NATIONAL TECHNICAL UNIVERSITY OF ATHENS
SCHOOL OF NAVAL ARCHITECTURE & MARINE ENGINEERING DIVISION OF SHIP DESIGN AND MARITIME TRANSPORT

Diploma Thesis

Modelling the Dry Bulk Shipping Market with the use of Recurrent Neural Networks

Fotios Sarantopoulos

Thesis Committee:

Supervisor:	**D. V. Lyrides,**	**Associate Professor NTUA**
Members:	**N. P. Ventikos,**	**Associate Professor NTUA**
	A. Tolis	**Associate Professor NTUA**

Athens, October 2021

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΝΑΥΠΗΓΩΝ ΜΗΧΑΝΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΤΟΜΕΑΣ ΜΕΛΕΤΗΣ ΠΛΟΙΟΥ ΚΑΙ ΘΑΛΑΣΣΙΩΝ ΜΕΤΑΦΟΡΩΝ

Διπλωματική Εργασία

Μοντελοποίηση Αγοράς Πλοίων Μεταφοράς Φορτίου Χύδην με χρήση Αναδρομικών Νευρωνικών Δικτύων

Φώτιος Σαραντόπουλος

Εξεταστική επιτροπή:

Επιβλέπων:	**Δ. Β. Λυρίδης,**	**Αναπληρωτής Καθηγητής ΕΜΠ**
Μέλη:	**Ν. Π. Βεντίκος,**	**Αναπληρωτής Καθηγητής ΕΜΠ**
	Α. Τόλης	**Αναπληρωτής Καθηγητής ΕΜΠ**

Αθήνα, Οκτώβριος 2021

Acknowledgements

After 5 truly amazing years, the time has come for me to say goodbye to the National Technical University of Athens and more specifically to the school of Naval Architecture and Marine Engineering. These past years have been a blessing to me, as I had the chance to meet a lot of new people and make strong friendships, join various organisations, attend conferences and competitions, travel around the world and to challenge myself in ways that allowed me grow as a person. I am certain that I will miss all these amazing experiences, but I now feel more than ready to move on to new adventures.

I would firstly like to thank my thesis advisor Associate Professor Dimitris V. Lyrides for his guidance and his trust during the writing of my thesis.

Special thanks go out to John Violos, Researcher at NTUA, for his mentoring on neural networks and for the encouragement he showed to me during the preparation of my Diploma Thesis.

In addition, I would like to thank my fellow classmates and friends, who constantly supported me from the beginning of my university years until now. These are Myrsini, Konstantinos, Aggelos, Thanasis and Stelios. Life in NTUA would have never been the same if it weren't for you and I am really grateful that I met you.

At this point, I would also like to express my gratitude to my friends outside the school of NAME, who were always standing next to me and advised me in the most challenging times, and who at the same time are the most fun people I know. This goes to Dimitra, Panagiotis, Lydia, Klairi, Christina, Katerina and Vaggelis. I certainly embrace the statement that "friends are the family one chooses".

Last but not least, I would like to thank my parents and my grandmother for always being there when I need them, for allowing me to go after my dreams and for wholeheartedly believing in me. Nothing would have been possible if it weren't for them.

Abstract

The shipping industry is an area that is subject to changes and strong fluctuations. The unstable nature of freights makes the process of predicting their value a very demanding problem that depends on multiple factors and the solution of which has the potential to provide significant financial profit margins. In this thesis, we will focus on the regression problem of predicting the price of the Baltic Dry Index with the use of several models, proving that Recurrent Neural Networks (LSTM, GRU, BiLSTM) are very efficient in time-series predictions.

Firstly, the basic concepts and the operation of the shipping market are presented, as well as the macroeconomy in general. Then, the theoretical background of neural networks is analysed in order to make them more understandable to the reader. Based on the operation of the shipping market, and more specifically the dry bulk shipping market, shipping and macroeconomic features are examined by using a set of feature selection techniques such as the feature correlation and the feature importance techniques.

In addition, we develop a set of machine learning models: LSTM, GRU, BiLSTM, Feedforward and ARIMA in order to predict the price of BDI in a time window of one month, which is a regression problem. The recurrent neural networks (LSTM, GRU and BiLSTM) have the best performance in the regression problem for a period of one month, with the LSTM being the most accurate and providing the lowest errors. In addition, the ARIMA statistical model showed a high accuracy as well, while the feedforward neural networks provided high error rates and not accurate results.

The results of this thesis confirm that the use of macroeconomic variables, other than the shipping ones, are very beneficial for predictions in the shipping market. Finally, they also establish the fact that Recurrent Neural Networks are precise decision-making tools, and can be applied to a large set of predictions of timeseries, highlighting their ability to predict prices with great accuracy.

Keywords: BDI, Baltic dry index, shipping market, machine learning, deep learning, recurrent neural networks, feature selection, feature correlation, LSTM, GRU, BiLSTM, ARIMA, Feedforward, time series, regression, price prediction.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Purpose

The purpose of this thesis is to model the shipping market by taking into account not only shipping variables but also macroeconomic variables, so that it can be successfully described even in cases of strong fluctuations. Taking into account the wide range of factors examined to be used by the neural networks, more accurate results are produced in comparison with other studies. Thus, this thesis will enrich the field of knowledge and deepen it in ways that previous studies have never done before.

Recurrent Neural networks (RNNs) will be used for the BDI forecasting and their results will be compared to those of an ARIMA statistical model and of a Feedforward neural network.

Several economists and researchers have worked on the subject of developing decision-making systems for economic issues. Of course, statistical forecasting is efficient, but machine learning can certainly provide more accurate results, with the use of more complicated algorithms.

Forecasting economic variables in the dry bulk shipping market is a topic of great interest especially in the times we are currently living in due to the global Covid-19 pandemic outburst and its worldwide consequences. This paper will be of great interest to the scientific community, as machine learning – and especially neural networks – are constantly evolving fields, which are widely used for forecasting.

1.2 Previous research

Many researchers have worked on the use of neural network techniques as tools to forecast the Baltic Dry Index. Various models have been developed, based on different structure and training of neural networks. The appropriate neural network for a prediction in each case is determined by continuous testing. Below we refer to several previous research works that have been conducted:

- Xin Zhang and Tianyan Xue (Zhang et al., 2019) compared the forecasting accuracy of two typical univariant econometric models and three artificial neural networks (ANNs)-based algorithms. They found that when using daily data, econometric forecasting models produce better one-step-ahead predictions than ANN-based algorithms. However, when forecasting weekly and monthly data, ANN-based algorithms produced fewer errors and a higher direction matching rate than econometric models.

- Imam Mustafa Kamal (Kamal, 2018) proposed the employment of Deep Neural Networks with 135 input layers and five hidden layers for BDI forecasting. Experiments with various DNN architectures and optimized hyperparameters were conducted, yielding an average RMSE of 0.1764109. This result outperforms his previous trials using Long Short-Term Memory (LSTM) with RMSE of 0.2312339.

- Min-Soo Han (Han & Yu, 2017) proposed the prediction of the BDI with the use of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM). In order for him to be able to do comparisons, he also trained a Multi Layer Perceptron (MLP) and an ARIMA model from 2009.04.01 to 2017.07.31. As a result, the LSTM recurrent neural network outperformed the other models.

- X. Zhang ((Zhang et al., 2020) created a novel hybrid approach to Baltic Dry Index forecasting based on a combined dynamic fluctuation network (DFN) and artificial intelligence method. The utilization of DFN with AI enables the non-linear, cyclical and dynamic features of the BDI to be extracted effectively and the prediction accuracy is not impacted by the length and time-scale of sample selection either in long-term or short-term forecasting. These advantages of DFN offset the well-known limitations of traditional AI-based algorithms and econometric models in BDI forecasting.

- Bekir Sahin and Samet Gurgen (Sahin & Gurgen, 2018) proposed an artificial neural network (ANN) approach for BDI forecasting. Data from January 2010 to December 2016 were used to forecast the BDI. Three different ANN models were developed: (i) the past weekly observation of the BDI, (ii) the past two weekly observations of the BDI, and (iii) the past weekly observation of the BDI with crude oil price. While the performance parameters of these three models are close to each other, the most consistent model is found to be the second one. Results show that the ANN approach is a significant method for modeling and forecasting the BDI.

1.3 Objectives

In the context of this thesis, the factors that can influence the BDI prices will be investigated and a set of machine and deep learning models will be developed, which will use the above data aiming at the most accurate forecasts possible. In particular, 3 recurrent neural networks will be developed, as well as a feedforward neural network. In addition, for comparison purposes with traditional statistical models, an ARIMA model will also be built. Forecasts will be made for one month time windows and the exact value of the BDI will be predicted, which is a regression problem. Finally, the model with the highest accuracy will be selected for BDI predictions.

1.4 Structure

In this introductory first chapter the reader is introduced to what is this thesis about. The purpose and objectives of the thesis, its structure and expected results are mentioned.

Chapter 2 discusses the basic concepts of the shipping market. It also describes several operations of the shipping market. Through this analysis, the rationale for the final selection of variables, to be used by the neural networks, is presented.

Chapter 3 presents the theoretical background of neural networks to such an extent that it is understandable to any reader who has not been exposed to them before. The deep and machine learning models developed, the evaluation metrics used and the theory behind various stages of data pre-processing are described in detail.

Chapter 4 firstly provides an analysis of all the features selected and their significance. The various feature selection techniques in the datasets are then developed to determine the features to be used in the monthly forecasts.

Chapter 5 firstly presents the stages of pre-processing the data obtained from the previous chapter in order for them to be used in the most efficient way by the machine learning algorithms in this chapter. Then the different models for the monthly forecasts are developed, their hyperparameter optimization is done and the results of their forecasts for the regression problem are presented.

Finally, Chapter 6 provides a summary of the methodology followed and the conclusions produced and suggests possible directions for future research.

1.5 Expected results

A view held by many researchers is that artificial neural networks, and especially recurrent neural networks, have the ability to adapt very well to unusual time series problems. Based on this view, and the outcomes of other research projects, the results are expected to be remarkable. More specifically, the recurrent neural networks will be able to show high accuracy in their Baltic Dry Index predictions. Macroeconomic data are also expected to be of great help in predictions, as they contain information about the general volatility of the market.

Chapter 2

Freight Market

2.1 The Dry Cargo Freight Market

In order to better understand the current state of the shipping market, it is necessary to examine freight rates and their fluctuations. The concept of the freight market can be defined as the system by which freight rates are determined (Giziakis, 2005).

The analysis of this system should include four main features: the area defining the market, the human capital involved in it, how it operates and an explanation of how they all interact with each other.

More specifically, the freight market is:

- The geographical area in which freight rates are set or transport is carried out.
- The set of natural and legal persons who behave and act in different ways, each expressing different interests, to fix freights.
- The system of interdependent people and situations that determine freight rates and maritime transport through economic mechanisms and procedures.

The analysis of the forms of freight market has to do with (Stopford, 1997): the type of ships, the type and form of cargo carried, the geographical distribution and the duration of the charter.

One category of cargo that will be dealt with in this thesis is bulk cargo. Bulk cargo is defined as "any cargo transported by sea in large batches, with the purpose of reducing the unit cost of transport" (Giziakis, 2005). This can be considered an economic approach.

The minimisation of transport costs of ships that are capable of fast loading and unloading, zero cost packaging and all the other advantages that characterise bulk transport, has resulted in an increase in fleet size, ship size and in the growth of free competition. The shipping market for bulk carriers is categorised according to the size of the vessels, the type of cargo they carry and the routes they take.

Bulk cargoes can be classified according to the following five categories (Giziakis, 2005):

- Liquid bulk cargoes
- Dry bulk cargoes
- Unit load cargoes
- Wheeled cargoes
- Refrigerated cargoes

In this thesis we will deal with dry bulk carriers that are on the charter market, and more specifically with the Baltic Dry Index. The ships we find in the dry bulk carriers category are (Gorton, 1999):

- Small vessels: 3,000-12,000 tons dwt with their own cargo loading/unloading equipment, operating worldwide.
- Handysize: 18,000-35,000 tons dwt with their own cargo loading/unloading equipment.
- Handymax: 35,000-50,000 tons dwt with their own cargo loading/unloading equipment.
- Ultra Handymax: 50,000-60,000 tons dwt.
- Panamax: 50,000-80,000 tons dwt with size suitable for passage through the Panama Canal, usually transporting grain, coal and iron ore and without their own loading/unloading equipment.
- Capesize: 80,000-120,000 tons dwt.
- Capebulkers: 120,000-175,000 tons dwt.
- Large Cape Bulkers: 175,000-200,000 tons dwt.
- Very Large Ore Bulkers: 200,000 tons dwt and above.

2.2 Baltic Exchange

The Baltic Exchange, which has a history of more than 270 years and is based in London, is an independent source of information regarding the trading and settlement of physical and derivative contracts in shipping. It has hundreds of members, including shipowners, charterers, and brokers, as well as financial institutions, shipping lawyers and insurers.

The diversity in the products transported, the ships used, the supply and demand conditions, etc., apart from the fact that they have dictated the need of different types of chartering, also indicate the need for the development of the appropriate shipping indicators that can contribute to the analysis and forecasting of the developments of the market under consideration. The most basic and most reliable indicators of this kind are produced by the Baltic Exchange and are used as financial instruments that directly measure the relationship between supply and demand.

Due to the importance of these indicators, the Baltic Exchange has established specific committees and groups, involving all types of market participants cooperating with each other, in order to ensure the quality and reliability of the indicators produced.

2.3 Indicators of the Baltic Exchange

2.3.1 Baltic Exchange Freight Index (BFI)

In 1985, the Baltic Exchange established the Baltic Freight Index (BFI) to reflect the change in freight rates on specific freight routes, on a daily basis. More specifically, it was calculated from the weighted average of 11 different dry cargo trade routes, based on data from specific brokerage houses. The BFI index was, at that time, the basis for the trading of the Baltic International Futures Exchange (BIFFEX), which was created to provide shipowners, charterers, large trading houses and brokers with the possibility of covering the risk of large and sudden fluctuations in the dry bulk market.

In practice, these contracts did not prove to be very effective tools for hedging risks, due to the fact that the BFI index is derived from a set of different routes and ships, and therefore the specificities of each case could not be taken into account. This weakness, together with the fact that constant changes occurred in the pattern of trade on the dry bulk shipping market, has forced the Baltic Exchange to modify the structure of the BFI index several times in the meantime.

2.3.2 Baltic Exchange Dry Index (BDI)

In 1988, the Baltic Exchange Dry Index (BDI) was introduced, specialising in the dry cargo market, providing a daily indication of the price of moving the main raw materials by sea. The BDI reflects freight rates, taking into account 23 main freight routes, for bulk carriers of the Handysize, Supramax, Panamax and Capesize categories. In practice, it is a combination of the individual indices BHSI, BSI, BPI and BCI, which are described in more detail below.

In essence, the BDI captures the supply of carrying capacity, in relation to demand, on dry bulk vessels, reflecting the trading conditions of these markets. Since supply is characterised by inelasticity, marginal increases in demand can push the index upwards smoothly, while large increases in demand can create sharp upward trends in the index. Conversely, marginal, or large decreases in demand can create downward pressure or sharp downward trends in the index respectively. For example, if 100 vessels compete for 99 cargoes, the index falls, while if 99 vessels compete for 100 cargoes, the index rises.

The changes in the BDI appear to be ahead of those of the stock market. Therefore, the BDI is considered a leading indicator for economic growth and production and is used to predict market trends, not only for the shipping industry but also for the market in general. For example, the BDI may predict higher interest rates, since more employment in shipping requires financing, thus increasing the demand for credit. The BDI also acts as a shaper of the shipping market in Forward Freight Agreements (FFA's).

2.3.3 Freight indicators by type of dry cargo vessel (BPI, BCI, BHSI, BSI)

Later, independent indicators for individual dry cargo vessel types were also created. Thus, in 1998 the Baltic Exchange Panamax Index (BPI) was first published on a trial basis, followed in 1999 by the Baltic Exchange Capesize Index (BCI) and the Baltic Exchange Handysize Index (BHSI). In 2000 the Baltic Exchange Handymax Index (BHI) was also published, which was replaced in 2005 by the Baltic Exchange Supramax Index (BSI).

This segmentation facilitated the analysis of time series to draw conclusions about the influence of changes in supply and demand on the markets, providing greater flexibility and transparency for participants, while allowing for increased activity in the future market as new products for freight derivatives were created. The representativeness of these indicators is evident by the fact that the Baltic Exchange handles 30-40% of the world's dry cargo freight.

2.4 The Charter and Liner markets

Maritime transport from an economic point of view is divided into two freight markets, Charter and Liner (Psaraftis, 2005).

2.4.1 The Charter market

In the Charter market, which is essential for the understanding of this paper, the entire vessel, with or without crew, is chartered for a single voyage or for a period of time that can range from a few days to even decades. The terms between the parties involved are laid down in charter contracts. Charter rates are determined by the laws of perfect competition where the relationship between supply and demand increases, decreases or keeps prices fixed. The type of vessels found in the charter market are specialized, such as Tankers, Bulk carriers, Chemical Tankers and OBO. The cargo being transported has low value and more emphasis is placed on limiting the ship's operating costs. Consequently, the vessels in this market operate at speeds around 15 knots.

2.4.2 The Liner market

The liner market is used for the transport of goods on specific routes. This enables the service providers to be organised in syndicates, controlling and setting the fares for these routes and giving them the power to repel potential competitors. The fleet of the liner market consists of general cargo ships, container ships, specialised Ro-Ros and passenger ships. The speeds at which they operate range from 20 to 25 knots. The goods carried are of high specific value and are rarely homogeneous. Contracts refer to the transport of a specific cargo at a specific time from one place to another. The main concern is the safety of the cargo against damage and the delivery time.

2.5 Basic concepts of the freight market

- **Supply**

When we refer to supply, we mean the quantity of goods or services offered. In the maritime transport sector, supply corresponds to the capacity of the fleet that carries cargo (Stopford, 1997).

Supply is less changeable and less adaptable to the changes in demand. The merchant fleet of ships represents the stable maritime transport capacity. Supply can meet demand with new shipbuildings. The supply of ships is controlled and influenced by the following groups:

- shipowners,
- shippers/charterers,
- the shipping banks,
- various regulatory authorities.

Shipowners are the key decision-makers in terms of orders for new ships and for the scrapping of older ships, while charterers/shippers influence the decisions of shipowners.

The sum of the deadweight (dwt) of ships is usually chosen as the unit of measurement, but other units such as total net tonnage, cargo volume, etc. may also be used. The supply is an important indicator of the number of the dry bulk vessel fleet, as far as this thesis is concerned.

The key variables for maritime supply are as follows:

- Decision makers
- World fleet capacity of merchant ships
- World fleet productivity
- Shipbuilding deliveries
- Ship scrapping & losses
- Expectations of freight rates

Supply will normally be adjusted to demand when shipowners correctly choose what the charter rate will be.

- **Demand**

Demand through the operations of various industries, creates the need for goods that require maritime transport. Demand is calculated by subtracting the deadweight of ships not carrying out transport from the total supply. This results in the term "surplus", which is explained below (Stopford, 1997).

The important variables of maritime transport demand are the following:

- World economy
- Sea borne commodity trades
- Average haul distance of sea routes
- Exogenous factors
- Transport costs

Exogenous factors are mainly unpredictable and random, such as natural ones, i.e. natural disasters and weather conditions. They can also be political, social and economic, influencing positively or negatively the demand for maritime transport and causing positive or negative changes in freight rates.

- **Surplus**

It is the total deadweight of vessels belonging to one of the following categories:

- **Slow steaming**: Vessels that undertake transportations at lower speeds in order to reduce operating costs.

- **Laid up**: Ships that cannot enter the market because the freight revenues do not cover the operating costs.

- **Idle**: Ships which for other reasons, such as repairs, do not carry out maritime transport.

- ## Freight - hire

The four main categories of chartering are (Alderton, 1995):

- **Single Voyage Rate, Voyage Charter**: It is a short term form of charter. The ship owner carries a certain amount of cargo to specific ports on behalf of the charterer. The charterer in turn has to compensate him according to the measurable per ton of cargo carried (US $ / ton of cargo). It is a volatile figure that typically shows features of economies of scale. Detailed contracts define the obligations of each party.

- **Time charter**: The shipowner grants the charterer the ship and its crew for a specified period of time, which may last from a few days or months up to 5 years. The charterer decides where the ship will move and is responsible for fuel, port, loading and other costs. The unit of measurement for time charter is the United States dollar per day (US$/day). A key factor affecting the price is the time period (months to years) for which the charterer wishes to charter the vessel.

- **Bareboat or Demise charter**: This is a form of long-term charter where the shipowner provides the charterer with the ship "naked" for a long period of time. The charterer takes over exclusively the whole ship, as if he were the shipowner, from manning to insurance and pays the shipowner at regular intervals.

- **Contract of Affreightment**: In this case we have a long term form of chartering where the shipowner satisfies the needs of the charterer for the transport of goods over a long period of time.

Freight rates, referring to single voyage chartering, is the payment by the charterer to the shipowner for the transport and delivery of goods to their final destination. Usually, the freight is paid in USD per tonne of cargo carried.

The following apply to chartering and freights (Giziakis, 2007):

- When it has been agreed that the freight is to be paid on delivery of the goods, the shipowner cannot claim it until the goods are delivered.

- The shipowner has the right to be paid the full freight even if the goods are delivered with damages.
- When the freight is calculated on the quantity loaded, the shipowner is entitled to the full freight even if he delivers an incomplete cargo.
- The freight shall not be paid if the goods are not in a marketable condition when delivered.
- The prepaid fare is not refundable even if the ship and the cargo are completely lost.
- The fare shall be paid in full even if not all the cargo but part of it is delivered.
- The pro rata freight shall be paid only when the shipowner is able and has the capacity to deliver the cargo to the destination.
- If the owner of the cargo transported requires it to be delivered to a port before the agreed destination, the freight shall be paid normally.

In types of charter, such as time charter, bareboat or demise charter and contract of affreightment, where the time element is involved, the payment to the shipowner for the charterer's use of his vessel is called hire. The hire is usually expressed in dollars per day. Of course, the methods of repayment vary depending on the agreement between the parties involved.

- **Order book**

These are the orders received by shipyards from shipowners for the construction of new ships. The order book can provide information on the dynamics of the market in the future.

- **Ship deliveries (future market)**

It is the number of the ships that are about to enter the world shipping market.

- **New building price**

It is the price at which a ship is ordered for construction at a shipyard. This price is indicative, as it may change depending on the size of the order from the shipowner or the delay in the construction of a ship by the shipyard. All relevant conditions are set out in a contract signed by the shipowner and the yard.

- **Second-hand price**

The second-hand cost is the indicative selling price of a second-hand ship that can be delivered in a relatively short period of time.

- **Demolition**

It is the number of ships that have reached the end of their lives and are taken to the scrap yard.

- **Demolition price, Scrapping**

At the end of its life a ship ends up in a scrapping yard. Initially the shipowner tries to make the best possible use of the ship's machinery and of a part of the equipment. The rest of the ship has value as raw material and is priced according to the weight of its metal structure. The purchase price of the ship from the yard is defined as its scrapping price.

- **OBO type vessels**

These are mixed cargo ships which, depending on the market situation, operate either in the liquid cargo or dry cargo market.

- **Ship losses**

These are the ships which are taken out of service due to an accident.

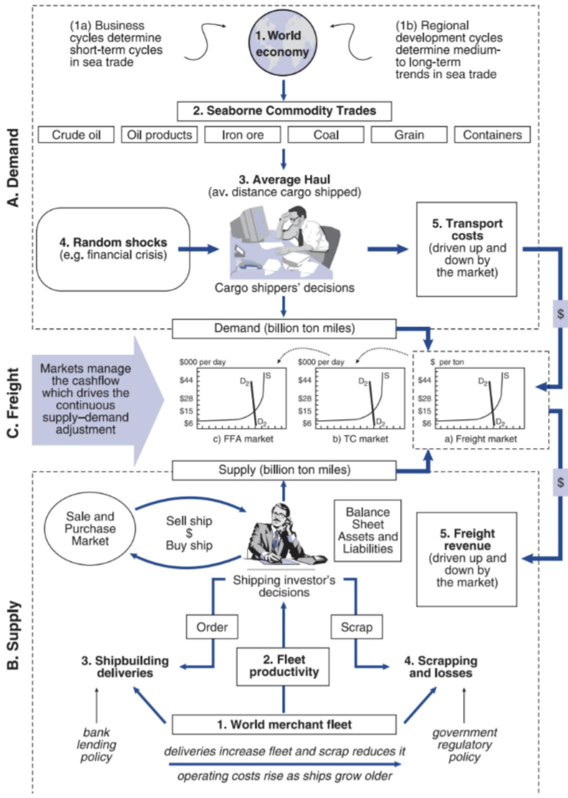


Figure 1: The shipping market supply and demand model

(Stopford, 1997)

2.6 Freight rate mechanism

2.6.1 Supply curve

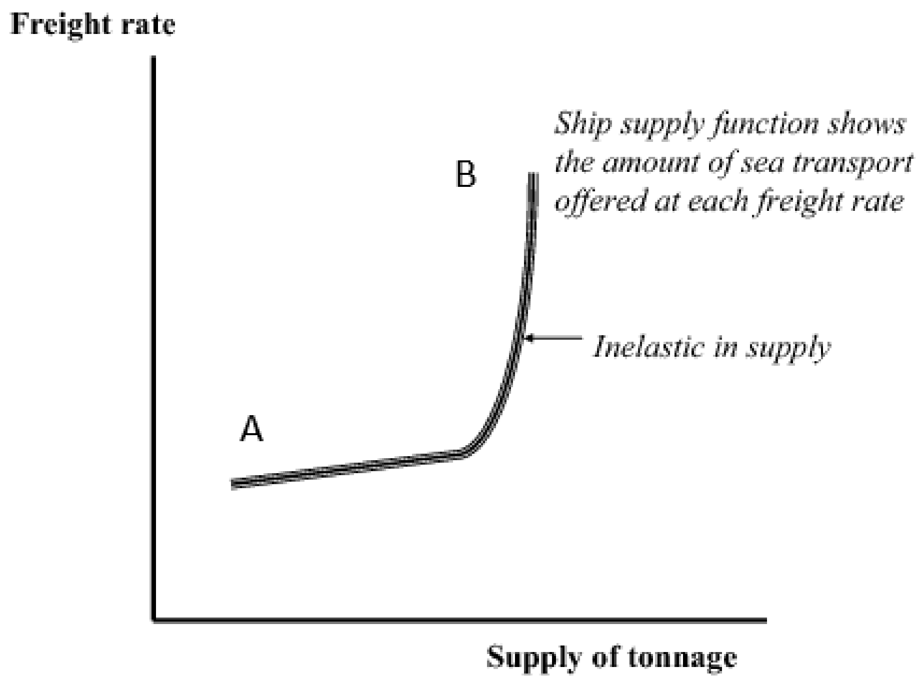


Figure 2: Supply curve

(Lun et al., 2010)

On the vertical axis of the supply curve (Figure 2) is the Freight rate (Psaraftis, 2005). A freight rate is a price at which a certain cargo is delivered from one point to another. After the calculation of the freight rate of the ships belonging to the specific route is studied, the above supply curve is drawn.

Area A: in this area we observe a low freight rate and it obviously refers to newly built vessels, which are able to enter the market at low fares.

Area B: the freight area is high and the ships in this area are old. In order to benefit the shipowner to charter such a vessel, the freight rate should be correspondingly high.

2.6.2 Demand curve

The demand curve is not easy to construct. In the present model that we are analysing, we can consider the demand to be inelastic in the short run. This is because the needs for cargo transportation in the world market are determined and alternative more economical modes of transportation on ship routes do not exist. Consequently, demand is not much affected by the freight rates.

2.6.3 Factors affecting the fares

The fares are subject to constant fluctuations and since it is a key term in shipping it is worth at this point to highlight the factors that disturb its equilibrium (Psaraftis, 2005).

- **External factors**
    - The course of the world economy. When the world economy is in a period of growth, the demand for the production of goods and therefore their transport is increased.
    - Military and political issues. During a war, production units are destroyed and regions are excluded from maritime trade, disturbing consequently its balance.
    - Natural causes. A severe drought can have a negative impact on grain production by increasing demand. A hurricane may also prevent ships from sailing.
    - Technological causes. The development of technology in the maritime transport sector, such as the introduction of automation in cargo handling, makes newer and technologically advanced ships more desirable than older ones.
- **Internal factors**

A typical example is a shipowner's decision to use one of his ships by reducing the operating costs. This is usually achieved by moving the ship at a lower speed (slow steaming). Thus, shipowners whose ships on the supply curve are in positions close to the demand curve but towards the right side (where the older ships are), seek to reduce their operating costs below the spot rate in order to enter the market. The number of ships in slow steaming is an important factor in assessing the state of the shipping market.

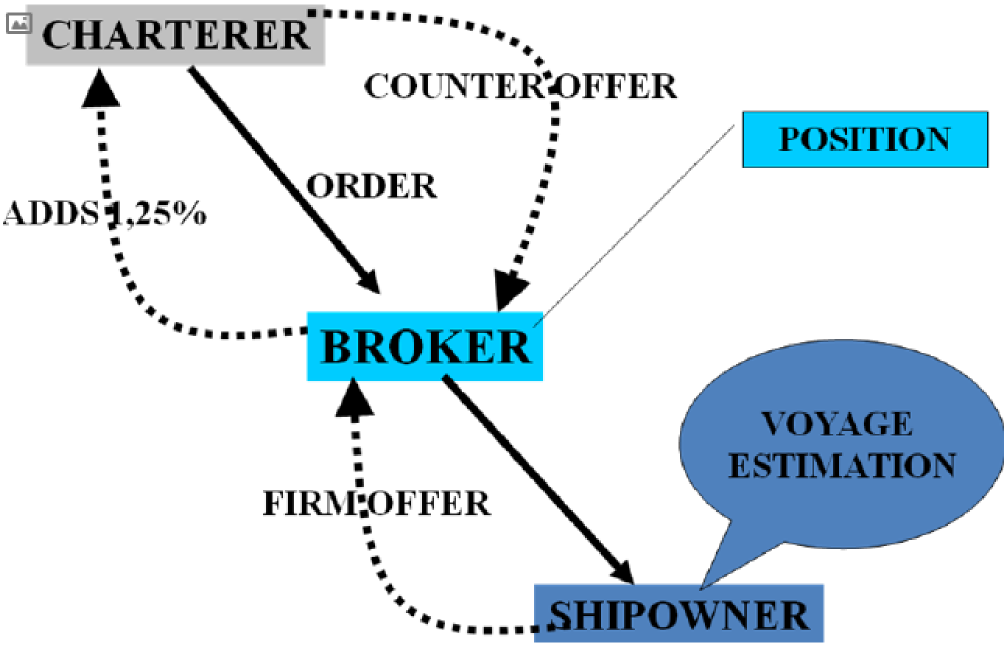


Figure 3: The chartering process

(Giziakis, 2007)

2.7 The Chartering process

Chartering is defined as the agreement for the commercial employment of a free boat between the ship owner and the charterer. The mediator of this agreement is the broker. There are two types of chartering contracts: the charterparty and the bill of lading. Both are discussed below.

