………………………………………………………………..29

Figure 9 – Bitcoin Closing Price After Normalization...………………………………………………..……29

Figure 10 - Bitcoin Volume After Normalization …….…………………………………………………..…30

Figure 11 - US Dollar Price After Normalization.……………………………………………………..…….30

Figure 12 – Output of ADF Test…...……………………..……………………………………………….....31

Figure 13- VAR Model Predictions………....…………………………...…………………………………..32

Figure 14 – Prophet Bitcoin Forecast Model...….………..………………………….………………………33

Figure 15 – Bitcoin Model Components..……………...…………………………………………………….33

Figure 16 – Prophet US Dollar Forecast Model …..…...…………………………………………………….34

Figure 17 – US Dollar Index Components ……………….……………………………………………….…34

Figure 18 - Prophet Bitcoin Volume Forecast.……….……….………………………………….………….35

Figure 19 - Bitcoin Volume Components ……………….…………………………………………………..35

Figure 20 - Neural Prophet Bitcoin Forecast …….……....................…………………………………….…36

Figure 21 – Neural Prophet Bitcoin Model Components ………………..………………………………….37

Figure 22 – Web Interface.……………………………….……………....………………………………….38

Figure 23 - Analyses Section.……………………………….……………………………………………….39

Table 1 - Correlation Matrix of Residuals ………………………………….………….……………...…….31

Table 2 - MAE and RMSE results for different train times..………………….……….…………………….37

### Introduction

##### Overview

Information systems are used in almost every field today. We touch information systems in everything we do. Information systems help us collect, process, use, and store information much more effectively. Modern information systems are rapidly replacing traditional information systems and making information systems more useful in every field. Integrating information systems into the financial system of the future enables more effective management of information, time, and money. Another popular field today is artificial intelligence and machine learning. Artificial intelligence has gained a place in many fields. Artificial intelligence and machine learning are used in many areas such as cyber security, voice assistants, navigation systems, health services, etc. In my project, I aim to predict the future price actions of cryptocurrencies by using machine learning and statistical analysis methods. Financial markets are one of the areas where machine learning is used most intensively. It is possible to analyze the future price with the help of algorithms by focusing only on rational data and examining the historical data of a currency. Artificial intelligence is a great need in financial markets where there is no room for emotion and hesitation.

##### Motivation

Cryptocurrencies are one of the rising trends of today. Many people have learned about cryptocurrencies and the technology that created them, but there are still many people who are new to cryptocurrencies, and these people often make wrong investments and suffer losses at the beginning of their crypto journey. In every financial market, people want to know the future of the stock they invest in. This is also true in the crypto market. To date, many methods have been used to predict the future price of bitcoin. These were traditional statistical or machine learning methods. Predicting the future price of bitcoin using sentiment analysis with the help of machine learning has been a very popular topic in recent years, but not everyone uses social media or expresses their opinions about bitcoin through social media. This suggests that making predictions with sentiment analysis is might not be a reliable method. The motivation for the preparation of this paper is to provide new crypto users with a simple interface at the start of their journey and to forecast future prices based on more rational foundations than sentiment analysis.

##### Contributions

This paper will show whether the correlation between the dollar index and bitcoin is negative or positive, and based on these foundations, it will predict future prices with machine learning techniques to guide new crypto users and will provide users with a basic interface.

**1.4 Development Tools**

The web interface of the project was prepared in Visual Studio Code environment using Html, CSS, and Javascript. The machine learning part of the project was prepared in the Google Colab environment using Python.

### Literature Review

##### Measuring Information Systems Success

Modern information systems have been growing continuously since their inception and their budgets continue to rise even in the face of potential economic downturns. However, fears about economic conditions and increased competition create pressures to reduce costs, resulting in the need for organizations to measure the benefits and costs of technology. Naturally, organizations are interested in knowing the return on these investments. The effects of IT are often indirect and are influenced by human, organizational and environmental factors. Early attempts to define information system success were ill-defined due to the complex, interdependent, and multi-dimensional nature of IS success. To address this problem, DeLone & McLean (1992) reviewed the research published during the period 1981–1987 and created a taxonomy of IS success based upon this review. In their 1992 paper, they identified six variables or components of IS success: system quality, information quality, use, user satisfaction, individual impact, and organizational impact. However, these six variables are not independent success measures but are interdependent variables [1, 2].

##### Artificial Intelligence in Finance

The concept of AI originates from a 1955 Dartmouth Summer [3]. Later evolutions of this cohort posited that ''every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it'' in an endeavor to ''find how to make machines use language, form abstractions, and concepts, solve problems now reserved for humans and improve themselves''. Turing tests, which assess the ability of machines to demonstrate intelligent behavior (Turing, 1995), show that machines often outperform humans in solving complex judgment-related problems, such as the identification of decision variables out of a very large number of candidates constrained within a high-dimensional space. Indeed, ML, when informed by deep learning, has been remarkably successful for a variety of tasks [4]. The emerging use of artificial intelligence (AI) and machine learning (ML) within financial systems has made artificial intelligence and machine learning an important actor in the financial sector and disrupting and transforming industries, and societies [5 ,6]. In recent years, the cryptocurrency market has experienced exponential growth in trading volüme and number of tradable coins, sparking the attention of governments, firms, investors, academics, and the general public alike [7].

##### Deep Learning in Finance

Deep learning (DL) is an advanced technique of machine learning (ML) based on artificial neural network (NN) algorithms. As a promising branch of artificial intelligence, DL has attracted great attention in recent years. Compared with conventional Machine Learning techniques such as support vector machine (SVM) and k-nearest neighbors (kNN), DL possesses advantages of unsupervised feature learning, a strong capability of generalization, and a robust training power for big data [8]. There are many reasons why deep learning in finance is important. The first is the availability of big data. Second, many finance applications depend on speed, and deep learning allows us to achieve fast response levels. Third, much of finance involves pattern recognition using data, where a wide variety of inputs are modeled to predict outputs.

**2.4 Algorithmic Trade**

Algorithmic trading refers to the use of algorithms to make better trade decisions. Usually, traders build mathematical models that monitor business news and trade activities in real-time to detect any factors that can force security prices to rise or fall. The model comes with a predetermined set of instructions on various parameters – such as timing, price, quantity, and other factors – for placing trades without the trader’s active involvement. Unlike human traders, algorithmic trading can simultaneously analyze large volumes of data and make thousands of trades every day. Machine learning makes fast trading decisions, which gives human traders an advantage over the market average. Also, algorithmic trading does not make trading decisions based on emotions, which is a common limitation among human traders whose judgment may be affected by emotions or personal aspirations. The trading method is mostly employed by hedge fund managers and financial institutions to automate trading activities [10, 11].

**2.5 Bitcoin’s Price Prediction**

Bitcoin is the first example of cryptocurrency. It uses computer software that provides complex mathematical solutions hence it helps to make business simple. However, Bitcoin's price is quite volatile and to estimate it machine learning techniques are presented by various researchers [12].

**2.6 Time Series Forecasting**

Generally, the time series analysis may be divided into main branches, that is, forecasting and dynamic modeling. Time series forecasting means that predicting the future values based on the understanding of previous data distribution, that is, mean, variance, skewness, and kurtosis. It implies that we do not test any type of hypothesis or economic theory in this branch of time series analysis. We simply try to understand the patterns of the time series through different statistics and predict the future. More clearly, we are not concerned with hypothesis testing, investigating the relationships, or building models in this line of research. We just try to extract the information from the past to predict the future [13]. Traditional time series models like SARIMAX have many stringent data requirements like stationarity and equally spaced values. Other time series models like Recurring Neural Networks with Long-Short Term Memory (RNN-LSTM) can be highly complex [14].

**2.7 Forecasting at Scale**

Facebook Prophet is an open-source algorithm for generating time-series models that uses a few old ideas with some new twists. It is particularly good at modeling time series that have multiple seasonalities and doesn’t face some of the above drawbacks of other algorithms. At its core is the sum of three functions of time plus an error term: growth g(t), seasonality s(t), holidays h(t), and error . [14].

### Methods

##### Overview

Three machine learning techniques were used in this project. For the models to work with their best performance, stationarity should be ensured in the data. Before applying the models to the data, normalization was performed to ensure stationarity. Before moving on to machine learning methods, we need to examine blockchain and cryptocurrency technologies.

##### Blockchain and Cryptocurrency

**3.2.1 Blockchain**

Blockchain is a shared, immutable ledger that simplifies the process of recording transactions and tracking assets in a business network. Assets can be tangible (house, car, cash, land) or intangible (intellectual property, patents, copyrights, brand). Almost anything of value can be tracked on a blockchain network. The importance of a blockchain network is that; business proceeds through information. The faster and more accurate information is received, the better. Blockchain is ideal for presenting information as it provides instantaneous, shared, and completely transparent information stored in a non-modifiable ledger that can only be accessed by authorized network members. A blockchain network can track orders, payments, accounts, production, and much more.

In recent years, interest in blockchain technology has increased. Increasing interest brings the spread of technology and its use in many different fields. The finance field is the field where blockchain technology is used most intensively. The most popular application of blockchain technology is bitcoin. To understand Bitcoin, we first need to understand the working logic of blockchain technology.

|  |
| --- |
|  |
| *Figure 1 Flow of Blockchain[15]* |

A blockchain consists of data sets which are composed of a chain of data packages (blocks) where a block comprises multiple transactions [15]. The blockchain is extended by each additional block and hence represents a complete ledger of the transaction history. Blocks can be validated by the network using cryptographic means. In addition to the transactions, each block contains a timestamp, the hash value of the previous block (‘parent’), and a nonce, which is a random number for verifying the hash. This concept ensures the integrity of the entire blockchain through to the first block (‘genesis block’). Hash values are unique and fraud can be effectively prevented since changes of a block in the chain would immediately change the respective hash value. If the majority of nodes in the network agree by a consensus mechanism on the validity of transactions in a block and on the validity of the block itself, the block can be added to the chain. Therefore new transactions are not automatically added to the ledger. Rather, the consensus process ensures that these transactions are stored in a block for a certain time before being transferred to the ledger. Afterward, the information in the blockchain can no longer be changed. In the case of Bitcoin, blocks are created by so-called miners who are rewarded with Bitcoins for validating the blocks. The example of Bitcoin illustrates that the principle of the blockchain cannot only change the process of money transactions. Using cryptography, people all over the world can trust each other and transfer different kinds of assets peer-to-peer over the internet [16].

**3.2.2 Cryptocurrency**

Cryptocurrency, a product of blockchain technology is a form of payment that can be exchanged online for goods and services. Unlike other investments, cryptocurrency is not tied to physical assets. The purpose of cryptocurrencies is to allow two people anywhere to exchange value directly. There is no central party controlling the network, there is no central bank that could shut down or arbitrarily raise or lower the value. The first cryptocurrency product was bitcoin and has been the market leader in cryptocurrencies since the first bitcoin was mined in 2009. After the bitcoin's genesis block, more than 15.000 altcoins have been created (Altcoin means alternative coin and is the name given to all cryptocurrencies produced after bitcoin). With the rise of cryptocurrencies, the number of centralized and decentralized exchanges where cryptocurrencies can be bought and sold has also increased rapidly. According to Coinmarketcap, today there are a total of 446 decentralized and centralized exchange platforms where crypto-assets can be bought and sold. Cryptocurrencies are divided into classes within themselves. Some cryptocurrencies have their blockchain network and could be minable and some of them do not. Some cryptocurrencies are called tokens developed on a blockchain network and some cryptocurrencies have a dApp(decentralized application) platform. Table of the qualitative factors of 30 selected cryptocurrencies to understand the difference between cryptocurrencies: [17- 19].