2.7.1 The Charterparty

A charterparty is a charter agreement that is made between a charterer/shipowner, who owns the ship, and the charterer, to whom the ship is made available. It sets out the rights and obligations of the parties. The terms of a charterparty are categorised into (Payne, 1989):

- Express and implied terms. Express terms are those that are explicitly described, while implied terms are those that are not written in the charterparty but are accepted by the parties.
- Representations. They have to do with the promises made in the negotiations. If they are shown to deviate from reality with the intention of misleading, the agreement can be cancelled.

- Conditions. These are conditions that if breached by one party of the agreement, then the other party is entitled to cancel the agreement and receive compensation. Examples of conditions are the position of the ship on the date of the charterparty, the time of sailing to the port of loading and the nationality of the ship.
- Warranties. They are important conditions whose breach entitles the aggrieved party to claim damages. Examples of warranties are the speed of the ship, the fuel at the time of delivery of the ship and the maintenance of the ship.
- Innominate terms. These are terms whose categorisation and treatment as conditions or warranties depends on the court's consideration of their consequences.

2.7.2 The Bill of Lading

When an agreement is made to transport cargo, the carrier, i.e. the one who transports the cargo, must issue a so-called bill of lading. The other party to the bill of lading is the shipper, i.e. the party who requests the transport of the cargo (Psaraftis, 2005).

A bill of lading includes: the description of the cargo, the name of the ship, the date signed by the captain, the name of the shipper and the recipient, the ports of loading and unloading, the freight rate and how it is to be paid, the number of bills of lading issued and the terms of the agreement.

The recipient must present the bill of lading issued to the captain in good time for the unloading to begin. Otherwise, the captain has the right not to deliver the cargo.

2.7.3 Duties of the shipowner and charterer

The duties of the decision-making teams, as far as ship supply and demand are concerned, are divided as follows (Giziakis, 2007):

- Duties of the Shipowner:
  - When loading starts and before the ship leaves the port of departure, the ship must be capable of completing the voyage and carrying the given cargo.
  - The ship must reach its destination without unjustified delays, otherwise the charterer shall be compensated. If the unjustified delay is extreme, the agreement may be cancelled.

- The route to be followed by the vessel must be the one agreed upon. Deviation from the course is justified only for the safety of the cargo and to protect human life.
- When an estimated date of arrival of the ship has been fixed, the ship is required to sail at the necessary speed to be there on the estimated time.
- The charterer has the right to cancel the agreement when the cancellation date arrives.

- Duties of the Charterer:
  - The cargo must not be dangerous, i.e. easy to cause damage or illegal.
  - The shipper is obliged to deliver the cargo.
  - If the agreement was made under illegal conditions the shipper has the right not to deliver cargo.
  - The charterer is obliged to bring the cargo alongside the ship for loading.
  - The charterer must load the full cargo.

2.8 The Shipping Cycle

The shipping industry is used to operate in cycles. It is constantly in a state of fluctuations with constant variation and almost no stability. The shipping cycle is defined as that mechanism designed to prevent imbalances between supply and demand. The 4 stages of the shipping cycle are the following (Stopford, 1997):

Trough: There is a surplus in the capacity of ships, which results in them being crowded in ports. The ships that carry cargo move with slower speeds to delay their arrival in full ports. Freight rates fall and inefficient ships are decommissioned. Low freight rates and low revenues, which in many cases do not exceed operating costs, lead shipowners to sell their ships at low prices. Older ships are being scrapped. The trough phase is the longest period in the shipping cycle and may last for several years.

Recovery: Freight rates increase slightly and ships are able to cover their operating costs. A part of the decommissioned ships is put on the market. The revenues of shipping companies are increasing. The prices of second-hand ships become higher. However, a feeling of uncertainty still exists in the market.

Peak: Supply and demand are in full balance as excess capacity has been absorbed. Fares are at very high levels and ships are moving at maximum speed. The market is euphoric. Banks and stock exchanges are financing the shipping market with great ease. Orders for newbuildings are increasing. Second-hand ships are experiencing a sharp rise in price and some of the younger ships are more expensive than the new ones. The psychological feeling in the market is very positive.

Contraction: Following the euphoric market climate and the many orders for newbuildings, supply exceeds demand. Fares are falling and ships are sailing at slower speeds. Older ships with higher operating costs are waiting for cargo to be transported. Shipping companies are in a state of confusion.

2.9 Economic crisis

The National Bureau of Economic Research defines recession as a significant decline in economic activity spreading throughout the economy. It lasts more than a few months, and is usually visible through the Gross Domestic Product (GDP), incomes, employment, industrial production, and wholesale-retail sales (Soros, 2009).

During recession periods, purchasing power is an issue for everyone, whether it is a new product or raw materials or transportation of goods. Everyone is looking for one thing and that is how to minimize costs. Entrepreneurs start to cut costs in times of recession. This can save a percentage of revenue that will perhaps be used in other services. Cutting costs can be the only way for a company to survive if it receives a long period of losses due to recession (Simitis, 2008). Generally, in times of recession, companies try to find out which parts of the company or which products of the company will be mostly affected.

Even if some people have enough money to buy goods, they usually do not do so because of the fear of unemployment or of the upcoming price increase, which results in the whole situation of not enough money circulating in the market, which in turn affects the business world with unemployment. Recession affects all businesses around the world in one way or the other.

2.9.1 Financial crisis of 2008

The financial crisis originated in the United States of America (Guina, 2019). This crisis was caused by multiple factors, including a dramatic change in the ability to create new lines of credit, which dried up the flow of cash and slowed down economic growth and the buying and selling of assets. This hurt a lot of individuals, businesses and financial institutions. Many financial institutions had asset-based mortgages whose value had declined sharply and did not yield sufficient value to repay the amount of money needed for the corresponding loans. This dried up their available cash and thus their ability to take out new loans.

There were other factors including subprime loans which made it very easy for individuals to buy houses or make other investments. Subprime credit created more money in the system and people wanted to spend that money. Unfortunately, many people wanted to buy the same things, which increased demand and caused inflation. Private equity firms leveraged billions of dollars of debt to buy companies and created hundreds of billions of dollars of wealth without creating anything of value. Later, speculation in oil prices and higher unemployment further increased inflation.

The American economy is based on credit. Credit is a great tool when used wisely. For example, credit can be used to start or expand a business, which can create jobs. It can also be used to purchase fixed assets, such as homes and cars. Again, more jobs are created and people's needs are met. But during 2000-2009, credit was been out of control.

Loan brokers, have only acted as middlemen, issuing loans, and then passing on the responsibility for those loans to others, in the form of mortgaging assets (after taking a fee for themselves). Exotic and risky loans became commonplace, and intermediaries approved these loans, relieving themselves of any liability by consolidating them with other loans and reselling them as 'investments'.

Thousands of people took out loans larger than they could repay, in the hope of selling the house to make profit or refinancing later at a lower interest rate and with a higher value of their house - which they could use to buy another 'investment' house. Many people became rich in a short period of time. All the bankers needed in order to provide housing loans was the word of the buyers that they could pay off their mortgage. In addition, realtors had no reason not to sell a house.

The collapse in the housing sector set off a chain reaction in the economy. Individuals and investors could no longer sell their houses for a quick profit, interest rates adjusted upwards, and mortgages were no longer affordable for many property-owners. Thousands of mortgages defaulted, burdening investors and financial institutions.

This caused huge losses in the securitisation of mortgage receivables and many banks and investment companies were running short of money. There was also a surplus of houses on the market, plummeting their prices and slowing down the building of new houses, putting thousands of builders and workers out of work. Low house prices caused further complications, as many properties were worth far less than the value of the mortgages and some owners chose not to pay their mortgage.

These huge losses forced banks to increase their lending requirements, but it was too late. Several banks and financial institutions merged with other institutions or were simply taken over. Others were fortunate enough to receive financial assistance from governments. Many companies, however, were shut down.

The financial crisis of 2008 could be described as a crisis of solvency and liquidity of the financial system. According to Mr. I. Stournaras, Professor and Scientific Director of IOBE, the causes of this crisis were the following (Stournaras,2019):

- Loose regulatory rules.
- The banking system was not adequately supervised.
- The supervisory authorities have not been able to coordinate at an international level.

- There was no transparency.
- Complex derivative products were developed.

2.9.2 The 2008 financial crisis and the shipping market

The financial crisis in the second half of 2008 hit the financial and banking sector (Petropoulos, 2009). As a result, banks could not finance the shipping sector and global trade was paralyzed. Bank losses and massive withdrawals by depositors, due to a sense of uncertainty about the market, led to a serious liquidity shortage. Banks did not trust each other and interbank transactions froze. The ground was no longer fertile for investment and demand fell significantly. The recession, inevitably affected also the demand for transport of liquid cargo, bulk cargo and containers. The result of the whole situation was a fall in freight rates, a reduction in the value of ships, the closure of shipping companies, the renegotiation of shipping contracts, pressure from banks to repay loans at high interest rates and the urgent search for financial resources to support shipping.

2.9.3 Dry bulk transport market and the 2008 financial crisis

Morgan Stanley in its Commodity Shipping Research on the 2/3/2009 publication reports that the DWT surplus of dry cargo increased by (Petropoulos, 2009):

- 6.9% in 2005
- 5.3% in 2006
- 6.8% in 2007
- 6.5% in 2008

The increase in supply was not as large as the increase in demand. The demand for transport of bulk cargo increased by:

- 2.3% in 2005
- 8.1% in 2006
- 12.5% in 2007
- 5.5% in 2008

The consequence of this situation was a rise in freight and ship prices and the ordering of a large number of newbuildings. This situation in the shipping market corresponds to the Peak stage of the shipping cycle, as I have analysed above. However, supply exceeded demand and with the onset of the financial crisis in the second half of 2008, the shipping market entered the stage of recession.

2.10 Conclusions

The history of global economic activity has shown that periods of strong economic activity and growth are usually followed by periods of downturns, recessions and financial crises. The alternation between periods of growth and recession is inevitable and occurs with a continuity creating economic cycles.

Economy is considered to enter a recession when the Gross Domestic Product has been in decline for two consecutive quarters. Economic cycles differ from one another in the duration and intensity, based on the fluctuations of the individual indicators of economic activity.

Central Banks and Governments have tools at their disposal, through their monetary and fiscal policies respectively, to lead to changes and anticipate situations both in their domestic economies and in the global economy.

Monetary policy: 1) Interest rate reductions

2) Quantitative easing

Fiscal policy: 1) Tax policy

2) Public expenditure policy

In order for policies to mature and to show their impact on economies, they need a reasonable period of time, which can be defined as five years, based on the study of past relevant situations. Political and governmental actions should not be judged immediately, as the immediate impact rarely coincides with the actual impact, which takes place over time.

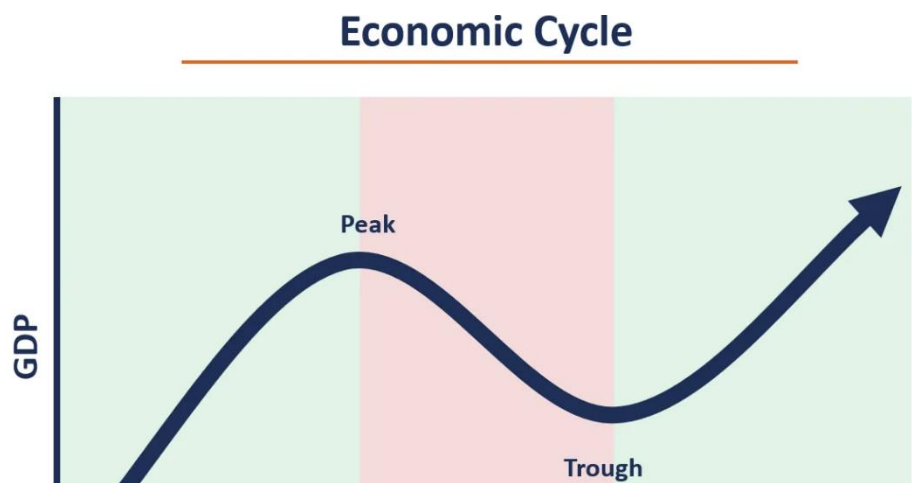


Figure 4: The economic cycle

(Corporate finance institution, 2019)

Moreover, it should be stressed out that the economic cycle will follow all the stages (recovery, peak, contraction, trough) without any possibility of preventing it in the long run. The smoothing of economic cycles should be the objective of both monetary and government authorities. Therefore, central banks and governments should not act in such a way as to prevent the possibility of a recession. Instead, their objective should be the rapid smoothing of the problems caused by the recession and a gradual return to recovery, promoting a long-term upward trend in economic activity with all the positive consequences for society. When an economy enters a period of recession, the aim is, once this period is over, to move straight into the recovery phase, with as little time as possible remaining at the bottom of the economic cycle. However, history has shown that mismanagement by supervisors can lead to a prolonged stay in the bottom of the economic cycle for a domestic economy (e.g. Japan - Lost decade) or even a global economy (e.g. Financial Crisis 2008), creating the conditions for entering the vortex of a vicious cycle.

Chapter 3

Theory of Neural Networks

3.1 Introduction

Artificial intelligence has been a prominent scientific field for many decades. Machine learning is a subset of artificial intelligence, while artificial neural networks are a type of machine learning. Artificial neural networks are based on the way the human brain works and have become popular for problems of classification, clustering, pattern recognition and regression. In turn, deep learning can be considered as a subset of machine learning. In deep learning, more complex ways of connecting the layers of the neural network are observed, more neurons are encountered, more computational power is needed, and even the possibility of automatic feature extraction is present. Figure 5 graphically illustrates the above correlations.

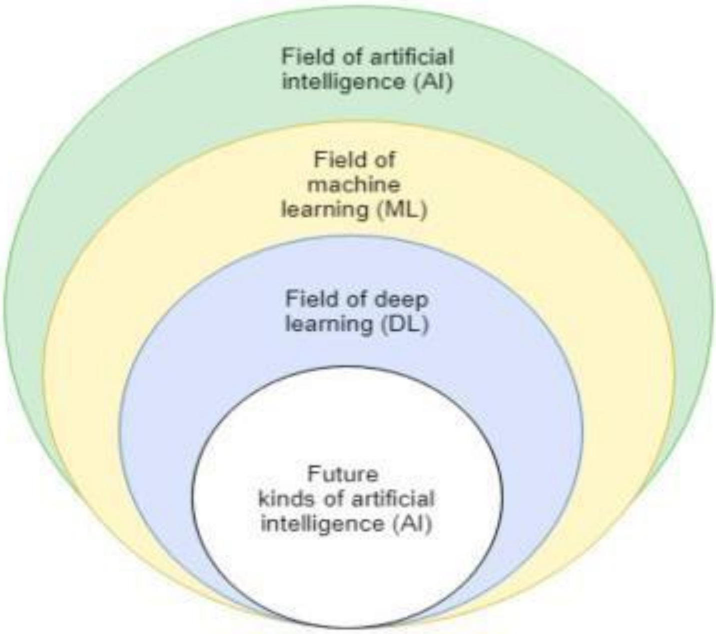


Figure 5: Correlation of key concepts

Deep machine learning algorithms find great application in big data management with notable success in voice recognition, pattern recognition, computer vision, natural language processing, recommendation systems, etc. (Abiodun et al., 2018).

3.1.1 Historical background of Neural Networks

Alan Turing observed in 1950 the need to create machines that "learn from experience" paving the way for what we now consider as machine learning (Harnad, 1950). The first developments in the field of machine learning date back to the 1960s, while the term machine learning was first used in the work (Samuel, 1959) of Arthur Samuel, a researcher at IBM.

Neural networks, a subfield of machine learning with many state-of-the-art results, have their theoretical basis in the invention of the perceptron by Rosenblatt (Rosenblatt, 1957). However, the limited ability of the perceptron to recognise non-linear patterns will limit interest in neural networks for several years until the emergence of multilayer perceptron networks, which have significantly greater computational power.

Some milestone inventions in the evolution of machine learning in the second half of the 20th century were the perceptron, a concept conceived in 1943 (McCulloch, 1943) and introduced in 1957 (Rosenblatt, 1957), the nearest neighbour classifier in 1967 (Cover, 1967), the first conception of the idea of the backpropagation algorithm in 1970 (Linnainmaa, 1976), Hopfield networks, a kind of recurrent neural networks in 1982 (Hopfield, 1982), the invention of the Q-learning algorithm in 1989 (Watkins, 1989) which paves the way for the development of the reinforcement learning field, the discovery of random decision forests in 1995 (Ho, 1995) and of support vector machines in the same year (Cortes, 1995), long-short-term memory (LSTM) networks in 1997 (Hochreiter, 1997) and convolutional neural networks during the eighties and nineties (Sahiner, 1996).

3.2 Time series - Statistical models

A time series is a sequence of numerical values in sequential order. A time series can be any value that changes over time, with typical examples being stock prices, demand and sales of products, etc. The problem of time series forecasting refers to the prediction of the future value of a time series at some time t + h using only data up to the current time t.

Statistical methods such as SMA (Simple Moving Average), Holt's exponential smoothing and ARIMA (Autoregressive Integrated Moving Average) are largely used to solve time series forecasting problems. More specifically, ARIMA is one of the most widely used statistical methods for time series forecasting due to its ease of use and high returns. An ARIMA model consists of the following three parts:

- Autoregression (AR): expresses the dependence of the time series value on its values in previous time periods.

- Integrated (I): refers to the number of successive value removals that must be made to make the time series static.

- Moving Average (MA): incorporates the dependence between an observation and a residual error from a moving average model applied to previous observations.

To develop the ARIMA model, the Python statsmodels library was used. The parameters p, d, q of the ARIMA model refer to as AR, I and MA respectively.

3.3 Neural Networks

Artificial neural networks or simply neural networks are an abstract algorithmic construct, which belongs to the family of machine learning, capable of "learning" to perform specific functions by reading examples of them, and without being specially programmed. The structure of the system is inspired by the neurons of the brain, hence its specific name. With artificial neural networks, the scientific community is trying to imitate and model this non-linear behaviour of the brain, which is directly related to decision-making. The methods used in these networks provide a liberated model, which can be adapted to a wide range of applications, being fault-tolerant and able to use parallel and distributed systems for faster extraction of the solution to the problem. They are essentially "black boxes" which "learn" the internal relationship of an unknown system without attempting to "guess" the functions governing the interconnections of that system. They are widely used in areas such as mathematical function approximation, problems of data classification, feature extraction from models, clustering, quantization of vectors, optimization, etc. They are one of the key methodologies in information processing and are applied in almost all areas of science and engineering.

As said above, the term "neural networks" comes from the neurons of the brain, as these algorithms "mimic" their functions in some way. A medical model of a brain neuron is shown in Figure 6. The basic idea of neurons is that they transmit messages through their neighbouring neurons, they have local control, and they are many in number (about 10^{11}), with even more synapses (about 10^{15}).

Figure 6 shows that the neuron has a core and synapses. More specifically, the information that "runs" in the nervous system passes through each neuron, where the nucleus decides whether or not to pass the signal to the next ones. This is how the messages are transmitted and ultimately the brain makes decisions. The neural network algorithm was based on this idea. Figure 7 shows the mathematical model of a neuron in a neural network.

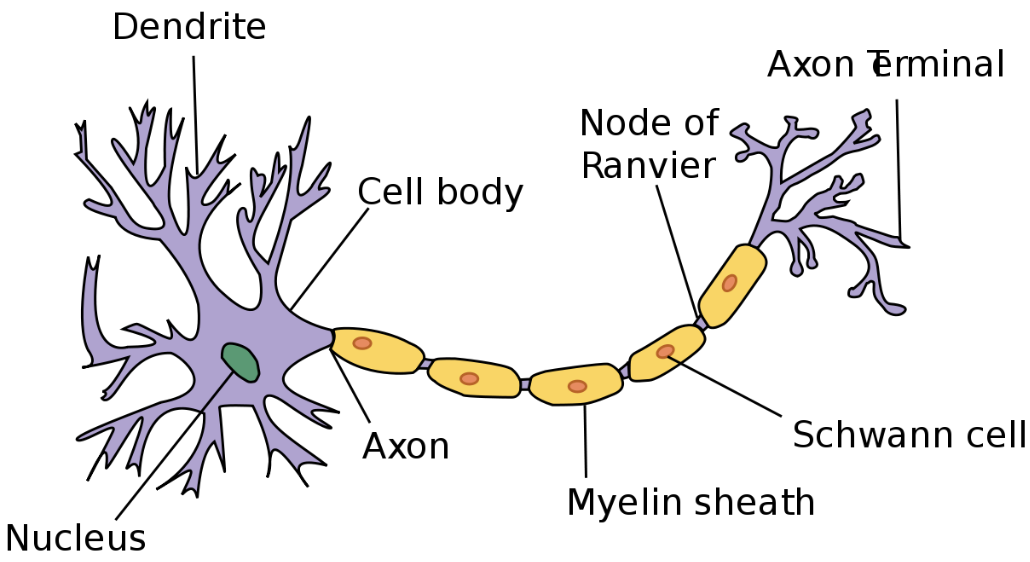


Figure 6: Medical model of a neuron

(Neuron.svg, 2019)

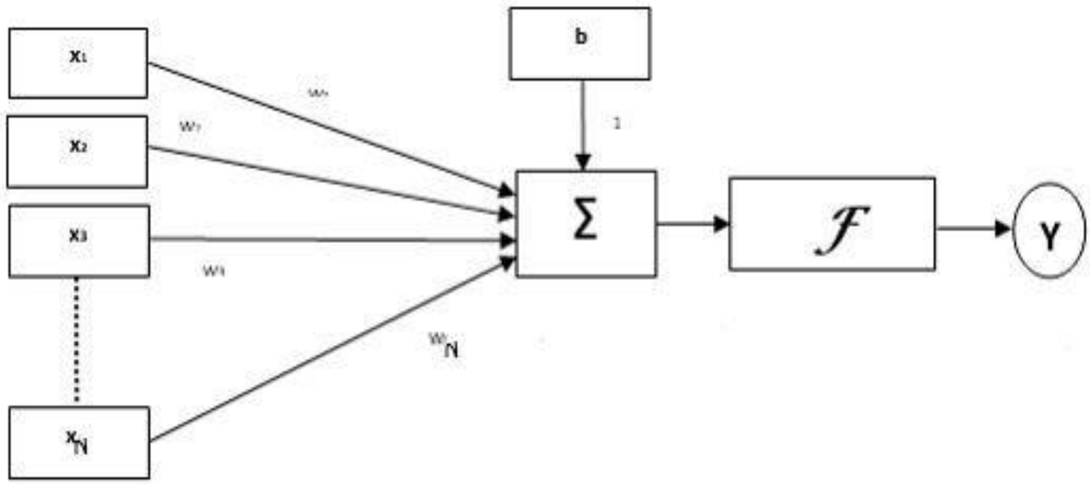


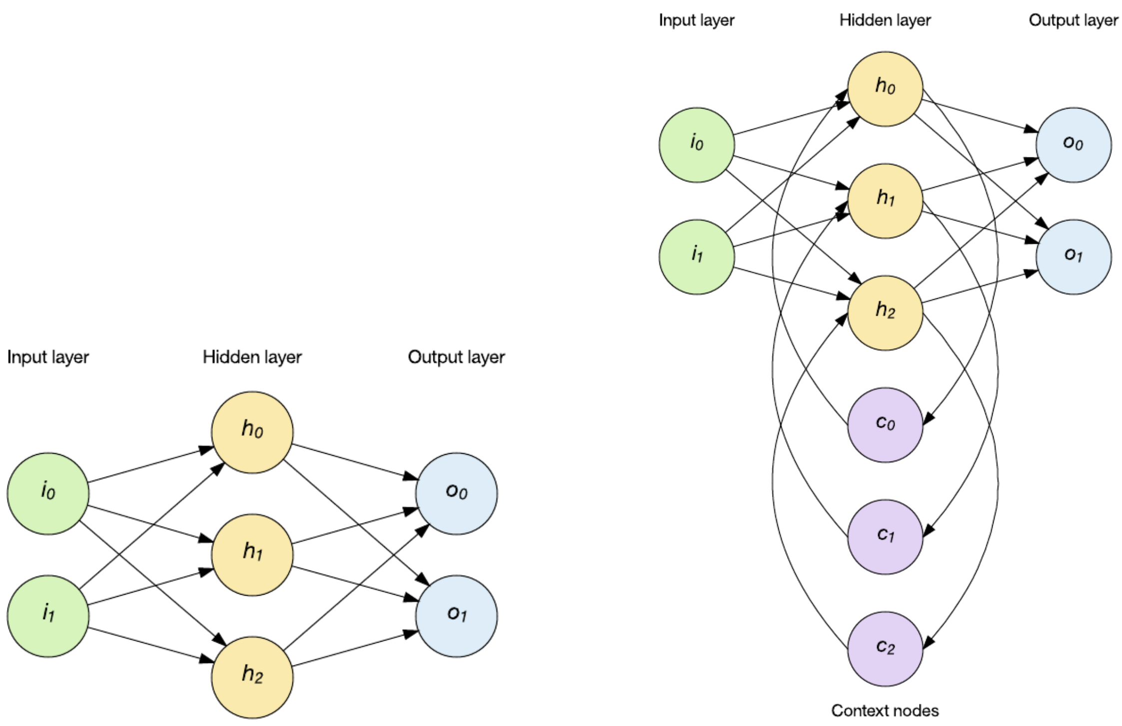
Figure 7: Artificial neuron used in neural networks, where b is the bias term and F is the activation function

(Supervised learning, 2021)

An artificial neural network (ANN) consists of one (in this case the ANN is called basic perceptron) or more neurons. Each neuron accepts an input $X = [x_1, x_2, \ldots, x_N]$ of size N. Each element of the vector X is multiplied by an element from the weight matrix $W = [w_1, w_2, \ldots, w_N]$. This process resembles the synapse function in the biological neuron.

Then, all the $x_i w_i$ products are summed and a term b called bias or threshold is added to the sum. The polarization term can also be added to the network in the form of an additional element w_0 in the weight matrix with a corresponding input to the vector X of unity. The above result is introduced as input to an activation function (activation functions are discussed in detail in a following subsection of this chapter) that acts as a filter that shapes the output (final response) of the neuron and usually restricts it to an interval of the form [0,1] or [-1,1].

The above process is formulated in a rigorous mathematical way with the expression $y=f(XW+b)$ where f is the activation function. The similarity between biological neurons and artificial neurons is evident in Fig. 8, with the dendrites being the place where the input is introduced and the neuroaxon terminals giving the output. The basic division that can be made in neural networks is into feedforward (Fig. 8a), recurrent neural networks (Fig. 8b), and hybrids of the two. Other popular topologies of neural networks are lattice networks, layered feedforward networks with lateral connections and cellular networks.



(a) feedforward network *(b) recurrent neural network*

Figure 8: Two main categories of neural networks

(Jones, 2017)

- **Feedforward neural networks** consist of one or more layers and the connections between neurons are in a unidirectional direction. The layers consist of the input, the internal or otherwise hidden layers and the output layer. In this kind of networks there is no connection between neurons located in the same layer as well as no feedback between layers. In a fully interconnected feedforward network, such as the one in Figure 8a, each node in one layer is connected to all nodes in the next layer.

- In **Recurrent neural networks** (Figure 8b) there is at least one feedback, either between nodes of the same level or between a node of the present and a node of a previous level. These networks will be studied in more detail below.

The operation of neural networks is divided into two stages, those of learning and recalling. The process of learning the system is usually done by using samples that fall within the general rule on which we want to train. In this way, the network parameters (the weight matrix W) are appropriately adjusted using some learning rules. The training involves the complete scanning of the data for some finite number of steps, called epochs, and the testing is done using some criteria. Once the network reaches the desired performance or the number of epochs set at the beginning of training has been completed, the learning stage is terminated, and the neural network can now be used for the recall process. In this second and final stage the neural network is able to completely replace the system on whose dynamics the learning was performed and to be used on new samples.

3.3.1 Types of Machine Learning

The field of Machine Learning develops three modes of learning, similar to the ways by which humans learn: supervised learning, unsupervised learning and reinforcement learning.

● In **Supervised Learning**, for each input vector, there is a desired output vector. The goal is for the system to construct a function f that maps given inputs x to known desired outputs y, with the ultimate goal being the generalization of this function to inputs with unknown outputs. This is achieved by continuously varying the system parameters until the minimum desired error is reached. Error is the difference between the desired value and the output predicted by the system. For better generalisation, the practice of training and testing the results on different subsets of the data set is usually chosen. This avoids overfitting, i.e. the case in which the model trains very well for one particular dataset but fails to generalise to another. Supervised learning problems are divided into classification problems and regression problems. In the first case, the data is categorized into discrete classes (e.g., given an input of a set of symptoms of a patient, the output is 0 for a healthy patient and 1 for a sick patient). In the second case, a continuous value prediction is made (e.g., given an input of some meteorological metrics, the temperature of the next day is predicted).