|  |
| --- |
|  |
| *Figure 2 Differences Between Cryptocurrencies[17-19]* |

##### Stationarity Tools

In order to the data fit the VAR model and get better prediction results, the data must be stationary and since we are dealing with time series data we must examine the components of time series data and minimize the effect of the components to make our data stationary. Time series data consist of 4 main components:

* Trend = Trend is the increase or decrease in the series over a period of time, it persists over a long period of time.
* Seasonality = Regular pattern of up and down fluctuations. It is a short-term variation occurring due to seasonal factors.
* Cyclically = It is a medium-term variation caused by circumstances, which repeat in irregular intervals.
* Irregularity = It refers to variations which occur due to unpredictable factors and also do not repeat in particular patterns.

###### Autocorrelation(ACF) and Partial Autocorrelation(PACF)

The coefficient of correlation between two values in a time series is called the autocorrelation function (ACF) For example the ACF for a time series is given by Equation (1):

|  |  |
| --- | --- |
|  | *(1)* |

This value of k is the time gap being considered and is called the lag. A lag 1 autocorrelation (k = 1 in the above) is the correlation between values that are one time period apart. More generally, a lag k autocorrelation is a correlation between values that are k time periods apart. The ACF is a way to measure the linear relationship between an observation at time t and the observations at previous times. If we assume an AR(k) model, then we may wish to only measure the association between and filter out the linear influence of the random variables that lie in between, which requires a transformation on the time series. Then by calculating the correlation of the transformed time series we obtain the partial autocorrelation function (PACF). The PACF is most useful for identifying the order of an autoregressive model. Specifically, sample partial autocorrelations that are significantly different from 0 indicate lagged terms that are useful predictors. To help differentiate between ACF and PACF, think of them as analogues to and partial values as discussed previously. Graphical approaches to assessing the lag of an autoregressive model include looking at the ACF and PACF values versus the lag. In a plot of ACF versus the lag, if you see large ACF values and a non-random pattern, then likely the values are serially correlated. In a plot of PACF versus the lag, the pattern will usually appear random, but large PACF values at a given lag indicate this value as a possible choice for the order of an autoregressive model. It is important that the choice of the order makes sense.

###### Transformation

Applying differencing or seasonal differencing log of the series should make the series stationary. Simply, stationarity removes trends from the dataset which can be extremely intrusive to our models. Basically, stationarity makes our models perform and predict better.

**3.3.3 Normalization**

Improving the quality of data has become one of the prime challenges in the field of machine learning (ML) as volume, velocity, and variety of data are increasing rapidly [20]. In practice, the data could be susceptible to numerous problems when acquired from either heterogeneous sources with a single method or homogeneous sources with different methods. The issues include noise, missing values, class imbalance, data inconsistency, presence of outliers, and dominant features which decreases the quality of data. Thus, the pre-processing of data becomes inevitable to deal with the low-quality of data before its evaluation [21]. Normalized data within statistics, often involves eliminating units of measurement from a dataset. As a result, this enables us to easily compare data with different scales and are measured from different sources. When training a machine learning model, we aim to bring the data to a common scale and so the various features are less sensitive to each other. In this case, we can utilize data normalization as a method of transforming our data, which may be of different units or scales. This allows our model to train using features that could lead to more accurate predictions.

**3.4 Vector Autoregressive Model (VAR)**

One of the most convenient and generally accepted methodologies to study the dynamic relationship among a group of variables is the Vector Auto Regression Model (VAR) introduced by Sims [22]. It is a natural extension of the univariate autoregressive model to dynamic multivariate time series. The VAR model has proven particularly useful for describing and predicting the dynamic behavior of economic and financial time series. It often provides predictions superior to those from univariate time series models and detailed theory-based simultaneous equation models [23]. Predictions from VAR models are quite flexible because they can be made dependent on the potential future paths of the variables specified in the model. As an extension of the AR (autoregressive) model, VAR is commonly applied to forecast systems for the inter-related time series as well as to analyze the dynamic impact for random disturbances on the variable systems. All variables in a VAR enter the model in the same way: each variable has an equation explaining its evolution based on its own lagged values, the lagged values of the other model variables, and an error term. The VAR model of order offers a way of producing the forecast as a weighted sum of historical time series not only at the target location but also from its surrounding sampled locations [24, 25].

Simply, the VAR model structure is that each variable is a linear function of past lags of itself and past lags of the other variables. As an example suppose that three different time series variables will be measured, denoted by and .

The vector autoregressive model of order 1, denoted as VAR(1), is as follows in Equation (2) and Equation (3):

|  |  |
| --- | --- |
|  | *(2)* |
|  | *(3)* |

Each variable is a linear function of the lag 1 values for all variables in the set. In a VAR(2) model, the lag 2 values for all variables are added to the right sides of the equations, In the case of three x-variables (or time series) there would be six predictors on the right side of each equation, three lag 1 terms and three lag 2 terms.

In general, for a VAR(p) model, the first p lags of each variable in the system would be used as regression predictors for each variable. VAR models are a specific case of more general VARMA models. VARMA models for multivariate time series include the VAR structure above along with moving average terms for each variable. More generally yet, these are special cases of ARMAX models that allow for the addition of other predictors that are outside the multivariate set of principal interests [23].

**3.4.1 ADF Test (Augmented Dickey-Fuller Test)**

The Augmented Dickey-Fuller test (ADF test) is a common statistical test used to test whether a given Time series is stationary or not. It is one of the most commonly used statistical tests when it comes to analyzing the stationary of a series. The ADF test belongs to a category of tests called the 'Unit Root Test'. It is the proper method for testing if a time series is stationary or not. The unit root is a characteristic of a time series that makes it non-stationary. Technically, a unit root is said to exist in a time series of the value of alpha = 1 in the below equation. Unit Root Test is as fallows in Equation (4):

|  |  |
| --- | --- |
|  | *(4)* |

is the value of the time series at time 't' and is an exogenous variable. The presence of a unit root means the time series is non-stationary. The ADF test is an 'augmented' type of the Dickey-Fuller test. Dickey-Fuller test is a unit root test that tests the null hypothesis that α=1 in the following model equation. alpha is the coefficient of the first lag on Y. Null Hypothesis (H0): alpha=1. Dickey-Fuller Test is as fallows in Equation (5):

|  |  |
| --- | --- |
|  | *(5)* |

where,

= lag 1 of time series and delta = first difference of the series at time (t-1)

The coefficient of is 1, implying the presence of a unit root. If not rejected, the series is taken to be non-stationary. The Augmented Dickey-Fuller test evolved based on the above equation. The ADF test expands the Dickey-Fuller test equation to include high order regressive process in the model. The ADF test is as fallows in Equation (6):

|  |  |
| --- | --- |
|  | *(6)* |

Only the term that makes more difference is added, while the rest of the equation remains the same. But the null hypothesis is still the same as the Dickey-Fuller test. Since the null hypothesis assumes the existence of a unit root, α=1, the p-value obtained must be less than the significance level (0.05) to reject the null hypothesis. From this, it is concluded that the series is stationary [26, 27].

**3.5 Facebook Prophet**

Understanding time-based patterns are important for any business. Questions such as how much inventory you should hold, how many visitors you expect in your store, how many people will travel with an airline, and what will be the future of a stock you invest in are important time series problems that need to be resolved. For such reasons, time series analysis is very important for companies, individuals, and data engineers working in this field. Many methods have been created, developed, and actively used for time series analysis. Undoubtedly, one of the most important methods that make predictions by making time series analysis is the open-source Prophet library created by Facebook. Of course, making highly accurate predictions is not an easy problem to solve, even for advanced machines. Facebook engineers focused on two main themes while developing Prophet. The two main themes they identified are as follows:

First, completely automatic forecasting techniques can be hard to tune and are often too inflexible to incorporate useful assumptions or heuristics. Second, the analysts responsible for data science tasks throughout an organization typically have deep domain expertise about the specific products or services that they support, but often do not have training in time series forecasting. Analysts who can produce high-quality forecasts are thus quite rare because forecasting is a specialized skill requiring substantial experience. The result is that the demand for high-quality forecasts often far outstrips the pace at which they can be produced. This observation is the motivation for the research we present here – we intend to provide some useful guidance for producing forecasts at scale, for several notions of scale [28].

The prophet is generally considered a competitor to ARIMA models and generally has higher accuracy rates than ARIMA models. Facebook designed the Prophet forecasting model for handling the characteristic features of business time series such as multiple strong seasonality, trend changes, outliers, and holiday effects. It is an additive model in which non-linear trends are fit with different seasonality and holidays [29].

While Facebook engineers were designing the prophet, it was not just for the use of experts; They paid attention to having a method and interface that can be easily used by people who do not have much knowledge and experience in time series analysis. So they designed the prophet in such a way that large numbers of people could use the data they had without training them to generate predictions. As a result, According to Facebook, Prophet offers a straightforward way to create a ‘reasonable and accurate forecast’. Here is how Facebook visualized its automated time series analysis:

|  |
| --- |
|  |
| *Figure 3 Human in the loop [28]* |

To interpret this image, Prophet automates the tasks of a data analyst specialized in the field of forecasting, unlike traditional time series methods. Now, in order to better understand the prophet, we must look at its working principles. Facebook uses a decomposable time series model with three main model components: trend, seasonality, and holidays [28]. They are combined in the following Equation (7):

|  |  |
| --- | --- |
|  | *(7)* |

* g(t) = Piecewise linear or logistic growth curve for modeling non-periodic changes in time series
* s(t) = Periodic changes (weekly/yearly seasonality)
* h(t) = Effects of holidays (user-provided) with irregular schedules
* = The error term εt provides the changes which are not accommodated by the model.

In the ARIMA model, the measurements must be regularly spaced, and it does not consider the outliers, whereas the Prophet model does include the outliers and data that need not be periodic. Hence, the Prophet model provides lesser error values in comparison with the ARIMA model.” Yenidogan” has compared the bitcoin price prediction done by ARIMA and Prophet models [30]. The Prophet model provided a prediction near the correct price with 94.5% precision, whereas the ARIMA model showed only 68% precision. The Prophet model performed better than the ARIMA model as it considers the seasonality and holiday effects [31].

Trend = Trend is modeled by fitting a piecewise linear curve over the trend or the non-periodic part of the time series. The linear fitting exercise ensures that it is least affected by missing data.

One of the most important issues that we should pay attention to while examining the trend and trend change, do we expect the target that creates our trend to show a continuous decrease or increase depending on the trend direction throughout the entire forecast period? Here, the analyst who will create the forecasting model comes into play. If there is a case of non-linear growth with a working maximum capacity, the analyst must define the varying capacity variable C(t) for subsequent time-series predictions. To give an example of this situation, if we are trying to predict the future trend of the number of active sim card users in a region, we should consider the following: the number of sim cards will always depend on the number of phone users in the region because there cannot be more active sim card than the number of phones in the region, so the number of sim cards and phones number will be in positive correlation.