● In **Unsupervised Learning**, a model is constructed for some sets of inputs in the form of observations without knowing the desired outputs. It is used in Association Analysis and Clustering problems. In Clustering, the goal is to organize samples into groups and the desired outcome is that the objects within a group are more similar to each other than to objects in other groups. The best-known unsupervised learning algorithm and more specifically Clustering algorithm is K-Means, which classifies samples into K classes. It does this in a series of steps. First, it initializes the K class/group centres, computes the distance between the sample and each class centre, classifies the sample into the class with the shortest distance from it, and updates the class centres by computing the mean of all vectors in each class. K-Means terminates after a certain number of iterations of the above process or when the class centres do not move enough in each iteration.

● In **Reinforcement Learning**, the algorithm tries to develop an autonomous agent that improves its performance through direct interaction with its environment. The evaluation is done through a reward function. The agent attempts to find a set of actions in its environment that maximises the reward function by trials and errors. It is mainly used in planning problems, such as robot motion control and optimisation of tasks in factories. It is also worth mentioning the very good effect of reinforcement learning in complex games such as chess.

3.4 Types of neural networks

3.4.1 Perceptrons or Feed-Forward Neural Networks

Feed-forward networks, as they are referred to in Greek literature, are the simplest neural networks. In these networks, information flows only in one direction (feed-forward), from the input layer to the hidden layers and from there to the output layer. There are no cycles or iterations in the network. More specifically, in feed-forward networks, the input layer is of equal dimensions to the input vector being fed into the network. Each input node is connected to all nodes in the next layer, while the output layer has equal dimensions to the output y. There can be multiple hidden layers connecting the input layer and the output layer, but in its simplest form, the perceptron connects the input directly to the output. The role of the hidden layers is to extract useful results, helping the network extract higher order information than that of the input data. When all nodes of each layer are connected to all nodes of the next layer, a fully connected neural network exists. Sometimes, the output signals of the intermediate neurons seem incomprehensible to humans, but they lead computers to better results.

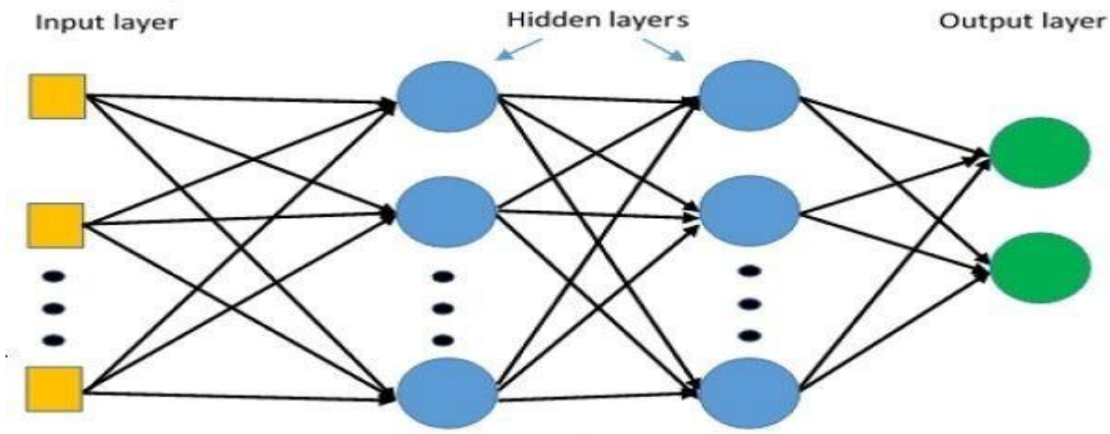


Figure 9: The structure of a fully connected neural network

(Machine Learning Generalization of Lumped Parameter Models for the Optimal Cooling of Embedded Systems - Scientific Figure on ResearchGate, 2021)

At this point, where a description of the simplest neural network architecture has been made, the way machine learning algorithms are trained should be analysed. In all these algorithms, the performance of the system is measured by cost function, which shall not only be observed but also minimized by changing the parameters of the network. This means that in each iteration, the direction in which the parameters (weights) of each layer must be modified is calculated in order to lead to the minimum cost. The main optimization techniques are summarized below:

1. **Gradient Descent:** It is known that the gradient of a function ∇f at some point is a vector perpendicular to the surface of f, which has a direction where f increases at a higher rate. Since our goal is to minimize the cost function, we want an opposite direction from ∇f, i.e. the weights will have the same direction as - ∇f. This is achieved by means of the formula W ′=W-a∇f, where f is the cost function and a is the learning rate and is essentially the step by which we go in the optimal direction.

2. **Stochastic Gradient Descent-SGD:** The most important problem of gradient descent is the convergence to local minimum, which prevents the finding of the overall minimum. To solve this, SGD uses random samples rather than the entire data set. Modern optimization techniques based on SGD that are proven to converge to the total minimum are Adaptive Gradient Algorithm (AdaGrad), Root Mean Square Propagation (RMSProp) and Adam.

3. **Backpropagation:** It is the method in which the partial derivatives of the cost function with respect to the weights of each neuron are computed by starting the computations from right to left, using the chain rule and storing the intermediate results. In simple words, we go backwards in the TND to find the partial derivatives of the cost function (quantifying the error) with respect to the weights and we calculate the new values of the weights by making use of the gradient descent algorithm as follows: $w_{ij}(l,k+1) = w_{ij}(l,k) - a\frac{dC}{dw_{ij(l,k)}}$, where $w_{ij}(l,k)$ is the

weight connecting neuron j of the previous level 1-1 to neuron i of the next level l after k steps of the algorithm, a is the training step and C is the cost function.

3.4.2 Recurrent neural networks

Recurrent neural networks (RNNs) are essential for sequential data prediction and therefore they find effective application in time series problems. The basic idea behind RNNs is that for an output to be produced, not only input data is used but also the previous outputs (Petnehazi et al., 2019). RNNs differ from traditional feedforward neural networks as they do not have neurons in a single direction. In other words, they can pass information both in the direction of a previous layer and in the same layer. In this way information does not flow only in one direction and as a result this enables us to have short term memory (Nelson et al., 2017). The RNN can be thought of as a Multi-Layer Perceptron (MLP) network with the addition of loops in its construction. The term MLP essentially refers to simple feedforward neural networks with at least three layers. In RNNs there is an input layer, a hidden layer and an output layer. This structure is similar to the MLP architecture except that the hidden layers are interconnected. In a vanilla (basic) RNN/LSTM model the nodes are connected in one direction. This type of architecture ensures that the output t=n depends on the inputs t=n, t=n-1, ..., and t=1. In other words, the output will depend on a sequence of data rather than on a data as shown in Figure 10 (Manaswi et al., 2018).

(Input1) → Output1

(Input2, Input1) → Output2

(Input3, Input2, Input1) → Output3

(Input4, Input3, Input2, Input1) → Output4

Figure 10: Correlation of inputs and outputs in a basic RNN/LSTM

The following diagram is indicative of the operation of a RNN, showing the correlation of inputs and outputs. The repetitive process allows information to pass from one step to the next.

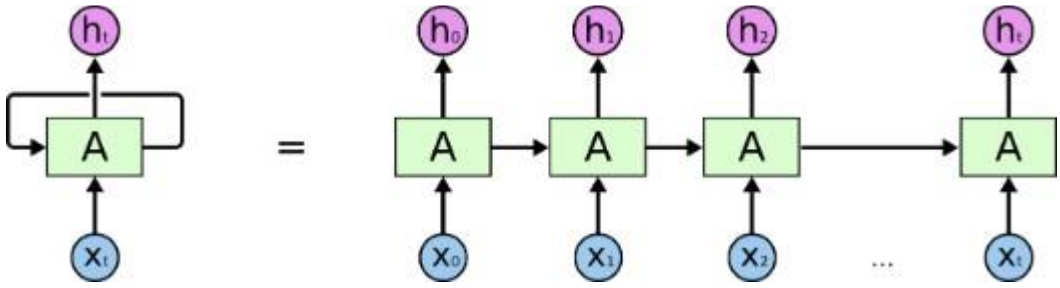


Figure 11: The operation of an RNN

(Roman, 2020)

Recurrent neural networks have the form of a chain of recurrent processes. In RNN the recurrent process is very simple as it consists of a simple tanh level (Olah et al., 2015).

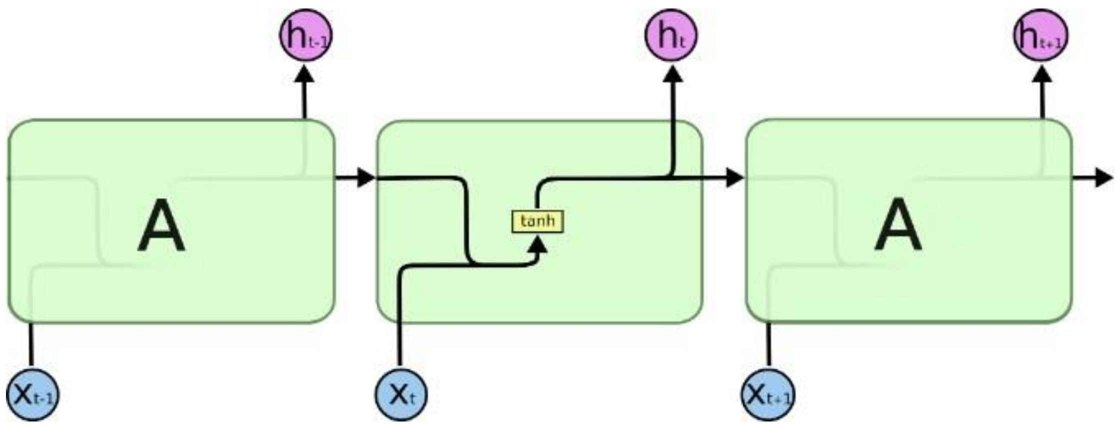


Figure 12: The repetitive process during the operation of an RNN

(Roman, 2020)

Based on how RNNs work, learning long-term dependencies is a great mathematical challenge. The main problem is that the gradients that need to be adjusted and are computed during the backpropagation of the error, they are propagated over so many stages and they tend to disappear most of the time and in fewer cases to increase excessively. Even if there is a way to prevent the gradient from growing excessively, the difficulty with long-term dependencies arises from the exponentially smaller weights given to long-term interactions compared to the weights given for short-term interactions (Goodfellow et al., 2018). This problem is present also in the use of RNNs with large data sequences and it is addressed by using a more complex neural network structure, the LSTM.

3.5 LSTM Recurrent Neural Networks (Long Short-Term Memory)

Long Short-Term Memory (LTSM) neural networks belong to the broader category of RNNs, but they are proven to be much more efficient in a large number of problems. Their key feature is that they can distinguish between recent and old data by giving them different weights, while they have the ability to "forget" information they consider irrelevant for predicting a next value. In this way, LSTMs are considered much more suitable for managing large sequences than other RNNs that can essentially only remember small sequences (Nelson et al., 2017).

More specifically, in LSTMs (Hochreiter and Schmidhuber, 1997) a unit is present which enables gated memory units, in neural networks. More specifically, an LSTM unit has three gates that manage the memory contents. These gates are essentially simple logarithmic weighted functions, where weights are assigned values through error backpropagation. Although this process is more complex, it is undoubtedly much more efficient, as through these gates the LSTM learns what it needs to learn, remembers what it needs to remember, and recalls what it needs to recall without requiring any special training or optimization. The three aforementioned gates are the input gate, the forget gate and the output gate. The input gate and the forget gate essentially form the cell state, which expresses the long-term memory. The output gate forms the output vector or hidden state, which is the memory on which the focus is made. This memory system enables the network to have memory for a long period of time, an element that was lacking in RNNs.

So, the LSTM has also the form of a chain, but this time the repetitive process is more complex. Particularly, instead of one neural network layer, there are four, as shown in Figure 13, which interact in a very special way.

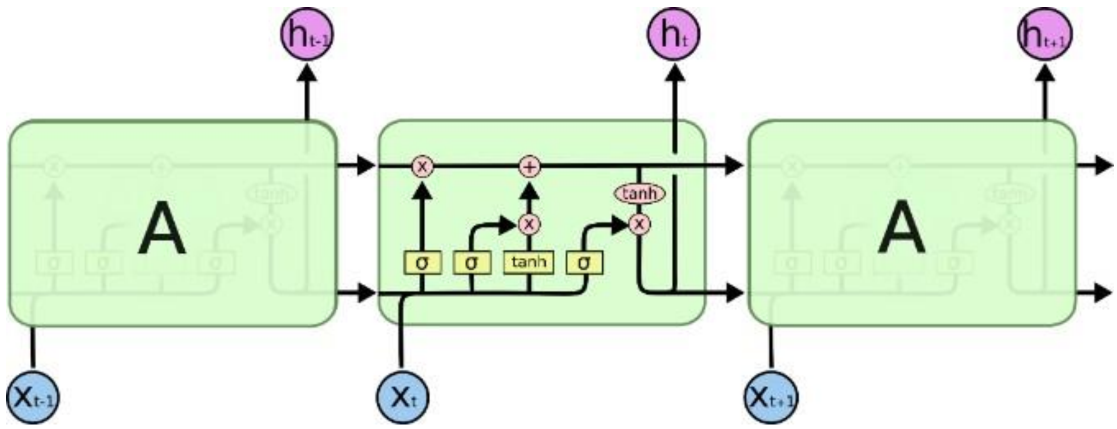


Figure 13: The LSTM as an extension of an RNN

(Roman, 2020)

As shown in Figure 13, each line carries an entire vector from the output of one node to the input of others. The round circles represent indicative functions, such as, for example, the vector addition, while the yellow boxes are levels or layers of neural networks. Furthermore, the merging of lines means concretion, while branching a line, results in two different copies. The key to the operation of LSTMs is the state cell, i.e. the horizontal line at the top of the diagram. This line is extended in the chain with small linear changes in its content (Figure 14).

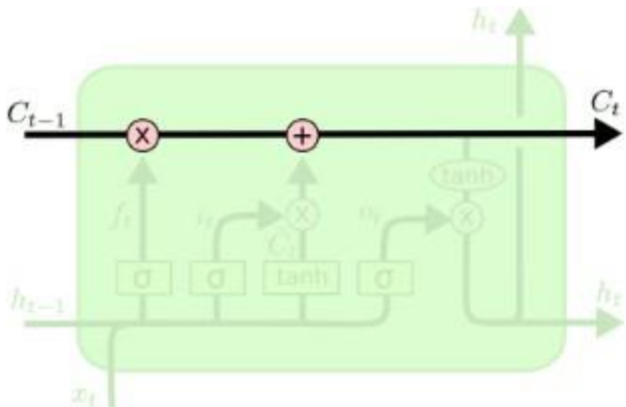


Figure 14: The status cell

(Roman, 2020)

The LSTM has the ability to remove or add information to the state cell, using structures called gates. The gates are a way to select the information, which will be transmitted. They consist of a sigmoid layer of neural network and a multiple multiplication function (Figure 15).

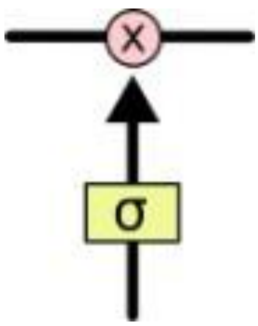


Figure 15: The shape of a gate inside the LSTM

(Roman, 2020)

Regarding the sigmoid layer, numbers from 0 to 1 are used as an output with the aim of describing "how much of each component must pass through". 0 means "let nothing pass" while 1 means "let anything pass". An LSTM has three such gates that protect and control the state cell.

The first step for the LSTM is to decide which information to remove from the state cell. This decision is made through a silicon layer called the forget gate layer. This gate sees h_{t-1} and x_t and gives a number from 0 to 1 at its output for each number of the state cell C_{t-1}. 1 means "keep it completely" while 0 means "forget it entirely".

The next step is to decide which new information to store in the state cell. This step consists of two parts. Firstly, a sigmoid layer called the "input gate layer" decides which values to update and then a tanh layer generates the vector of new candidate values C_t that are intended to be added to the state cell. These first two steps are illustrated in Figure 16.

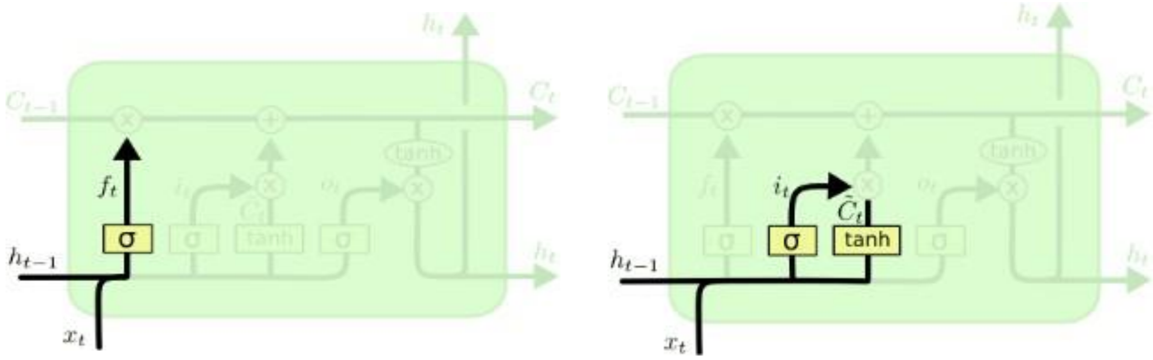


Figure 16: The first steps in the internal operation of an LSTM

(Roman, 2020)

The resulting correlations in Figure 16 are the following:

$$f_t = \sigma\left(W_f * [h_{t-1}, x_t] + b_f\right)$$

$$i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i)$$

$$C_t = tanh(W_c * [h_{t-1}, x_t] + b_c)$$

It is therefore, time to update the state of the cell from the old one, C_{t-1}, to the new one, C_t. The previous steps have already decided what will happen and what remains is its implementation. The old state is multiplied by $\boldsymbol{f}_t$, omitting the elements that it was decided beforehand to be removed. Then the product $\boldsymbol{i}_t$ * C_t - the new candidate values- are added, which are normalized according to how much it was decided to update each state cell value (Figure 17).

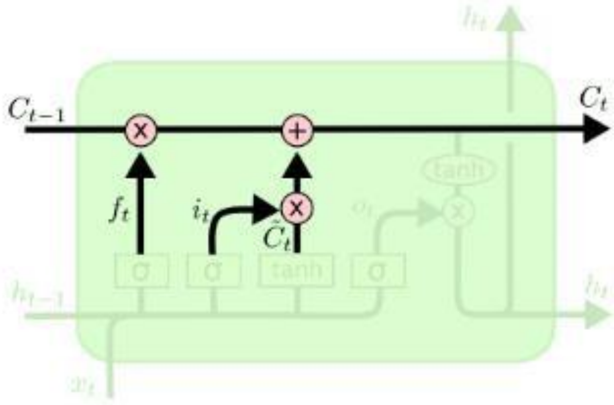


Figure 17: The renewal of the status cell

(Roman, 2020)

This renewal is expressed as follows:

$$C_t = f_t * C_{t-1} + i_t * C_t$$

Finally, we need to decide what the output $\boldsymbol{h}_t$ will be. The output will be based on the state cell, but it will be a filtered version of it. Firstly, a sigmoid layer is used which decides which parts of the state cell will be forwarded to the output. Next, the tanh function is applied to the state cell values so that they move in the range from -1 to 1, and finally it is multiplied with the output of the sigmoidal layer so that only the selected parts end up at the output (Figure 18) (Olah et al., 2015).

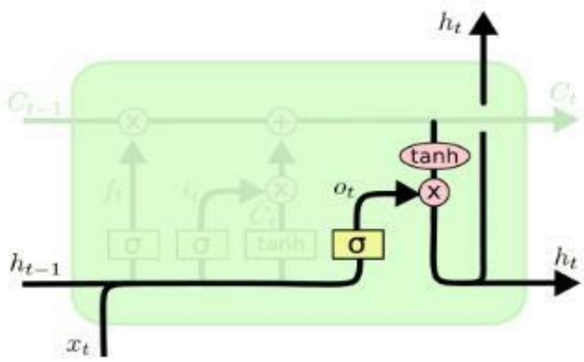


Figure 18: The output of the LSTM

(Roman, 2020)

The relationships described by Figure 18 are the following:

$$o_t = \sigma(W_o * [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh(C_t)$$

3.5.1 Time series analysis with LSTM

For Time Series Forecasting using Long Short-Term Memory (LSTM) Neural Networks we try to provide momentum indicators of the prices of unique items. Before we dive into forecasting and LSTM, there are some important parameters and arguments that need to be explained.

Input Timesteps (Lag)

Traditional neural networks take in a stand-alone data vector each time and have no concept of memory on data. LSTM networks keep a context of memory within their pipeline and thus become powerful at tackling sequential and temporal problems without the issue of the vanishing gradient affecting their performance. The memory our network keeps is the lag. For each forecast, we feed a lag of sequential information to our network, for it to learn from.

Forecasting Sequence

If the lag is the input, forecasting sequence is the output. It is the time-period window our model will try to forecast.

Train / Test Split

Just like in ANN 's, using an LSTM we have to split our dataset to a train set and a test set. There are a few differences though:

- Splitting cannot be on a shuffled dataset. We have to decide beforehand how many time-periods we will use as training.
- Lag cannot be higher than the train dataset.
- Train and test splits are sequential, just like before the split and train split precedes the test split.

Normalization

We decided to take each n-sided window of training /testing data and normalize it to reflect percentage changes from the start of that window.

Epochs

An epoch is simply one forward pass and one backward pass of all the training examples. Training for more epochs makes our model better but also more prone to overfitting.

Batch Size

Batch size is the number of training examples in one forward/backward pass. The higher the batch size, the more memory space we need. It has been observed in practice that when using a larger batch there is a significant degradation in the quality of the model, as measured by its ability to generalize.

Dropout

Dropout is a regularization method that approximates training a large number of neural networks with different architectures in parallel.

During training, some number of layer outputs are randomly ignored or dropped out. This has the effect of making the layer look like and be treated like a layer with a different number of nodes and connectivity to the prior layer. In effect, each update to a layer during training is performed with a different view of the configured layer. For example, a Dropout layer with a rate of 0.2 has a 20% chance to drop each neuron.

Loss Functions

A loss function is used to optimize the parameter values in a neural network model. Loss functions map a set of parameter values for the network onto a scalar value that indicates how well those parameters accomplish the task the network is intended to do. It is essentially a mathematical way of measuring how wrong our predictions are. Loss is that measure.

Optimizer

During the training process, we change the parameters of our model to try and minimize the loss function and make better, more accurate predictions/forecasts. Optimizers tie together the loss function and the model parameters by updating the model in response to the output of the loss function (Keras, 2021).

Activation Function

The activation function of a node defines the output of that node, or "neuron," given an input or set of inputs. This output is then used as input for the next node and so on until a desired solution to the original problem is found. If we do not apply an Activation function then the output signal would simply be a simple linear function.

HyperParameters

A hyperparameter is a parameter whose value is set before the learning process begins. By contrast, the values of other parameters are derived via training. Different model training algorithms require different hyperparameters. The hyperparameters of an LSTM neural network are on Table 1.

- Input timesteps
- Forecasted sequence
- Train / test split
- Epochs
- IQR high quantile threshold
- Loss
- Optimizer
- Learning rate for SGD
- Neurons of different LSTM and Dense layers
- Activation Function
- Dropout rate

Table 1: Hyperparameters

3.5.2 Activation functions

Activation functions are really important for an Artificial Neural Network to learn and make sense of something really complicated, as we can observe non-linear complex functional mappings between the inputs and the response variable. They introduce non-linear properties to our Network. Their main purpose is to convert an input signal of a node in an ANN to an output signal. That output signal now is used as an input in the next layer in the stack.

Specifically in ANN we do the sum of products of inputs(X) and their corresponding Weights(W) and apply an Activation function f(x) to it to get the output of that layer and feed it as an input to the next layer.

If we do not apply an Activation function then the output signal would simply be a simple linear function. A linear function is just a polynomial of one degree. Linear equations may be easy to solve but they are limited in their complexity and have less power to learn complex function mappings from data.

Non-linear functions are those which have degree more than one and they have a curvature when plotted. Hence, we need to apply an Activation function f(x) so as to make the network more powerful and add to it the ability to learn something complex and complicated from data and represent non-linear complex arbitrary functional mappings between inputs and outputs. In the figure 19, a linear activation function graph is illustrated.

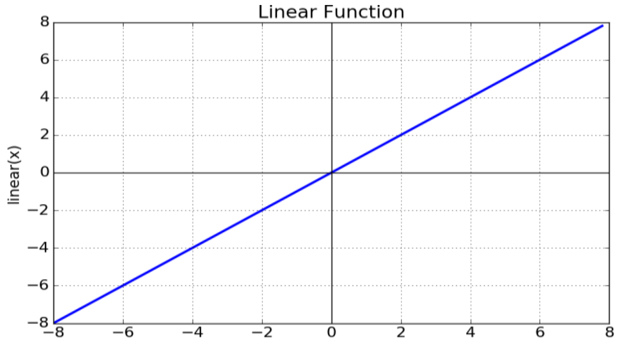


Figure 19: Linear Activation Function

("Applications of Linear Functions | Boundless Algebra", 2021)

Except from the **Linear** activation functions, we also experimented with the following ones:

Swish

With the function f(x) = x · sigmoid(x), the swish is better than ReLU on deeper models across a number of challenging data sets.

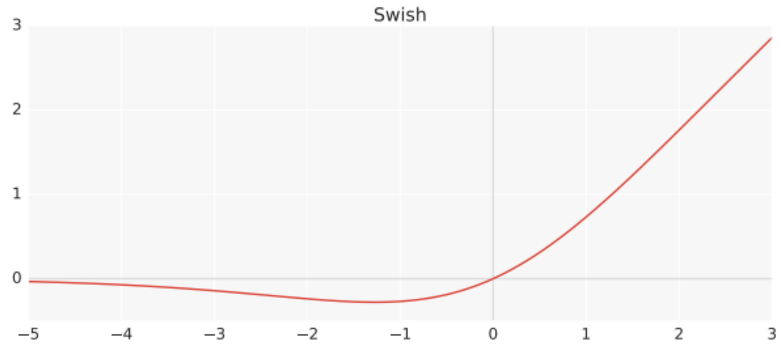


Figure 20: Swish Activation Function

(Singh, 2021)

Swish is a smooth, non-monotonic function that consistently matches or outperforms ReLU on deep neural networks applied to a variety of challenging domains such as Image classification and Machine translation. It is unbounded above and bounded below & it is the non-monotonic attribute that actually creates the difference. With self-gating, it requires just a scalar input whereas in multi-gating scenario, it would require multiple two-scalar inputs.

Sigmoid

t is an activation function of form f(x) = 1 / 1 + exp(-x). Its Range is between 0 and 1. It is a S— shaped curve. It is easy to understand and apply but it has major reasons which have made it fall out of popularity.

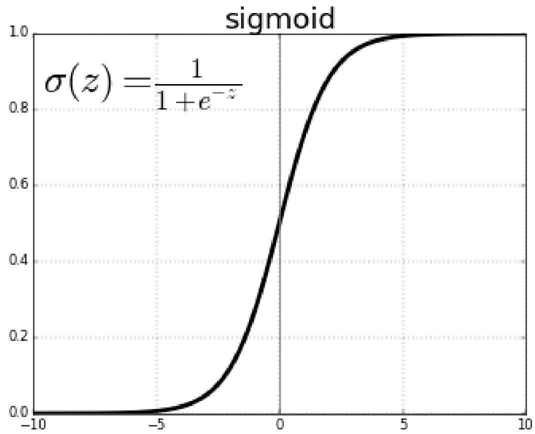


Figure 21: Sigmoid Activation Function

("Activation functions and their derivatives - Master Data Science", 2018)

The Vanishing gradient problem in Sigmoids:

- Its output is not zero centered. It makes the gradient updates go too far in different directions. 0 < output < 1, and it makes optimization harder.
- Sigmoids saturate and kill gradients.
- Sigmoids have slow convergence.

Softmax

The sigmoid functions are not suitable for classification problems, and for that we needed a new function. The softmax function is a more generalized logistic activation (sigmoid) function which is used for multiclass classification. If the problem was on binary classification, the sigmoid function would work just as well.

Rectified Linear Unit (ReLU)

A function that is very much used in neural networks especially in convolutional networks is ReLU, which is computationally efficient and significantly speeds up the training of the network. Its mathematical formula is: f(x)= x+=max(0,x)

However, looking at the graph in Figure 22 one can see that for zero or negative input values the slope of the function becomes zero. This results in the neurons whose inputs are in the negative to be considered as "dead" and in their output to be continuously zero, rendering them useless. This phenomenon is known as the dying ReLU problem and is one of the major drawbacks of this function as, due to the nature of ReLU, a "dead" neuron is difficult to recover. This carries the risk that a large part of the network, given the right conditions, may become inactive, significantly reducing its efficiency.

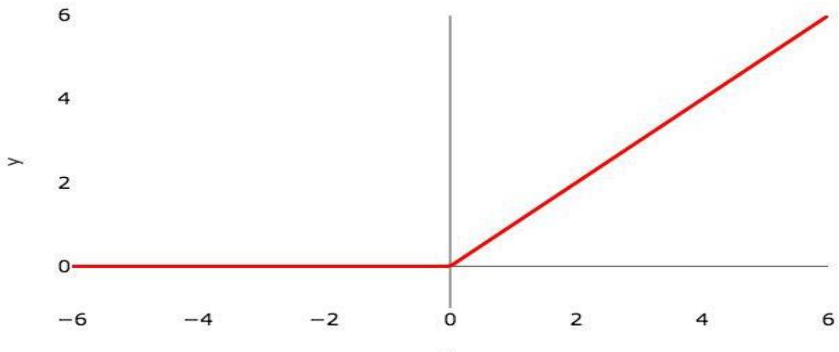


Figure 22: Rectified Linear Unit (ReLU)

("Activation functions and their derivatives - Master Data Science", 2018)

3.5.3 Feature scaling

Feature scaling is a method used to standardize the range of independent variables or features of data. In data processing, it is also known as data normalization and is generally performed during the data preprocessing step. Since the range of values of raw data varies widely, in some machine learning algorithms, objective functions will not work properly without normalization. For example, the majority of classifiers calculate the distance between two points by the Euclidean distance. If one of the features has a broad range of values, the distance will be governed by this particular feature. Therefore, the range of all features should be normalized so that each feature contributes approximately proportionately to the final distance. For scaling we experimented with 2 different scalers: Min-max scaler and Standard Scaler.