Changepoints = Another topic to be aware of is whether the time series encounters any fundamental changes in the phenomena; a new product launch, unforeseen disaster, pandemics, sudden economic crises, etc. At such points, the growth rate is allowed to change. These change points are automatically selected. However, a user can also set change points himself if needed.

Seosanility = In order to forecast the effects of seasonality, the prophet relies on the Fourier series to provide a flexible model [29]. Seasonal effects s(t) are calculated with the following Equation (8):

|  |  |  |
| --- | --- | --- |
| |  |  | | --- | --- | |  | *(8)* | |
|  |

P is the period (365.25 for yearly data and 7 for weekly data)

Parameters [, , ….., , ] need to be estimated for a given N to model seasonality.

The Fourier order N that defines whether high-frequency changes are allowed to be modeled is an important parameter to set here. For a time series, if the user believes the high-frequency components are just noise and should not be considered for modeling, he/she could set the values of N from a lower value. If not, N can be tuned to a higher value and set using the forecast accuracy.

Holidays and events = Holidays and events cause predictable changes in a time series. For example, a country's national holiday may cause unusual effects on that day. Prophet allows the analyst to present a custom list of past and future events. In this way, special days can be evaluated differently with additional parameters [32].

**3.6 Neural Prophet**

"Neural Prophet" is built on PyTorch by Facebook and heavily inspired by Facebook Prophet and AR-Net libraries [33, 34]. It is simply a python library created for modeling time series data based on neural networks. According to Facebook, with the proliferation of time series data, explainable forecasting remains a challenging task for businesses, and operational decision-making hybrid solutions are needed to bridge the gap between interpretable classical methods and scalable deep learning models [33]. Prophet was designed for this purpose, but due to a lack of local context, it was not very successful in predicting the near future. Therefore, to solve this problem, Prophet's successor hybrid solution "Neural Prophet" was designed. There is only one difference between Neural Prophet and Prophet, but this difference provides a much higher success rate than Prophet in near-term predictions. The local context that Prophet lacks is introduced with automatic regression and covariate modules that can be configured as classical linear regression or Neural Networks. Other than that, Neural Prophet maintains Prophet's design philosophy and provides the same core model components. For short to medium-term forecasting, Neural Prophet improves forecast accuracy by 55 to 92 percent compared to Prophet [33]. Forecasting is not a new term and traditional statistical techniques were used for forecasting before machine learning became popular. For example, models such as ARIMA, SARIMA, ETS are models that have been used for years and have had a lot of work and research. However, due to the weaknesses of these models, for example, their restrictive assumptions and parametric structures have led experts to seek and create different models and techniques. As a competitor to traditional statistical methods, machine learning models did not perform well at first. In fact, the neural network was labeled as not competitive for forecasting [35]. But later, with the increase in the availability of large-scale time series, neural network-based techniques gained their popularity in the field of forecasting. Neural Prophet's contribution is described as follows: Inspired by Prophet's impact on the forecasting community, it is our objective to continue the democratization of forecasting tooling, by making hybrid models as accessible as Prophet-made classic models [33]. Neural Prophet combines classical time series components into a hybrid model that allows it to conform to nonlinear dynamics. These components are automatic regression and covariate modules. Unlike Prophet, Neural Prophet uses PyTorch as its backend, so Neural Prophet can be updated with the latest innovations in deep learning. A core concept of the Neural Prophet model is its modular composability [33]. The working principle of the model is explained as follows:

The model is composed of modules which each contribute an additive component to the forecast. Most components can also be configured to be scaled by the trend for a multiplicative effect. Each module has its individual inputs and modeling processes. However, all modules must produce h outputs, where h defines the number of steps to be forecasted into the future at once. These are added up as the predicted values for the time series future values. If the model is only time-dependent, an arbitrary number of forecasts can be produced. In the following descriptions, that special case will be treated mathematically equivalent to a one-step-ahead forecast with h = 1. Neural Prophet function is as fallows in Equation (9):

|  |  |
| --- | --- |
| = T(t) + S(t) + E(t) + F(t) + A(t) + L(t) | *(9)* |

* T(t) = Trend at time t
* S(t) = Seasonal effects at time t
* E(t) = Event and holiday effects at time t
* F(t) = Regression effects at time t for future-known exogenous variables
* A(t) = Auto-regression effects at time t based on past observations
* L(t) = Regression effects at time t for lagged observations of exogenous variables

All model component modules can be individually configured and combined to compose the model. If all modules are switched off, only a static offset parameter is fitted as the trend component. By default, only the trend and seasonality modules are activated.

### Experimental Results

##### Overview

This chapter includes the results of machine learning models applied to time series data. Each model serves a different purpose. The VAR model makes short-term predictions, the Neural Prophet model makes medium-term predictions, and the Prophet model makes long-term predictions.

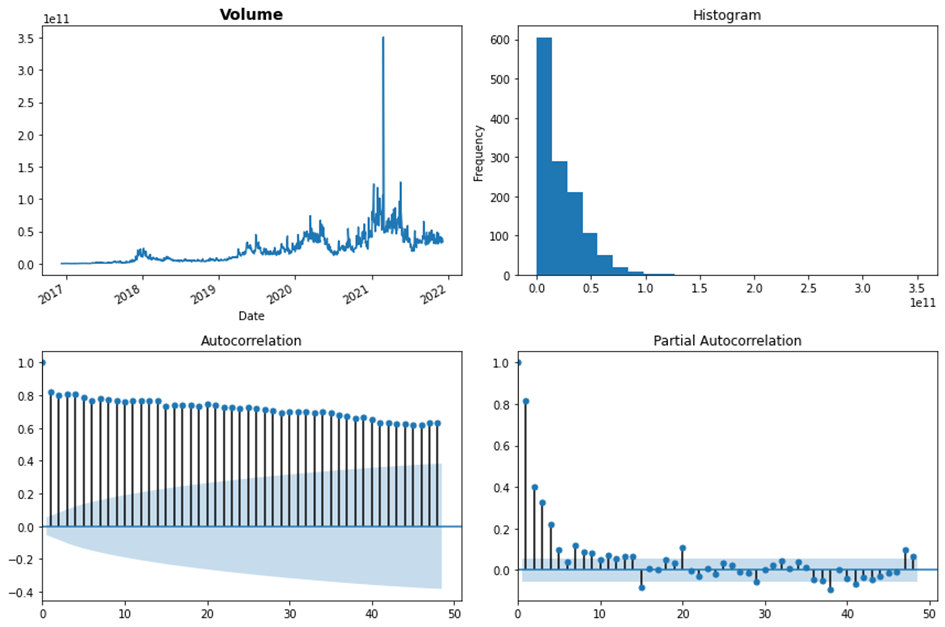
##### Data Preprocessing and Normalization

Time series data required for the model were collected from Yahoo Finance and Investing.com. I worked on 5 years of bitcoin and dollar index data from 2016-12-08 to 2021-12-07. I used the closing and volume columns for Bitcoin and only the closing column for the dollar index. But a problem arose while outputting the data. While there were 1826 rows of data for Bitcoin's closing and the volume column, there were 1305 data for the dollar index. This is because the cryptocurrency markets are uniquely open 24/7, while the same is not true for the dollar index. The dollar index is traded 5 days a week. So I needed to sync row data of bitcoin and dollar index to fit the data to the VAR model. After performing the synchronization, Graphical outputs of the data.

|  |
| --- |
|  |
| *Figure 4 Time Series Data* |

Close represents the daily closing price of bitcoin, the volume represents the daily volume data of bitcoin, and the price represents the daily closing price of the Dollar index. In order to interpret the graphics in more detail, we need to examine the price, histogram, autocorrelation, and partial autocorrelation values.

|  |
| --- |
|  |
| *Figure 5* *Bitcoin Closing Price Before Normalization* |



*Figure 6 Bitcoin Volume Before Normalization*

|  |
| --- |
|  |
|  |
| *Figure 7 US Dollar Index Before Normalization* |
|  |

These graphs showed that the data needed normalization to be stationary and fit the VAR model. After converting the data to a logarithmic scale, the min-max normalization method was used for the normalization process. Graphs obtained as a result of the normalization process:

|  |
| --- |
|  |
| *Figure 8 Normalization Process* |

These graphs showed that the data needed normalization to be stationary and fit the VAR model. After converting the data to a logarithmic scale, the min-max normalization method was used for the normalization process. Graphs obtained as a result of the normalization process:

|  |
| --- |
|  |
| *Figure 9 Bitcoin Closing Price After Normalization* |

|  |
| --- |
|  |
| *Figure 10 Bitcoin Volume After Normalization*   |  | | --- | |  | | *Figure 11 US Dollar Index After Normalization* | |

After the normalization process, we will run a statistical test (ADF test) to determine how well the time series was transformed to be stationary. This is a test that outputs certain statistical patterns that we can use to judge whether each parameter is stationary.

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

*Figure 12 Output of ADF Test*

According to the results of the ADF test, we see that the stationarity condition is met for all three-time series. Here we have to look specifically at whether the P-value is less than 0.05. If the P-value is less than 0.05, we can say that our time series is stationary.

**4.3 Vector Auto Regression**

"VAR" model results:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **BTC Close** | **BTC Volume** | **US Dollar Price** |
| **BTC Close** | 1.000000 | 0.103189 | -0.058065 |
| **BTC Volume** | 0.103189 | 1.000000 | 0.009440 |
| **US Dollar Price** | -0.058065 | 0.009440 | 1.000000 |

*Table 1* *Correlation Matrix of Residuals*

As seen in the table, there is a negative correlation between Bitcoin Closing Price and US Dollar price according to the Matrix results. This means that when the price of Bitcoin rises, the Dollar index price may tend to fall. When the price of the dollar index rises, the price of bitcoin may tend to fall. There is a positive correlation between the volume and price of Bitcoin. When Bitcoin's volume increases, its price tends to increase as well.

These results show us that before investing in Bitcoin, it may be useful to look at the future forecast of the dollar index and bitcoin volume. We will obtain these predictions with machine learning models. The VAR model will be used for the 7-day forecast, the Neural Prophet for the 3-month forecast, and the Facebook Prophet for the 1-year forecast. The reason for determining the estimation times in this way is; According to the literature research, these time intervals are the periods when the models are most effective.

More detailed results of the VAR model can be found at the link below.

[VAR Results.txt](file:///C:\Users\hyrcn\Downloads\VAR%20Results.txt)

|  |
| --- |
|  |
| *Figure 13 VAR Model Predictions* |

The Mean Square Error value of the VAR was 1.80.