Min-Max Scaler

In this we subtract the Minimum from all values – thereby marking a scale from Min to Max. Then divide it by the difference between Min and Max. The result is that our values will go from 0 to 1. This is quite acceptable in cases where we are not concerned about the standardisation along the variance axes e.g. neural networks expecting values between 0 to 1.

The downside however is that because we have now bounded the range from 0 to 1, we will have lower standard deviations and it suppresses the effect of outliers.

Standard Scaler

We standardize features by removing the mean and scaling to unit variance. The standard score of a sample x is calculated as:

$$z = \frac{x - u}{s}$$

where u is the mean of the training samples, which is 0 if with_mean=False, and s is the standard deviation of the training samples, which is 1 if with_std=False.

Centering and scaling happen independently on each feature by computing the relevant statistics on the samples in the training set. Mean and standard deviation are then stored to be used on later data using the transform method.

3.5.4 Hyperparameter selection in recurrent neural networks

In addition to selecting the best recurrent neural network model for each application, there are some key parameters that need to be defined and it is these that determine to a greater extent the success or failure of the training. Most of them have already been mentioned in this thesis, but we have not discussed how to choose the appropriate values for them. The most basic ones are the number of hidden layers in the network, the number of nodes in each layer and some others.

Number of hidden layers

The number of hidden layers in the network depends basically on the problem we want to simulate. There is a general principle that says that if the data of the problem solution is linearly separable, then there is no need to use any hidden layer, since the activation function can be applied to the input layer, which combined with that of the output is enough to model such problems. However, in the case of problems dealing with complex nonlinear systems, one or two hidden layers are usually used in addition to that of the input. More specifically, one hidden layer is used in the case of some nonlinear functions that connect data from one finite space to another, while two hidden layers can represent any arbitrary decision criterion with arbitrary precision and any linear or nonlinear relation with any precision (as small as desired). If accuracy is the only concern of this approach, a third hidden layer can be used, but this technique will greatly increase the time needed for the training of the system and is not generally appropriate, since one or two hidden layers are sufficient to solve any complex nonlinear problem. In conclusion, most researchers recommend using as few hidden layers as possible, always according to the application, to save computational resources, training time and solution complexity (Karsoliya, 2020)

Number of hidden layer nodes

Calculation is also needed to select the number of hidden layer nodes in the network. When it comes to the input and output layers, the number of nodes is determined by the input-output data of the application, but when it comes to the best choice of the number of nodes of the inner layers there is no fixed policy that guarantees the optimal result. A key concern is to avoid events such as overfitting and underfitting in the training data, which occurs when the number of nodes is too large or too small respectively with respect to the size and complexity of the problem. Other considerations to be taken into account when choosing the number of nodes are minimizing the training time and improving the accuracy of the output.

The most popular methods used for selecting the number of nodes of hidden layers are listed below (Stathakis, 2009). Perhaps the best-known method is that of Kolmogorov, which claims that any continuous function can be represented by a neural network containing exactly 2n + 1 nodes in a single hidden layer, with n being the number of input nodes. This theory, however, has been proven to be inaccurate because it is only valid for a specific node activation function that is considerably more complex than the sigmoidal ones, which are most often used in practice. Otherwise, the number of nodes needed to maintain only one hidden layer may even reach the number of training samples. As an extension of this theorem, Huang proved in 2003 that with two hidden layers, m neurons in the output layer and a number of samples for training, the number of nodes in the hidden layers should not exceed $2\sqrt{(m+2)N}$ nodes. More specifically, he proposes that the number of nodes should not exceed:

In the first hidden layer:

$$\sqrt{(m+2)N} + 2\sqrt{\frac{N}{m+2}}$$

In their second hidden level:

$$m\sqrt{\frac{N}{m+2}}$$

Other methods of selecting the optimal number of nodes in the internal layers are by conducting exhaustive research that tests all possible topologies of the system, but is not at all practical because of the time required to implement it, and by testing a random number of nodes and then correcting for error. Other methods used are the heuristics, which make use of the empirical knowledge accumulated from solving previous problems with neural networks and are mainly used as a starting point of research for the previous techniques. Some of them are the following: the number of hidden layer nodes should be in the range of 70-90% of the number of input nodes, or the total number of hidden layer nodes should not exceed twice the number of input nodes, or the size of each inner layer should be between those of the input and output.

Another method worth noting is the one that, with a given number of nodes, successively adds or removes connections between them until the best result is obtained. Such techniques work very efficiently, but their complexity contributes to the fact that they are not often used. Finally, there may be no clear way to optimally select the nodes of the inner layers, but it has been shown experimentally that methods that use the same or approximately the same number of nodes in all hidden layers have better results.

3.6 GRU Recurrent Neural Networks (Gated Recurrent Unit)

GRU networks were first proposed by Cho et al. (2014) and are a simplified version of LSTMs as they contain a smaller number of gateways.

Moreover, there is no concept of cell state as it is embedded in the hidden state and the information flow is controlled by the reset and update gates which determine what information should be omitted and what information should be passed to the next unit respectively. The structure of the GRU module and the equations that define its operation are shown in Figure 23.

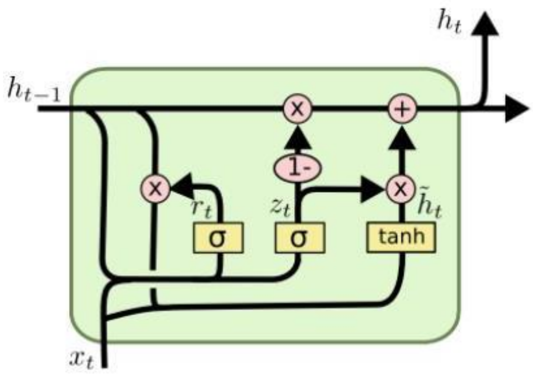


Figure 23: GRU unit structure and equations

(Roman, 2020)

Due to its simpler form, the GRU has a smaller number of parameters than the LSTM resulting in a generally shorter training time. Many studies have been conducted comparing the performance of these networks such as that of Karpathy et al. with the results showing that neither form is clearly better than the other as the networks present very close results.

3.7 Bidirectional LSTM Recurrent Neural Networks (BiLSTM)

The idea of Bidirectional Recurrent Neural Networks (RNNs) (Brownlee, 2017) involves duplicating the first recurrent layer in the network so that there are now two layers side-by-side, then providing the input sequence as-is as input to the first layer and providing a reversed copy of the input sequence to the second. More specifically, the first model learns the sequence of the input provided, and the second model learns the reverse of that sequence.

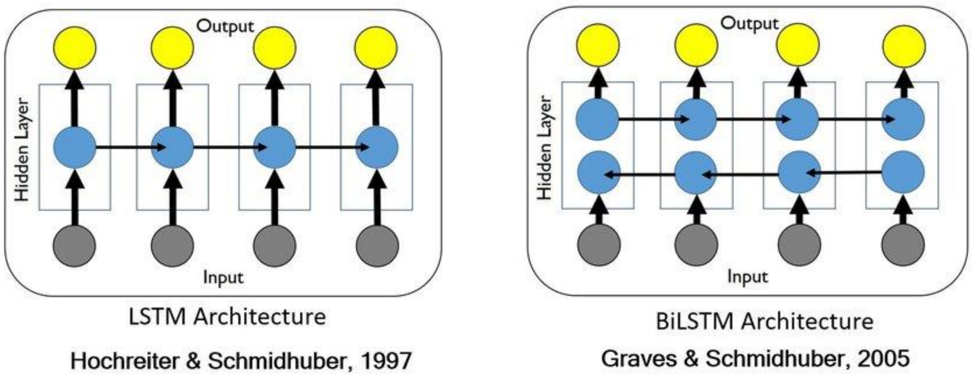


Figure 24: BiDirectional LSTM architecture

Since we do have two models trained, we need to build a mechanism to combine both. Bidirectional LSTMs are supported in Keras via the bidirectional layer wrapper. This wrapper takes a recurrent layer (e.g. the first LSTM layer) as an argument. It also allows you to specify the merge mode, which is how the forward and backward outputs should be combined before being passed on to the next layer. The options are:

- Sum: The outputs are added together.
- Multiplication: The outputs are multiplied together.
- Averaging: The average of the outputs is taken.
- Concatenation: The outputs are concatenated together (the default), providing double the number of outputs to the next layer.

$$z_t = \sigma(W_Z * [h_{t-1}, x_t])$$

$$r_t = \sigma(W_r * [h_{t-1}, x_t])$$

$$\hat{h}_t = tanh(W * [r_t * h_{t-1}, x_t])$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \hat{h}_t$$

The default mode is to concatenate, and this is the method often used in studies of bidirectional LSTMs.

3.8 XGBoost

XGBoost is an implementation of the Gradient Boosting algorithms first proposed by Chen et al. (2016) and aims to reduce the training speed, while increasing the performance of the above algorithms. It belongs to the broader class of algorithms based on the Ensemble learning method and in recent years has become the state-of-the-art method in solving a large number of classification and regression problems. Like all gradient boosting algorithms, XGBoost bases its operation on the generation of simple models ("weak learners"), which in this case are decision trees. It starts with one tree which generates the initial predictions and gradually tries to reduce the prediction error by creating one more tree at a time and summing up the individual predictions, up to a specified number of trees or until a desired error level is reached. The final model is a combination of the trees constructed and its predictions are the sum of all the individual predictions. Figure 25, taken from the work of Chen et al., shows a simple example of a combination of two decision trees.

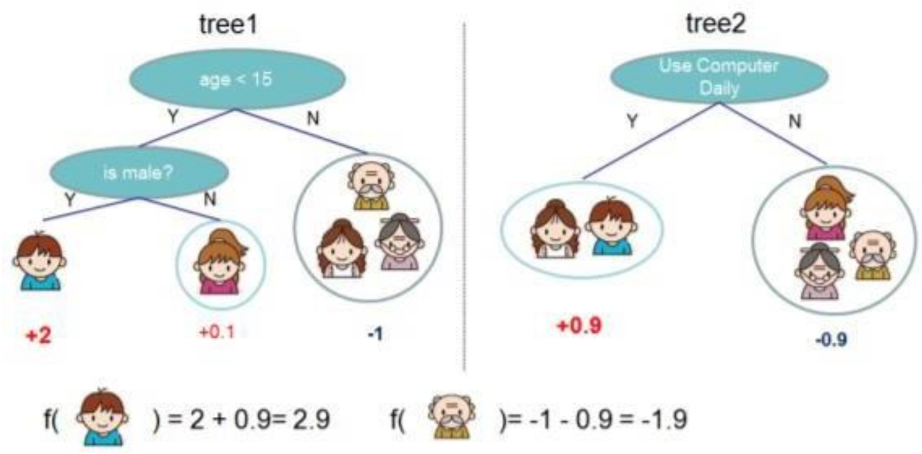


Figure 25: Combination of predictions of individual trees

(Patel, 2020)

XGBoost is used in order to estimate the importance of the selected features for the neural networks, and the diagram will be presented on Chapter 4.

3.9 Back Propagation Algorithm

Backpropagation of error algorithm is one of the most widely used training algorithms in the case of neural networks. It is capable of training feedforward networks of any size and number of layers. Its objective is to calculate the values of weights and biases, which are often referred to as learnable parameters of an ANN, in order to minimize the loss function of the neural network. Given a neural network architecture and a loss function, back- propagation calculates the gradient of the loss function, ∇E with respect to all weights in order to achieve a local minimum using dynamic programming. This is a mathematical process called gradient descent, which is a first-order iterative optimization algorithm for finding the minimum of a function. This is, given a neural network and a loss function, a back- propagation algorithm that propagates the loss at the output layer backward, so that the gradient at all hidden layers can be calculated using the chain rule in order to adjust the weights at each neuron, as described below.

Assuming the output of the j-neuron is o_j, then:

$$o_j = f(net_j)\ then\ net_j = \sum_{j}^{i} w_{ij} o_i + \theta_j$$

Then the error produced in the output of the neuron j is:

$$e_j = t_j - o_j$$

Summing the errors of all the neurons in the output layer:

$$E_p = \frac{1}{2} \sum_{j-output} \left(t_{pj} - o_{pj}\right)^2$$

The weights and biases are updated as following:

$$\Delta_p w_{ji} = -a \frac{dE_p}{dw_{ji}}$$

$$\Delta_p \theta_j = -a \frac{dE_p}{d\theta_j}$$

where *a* describes the learning rate parameter.

The calculation of the partial derivative of the objective function with respect to a weight w_{ij} is done, using the chain rule:

$$\frac{dE}{dw_{ij}} = \frac{dE}{do_j}\frac{do_j}{dnet_j}\frac{dnet_j}{dw_{ij}}$$

Similarly, the partial derivative of the object function with respect to the biases is calculated.

As it can be seen from the Figure 26 below, the loss function is calculated by considering the squared difference between the target and predicted output and summing these values for all output neurons. Then, the derivative of the loss function with respect to all weights is calculated using the chain rule, and by propagating the error backwards, the values of all weights are adjusted accordingly.

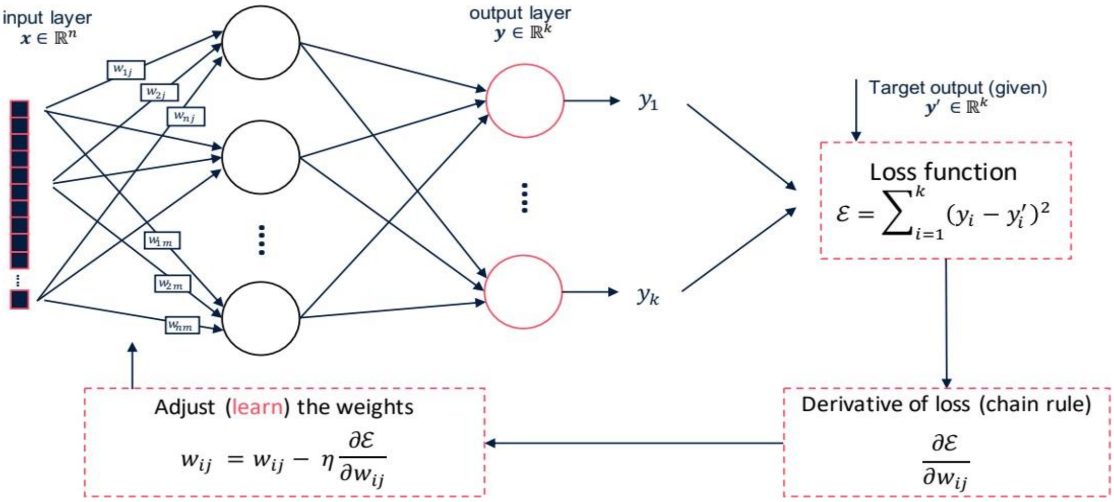


Figure 26: Schematic Representation of Backpropagation Algorithm

("Artificial neural network ensemble modeling with exploratory factor analysis for streamflow forecasting - Scientific Figure on ResearchGate", 2021)

3.10 Optimization of Backpropagation Algorithm

3.10.1 Batch Normalization

Batch normalization is a technique applied in artificial neural networks in order to improve their efficiency and stability. While training a neural network, the values of the weights of the layers constantly change, therefore the distribution of the inputs also changes. This makes training slower, especially in cases of very deep networks. This problem is known as *internal covariate shift*. More specifically, as long as the statistical distribution of the input keeps changing after some iterations, the hidden layers will keep trying to adapt to the new distribution, therefore, making convergence slower. So, the batch normalization algorithm is proposed in order to solve that issue. It involves a normalization of the inputs to a layer with zero mean value and unity standard deviation. This makes each layer in the network to learn faster, and also independently of the other layers. Furthermore, Batch Normalization makes neurons work in their linear regions of their activation functions, which improves learning performance. It also prevents the *vanishing gradient problem*, which sigmoid and tanh functions are dealt with.

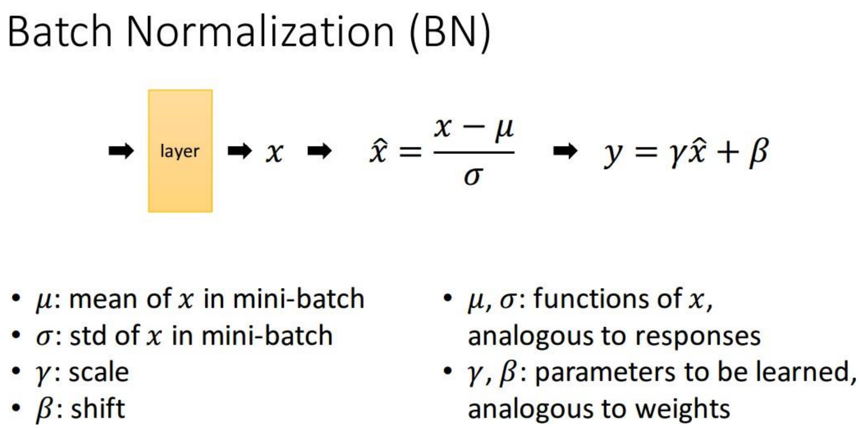


Figure 27: Batch Normalization (BN) applied to the inputs

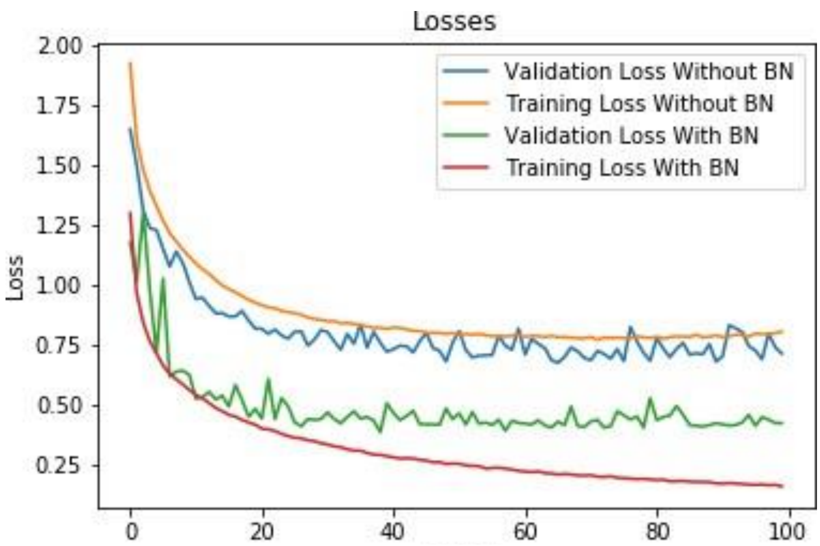


Figure 28: Average loss with and without Batch Normalization

(Nayak, 2018)

3.10.2 Adaptive Learning Rate

Learning rate in artificial neural networks is defined as a parameter which indicates how fast the network adapts the values of its weights, with respect to minimizing the loss function. The lower the learning rate is, the slower it will take the ANN to converge. However, if its value gets too great, the network might overshoot the minimum of the loss function, and it might cause divergence. The weight update formula is as following:

$$w_{ij} = w_{ij} - \eta \frac{dE}{dw_{ij}}$$

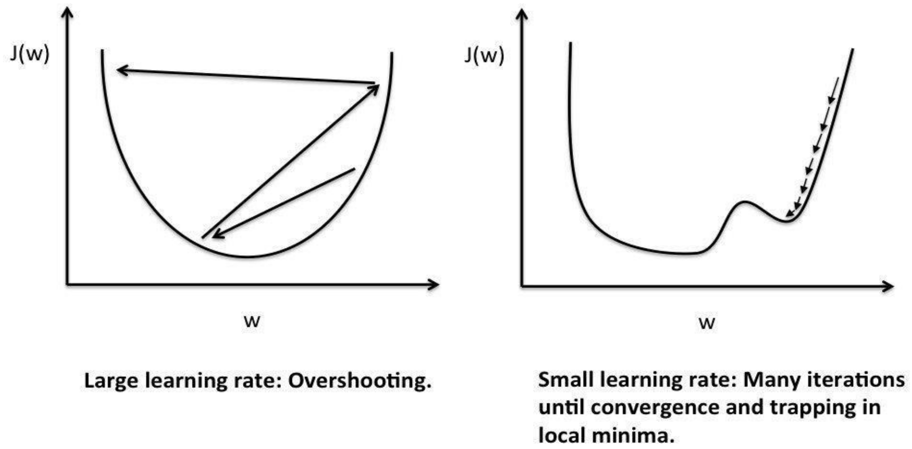


Figure 29: Large and small learning rate issues

(Kulkarni, 2018)

3.10.3 Momentum Rate

In artificial neural networks, the error function comprises of local minimums, which the system can get stuck to, thinking it has reached the global minimum. To overcome such situations, a momentum term is considered in the objective function, which is a value between 0 and 1 that increases the size of the steps that are taken towards the minimum. Therefore, it helps the system jump from local minimums, and move towards the global minimum. A large momentum rate implies faster convergence, but the global minimum point might be skipped. In cases that the learning rate is large, momentum rate should be kept at low values.

$$\Delta w_{ij} = \eta \frac{dE}{dw_{ij}} + \gamma \Delta {w_{ij}}^{t-1}$$

where η is the learning rate and γ is the momentum rate

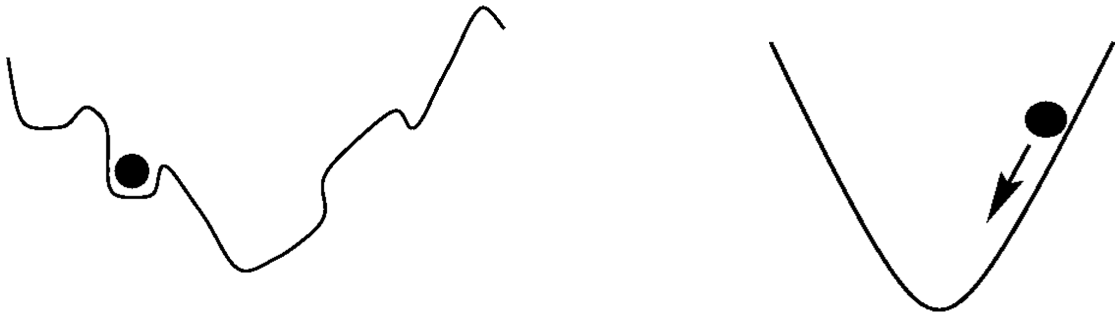


Figure 30: Neural network stuck into local minimal

("Momentum and Learning Rate Adaptation", 2021)

3.11 Issues in Artificial Neural Networks

Thanks to their huge number of parameters, artificial neural networks can be easily modified to fit almost any complex dataset. This ability has allowed them to find applications in many areas, in which it has been difficult to make progress, such as image recognition, natural language processing, etc. However, sometimes, the complexity of these networks may become a potential weakness, since it can lead to overfitting or underfitting.

3.11.1 Overfitting

This situation occurs when the neural network is so closely fitted to the training set that it is difficult to generalize and make predictions for new, unseen data. In practice, detecting that a trained model is overfitting is a difficult task, so it is necessary that some steps should be followed during training. Specifically, it is advisable that a dataset should be divided into three parts – training set, cross-validation set and test set. The model learns by only considering data from the first part, while cross-validation set is used to track progress of training and optimize the model. Furthermore, at the end of the training process, the test set is used, in order to evaluate the performance of the trained network.

Although until recently the most frequently recommended division of dataset would be: 80% training set and 20% test set, when a dataset comprises of millions of entries, these proportions are no longer appropriate. In short, everything depends on the total size of the dataset, and in cases of millions of data, it could even be better to divide the set in 98/2 ratio.

3.11.2 Underfitting

This situation occurs when the trained model can neither fit the training data, nor generalize to new unseen data. Detecting an underfit model is obvious, since it will have poor performance on training data, and it is, of course, an unsuitable.

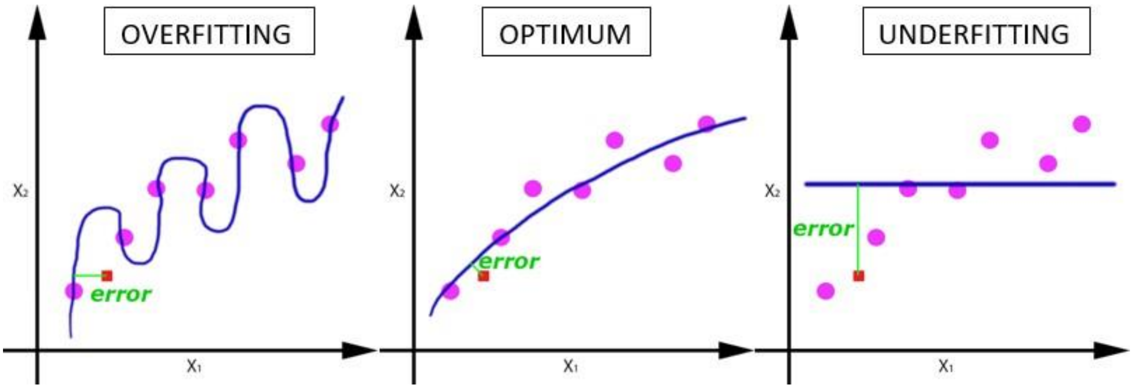


Figure 31: Description of Overfitted, Optimum and Underfitted model

("Overfitting vs Underfitting", 2019)

3.12 Regression evaluation metrics

The evaluation of the developed models was carried out using a set of performance metrics to provide a more comprehensive picture of the performance of each model.

Mean Squared Error (MSE)

Expresses the mean squared deviation of the model's predictions from the actual values. Due to the fact that errors are squared this indicator is significantly affected by outliers and should not be chosen in cases where the data set is characterized by many such values. It is calculated by the formula:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Root Mean Squared Error (RMSE)

It is one of the most used metrics in regression problems and like the MSE is significantly dependent on outlier values, however to a lesser extent due to the root. It is calculated by the formula:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Mean Absolute Error (MAE)

The above metrics are very good indicators of a model's performance for a particular variable; however, they cannot be used to compare model predictions for different variables as they depend on the range of values of the variables. In this case the MAE metric can be used, which expresses the average deviation of the predictions from the actual values. With the MAE metric, in addition to comparisons between forecasts for different variables, a more intuitive picture of the range of error is provided as it is given over the true value. It is calculated by the formula:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

3.13 TensorFlow Deep Learning Framework

TensorFlow (Tensorflow, 2015) is an end-to-end open-source platform for machine learning. It was developed by researchers and engineers from Google to conduct machine learning and deep neural networks research. It employs Python for a convenient front-end API used to build applications, while executing them in high-performance C++.

It allows developers to represent computations without performing them until asked. This happens by creating dataflow graphs, like the one in figure 31, that describe how data moves through a graph, or a series of processing nodes. Nodes in the graph represent a mathematical operation, and connections between nodes represent multidimensional data arrays, or tensors. The computations of the graph are executed in a session that places the graph operations on hardware and provides methods to execute them. This study employs TensorFlow and its high-level API keras to create deep Recurrent neural networks used for the BDI prediction.

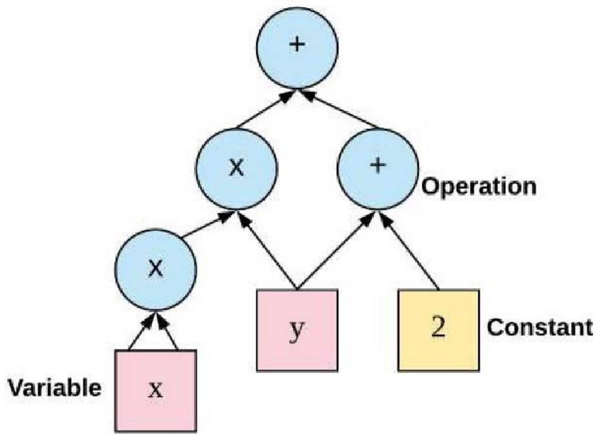


Figure 32: TensorFlow dataflow graph

("How do you build computational graph in TensorFlow?", 2019)

Chapter 4

Feature selection

4.1 Methodology

Data collection. The first step in the present research was the collection of all those macroeconomic and shipping factors which, according to the experience to date, were considered most likely to contribute to the formation of the BDI, the Baltic Dry Index.

Data processing. A more theoretical approach is followed, in which from the variables we have collected, we select those that are mostly correlated to the Baltic Dry Index. The data we have available are subjected to many different types of processing before being used by the neural networks. In particular, their correlation with the variable to be predicted is firstly checked. Also, the data which will be used as input to the neural network will be divided to training, validation and testing data.

4.2 Data collection

By researching past events of the economy, one can observe that the viability of businesses, in whatever sector they operate, depends on the conditions prevailing in the macroeconomy determining their level of prosperity or the probability of default. The critical macroeconomic factors that have been selected are described below. Many of these factors are directly relevant to the US economy, as it is the leader among the developed economies of the world and so it determines the general economic development globally. Other emerging economic powers such as China or India are in fact interrelated with the US economy, and the evolution of their main macroeconomic characteristics has been incorporated into the evolution of the US economy. In addition, by studying the major economic crises that took place in the 20th and 21st centuries, one can conclude that the majority of the crises that eventually affected the shipping industry were those that spread around the world, negatively affecting international trade. The majority of these global economic crises originated in the United States, before affecting the global economy through a domino process. Other financial crises that occurred in major economies such as Sweden and

Japan during the 1990s did not have a strong influence on the global economy and the shipping industry.

The macroeconomic time series available for processing are the following (Manos, 2019):

1) Jobless claims

Jobless claims (U.S. Department of Labor, 2021) are a way of measuring labor market dynamics by counting the claims of those who first attempt to enter into the unemployment insurance system. The fewer people who apply for unemployment insurance, the more people are expected to be employed, which is a positive sign for the health of the economy. The Employment and Training Administration of the U.S. Department of Labour is responsible for compiling and publishing this indicator on a weekly basis, specifically every Thursday.