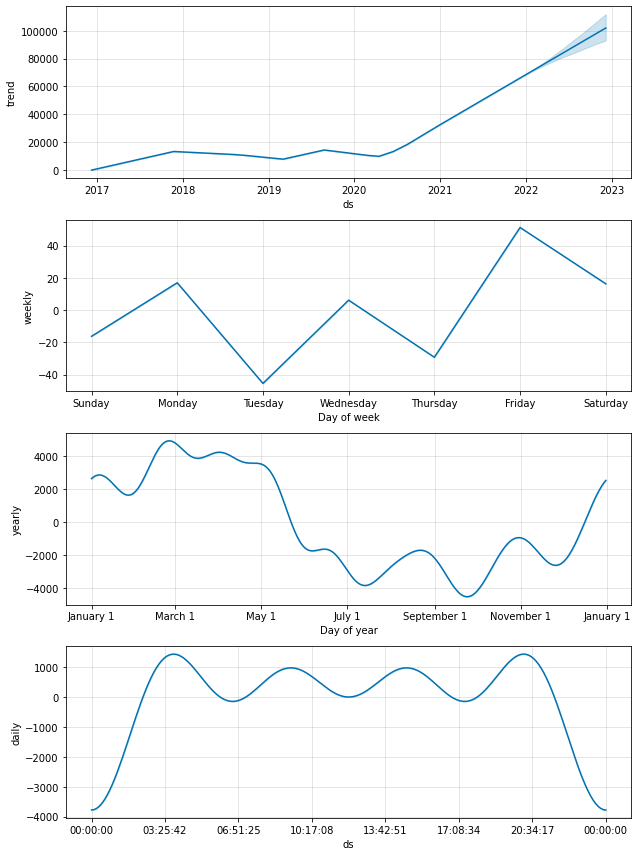
According to our VAR model results, we have defined the negative correlation between Bitcoin and Dollar index. Now we will examine bitcoin and the dollar index together with our next machine learning methods.

**4.4 Prophet**

After transforming the data as Prophet wanted, graphic outputs were taken.

|  |
| --- |
|  |
| *Figure 14* *Prophet Bitcoin Forecast Model* |

Prophet displays its predictions both as a line and as an interval. This means, for example; bitcoin could be between $90,000 and $105,000 in early 2023. This demonstration is ideal for new cryptocurrency users to understand easily. Complex graphics can be difficult to parse. Components show us the general trend, weekly, daily and annual seasonality. You can see the Bitcoin components in the figure below.



*Figure 15 Bitcoin Model Components*

We will get the same graphic outputs for the dollar index and bitcoin volume. The dollar index graph is given in the figure below.

|  |
| --- |
|  |
| *Figure 16* *Prophet US Dollar Forecast Model*  You can see the US Dollar Index components in the figure below.   |  | | --- | |  | | *Figure 17 US Dollar Index Components* | |
|  |
| *Figure 18* *Prophet Bitcoin Volume Forecast* |
|  |
| *Figure 19* *Bitcoin Volume Components* |

**4.5 Neural Prophet**

As mentioned earlier, Neural Prophet is a hybrid model. So Neural Prophet is a mix of Facebook Prophet and neural networks. While applying this model to our data, we will train our time series data in accordance with neural network principles. Since the working principle is the same as Prophet, the same conditions are required for the data in order to fit the model. Neural Prophet outperforms Prophet in short to medium time forecasts. So we will use Neural Prophet for the 120-day forecast.

|  |
| --- |
|  |
| *Figure 20* *Neural Prophet Bitcoin Forecast* |

|  |
| --- |
|  |
| *Figure 21* *Neural Prophet Bitcoin Model Components* |

|  |  |  |
| --- | --- | --- |
|  | **MAE** | **RMSE** |
| **10 Training** | 3.72e+3 | 5.14e+3 |
| **100 Training** | 3.36e+3 | 4.46e+3 |
| **1000 Training** | 3.39e+3 | 4.43e+3 |

*Table 2* *MAE and RMSE results for different train times*

### Implementation of Project

##### Overview

The purpose of this paper, a website has been created for new cryptocurrency users. The website "https://developers.coinranking.com/api/documentation" works with an API from this address. Through the website created, users can access the details of the projects in the top 50 in terms of market value in the cryptocurrency market. They can view the market rankings, symbols, icons, market values of cryptocurrencies, and access future price predictions with the analysis tab. Users who want to get more detailed information about any cryptocurrency project can connect to https://coinranking.com with the "Coin Detail" link next to each coin.

##### Project’s Web Interface

The backend of the website was designed with Javascript and the Frontend is designed with HTML and CSS.

|  |
| --- |
|  |
| *Figure 22* *Web Interface* |

|  |
| --- |
|  |
| *Figure 23 Analyses Section* |

### Results and Discussion

In this chapter, the results of the project and the developments that can be made in the future are mentioned.

##### Results

As a result of this study, it was proved that there is a negative correlation between Bitcoin and Dollar index based on the VAR model results. We should not ignore the dollar index while predicting the future price of Bitcoin. Facebook Prophet was used for long-term forecasts. Prophet was used because it was very popular, maintainable, and easy to use for non-experts as well. Prophet, who was very successful in long-term predictions, could not achieve the same success in short-term predictions due to the lack of local context. For this reason, Prophet's hybrid brother, which was created to make up for Prophet's shortcomings in short and medium-time predictions, was used for middle-time predictions. As a result of all these, machine learning and deep learning methods that serve three different time intervals with three different machine learning methods were used.

##### Discussion

With developments which can be made in the future, the website can be made more dynamic. Other items that new users may need can be identified and added to the website. With a simple interface, an integrated exchange can be created on the website, where new users can trade without getting confused. On the other hand, since the main purpose of this project is to serve new users and prevent them from being harmed, no plugins that could endanger them are acceptable.

### References

[1] Petter, S., DeLone, W. & McLean, E. Measuring information systems success: models, dimensions, measures, and interrelationships. Eur J Inf Syst 17, 236–263 (2008)

[2] William H. DeLone, Ephraim R. McLean, Information Systems Success: The Quest for the Dependent Variable. Information Systems Research 3(1):60-95, 1992.

[3] McCarthy, J., Minsky, M.L, Ronchester , N., Shannon C.E., 2006. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. AI Mag . 27 (4),12.

[4] Dixon, M.F., Halperin, I., Bilokon, P., 2020. Machine Learning in Finance : From Theory to Practice. Springer International Publishing.