Labour is intertwined with personal income which provides purchasing power to consumers who in turn help sustain the growth of the economy. By monitoring this indicator, investors can draw useful conclusions for timely decision-making. Data on new jobless claims are published on a weekly basis with any increase or decrease implying a deterioration or improvement in the labour market accordingly. The following example illustrates possible developments in the economy from the continuous improvement of the indicator (Figure 33).

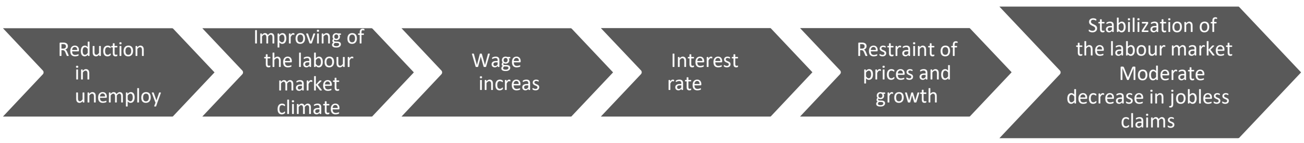


Figure 33: Potential long-term consequences of sustained reductions in unemployment claims

However, a serious drawback of this indicator should be mentioned: the fact that it does not provide information on unemployed people who, during their stay in the Unemployment Fund, did not manage to secure a job until their definitive removal from the Fund. This means that while some people remain unemployed, they are removed from the Unemployment Fund without having the right to re-register, and consequently this important element is not recorded anywhere. This information gap is mitigated by using as many factors as possible as 'input data' in this survey.

In figure 34, the evolution of this indicator over the last 30 years, which is the period of time on which the survey is based, is presented. The graph clearly shows the sharp rise in the index during the 2008 financial crisis and the Covid-19 health crisis in 2020.

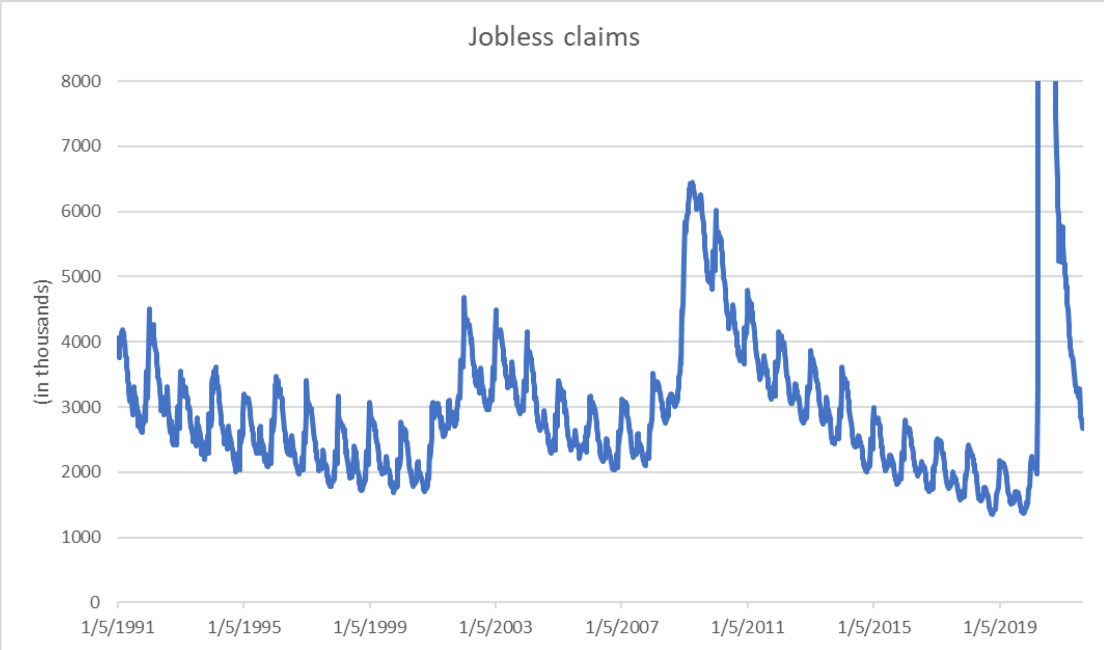


Figure 34: Thirty-year trend in Jobseeker's applications

2) Consumer Price Index (CPI) United States

The Consumer Price Index (U.S. Bureau of Labour statistics, 2021) is a time series that reflects the level of prices of consumer goods and services. The Bureau of Labour Statistics (BLS) of the U.S. Department of Labour is responsible for developing and publishing the Consumer Price Index on a monthly basis. The Consumer Price Index is the most widely used monthly inflation indicator and refers to a basket of goods and services, essentially reflecting the cost of living rather than the general level of prices. Consumer goods account for 40% of the index, while the remaining 60% relates to the change in service prices. The base period for the calculation of the current US CPI is 1982.

Depending on the changes in the rate of inflation and on the expectations of its future path, the Federal Reserve adjusts monetary policy with corresponding increases in the key interest rate. These changes have important effects on equity, bond, and commodity prices. An example (Figure 35) of the measures taken in the event of successive increases in the CPI is given below. Similar policies are followed for sharp declines in the US CPI.

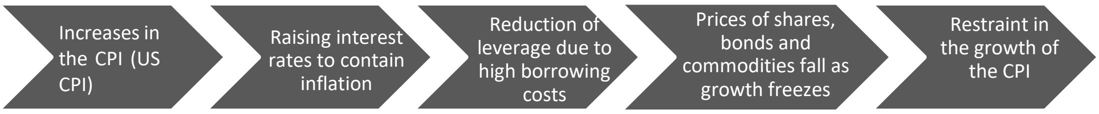


Figure 35: Case of successive CPI increases and reaction policies

In figure 36, the 30-year trend of the consumer price index, which shows a steady upward trend, is presented.

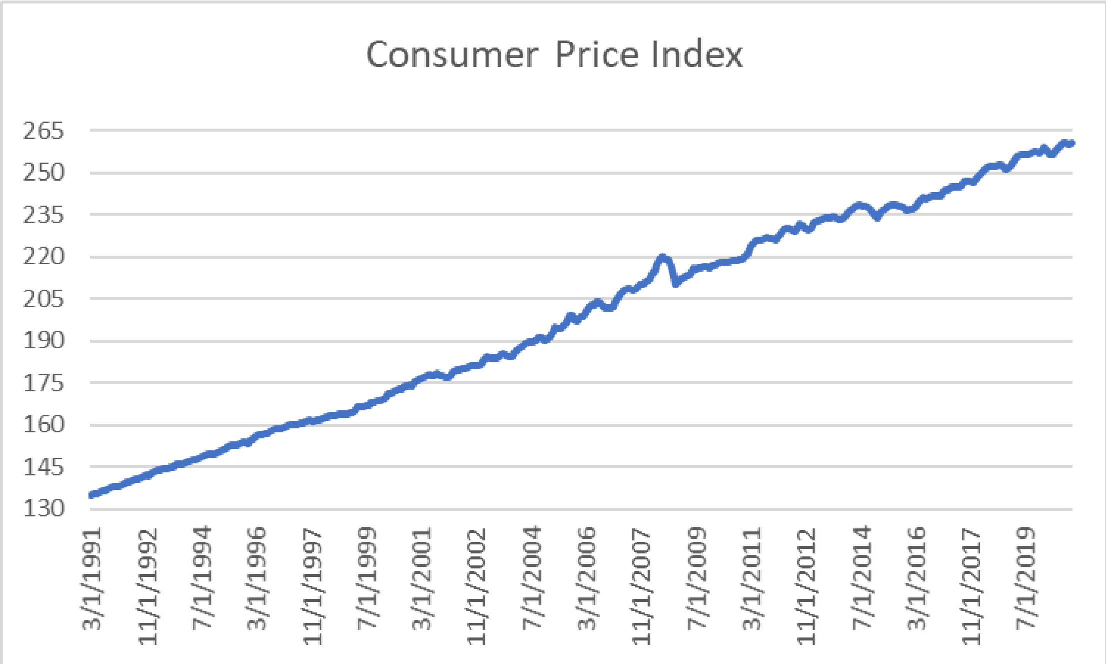


Figure 36: Thirty-year trend in the Consumer Price Index

3) Producer Price Index (PPI)

The Producer Price Index (PPI) calculates, on a monthly basis, the price level of a specific "basket" of products (raw materials and industrial products) at the production stage. In other words, the PPI measures the prices of products at the producer level before they are made available to consumers. Since this index measures inflation at the production stage, it follows that a proportion of the change in prices is passed on to the Consumer Price Index (CPI). The importance of the evolution of the consumer price index in the formulation of monetary policy by means of a hike in the key interest rate was mentioned earlier. The producer price index is, however, the harbinger of inflation developments, providing early indications of what is to follow at the level of consumer goods and services.

The Producer Price Index (PPI) in the US has three versions and refers to:

- Finished goods ready for wholesale
- Processed goods and materials ready to cover some intermediate stage of production
- Raw material requiring further processing

The PPI relating to finished products is the one that will be used in this investigation as the other two indicators, although very useful for the information they provide at an early stage, show strong fluctuations.

In figure 37 (below), the thirty-year path of the producer price index, which follows a similar path to that of the consumer price index, is shown. In the financial crisis of 2008, we have seen a downturn caused by a sharp fall in demand resulting in a fall in prices, as in the case of the consumer price index, which in fact absorbs a large proportion of the evolution of the producer price index.

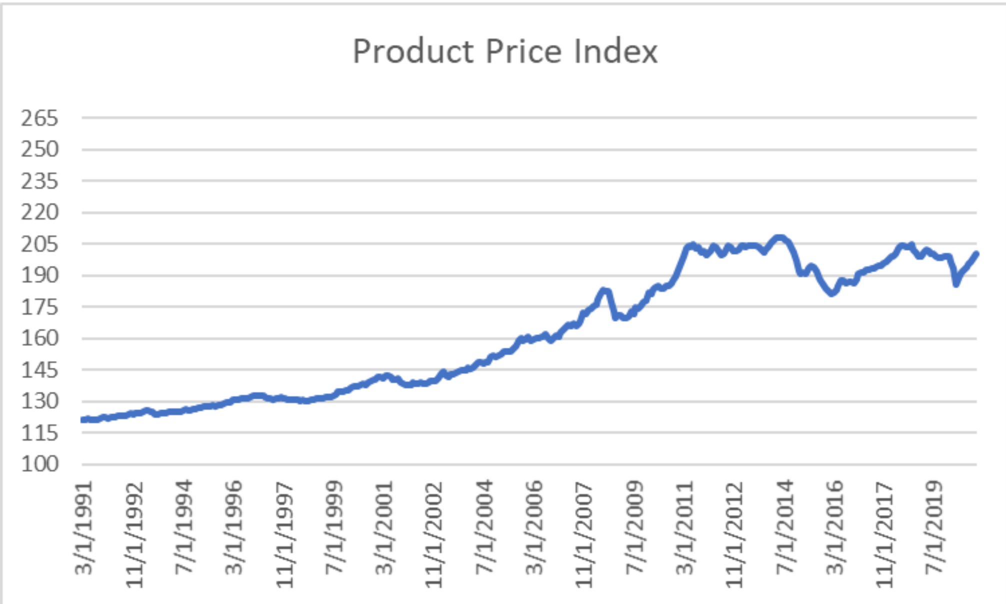


Figure 37: 30-year path of the Producer Price Index

4) Gross Domestic Product (GDP)

GDP represents the total value of domestic output for a given period of time and consists of purchases of domestically produced goods by individuals, firms, foreigners and public enterprises. The Bureau of Economic Analysis (BEA) of the U.S. Department of Commerce is responsible for compiling and publishing this measure on a quarterly basis (usually in the fourth week of the relevant month).

The data are available in nominal and real (deflated) value, as well as in index form. GDP is the most comprehensive indicator of economic activity. The individual aspects of gross domestic product, i.e. consumer spending, business investment and investment in real estate, are the foundation of the economy. The graph below reveals how investors typically react to some larger than expected GDP growth. Similar effects will occur in cases of falling or anaemic GDP growth.

GDP also determines the phases of a business cycle. The four main categories of Gross Domestic Product are:

- Personal Consumption Expenditure
- Investments
- Exports
- State

Each entrepreneur-investor must consider each individual category separately in order to be able to make rational decisions. Finally, GDP growth through investment is considered healthier than GDP growth through consumption as it stimulates productive capacity without stimulating strong inflationary pressures.

In figure 38, we present the evolution of the index over the last 30 years, which is the time period on which the survey is based. The Gross Domestic Product index shows the percentage difference in GDP compared to the previous year.

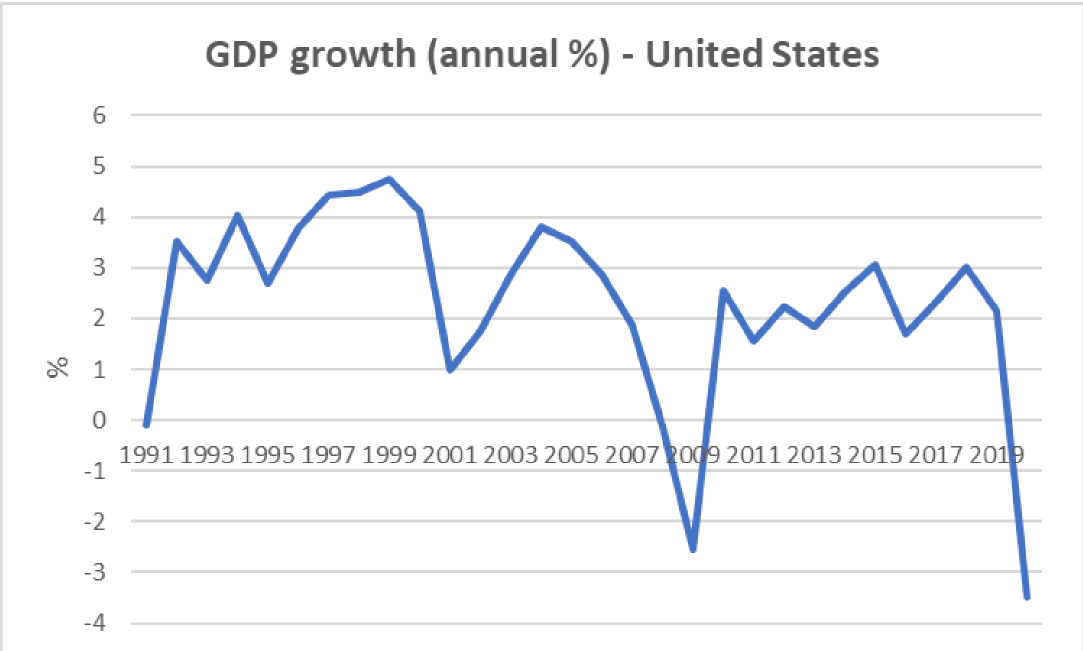


Figure 38: Thirty-year GDP trend

5) Retail Sales

Retail sales (SSSD, 2021) calculate the total receipts from shops selling durable and non-durable consumer goods. The Bureau of the Census of the U.S. Department of Commerce prepares the retail sales report on a monthly basis, with a one-month lag (usually the 2nd or 3rd week of the month). This index is calculated each time at the current dollar value without deflation. The importance of this indicator is particularly significant when one considers that retail sales account for almost half of consumer expenditure which in turn determines two-thirds of GDP. Retail sales are published at national and local level and are broken down into categories of product type on the basis of registered retail establishments. Therefore, the available retail sales data provide us with both the overall retail sales trend and the trends for the individual categories of traders, thus providing essential information on economic growth.

In figure 39, the development of the index over the last 30 years, which is the time period on which this research is based, is presented.

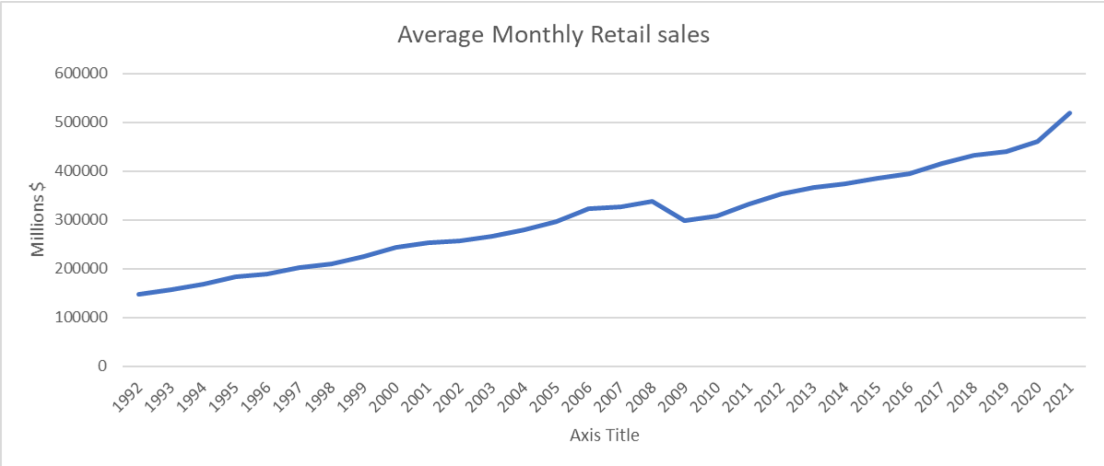


Figure 39: 30-year history of retail sales

6) Factory Orders - Durable Goods Orders

Durable goods orders represent new orders placed with domestic manufacturers for the immediate and future delivery of durable goods of all kinds. The data refer not only to consumer durables (cars, refrigerators, etc.) but also to capital durables (heavy machinery, trucks, electrical appliances, computers, etc.).

On the other hand, industrial orders represent a more comprehensive indicator as they reflect new orders for both durable and non-durable goods. Therefore, this survey focuses on industrial orders rather than on orders of durable goods.

The industrial orders index is a Leading Indicator of industrial production and capital investment and is essentially an indicator of the 'productive' sector of the economy.

Orders of goods are a sign of how busy factories will be in order to fulfil those orders at any given time. Useful information can also be obtained from unfilled orders, which essentially reveal the backlog of production to be produced immediately, shipments of goods attributable to sales, and inventories, which give a sense of the dynamics of current and future production.

In figure 40, we present the evolution of the index over the last 30 years, which is the period on which this research is based. The industrial orders index to be used reflects the percentage difference in relation to the previous month.

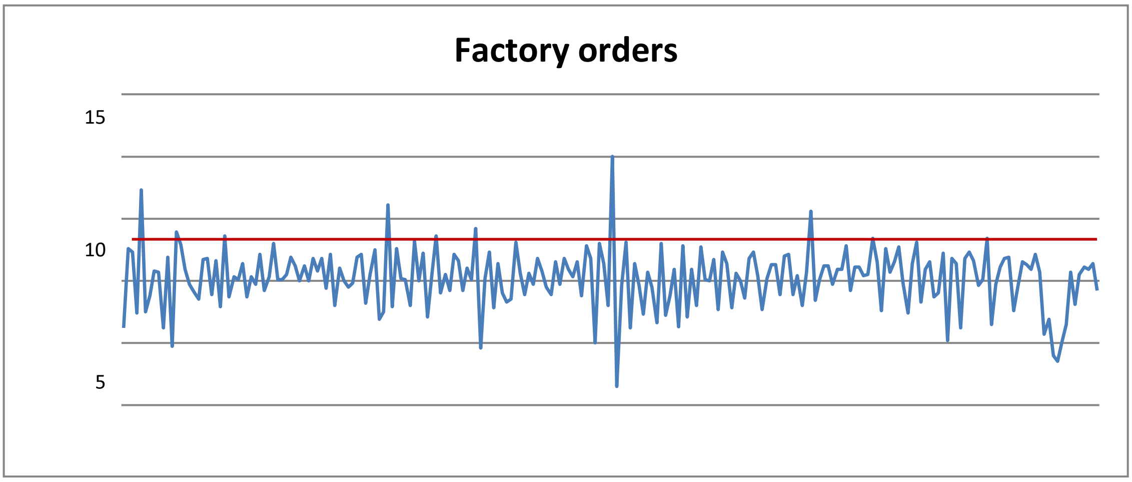


Figure 40: Thirty-year trend in industrial orders

7) ISM Manufacturing Index (ISM Mfg Index)

The ISM Mfg Index is a comprehensive composite index that encloses information on the whole economic activity. In order to create this index, the Institute for Supply Management monitors over 300 industries on a monthly basis in terms of employment, production, new orders, deliveries of supplies and stocks. Using the information collected, it compiles this index which is created so that when it is above or below 50% it indicates a booming or declining economy respectively. Although the Institute also calculates orders from abroad, orders for imports, back orders and cost prices of raw materials and semi-finished products, it does not count them in the general index.

In figure 41, the development of the index over the last 30 years, which is the time period on which the present investigation is based, is presented.

Figure 41: 30-year history of the ISM Manufacturing Composite Index

8) Personal Income

Personal income is the absolute dollar value of the total income received by individuals (wages, indirect income such as employer contributions to private pension plans, rent, dividends, interest, and benefits including unemployment benefits). Personal social security contributions are deducted from personal income. The Bureau of Economic Analysis of the U.S. Department of Commerce is responsible for compiling the personal income index on a monthly basis for the previous month.

Total personal income is, in other words, the total taxable income of each individual. If the tax due is deducted, the result is the disposable income that each individual can spend on consumption or savings. Further, if it is deflated, then the actual increase in income can be ascertained.

The following are the effects on the economy of increases in real personal income. Similar effects with opposite results are observed in the case of a decrease in real personal income (Figure 42).

Figure 42: Personal Income - Growth and Impact

In Figure 43, the evolution of the indicator over the last 30 years is illustrated, which is the time period on which this research is based. The personal income indicator to be used reflects the percentage difference from the value of the previous month.

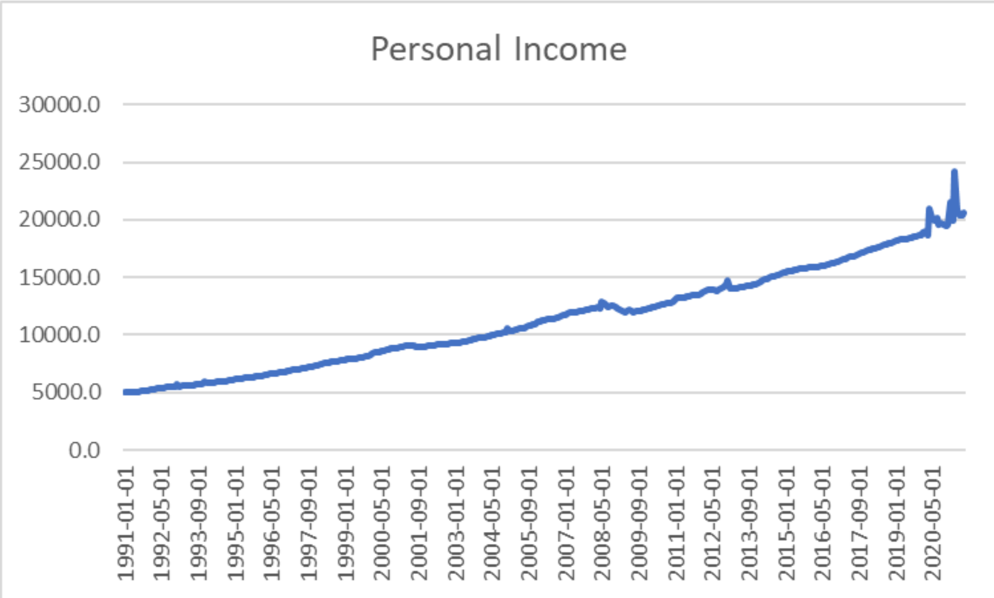


Figure 43: 30-year trend of the personal income index

9) Personal Consumption Expenditure (PCE)

Personal Consumption Expenditure is an indicator that captures the monthly expenditure of consumers, taking into account expenditure on durable goods, consumer goods, services and interest charges. Consumer expenditure consists of expenditure on durable goods, nondurable goods and services. These categories of expenditure are, by analogy, the monthly consumer expenditure included in the quarterly GDP. Therefore, knowing on a monthly basis a proportion of the data that determine the GDP, corresponding expectations for its evolution can be developed and corresponding actions can be taken.

Consumer spending accounts for two thirds of total economic activity and indirectly influences capital expenditure, imports, and investment in inventories. The following is an example of the impact on the economy of the increases in consumer spending.

Similar effects with opposite results can be observed in the case of a reduction in consumer spending, the prolonged duration of which leads to a recession (Figure 44).

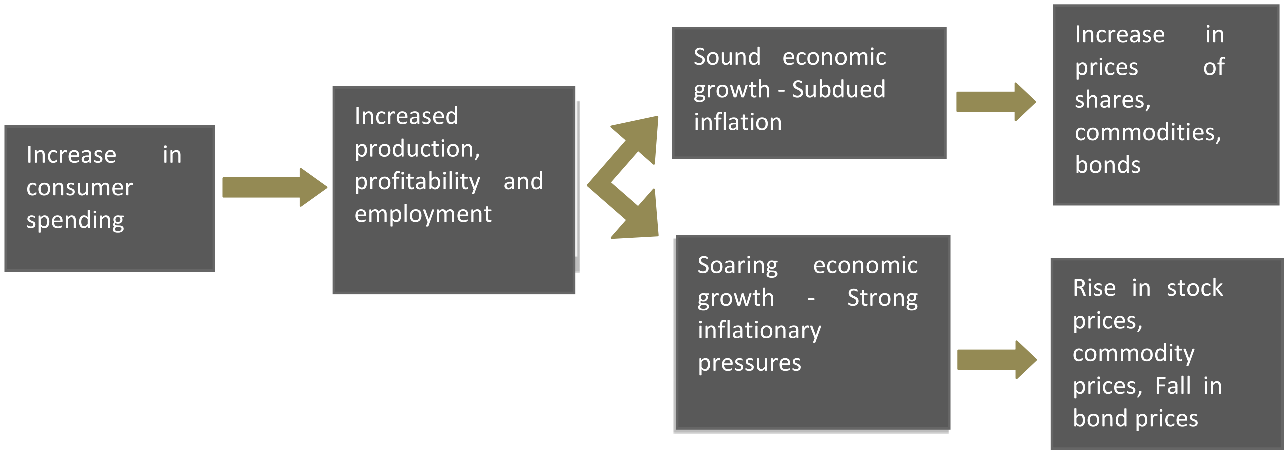


Figure 44: Consequences of an increase in Personal Consumption Expenditure

In Figure 45, we present the evolution of the personal consumption expenditure over the last 30 years, which is the time period on which this research is based.

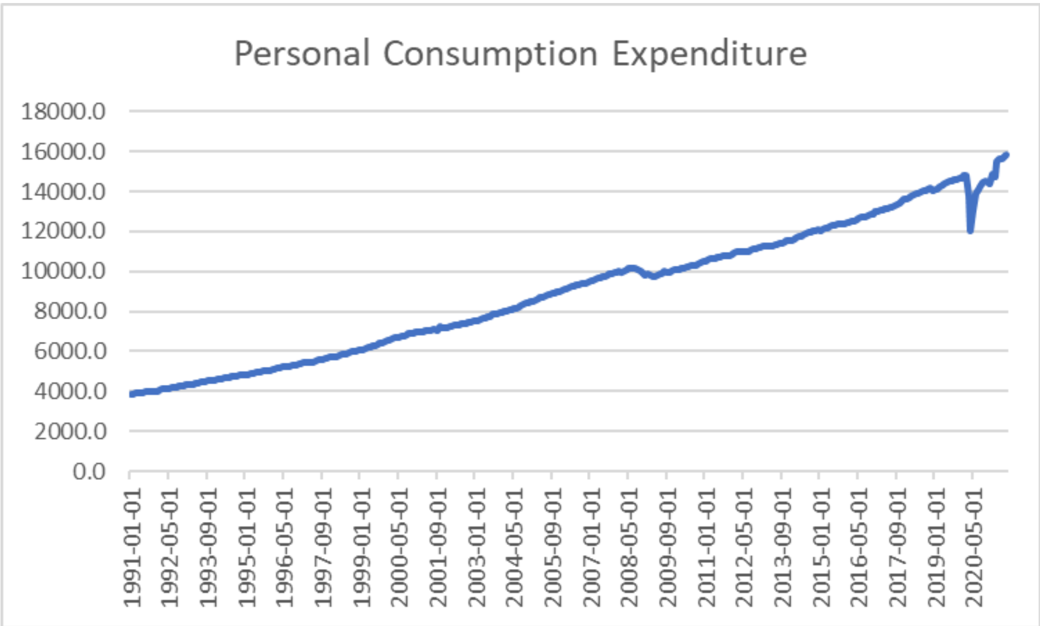


Figure 45: Personal Consumption Expenditure – 30-year evolution

10) Consumer Sentiment or Consumer Confidence

Consumer sentiment and consumer confidence are two indicators that measure consumer behaviour. The following are the effects on the economy of an increase in consumer confidence. Similar effects with opposite results are observed when consumer confidence is weakened (Figure 46).

Figure 46: Consumer Confidence - Strengthening and Impact

These two indicators are inextricably linked to the ability and willingness to spend. However, their paths are not identical as propensity to consume precedes consumption expenditure.

Consumer Sentiment

The University of Michigan surveys five hundred households on a monthly basis regarding their financial situation and their attitude towards the market and the economy in general.

Consumer Confidence

The global non-profit business organisation "Conference board" produces on a monthly basis a survey and a related index of consumer attitudes towards current conditions and expectations for the future.

In Figure 47, we present the evolution of the consumer confidence over the last 30 years, which is the time period on which this research is based.

Figure 47: Consumer Confidence Index - 30-year evolution

11) Consumer Credit

Consumer credit refers to the total amount of existing consumer credit in the US and is valued in dollars. Fluctuations in the consumer credit balance the corresponding trend in consumer spending. The following are the effects on the economy of the strengthening of consumer credit.

An increase in consumer credit has a positive impact according to the above graph for the whole economy, provided that it does not exceed a certain threshold. According to banking practice, this threshold is set at 40% of total personal income of consumers. Borrowing beyond consumers' means puts a strain on future consumption as most of the income is devoted to repaying existing debts. It was overborrowing that led to the recent financial crisis, which started with mortgage credit and spread to all sectors of the economy.

This indicator, although published two months behind schedule, provides useful information on the potential evolution of consumer spending and on the willingness of banks and monetary authorities to grant loans.

In figure 48, the evolution of the consumer credit over the last 30 years is evident.

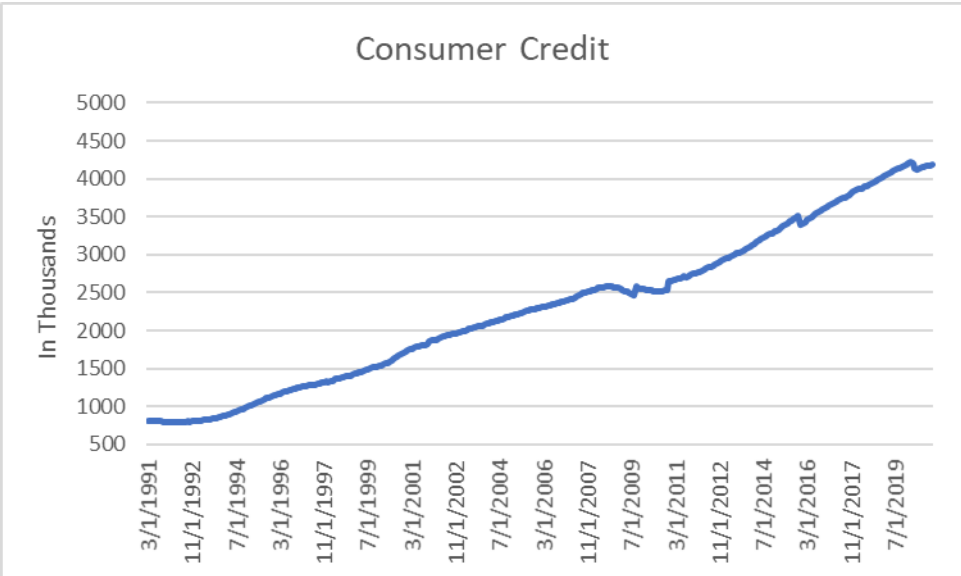


Figure 48: Consumer Credit - 30-year evolution

12) Standard & Poor's 500 stock index (S&P 500)

The S&P 500 stock market index (Finance.yahoo.com, 2021) is a market capitalisation-weighted floating index that reflects the prices of the common shares of the 500 largest companies (large-cap) traded in the United States (S&P 500®; S&P Dow Jones

Indices). The stocks included in the S&P 500 are traded on the two largest U.S. stock markets and are companies in the energy, materials, industrials, health care, manufacturing, materials, financials, technology, telecommunications and services sectors.

The S&P 500 is considered the mirror of the US economy and is a leading indicator of economic activity. At the same time, according to other views, stock prices are not just an important leading indicator but a major component of economic activity (Greenspan, 2009). Moreover, if we consider that the US market capitalisation represents 42% of the global market capitalisation then we can argue that the S&P 500 is an index in which global economic prosperity is reflected.

Figure 49 showcases the 30-year history of this index, in which it is possible to distinguish how the crises affect the capitalisation of companies each time.

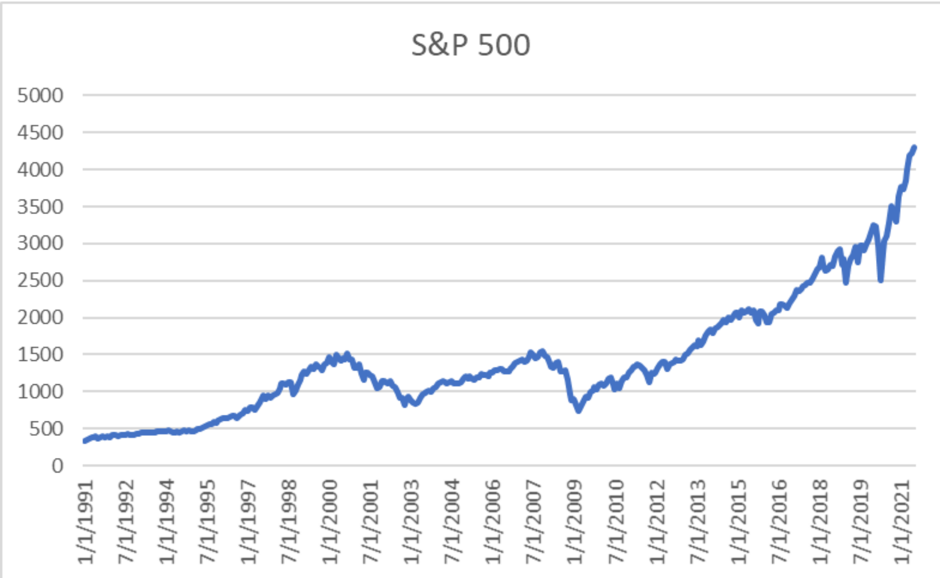


Figure 49: 30-year evolution of the S&P 500 index

13) MSCI World Stock Index

In this research, the main US economic indices were used, since, as mentioned above, the US economy is the dominant economy and shapes the general trends around the world. However, country-specific stock market indices may be subject to increased volatility due to not only economic but also socio-political triggers stemming from the country's stock exchange.

Therefore, in addition to the S&P 500 index, the use of the MSCI World Index was chosen, which is a global stock market index that tracks the performance of stocks for the financial markets of 24 developed countries around the world: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Hong Kong, Ireland, Israel, Israel, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Singapore, Spain, Sweden, Switzerland, United Kingdom, United States and Spain (www.mscibarra.com-MSCIM, 2021).

The 30-year evolution of the MSCI World Stock Index can be seen in Figure 50.

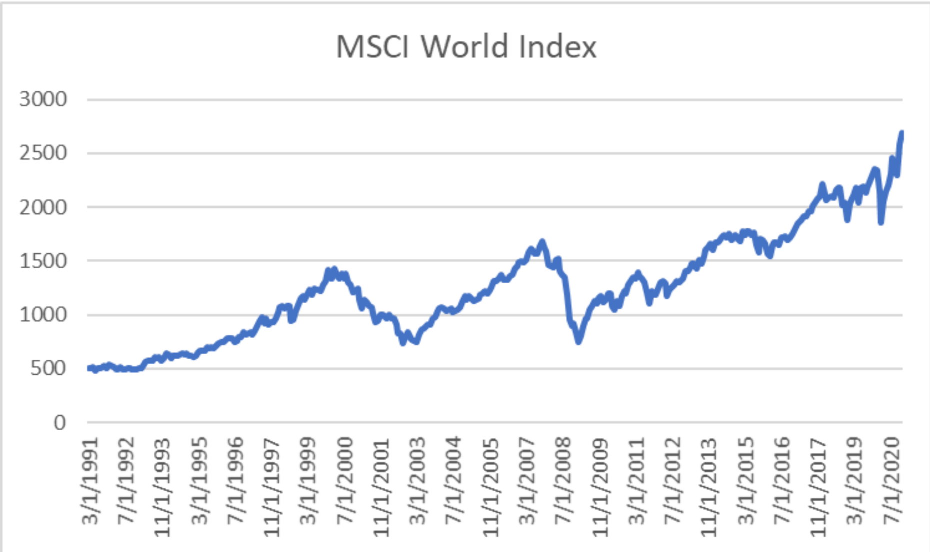


Figure 50: 30-year evolution - MSCI World Stock Index

14) Spot price of gold per ounce

The price of gold per ounce is an important economic indicator as it reveals the level of confidence of investors in the fiat currencies and the economies in general. The role of gold has always been important and has gone through various stages before it reached the point of becoming a separate economic indicator.

Figure 51 depicts the 30-year trend of the price of gold per ounce, up to August 2021, which shows how crises have each time affected the treatment of gold as a 'safe haven', resulting in an increase in demand and hence in its price. In the 2008 crisis in particular, the increase in the price of gold was sharp and sudden, reaching as high as USD 1,400 today.

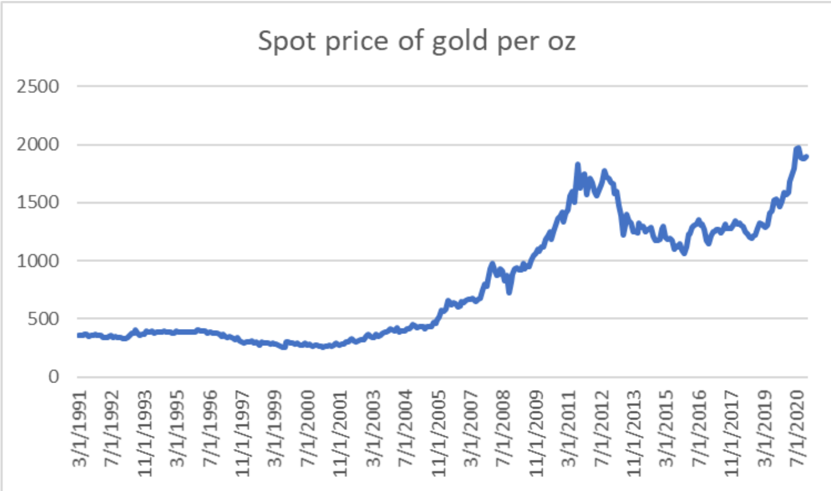


Figure 51: 30-year evolution – Spot price of gold per ounce

15) Spot price of Copper USD / tonne

Copper is considered a key leading indicator that reveals the future growth path of an economy (Rao, 2016). This feature of copper comes from its actual use as a raw material in the production process for the manufacture of electrical wires, piping and electronics. The consequence of this is that any increase in demand of copper indicates that a recovery in manufacturing, construction and the economy as a whole is coming (Sinn, 2010).

It should be noted here that a disadvantage of this key leading indicator is that a possible over-stocking of copper will correspond an increase in its price which will not be linked to any increase in demand, thus this might be misleading for the real economy who only observes the price of copper (Downey, 2009). However, in this research this indicator is used in conjunction with other key leading indicators, eliminating such risks that could distort the results.

The 30-year evolution of the spot price of copper is illustrated in Figure 52.

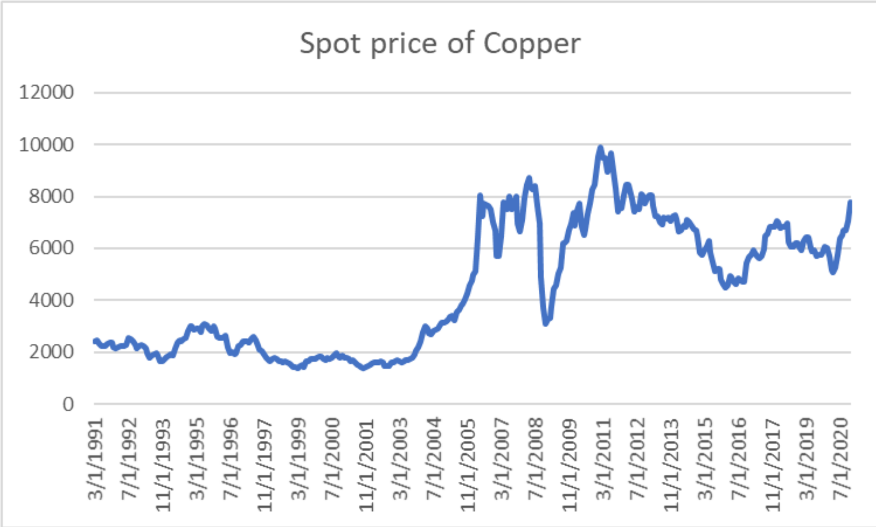


Figure 52: 30-year evolution – Spot price of copper per ounce

16) Weekly U.S. Ending Stocks of Crude Oil and Petroleum Products (Thousand Barrels)

In this survey it was chosen to use stocks instead of the current oil price as they reflect the actual situation of oil demand. The use of the current oil price (e.g. the current price of West Texas Intermediate light) would compromise the objectivity of the research as the price of this commodity is often influenced by geopolitical reasons in addition to the laws of supply and demand. This means that increases in the price of oil do not necessarily signal economic growth conditions. In other words, in the past, oil prices have risen as a result of unrest or wars in oil-producing countries or as a result of a sustained build-up of reserves. At the same time, the price of oil may be kept low by the policies of OPEC (Organization of the Petroleum Exporting Countries) or due to increased production, so it is evident that the price of oil in many cases is influenced by non-economic factors.

The level of ending stocks reflects the actual situation of demand for crude oil. In periods of high economic growth, demand strengthens and prices rise if stocks cannot meet demand. This pattern is more important for our model than the price of crude oil, as crude oil prices are often influenced by geopolitical factors rather than the fundamental economic principle of supply and demand that we are concerned with.

Figure 53 shows the 30-year trend in US oil and derivatives stocks, including Special Purpose Reserves (SPRs), which are provided by the Energy Information Administration (EIA).

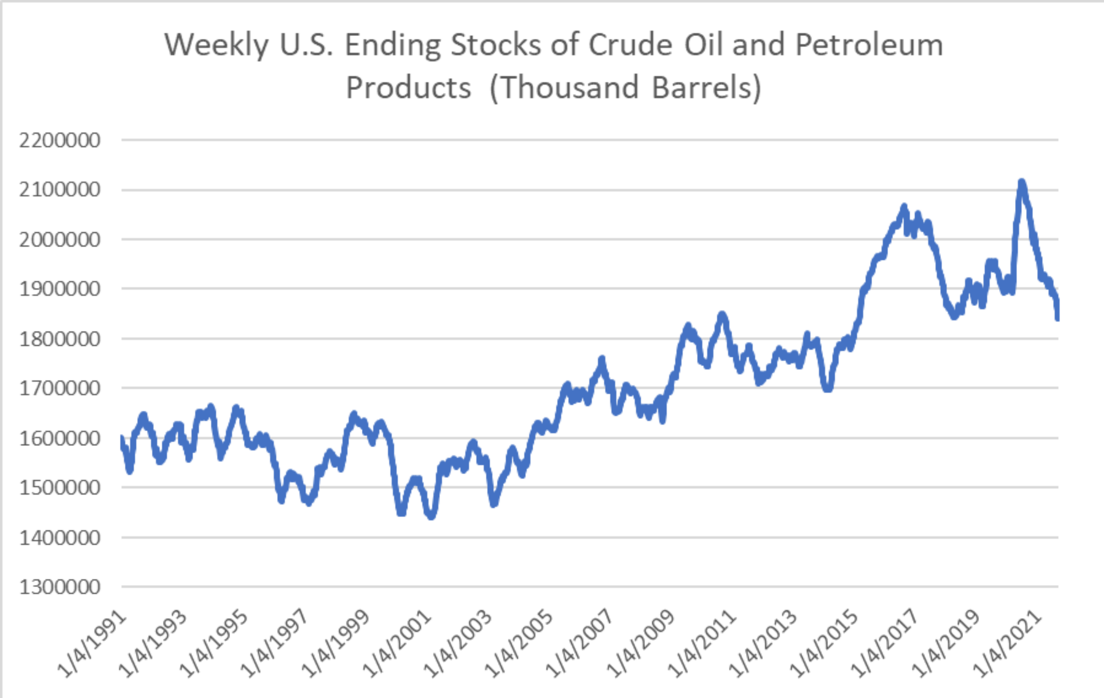


Figure 53: 30-year evolution – Weekly U.S. Ending oil stocks (in thousand barrels)

17) World steel production in thousand metric tons

Steel is considered one of the main raw materials used in industry, playing an essential role in world economic activity. It has been observed that during periods of economic growth, the demand for steel and therefore its production increases. Similarly, in periods of economic downturn, demand for steel falls and consequently, its production decreases. In addition, this variable is of particular importance for our research as the most common way of transporting steel is by bulk carriers. It therefore plays an essential role in the demand for capacity expressed in terms of bulk carriers, influencing freight rates and thus the BDI.

Figure 54 shows the evolution of world steel production over the last 30 years. Analysing the evolution of the development of steel production in combination with the prevailing economic circumstances, it can be observed that the greatest impact on steel production took place during the global economic crisis that started in October 2007.

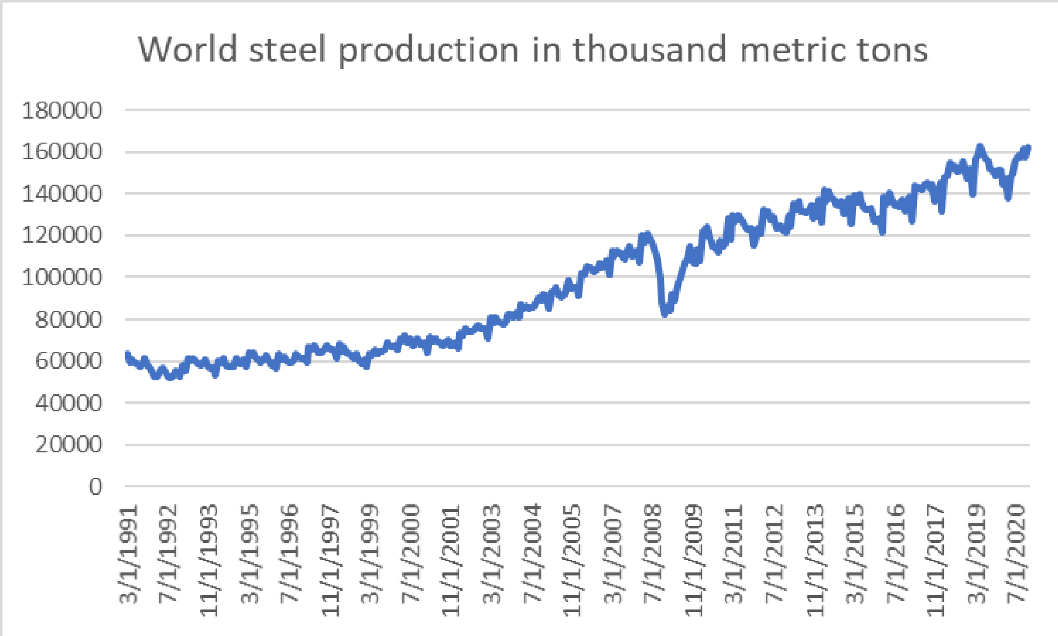


Figure 54: 30-year evolution – World steel production (in thousand tonnes)

18) Volatility Index (VIX)

The volatility index (CBOE Volatility Index, 2021) was created in 1993, when Professor Robert E. Whaley of Duke University first presented it in his scientific paper. On 26 March 2004, the first contracts relating to the index were negotiated on the Chicago Board Options Exchange (CBOE) (Vix-Index-Chicago Board Options Exchange). Since then, various derivative products have been created to track the VIX.

The VIX is considered the most famous measure of market volatility and is also known as the 'fear index'. The value of the VIX moves upwards in cases of high market volatility and also in cases of large stock market falls.

Figure 55 shows the 30-year history of this index. The index fluctuates sharply, with an increase in its price during periods of stock market crises. Particularly interesting is the reaction of the index to the onset of the global financial crisis of 2008 and to the covid-19 pandemic, when it jumped very sharply to a value around 60 points. This value is particularly large given that under normal circumstances the index value ranges between 10 and 20 points.

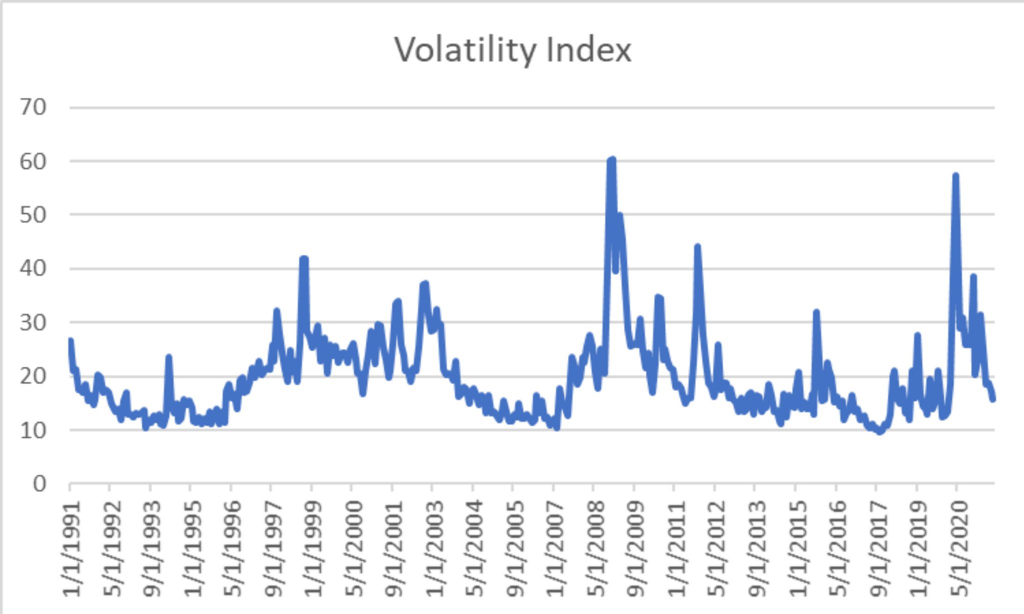


Figure 55: 30-year evolution – VIX volatility index

19) Exchange rate of Special Drawing Rights (SDRs) against the dollar

This currency was created by the International Monetary Fund to reduce pressure on gold and the US dollar in international transactions. It was established in the late 1960s. In addition, it was created to enhance the stability of the official reserves of the IMF member countries and to support the system of fixed exchange rates (Vamboukas, 1999; Ferguson, 2008; International Monetary Fund).

The value of SDRs (Special Drawing Rights) is based on the four major currencies (US dollar, British pound, Japanese yen and Euro) which are considered global currencies because of the size of the economies they represent.

From the above descriptions it can be seen that the exchange rate between the SDR and the US dollar essentially reflects the strength of the dollar and the US economy in general relative to the rest of the economies. The graph below illustrates the 30-year path of the SDR exchange rate against the US dollar.

Figure 56 shows that in times of economic crises, the dollar strengthens as investors, at least until now, have viewed the US currency and the US economy as a "safe haven".

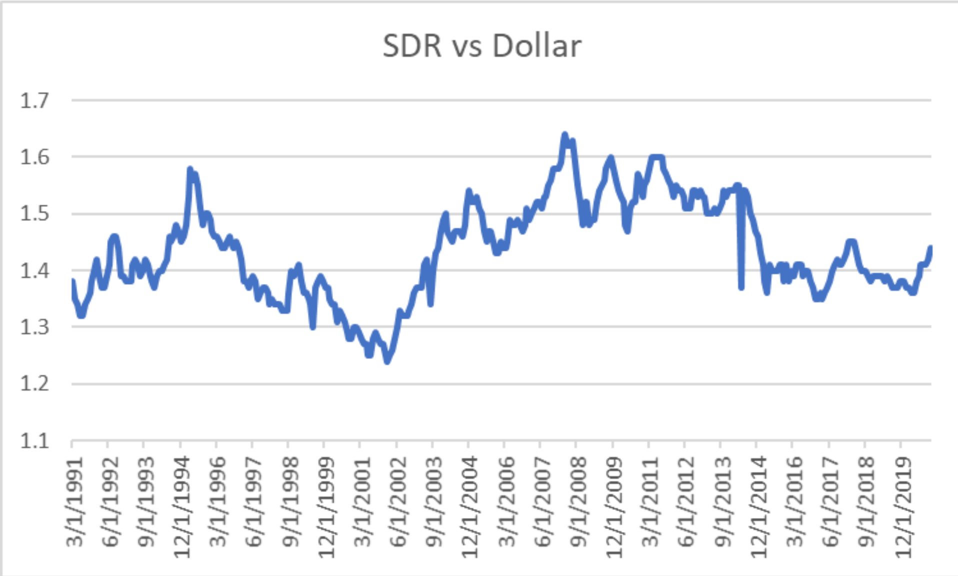


Figure 56: 30-year evolution – Exchange rate of Special Drawing Rights against the US dollar

20) Baltic Dry Index

The Bulk Dry Index (BDI) is the variable to be predicted in the present research. It is a leading indicator of economic activity (the successor to Biffex) and is compiled on a daily basis by the Baltic Exchange, which is based in London. It represents the average cost of chartering dry cargo vessels for 20 different routes covering the Handymax, Panamax and Capesize dry cargo vessel categories. In times of economic growth, there is an increase in demand for the transport of commodities such as iron, coal, cement, fertilisers, grain, steel etc. leading to an increase in the price of the BDI. Similarly, in times of economic contraction, the demand for commodities and hence for their transportation decreases, resulting in a decrease in the price of the BDI.

However, in addition to the course of world economic activity, which plays an essential role in the formation of the index, there are several exogenous factors that may occasionally affect the price of the BDI, namely:

- Natural causes (e.g. crop disasters due to natural causes may lead to a fall in the index without being due to economic causes)
- Political causes (e.g. policies to restrict exports or imports by imposing tariffs may also affect the index in one direction or the other)
- Supply of ships (e.g. a large supply of ships, which may result from an increase in orders for new ships as a result of an improving economy, may lead to a fall in the index without there being any particular problem in the general economic activity)

The evolution of the Baltic Dry Index the last 30 years is illustrated in Figure 57.

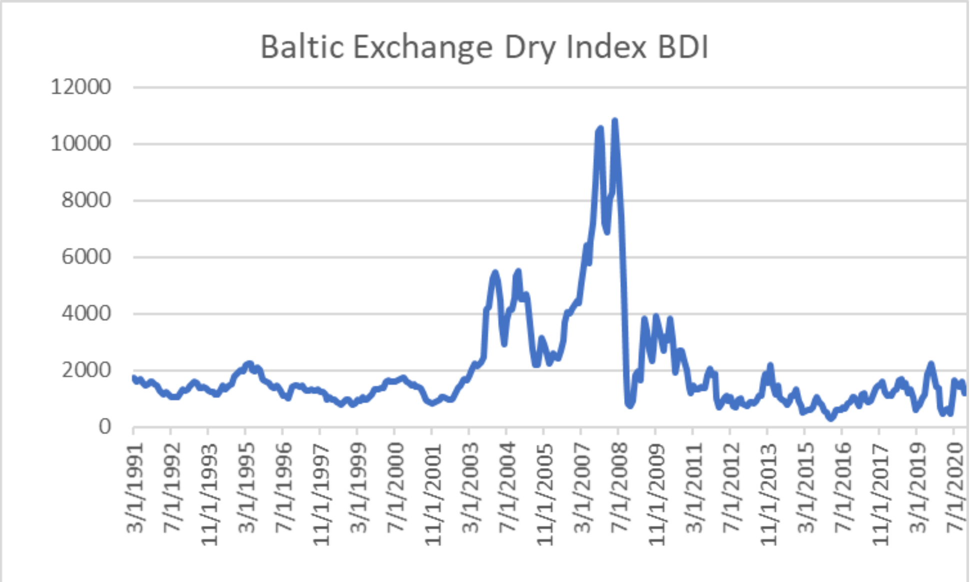


Figure 57: 30-year evolution of the Bulk Dry Index - BDI

The **Shipping time series** selected as possible factors influencing the BDI are the following:

1. Fleet development in million dwt for Panamax, Capesize, Handymax, Handysize dry cargo vessels.

2. Fleet removals, demolitions and losses in mil dwt for Panamax, Capesize, Handymax, Handysize dry-cargo carriers.

3. Order book of bulk fleets for Panamax, Capesize, Handymax and Handysize dry cargo vessels.

4. Newbuilding Prices (Newbuilding Prices) for Panamax, Capesize, Capesize, Handymax, Handysize dry-cargo vessels in millions of dollars.

5. 5-year Second- hand Vessels prices in $ million for Panamax, Capesize, Handymax, Handysize dry-cargo vessels.

6. Scrap Value in million $ for Panamax, Capesize, Handymax, Handysize dry-cargo vessels.

7. Voyage Rates in $ per day for Panamax, Capesize, Handymax, Handysize dry cargo vessels.

8. Time charter rates in $ per day for Panamax, Capesize, Handymax, Handysize dry-cargo vessels.

9. Bulkcarrier Newbuilding Price Index Year/year change (%yr/yr)

In an effort to create a manageable, comprehensive and efficient set of data for the input level of the neural networks, three indices were selected:

1) The first index is the **Panamax scrap value**, whose evolution over the last 30 years is illustrated in figure 58.

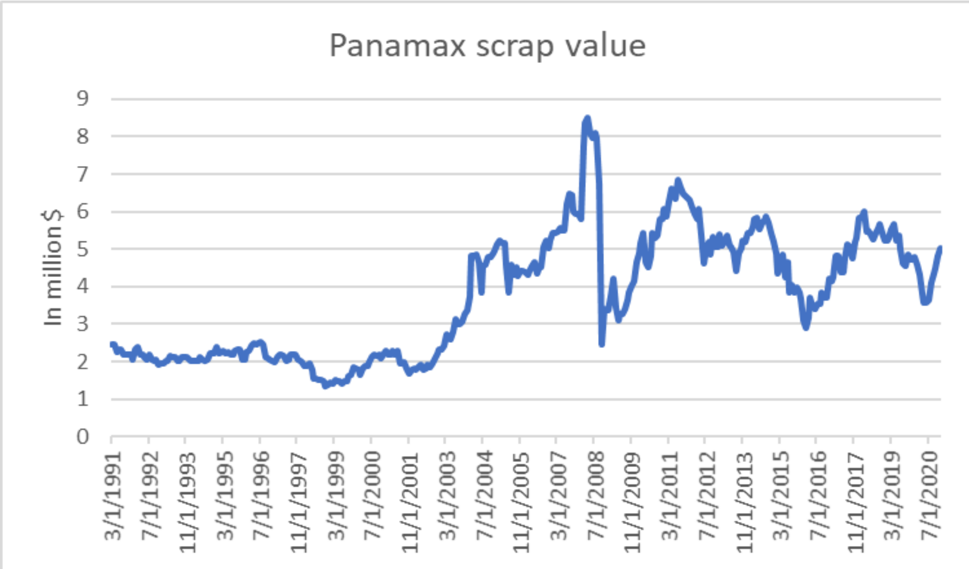


Figure 58: 30-year evolution of the Panamax scrap value

2) The second index is the **6-month time charter rate for 32,000 ton dwt ship**, whose evolution over the last 30 years is showcased in figure 59.