[5] Li,Xiaojun,Tang, P.., 2020, Stock Index Prediction Based on Wavelet Transform and FCD-MLGRU. J. Forecast. 39(8), 1229-1237

[7] Kennard Fung , Jiin Jeong , Javier Pereira , More to Cryptos than Bitcoin: A GARCH Modelling of Heterogeneous Cryptocurrencies, Finance Research Letters (2021)

[6] John W. Goodell, Satish Kumar, Weng Marc Lim, Debidutta Pattnaik, Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis, Journal of Behavioral and Experimental Finance, Volume 32, 2021.

[8] Huang, J., Chai, J. & Cho, S. Deep learning in finance and banking: A literature review and classification. Front. Bus. Res. China 14, 13 (2020).

[9] Culkin R. And Das S.R., 2017. Machine Learning in Finance: The Case of Deep Learning for Option Pricing. [ebook] p.15. Available at: < https://srdas.github.io/Papers/BlackScholesNN.pdf.

[10] G. Nuti, M. Mirghaemi, P. Treleaven and C. Yingsaeree, "Algorithmic Trading," in Computer, vol. 44, no. 11, pp. 61-69, Nov. 2011, doi: 10.1109/MC.2011.31.

[11] Hendershott, T. and Riordan, R., 2011. Algorithmic Trading and Information. [ebook] p.41. Available at: <http://faculty.haas.berkeley.edu/hender/ATInformation.pdf.

[12] Sharma, R., Menghani, B., Kumar, M., Jangir, S.K., 2021. Bitcoin Exchange Rate Price Prediction Using Machine Learning Techniques: A Review. [ebook] p.8.

[13] Abdul Jalil, Nasir Hamid Rao, Chapter 8 - Time Series Analysis (Stationarity, Cointegration, and Causality), 2019, Pages 85-99.

[14] Krieger, M. 2021. Time Series Analysis with Facebook Prophet: How it works and How to use it. https://towardsdatascience.com/time-series-analysis-with-facebook-prophet-how-it-works-and-how-to-use-it-f15ecf2c0e3a, November 12, 2021.

[15] Nofer, Michael, et al. "Blockchain." Business & Information Systems Engineering 59.3 (2017): 183-187.

[16] M. Di Pierro, "What Is the Blockchain?," in Computing in Science & Engineering, vol. 19, no. 5, pp. 92-95, 2017.

[17] Chuen, David LEE Kuo, Li Guo, and Yu Wang. "Cryptocurrency: A new investment opportunity?." The Journal of Alternative Investments 20.3 (2017): 16-40.

[18] Milutinović, Monia. "Cryptocurrency." Економика-Часопис за економску теорију и праксу и друштвена питања 1 (2018): 105-122.

[19] Larisa Yarovaya, Damian Zięba, Intraday volume-return nexus in cryptocurrency markets: Novel evidence from cryptocurrency classification, Research in International Business and Finance, 2022.

[20] B. Saha, D. Srivastava, Data quality: the other face of big data, in: IEEE 30th International Conference on Data Engineering, 2014, pp. 1294–1297.

[21] Dalwinder Singh, Birmohan Singh, Feature wise normalization: An effective way of normalizing data, Pattern Recognition, Volume 122, 2022, 108307, ISSN 0031-3203.

[22] Christopher A, Sims “Macroeconomics and Reality, Econometrica, 1980,Vol. 48, No.1, Jan, 1-48.

[23] Lütkepohl, Helmut. "Vector autoregressive models." Handbook of research methods and applications in empirical macroeconomics. Edward Elgar Publishing, 2013.

[24] T. Li, X. Li and X. Zhang, "The Design and Implementation of Vector Autoregressive Model and Structural Vector Autoregressive Model Based on Spark," 2017 3rd International Conference on Big Data Computing and Communications (BIGCOM), 2017, pp. 386-394, doi: 10.1109/BIGCOM.2017.46.

[25] H. Lütkepohl, New Introduction to Multiple Time Series Analysis, New York: Springer-Verlag, 2005.

[26] Prabhakaran, S. 2021. Augmented Dickey-Fuller (ADF) Test - Must Read Guide - ML+. Machine Learning Plus. https://www.machinelearningplus.com/time-series/augmented-dickey-fuller-test/, December 23, 2021.

[27] J.Humberto Lopez, The power of the ADF test, Economics Letters, Volume 57, Issue 1, 1997, Pages 5-10, ISSN 0165-1765.

[28] Taylor, Sean J., and Benjamin Letham. "Forecasting at scale." The American Statistician 72.1 (2018): 37-45.

[29] Weytjens Hans, Lohmann Enrico, Kleinsteuber Martin. Cash flow prediction: MLP and LSTM compared to ARIMA and Prophet. Electron Commerce Res 2019: 1–21.

[30] Yenidogan ˘ Is¸il, et al. Bitcoin forecasting using ARIMA and prophet. In: 3rd international conference on computer science and engineering (UBMK). 2018. IEEE.

[31] Ahmed Ibrahim, Rasha Kashef, Liam Corrigan, Predicting market movement direction for bitcoin: A comparison of time series modeling methods, Computers & Electrical Engineering, Volume 89, 2021.

[32] Harvey, A. C. & Shephard, N. (1993), Structural time series models, in G. Maddala, C. Rao & H. Vinod, eds, ‘Handbook of Statistics’, Vol. 11, Elsevier, chapter 10, pp. 261–302.

[33] Triebe, Oskar, et al. "NeuralProphet: Explainable Forecasting at Scale." arXiv preprint arXiv:2111.15397 (2021).

[34] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., & Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alche Buc, E. Fox, & R. Garnett (Eds.), Advances in Neural Information Processing Systems 32 (pp. 8024–8035). Curran Associates, Inc.

[35] Hyndman, R. (2018). A brief history of time series forecasting competitions. URL: robjhyndman.com/hyndsight/ forecasting-competitions.