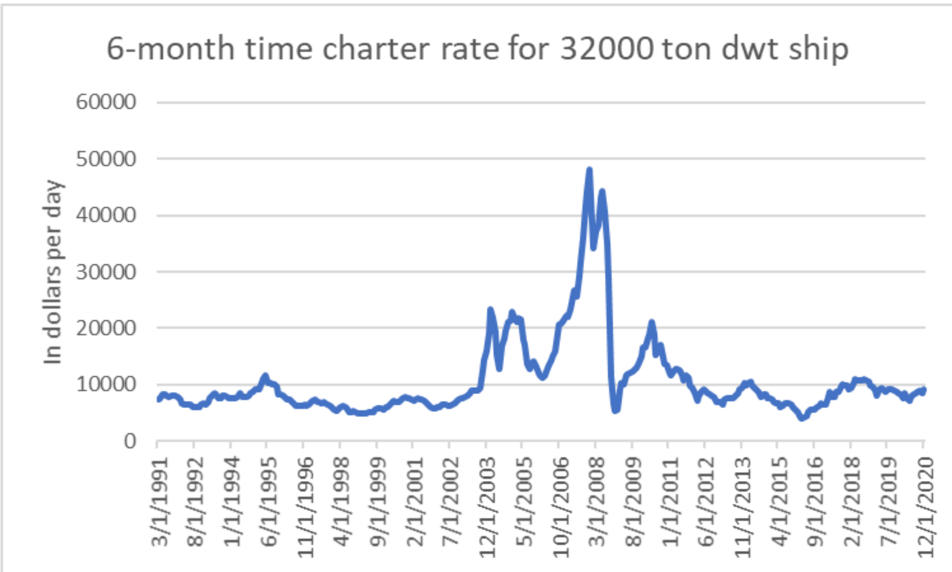


Figure 59: 30-year evolution of the 6-month time charter rate for 32000 ton dwt ship

3) The third index is the **Bulkcarrier Newbuilding Price Index Year/year change (%yr/yr)** and we can see its evolution during the last 30 years in figure 60.

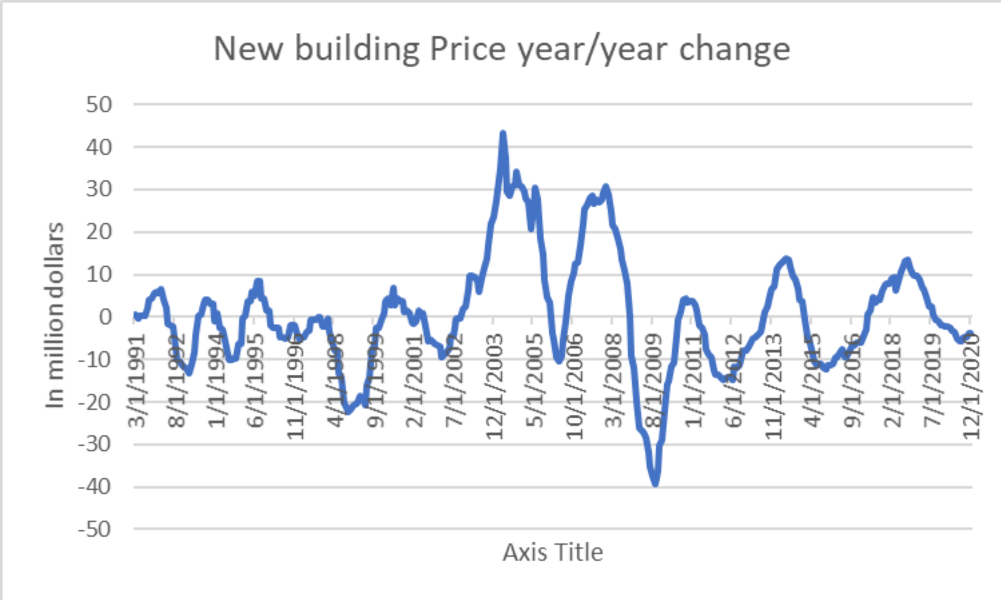


Figure 60: 30-year evolution of the Newbuilding Price Index Year/year change

In order to proceed with the research, the time series were processed to provide monthly data from March 1991 to August 2021 for each of the individual factors outlined above. The different units of measurement did not pose a problem, as neural networks have the ability to isolate the influence of each factor on the evolution of the dry bulk shipping index (BDI).

4.3 Feature Selection

Feature selection is one of the most important stages of data preprocessing as it can provide the following important benefits when developing the machine learning model:

- Increasing the performance of the model due to a reduction in its complexity
- Reduction of the model training time
- Reduction of the overfitting

In the context of this work, several feature selection techniques were used in order to find the optimal dataset. The first one is to eliminate unnecessary features while the following one will highlight those that are likely to significantly affect the final prediction. Finally, these features will also be tested under normal forecasting conditions to see if they do indeed have a positive contribution as the "Feature importance" metrics will relatively measure the influence over the final forecast.

4.3.1 Feature correlation

It is well known that two variables x, y are linearly correlated if they are linked by a relationship of the form:

$$y = ax + b$$

where a,b are constant factors. The existence of linearly correlated features in a dataset does not contribute positively to the performance of the algorithm, often leading to worse performance. For this reason, the degree of correlation of the features was calculated through the Pearson correlation coefficient and more specifically through the formula:

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{(\sum X^2 - \frac{(\sum X)^2}{N}) * (\sum Y^2 - \frac{(\sum Y)^2}{N})}}$$

The Pearson correlation of all the factors presented in the chapter 4.2 is illustrated in Figure 61.

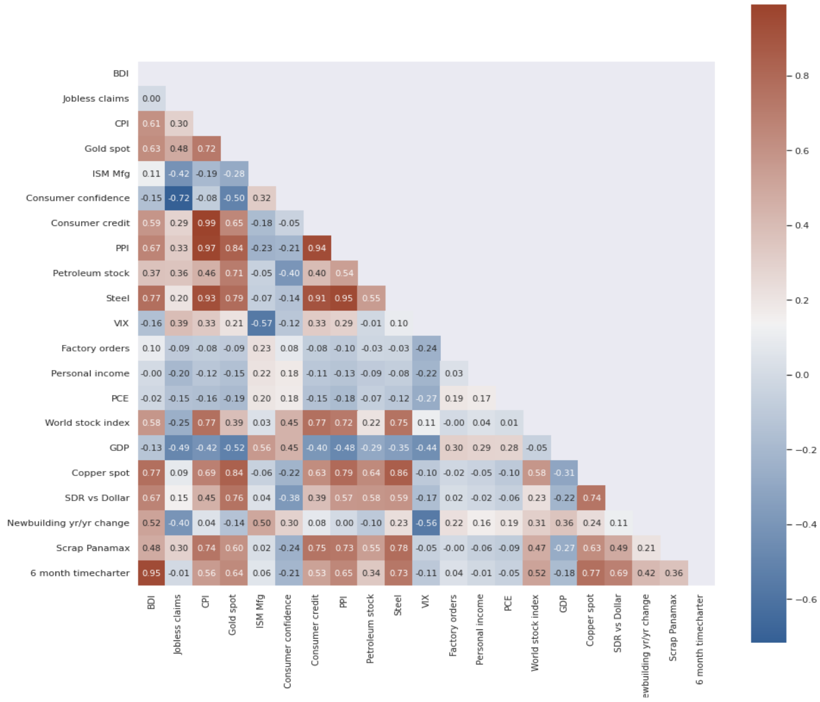


Figure 61: Pearson correlation of all features

The values of the correlation coefficient are between -1 for a perfect negative correlation, 0 for fully independent variables and +1 for a perfect positive correlation between variables. The variables selected for inclusion in the neural network were those with a greater than 0.55 correlation with the BDI, which will be forecasted. In addition, the factors selected by the above procedure should not have a high correlation between them (close to +1), as in this case it means that they will provide the same information to the neural network (Watsham & Parramore, 1997).

The factors that were selected are the following, with their correlations being illustrated in figure 62.

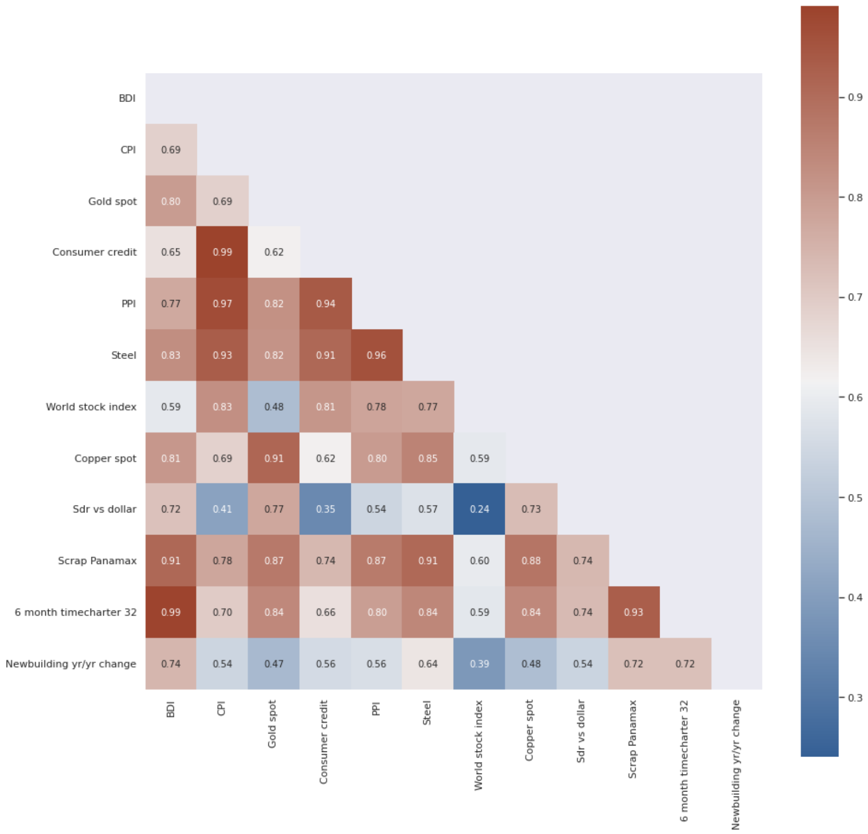


Figure 62: Pearson correlation of selected features

The factors that show a high correlation between them (close to +1) are:

- BDI and the 6-month time charter rate for 32,000 ton dwt ship
- Spot prices of gold and copper
- Producer Price Index and World Steel Production
- Consumer Price Index, Consumer Credit, Producer Price Index, World Steel Production

However, it is chosen to use the set of variables that showed a correlation with the variable to be predicted greater than 0.55, as neural networks have the ability to identify variables with high correlation and discard them when they are deemed to provide the same information.

Therefore, the following 12 time series will be used to forecast the Blatic Dy Index:

BDI, Consumer Price Index, Spot price of Gold per Ounce, Consumer Credit, Producer Price Index, World Steel Production, World stock market index, Spot price of Copper, Special drawing rights exchange rate against the dollar, Panamax scrap value, 6-month time charter rate for 32,000 ton dwt ship, Bulk carrier Newbuilding Price Index Year/year change (%yr/yr)

4.3.2 Feature Importance

Algorithms that are based on decision trees allow the calculation of an importance factor for each feature of the dataset on which they are trained. This factor relates to how important each feature was in the formation of the decision trees used by the above algorithms and can provide important information about the influence a feature may have on the final forecast. The more the feature was used in important decisions of the algorithms, the higher the coefficient.

The importance coefficients were calculated using the XGBoost algorithm, as it is one of the most efficient and widely used decision tree algorithms. The calculation was performed at the level of monthly forecast of the BDI value and the results are presented in Figure 63.

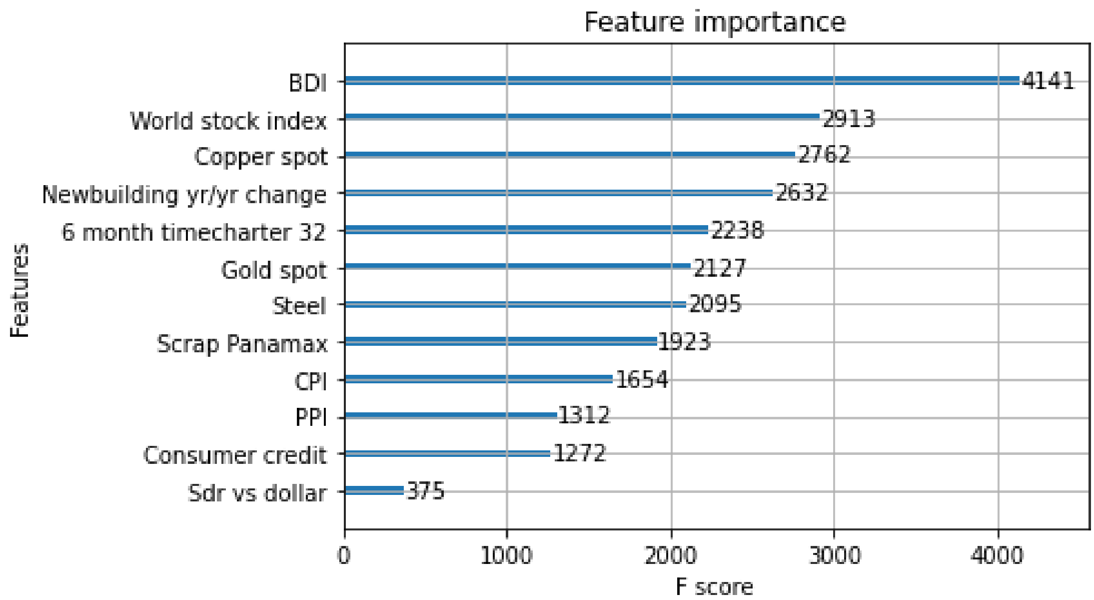


Figure 63: Feature importance diagram

Chapter 5

Data pre-processing and the development of the machine learning algorithms

In this chapter we will discuss the different methods of preprocessing the dataset that was formed in the previous chapter, in order to improve the performance of the machine learning algorithms that will be applied.

Afterwards, the algorithms of LSTM, GRU, BiLSTM and Feedforward neural networks will be developed, as well as the ARIMA statistical model. Finally, the prediction results of the five different algorithms for the regression problem will be presented and compared.

5.1 Data preprocessing

One of the most important factors that affect the final performance of machine learning algorithms is the proper preprocessing of the data they use, as their performance depends significantly on the quality of the dataset. In the previous chapter, preprocessing techniques were developed to remove redundant or even incorrect information from the data. In this one we will focus on transforming the data into a format that can be optimally exploited by machine learning algorithms. The following data preprocessing steps will be performed:

- Separation of dataset into train, validation and test set
- Feature scaling
- Determination of the past timesteps
- Data transformation into a consistent form for the machine learning algorithms

5.1.1 Separation of dataset into train, validation and test set

80% of the data (282 records) was used as the train set and the remaining data was split into the validation 10% (36 records) and test 10% (35 records) sets. A point worth emphasizing is that the splitting was done keeping the time order of the data which is necessary due to the nature of the problem. More specifically, the first 282 records constituted the train set, the next 36 the validation set and the last 35 the test set.

5.1.2 Feature scaling

Feature scaling is one of the necessary steps during preprocessing as it leads to an increased convergence speed of the algorithms. Moreover, in cases where different features range over very different value ranges a large number of machine learning algorithms (especially those that compute distances between features) do not produce correct results as they assume that variables that have a larger value range have more power over the prediction, which is untrue.

Two of the most well-known normalization methods for the feature scaling are the Min-Max scaler and Standard scaler, also mentioned in chapter 3. The first method transforms the data in the interval [0, 1] while the second transforms the distribution of each attribute in such a way that it has zero mean and variance equal to 1.

- Min-Max scaler: $x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}}$

- Standard scaler: $x_{new} = \frac{x - \mu}{\sigma}$

Both normalization methods were tested and it was observed that the first one showed slightly better results, thus it was chosen.

5.1.3 Determination of past timesteps

One of the most important features of the LSTM, BiLSTM and GRU recurrent neural networks is their ability to discover both long and short term dependencies in the data, which helps to improve the performance of their future predictions. It is therefore particularly important to define the time window that these networks use to optimize their forecasts.

To find the appropriate number of timesteps to use, the autocorrelation of BDI values was studied, i.e. the extent to which the value at a given time is related to values at past times. The autocorrelation plot function of the pandas library was used to find the autocorrelation and the result is shown in Figure 64.

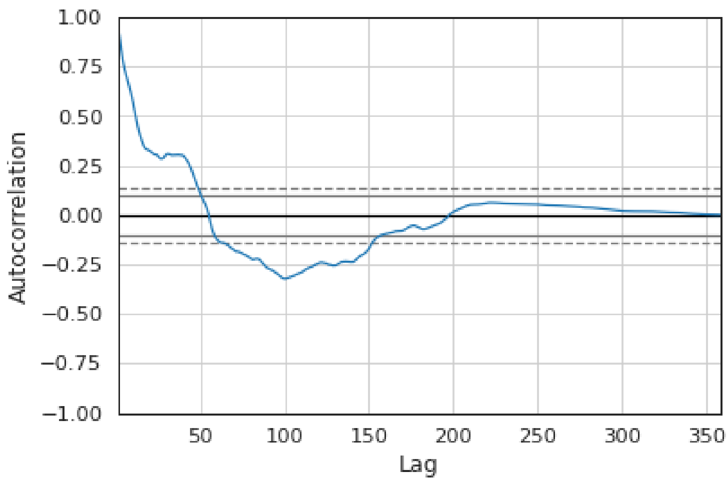


Figure 64: Autocorrelation plot of BDI

The statistically significant past values are those whose autocorrelation coefficient is above the horizontal dashed line. In this particular case we see that about 1-10 months ago the BDI values are statistically significant. For this purpose, all values from 1 to 10 were tested as timesteps and showed particularly close results, with the value of 3 having the best results. For this reason, the timestep value was set to 3.

5.1.4 Transforming the dataset into its final form

Once the necessary past time windows were determined, the data was transformed to its final form to conform to the algorithm templates. These networks require the data to be in the three-dimensional format of Figure 65. Thus, we came up with the following datasets for the monthly forecasts:

- X_train (282, 5, 11) y_train (282, 1)
- X_validation (36, 5, 11) y_validation (36, 1)
- X_test (35, 5, 11) y_test (35, 1)

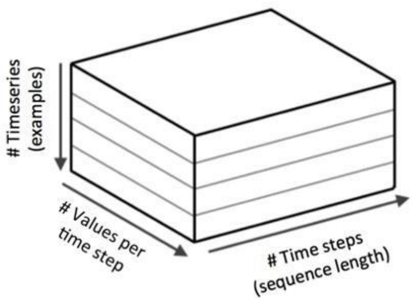


Figure 65: The final form of the input dataset

5.2 Development of the Recurrent Neural Networks (RNN)

In this section we will present the development of the 3 types of Recurrent Neural Networks, the LSTM, BiLSTM and GRU neural networks, that were developed with the use of the tensorflow.keras libraries. Furthermore, the results of their predictions for the regression problem will be presented, as well as the accuracy metrics for each algorithm.

Forecasting the next month's price for a time series is a particularly important problem that has been extensively studied in the literature as this knowledge can provide significant benefits both at economical level (e.g. for stock value forecasting) and in decision making, by organising optimally future actions.

5.2.1 LSTM neural networks (Long Short Term Memory)

Apart from the correct pre-processing of the data, the performance of neural networks and hence of LSTMs depends on their hyperparameters. To develop the optimal model, the hyperparameters that are stated in Table 2 were taken into consideration.

• number of LSTM layers	• number of hidden layers
• number of LSTM units per layer	• number of epochs
• batch size	• optimizer
• dropout per layer	

Table 2: Important hyperparameters for Recurrent Neural Networks

The mean squared error (mse) was used as loss function and tanh and relu as activation functions for the LSTM and Dense layers respectively. Finding the appropriate set of hyperparameters was done by running tests for each hyperparameter and also by the grid search method, calculating the grid performance over the validation set. The optimal hyperparameters set is shown in Table 3.

Hyperparameters	Potential	Optimal
LSTM layers	[1, 2]	2 layers
Dense layers	[1, 2]	1 layer with 1 unit
LSTM units per layer	[16, 32, 64, 128, 256]	256, 128 respectively
Dropout per layer	[0.1, 0.2, 0.3]	0.2 in all layers
Epochs	[20, 30, 50, 80, 100]	50
Batch size	[4, 8, 16, 32]	4
Optimizer	[“rmsprop”, “adam”]	rmsprop

Table 3: Optimal hyperparameters for LSTM neural network

Figure 66 depicts the architecture of the developed LSTM network. The same logic is followed in the architecture of all the RNNs, taking into account the peculiarities of each one.

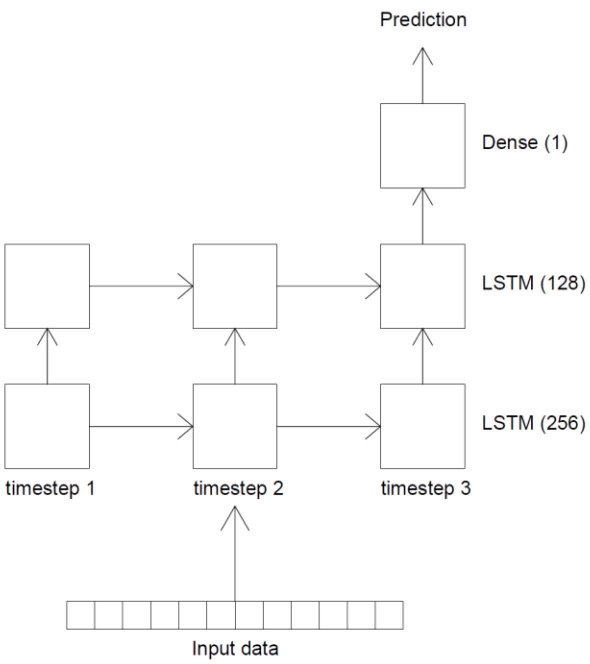


Figure 66: Architecture of LSTM Neural network

Figure 67 displays part of the python code, where the function for creating the LSTM neural network and the hyperparameters are defined.

```
DROPOUT = 0.2
# 20% Dropout is used to control over-fitting during training
WINDOW_SIZE = SEQ_LEN - 1
# Create GRU model
def create_lstm(units):
    model = Sequential()
    # Input layer
    model.add(LSTM (units = units, return_sequences = True,
    input_shape = [X_train.shape[1], X_train.shape[2]]))
    model.add(Dropout(0.2))
    # Hidden layer
    model.add(LSTM(units = 128))
    model.add(Dropout(0.2))
    model.add(Dense(units = 1))
    #Compile model
    model.compile(optimizer='rmsprop',loss='mse')
    return model
model = create_lstm(256)
```

```
def fit_model(model):
    early_stop = keras.callbacks.EarlyStopping(monitor = 'val_loss',
                                               patience = 10)
    history = model.fit(X_train, y_train, epochs = 50,
                        validation_split = 0.1,
                        batch_size = 4, shuffle = False,
                        callbacks = [early_stop])
    return history
history = fit_model(model)
```

Figure 67: Part of the LSTM neural network code

LSTM outcomes

The predictions of the LSTM neural network are depicted in the following figure 68. It isobvious by the figure that the predictions of the BDI by the neural network are really close to the actual values, therefore it seems that it has gained a good understanding of the distribution of its price, and that it is able to predict accurately future prices.

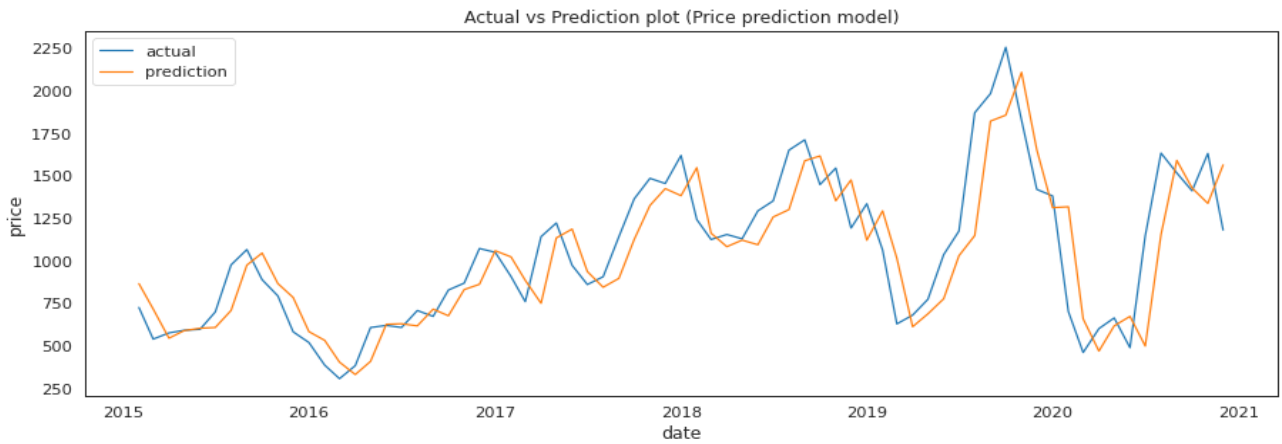


Figure 68: Performance visualization of the predictions made by the LSTM neural network

Finally, the accuracy metrics for the regression problem of the predictions of BDI by the LSTM neural network, are in the following Table 4.

MAE	MSE	RMSE
0.016352	0.000463	0.021515

Table 4: Accuracy metrics LSTM

5.2.2 GRU neural networks (Gated recurrent unit)

The optimal hyperparameters for the GRU neural network were found by following the same procedure we followed for the LSTM, i.e. by running tests for each hyperparameter and by the grid search method, calculating the grid performance over the validation set. The optimal hyperparameters set is shown at Table 5.

Hyperparameters	Potential	Optimal
GRU layers	[1, 2]	2 layers
Dense layers	[1, 2]	1 layer with 1 unit
Units per GRU layer	[16, 32, 64, 128, 256]	64 in both layers
Dropout per layer	[0.1, 0.2, 0.3]	0.2 in all layers
Epochs	[20, 30, 50, 80, 100]	30
Batch size	[4, 8, 16, 32]	4
Optimizer	[“rmsprop”, “adam”]	rmsprop

Table 5: Optimal hyperparameters for GRU neural network

Figure 69 depicts the architecture of the developed GRU neural network, which is similar to the LSTM’ s architecture, but with the correct GRU units per layer.

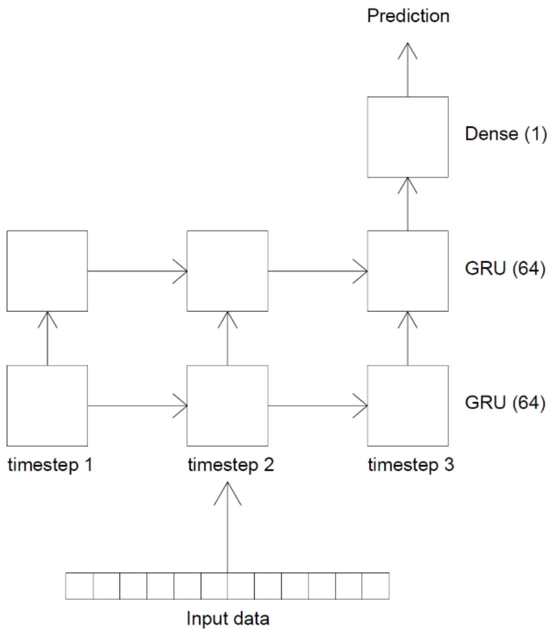


Figure 69: Architecture of GRU Neural network

Figure 70 displays part of the python code, where the function for creating the GRU neural network and the hyperparameters are defined.

```
DROPOUT = 0.2
# 20% Dropout is used to control over-fitting during training
WINDOW_SIZE = SEQ_LEN - 1
# Create GRU model
def create_gru(units):
    model = Sequential()
    # Input layer
    model.add(GRU (units = units, return_sequences = True,
    input_shape = [X_train.shape[1], X_train.shape[2]]))
    model.add(Dropout(0.2))
    # Hidden layer
    model.add(GRU(units = units))
    model.add(Dropout(0.2))
    model.add(Dense(units = 1))
    #Compile model
    model.compile(optimizer='rmsprop',loss='mse')
    return model
model = create_gru(64)
```

```
def fit_model(model):
    early_stop = keras.callbacks.EarlyStopping(monitor = 'val_loss',
                                               patience = 10)
    history = model.fit(X_train, y_train, epochs = 30,
                        validation_split = 0.1,
                        batch_size = 4, shuffle = False,
```

Figure 70: Part of the GRU neural network code

GRU outcomes

The predictions of the GRU neural network are depicted in the following figure 71.

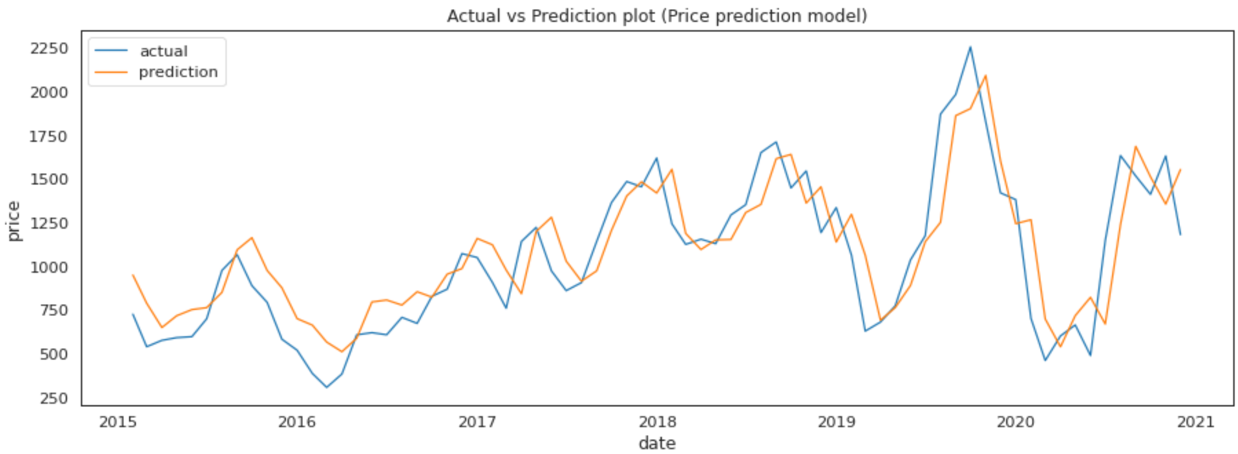


Figure 71: Performance visualization of the predictions made by the GRU neural network

In this case too, we see that the network makes very accurate predictions, managing to closely approximate the actual BDI values. The format of the predicted time series is to a considerable extent similar to that of the LSTM which makes sense given the many similarities between the networks.

Finally, the accuracy metrics for the regression problem of the predictions of BDI by the GRU neural network, are depicted on Table 6, by which it is evident that they also show high performance in the regression problem.

MAE	MSE	RMSE
0.018025	0.000487	0.022071

Table 6: Accuracy metrics GRU

5.2.3 BiLSTM neural networks (Bidirectional Long Short Term Memory)

A Bidirectional LSTM, or BiLSTM, is a sequence processing model that consists of two LSTMs: one taking the input in a forward direction, and the other in a backwards direction. BiLSTMs effectively increase the amount of information available to the network, improving the context available to the algorithm.

The optimal hyperparameters set for the BiLSTM neural network is shown at Table 7.

Hyperparameters	Potential	Optimal
BiLSTM layers	[1, 2]	2 layers
Dense layers	[1, 2]	1 layer with 1 unit
Units per LSTM layer	[16, 32, 64, 128, 256]	128, 64 respectively
Dropout per layer	[0.1, 0.2, 0.3]	0.2 in all layers
Epochs	[20, 30, 50, 80, 100]	30
Batch size	[4, 8, 16, 32]	16
Optimizer	[“rmsprop”, “adam”]	rmsprop

Table 7: Optimal hyperparameters for GRU neural network

Figure 72 depicts the architecture of the developed GRU neural network, which is similar to the LSTM’ s architecture, but for each 1 LSTM we now have 2 (1 in the forward directions and 1 in the backwards direction).

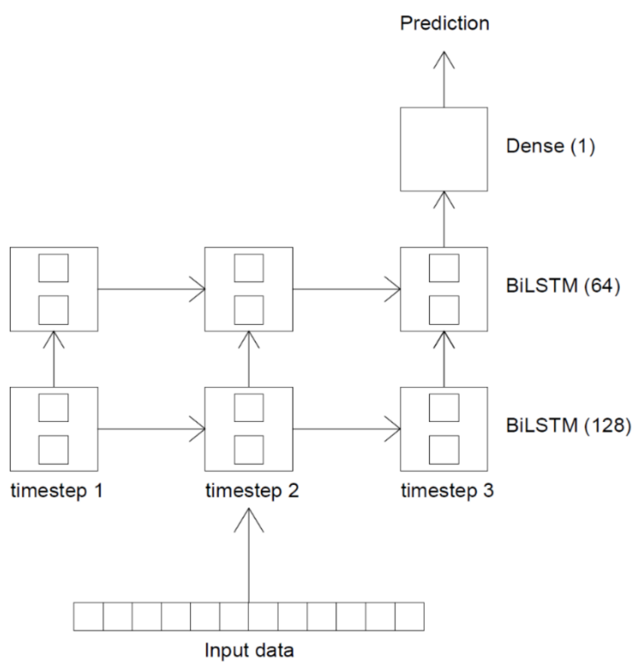


Figure 72: Architecture of BiLSTM Neural network

Figure 73 displays part of the python code, where the function for creating the BiLSTM neural network and the hyperparameters are defined.

```
# Create BiLSTM model
def create_bilstm(units):
    model = Sequential()
    # Input layer
    model.add(Bidirectional(
              LSTM(units = units, return_sequences=True),
              input_shape=(X_train.shape[1], X_train.shape[2])))
    # Hidden layer
    model.add(Bidirectional(LSTM(units = 64)))
    model.add(Dense(1))
    #Compile model
    model.compile(optimizer='rmsprop',loss='mse')
    return model
model = create_bilstm(128)
```

```
def fit_model(model):
    early_stop = keras.callbacks.EarlyStopping(monitor = 'val_loss',
                                               patience = 10)
    history = model.fit(X_train, y_train, epochs = 30,
                        validation_split = 0.1,
                        batch_size = 16, shuffle = False,
                        callbacks = [early_stop])
    return history
history = fit_model(model)
```

Figure 73: Part of the BiLSTM code

BiLSTM outcomes

The predictions of the BiLSTM neural network are depicted in the following figure 74.

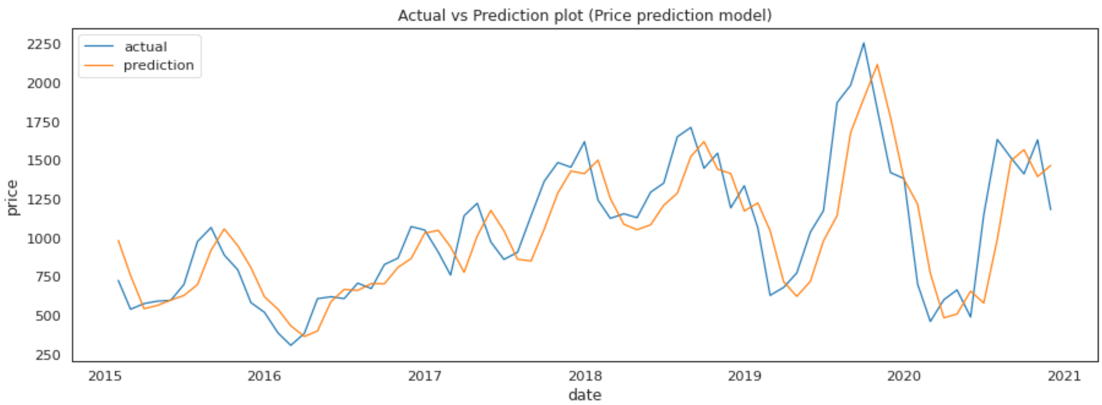


Figure 74: Performance visualization of the predictions made by the BiLSTM neural network

It is obvious that BiLSTM neural networks are also able to makes very accurate predictions of the actual BDI values, as the rest recurrent neural networks. The format of the predicted time series is to a considerable extent similar to that of the LSTM, which makes sense given the many similarities between the networks.

Finally, the accuracy metrics for the regression problem of the predictions of BDI by the BiLSTM neural network, are depicted on Table 8. It seems that they also show high performance in regression problem.

MAE	MSE	RMSE
0.018319	0.000525	0.022917

Table 8: Accuracy metrics BiLSTM

5.3 ARIMA

This model is the most common statistical model used in time series forecasting due to its increased performance and ease of use. ARIMA does not use additional features like the other machine learning algorithms and the only requirement it has from the input time series is that the time series is static, hence different preprocessing was done on the data compared to the other algorithms.

Then, an ARIMA model was developed and after testing and grid search, its parameters were set to the values:

- p = 3, check in the interval [0 - 5]
- d = 1, only the value 1 was checked due to the nature of the dataset
- q = 0, check in the interval [0 - 2]

Figure 75 displays part of the python code, where the function for creating the ARIMA model and the parameters are defined.

```
training_data = y_train
test_data = y_test
print
history = [x for x in training_data]
model_predictions = []
N_test_observations = len(test_data)
for time_point in range(N_test_observations):
    model = ARIMA(history, order=(3,1,0))
    model_fit = model.fit(disp=0)
    output = model_fit.forecast()
    yhat = output[0]
    model_predictions.append(yhat)
    true_test_value = test_data[time_point]
    history.append(true_test_value)
```

Figure 75: Part of the ARIMA code

ARIMA outcomes

The predictions of the ARIMA model are depicted in the following figure 76.

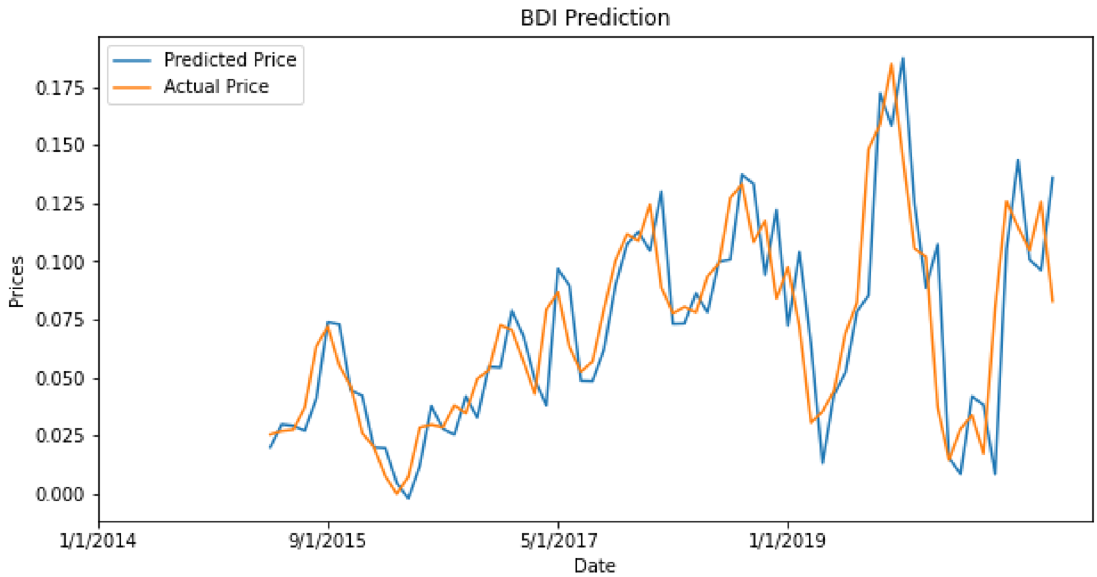


Figure 76: Performance visualization of the predictions made by the ARIMA model

The predictions for the regression problem made by the ARIMA model seem accurate as well, as the predicted price is close to the actual price during the whole testing phase. However, according to the following accuracy metrics on Table 9, the recurrent neural networks seem to have a higher accuracy level and less errors.

MAE	MSE	RMSE
0.017316	0.00055687	0.023598

Table 9: Accuracy metrics ARIMA

5.4 Feedforward neural networks

The performance of feedforward neural networks depends largely on the values of their hyperparameters. To construct the optimal model, the following hyperparameters of Table 10 were taken in consideration.

• number of Dense layers	• number of hidden layers
• number of units per Dense layer	• number of epochs
• dropout per layer	• optimizer

Table 10: Important hyperparameters for Feedforward Neural Networks

As loss function, we used the Mean squared error (mse) and as activation function the “relu”, as we only have Dense layers. Finding the appropriate set of hyperparameters was done by testing for each hyperparameter and the optimal set is shown Table 11.

Hyperparameters	Potential	Optimal
Dense layers	[1, 2, 3]	3
Units per Dense layer	[16, 32, 64, 128, 256]	128, 64, 32 respectively
Dropout per layer	[0.1, 0.2, 0.3]	0.2 in all layers
Epochs	[20, 30, 50, 80, 100, 120]	100
Optimizer	[“rmsprop”, “adam”]	adam

Table 11: Optimal hyperparameters for Feedforward neural network

Figure 77 displays part of the python code, where the function for creating the Feedforward neural network and the hyperparameters are defined.

```
# Flatten input (to support multivariate input)
n_input = trainX.shape[1] * trainX.shape[2]
trainX = trainX.reshape((trainX.shape[0], n_input))
n_input = testX.shape[1] * testX.shape[2]
testX = testX.reshape((testX.shape[0], n_input))
# Create multilayered FFNN model
model = Sequential()
model.add(Dense(128, activation='relu', input_dim=trainX.shape[1]))
model.add(Dropout(0.2))
model.add(Dense(64, activation='relu'))
model.add(Dropout(0.2))
model.add(Dense(32, activation='relu'))
model.add(Dense(trainY.shape[1]))
model.compile(loss='mean_squared_error', optimizer='adam')
model.summary()
# Fit model
history = model.fit(trainX, trainY, epochs =50, verbose =1)
# Predict the test set
predictions = model.predict(testX)
```

Figure 77: Part of the Feedforward neural network code

Feedforward neural networks outcomes

The predictions of the Feedforward neural networks model are depicted in the following figure 78.

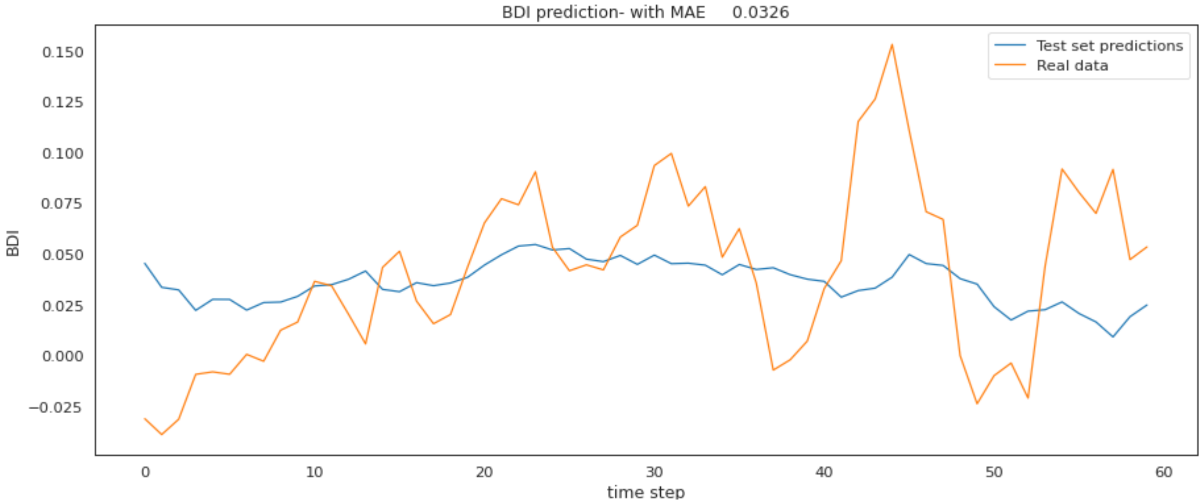


Figure 78: Performance visualization of the predictions made by the Feedforward neural network

The predictions for the regression problem made by Feedforward neural networks seem inaccurate, as the predicted BDI curve is not close to the curve of the actual BDI.

However, this result was expected due to the fact that feed forward neural networks are mostly used in classification problems, rather than regression ones, as they are more accurate in the first ones mentioned. Therefore, recurrent neural networks are preferred for regression problems, while feed forward neural networks are avoided.

Moreover, according to the following accuracy metrics of Table 12, the accuracy levels of both recurrent neural networks and the ARIMA model are better that those of the feedforward neural network. Consequently, it can be inferred that for regression problems, RNNS or ARIMA should be the ones chosen for predictions of timeseries and especially of the BDI, comparing to Feedforward neural networks.

MAE	MSE	RMSE
0.032614	0.001690	0.041109

Table 12: Accuracy metrics Feedforward neural networks

5.5 Summary of the accuracy metrics of the Recurrent Neural Networks

Table 13 below presents the aggregated results of the recurrent neural network predictions, since they are the main focus of this thesis. Based on the summarized accuracy metrics, it is facilitated to annotate and draw conclusions about the 3 types of RNNs and their performance in timeseries predictions.

RNN Models	MAE	MSE	RMSE
LSTM	0.016352	0.000463	0.021515
GRU	0.018025	0.000487	0.022071
BiLSTM	0.018319	0.000525	0.022917

Table 13: Summary of accuracy metrics of the Recurrent Neural Networks

The RNNs that will be compared are the LSTM, GRU and BiLSTM, based on the regression accuracy metrics, as it can be seen in the above table. It is evident that all 3 neural networks appear to have gained knowledge about the path of the predicted time series and therefore they present high accuracy and low errors. In particular, LSTM has the highest accuracy in the regression problem, followed by the GRU. The performance of the BiLSTM is equally good, but it is not as high as the performance of the other two recurrent neural networks.

5.6 Summary of the accuracy metrics of all models

A summary of the accuracy metrics for all 5 models examined can be found on the following Table 14.

Models	MAE	MSE	RMSE
LSTM	0.016352	0.000463	0.021515
GRU	0.018025	0.000487	0.022071
BiLSTM	0.018319	0.000525	0.022917
ARIMA	0.019316	0.000557	0.023598
Feedforward	0.032614	0.001690	0.041109

Table 14: Summary of accuracy metrics of all models

It can be understood that in regression problems, and more specifically in the time series prediction of the Baltic Dry Index, the recurrent neural networks have the highest accuracy and the lowest errors. The LSTM, GRU and BiLSTM models appear to have a better performance than the remaining two models, as MAE, MSE and RMSE are lower. ARIMA also presents a high performance in the regression problem, despite its less compicated algorithm. Moreover, the Feedforward neural networks showed two to three times less accuracy than the RNNs and ARIMA, thus it would not be preferred to be utilised for such a problem. Finally, as the LSTM recurrent neural network presents the highest accuracy, it is the most suitable for predictions of future prices of the BDI.

Chapter 6

Conclusion – Future work

6.1 Conclusion

Predicting the value of the Baltic Dry Index (BDI) is a particularly challenging problem for researchers because of the high volatility of the BDI, which in turn is due to the complexity and the large number of the its influencing factors. In the context of this thesis, in addition to the use of shipping variables, macroeconomic variables were used in order to model the dry bulk shipping market. These variables were utilized by multiple neural networks and statistical models that differ in type, architecture, hyperparameters and training function.

The potential factors that can influence the BDI price for monthly forecasts were thoroughly examined. In particular, in order to find out how much each variable affects the prediction of the index, we studied the different factors' correlation with the BDI and afterwards calculated the feature importance of each one, in relation to the variable to be predicted. To be exact, 30 possible features (10 shipping and 20 macroeconomic) were considered, from which we eventually chose 12 features that have a correlation greater than 0.55 with the BDI, 4 of which are shipping variables and 8 of which are macroeconomic variables. Lastly, we created an appropriate dataset with 30-year monthly values of the 12 features, dating from 1991 to 2021.

The 12-time series that were selected in the forecasting of the Baltic Dry Index are the following:

BDI, Consumer Price Index, Spot price of Gold per Ounce, Consumer Credit, Producer Price Index, World Steel Production, World stock market index, Spot price of Copper, Special drawing rights exchange rate against the dollar, Panamax scrap value, 6-month time charter rate for 32,000 ton dwt ship, Bulkcarrier Newbuilding Price Index Year/year change (%yr/yr).

From these features, according to the XGBoost algorithm that was used to estimate the feature importance, we reached the conclusion that the most important ones are ranked with the following order:

1) BDI
2) World stock market Index
3) Spot price of Copper
4) Bulkcarrier Newbuilding Price Index Year/year change (%yr/yr)
5) 6-month time charter rate for 32,000 ton dwt ship
6) Spot price of Gold per Ounce

It is worth mentioning that the autocorrelation (i.e. the correlation of a feature with itself) in a timeseries prediction is of great significance.

Furthermore, the data preprocessing took place in order for the dataset to be suitable for use by a machine learning algorithm and for the improvement of its performance. The following data preprocessing steps were performed:

- Separation of dataset into train, validation and test set
- Feature scaling
- Determination of the past timesteps
- Data transformation into a consistent form for the machine learning algorithms

After the data was correctly preprocessed, the development of 3 types of Recurrent Neural Networks took place, the LSTM, BiLSTM and GRU neural networks, with the use of python and more specifically of the tensorflow.keras libraries. The optimal hyperparameters for each neural network were selected to produce the most accurate results with the lowest errors.

For comparison purposes, the development of the ARIMA statistical model and of a Feedforward neural network seemed necessary. The comparison of the 5 developed models was based on 3 evaluation metrics of regression problems:

- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- Mean Absolute Error (MAE)

The RNNs showed the highest accuracy in the regression problem, with the LSTM model presenting the best performance, followed by the GRU model. The ARIMA model also showcased high accuracy and low errors in its predictions, however the RNNs due to their more advanced algorithms, presented better results. Finally, the Feedforward neural network model provided inaccurate results, as it was expected, due to the fact that feed forward neural networks are mostly used in classification problems where they are mostly accurate in, rather than regression ones.

Consequently, the obtained results verify the ability of Recurrent Neural Networks to adapt to complex regression problems and prove as well, that the study of macroeconomic variables, besides the shipping ones, is essential for decision making in the shipping industry.

This thesis can contribute positively to future research both through the methodology followed and the conclusions reached. More specifically, it provides the following contributions:

- The influence of the BDI price was examined through factors belonging to a wide range of fields, some of which were considered relevant for the first time (Bulkcarrier Newbuilding Price Index Year/year change).
- A framework was proposed for selecting the most highly correlated of the above factors with the BDI. Afterwards, the importance of each one of the factors was examined separately.
- It was found that the most important factors for the formation of the price of the BDI in monthly forecasts are the MSCI world stock index, the spot price of copper (USD/tonne) and the Bulkcarrier Newbuilding Price Index Year/year change. In addition to the above-mentioned, the spot price of gold per oz and the World steel production in thousand metric tons have a significant influence on the BDI as well.
- The proposed LSTM model showcased the best performance in the regression problem, followed by the other 2 types of Recurrent Neural Networks.
- The ARIMA model presented a high accuracy as well, thus it seems as a good alternative to be used for regression problems and timeseries forecasting.
- The Feedforward model had inaccurate results in the specific regression problem, thus the theory that they are mostly used for classification problems seems valid.

6.2 Future work

In this work, the influence of a large set of factors on the BDI was examined and models were developed to predict its future value with high accuracy. However, the developed models cannot be used directly for commercial purposes as the study was carried out for a specific period of time. In future research, the developed system can be modified to be interactive and make real-time predictions using updated data. Such a scenario is feasible, as we have seen that the training of the neural network models is very fast, therefore we can retrain the model at regular time periods in order to make use of the latest data.

Furthermore, the influence of additional factors on the BDI can be explored and predictions can be made over a longer time range. Finally, an interesting direction is to train the models with hourly and daily data and explore the factors that influence the forecasts in these cases.

Bibliography

1. Abiodun, O. I., Jantan, A., & Omolara, A. E. (2018). Kemi Victoria Dada, Nachaat AbdElatif Mohamed, and Humaira Arshad. State-of-the-art in artificial neural network applications: A survey. Jurnal Heliyon Elsevier, 4.

2. *Activation functions and their derivatives - Master Data Science*. Master Data Science. (2018). Retrieved 6 October 2021, from http://datahacker.rs/activation-functions-and-derivatives/.

3. Albert Willem Veenstra and Philip Hans Franses (1996), A co-integration approach to forecasting freight rates in the dry bulk shipping sector, Department of Regional Transport and Port Economics, Erasmus University, Rotterdam, The Netherlands.

4. *Applications of Linear Functions | Boundless Algebra*. Courses.lumenlearning.com. (2021). Retrieved 6 October 2021, from https://courses.lumenlearning.com/boundless-algebra/chapter/applications-of-linear-functions/.

5. *Artificial neural network ensemble modeling with exploratory factor analysis for streamflow forecasting - Scientific Figure on ResearchGate*. Reserchgate.net. (2021). Retrieved 6 October 2021, from https://www.researchgate.net/figure/Backpropagation-training-algorithm-of-MFNN_fig2_281689530.

6. Brownlee, J. (2018). How to Develop a Bidirectional LSTM For Sequence Classification in Python with Keras. Machine Learning Mastery. Retrieved 4 October 2021, from https://machinelearningmastery.com/develop-bidirectional-lstm-sequence-classification-python-keras/.

7. Carling, A. (1992). Introducing Neural Networks. Wilmslow, UK: Sigma Press

8. Chen, T., & Guestrin, C. (2016). XGBoost. *Proceedings Of The 22Nd ACM SIGKDD International Conference On Knowledge Discovery And Data Mining*. https://doi.org/10.1145/2939672.2939785

9. Cho, K. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation (pp. 1724–1734).

10. Christos Stergiou and Dimitrios Siganos, Neural Networks, paper, Imperial college, Department of Computing.

11. Cortes, C., & Vapnik, V. (1995). Support-vector networks (pp. 273-297).

12. Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification.

13. *DEEP LEARNING BASED SEED QUALITY TESTER - Scientific Figure on ResearchGate*. (2021). Retrieved 6 October 2021, from https://www.researchgate.net/figure/Rectified-Linear-Unit-ReLU-Activation-Function_fig1_337485695.

14. Diamantaras Konstantinos (2007), Artificial Neural Networks, Keydarithmos Publications, Athens.

15. Donald F. Specht (1988), Probabilistic Neural Networks.

16. Economic Cycle. Corporate Finance Institute. (2021). Retrieved 4 October 2021, from https://corporatefinanceinstitute.com/resources/knowledge/economics/economic-cycle/.

17. Giziakis, Papadopoulos, Plomariou (2002), Introduction to chartering, Ath. Stamoulis Publisher.

18. Giziakis, Papadopoulos, Plomaritou (2002), Introduction to Shipping, publications of Ath. Stamoulis.

19. Glynos Nikolaos, Georgiou B. Harris, Melissovas V. Spyros, Papadopoulos S. Dimitris (1995) , Training Methods and Technical Neural Network Models, University of Ioannina, Department of Computer Science.

20. Goodfellow, I., Bengio, Y., & Courville, A. (2018). Deep Learning. Frechen: MITP.

21. Gorton L. , Ihre R. , Sandevarn A. (1999), Shipbroking and Chartering Practice 5th Edition, London: Lloyd's of London.

22. Guina, R. (2019). *The 2008-2009 Financial Crisis – Causes and Effects*. cashmoneylife. Retrieved 4 October 2021, from https://cashmoneylife.com/economic-financial-crisis-2008-causes/.

23. Guina, R. (2021). The Great Recession - Causes and Effects of the 2008-2009 Financial Crisis. Cash Money Life | Personal Finance, Investing, & Career. Retrieved 4 October 2021, from https://cashmoneylife.com/economic-financial-crisis-2008-causes/.

24. Han, M., & Yu, S. (2017). Time-Series Prediction of Baltic Dry Index (BDI) Using an Application of Recurrent Neural Networks. Retrieved 5 October 2021, from https://www.koreascience.or.kr/article/CFKO201716269584003.page.

25. Harnad, Stevan (2006) The Annotation Game: On Turing (1950) on Computing, Machinery, and Intelligence.

26. Haykin Simon (1999) , Neural Networks: A Comprehensive Foundation, Prentice Hall

27. Ho, T. (1995). Random decision forests (pp. 278–282).

28. Hochreiter, S., & Schmidhuber, J. (1997). Long short term memory (pp. 1735-1780).

29. Hopfield, J. (1982). Neural networks and physical systems with emergent collective computational abilities (8th ed., pp. 2554-2558).

30. *How do you build computational graph in TensorFlow?*. i2tutorials. (2019). Retrieved 6 October 2021, from https://www.i2tutorials.com/how-do-you-build-computational-graph-in-tensorflow/.

31. J.M. Mendel and R.W. McLaren (1970), Adaptive, Learning and Pattern Recognition Systems: Theory and Applications, Academic Press, New York.

32. Jones, T. (2017). *Recurrent neural networks deep dive*. IBM Developer. Retrieved 6 October 2021, from https://developer.ibm.com/articles/cc-cognitive-recurrent-neural-networks/.

33. Kamal, I. (2018). Forecasting the Baltic Dry Index by using an artificial neural network approach. TURKISH JOURNAL OF ELECTRICAL ENGINEERING & COMPUTER SCIENCES, 26(3). https://doi.org/10.3906/elk-1706-155

34. Konstantinos Giziakis, Athanasios Karlis (2007), Course Notes: Introduction to Shipping, University of Piraeus, Department of Maritime Studies.

35. Konstantinos, G., & Athanasios, K. (2007). Introduction to chartering.

36. Kostas Siriopoulos (1998), Analysis and Controls of Univariate Financial Time Series

37. Kulkarni, C. (2018). *Learning Rate Tuning and Optimizing*. Medium. Retrieved 6 October 2021, from https://medium.com/@ck2886/learning-rate-tuning-and-optimizing-d03e042d0500.

38. Lai Tingjia (2009), Prediction model for Dry Bulk Secondhand Ship Market, National Cheng Kung University.

39. Linnainmaa, S. (1976). Taylor expansion of the accumulated rounding error,.

40. Lippmann Richard P. (1987), An Introduction to Computing with Neural Nets

41. Lun, Y., Lai, K., & Cheng, T. (2010). *Shipping and logistics management*. Springer.

42. Lun. (2010). Shipping and logistics management (1st ed.). Springer.

43. Machine Learning Generalization of Lumped Parameter Models for the Optimal Cooling of Embedded Systems - Scientific Figure on ResearchGate. Retrieved 6 October 2021, from: https://www.researchgate.net/figure/Deep-Learning-Sequential-Neural-Network-with-two-hidden-layers_fig2_342641657.

44. Manaswi, N. (2018). RNN and LSTM In Deep Learning with Applications Using Python (pp. 115-126). Apress.

45. Mcculloch, W., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. Univ. Press.

46. Minsky and Papert (1969), Perceptrons, An introduction to computational geometry, MIT press, expanded edition.

47. *Momentum and Learning Rate Adaptation*. Cnl.salk.edu. (2021). Retrieved 6 October 2021, from https://cnl.salk.edu/~schraudo/teach/NNcourse/momrate.html.

48. Nayak, S. (2018). *Batch Normalization in Deep Networks*. LearnOpenCV. Retrieved 6 October 2021, from https://learnopencv.com/batch-normalization-in-deep-networks/.

49. Nelson, D., & Pereira, M. (2017). Stock market's price movement prediction with LSTM neural networks.

50. Neural Networks: A Requirement for intelligent systems, http://www.learnartificialneuralnetworks.com/

51. *Neuron.svg - Wikimedia Commons*. Commons.wikimedia.org. (2019). Retrieved 6 October 2021, from https://commons.wikimedia.org/wiki/File:Neuron.svg.

52. Nikolaos D. Manos, PhD Thesis - "Methods of shielding shipping companies and Banks against financial risks".

53. Olah, C. (2015). Understanding lstm networks.

54. Omri Weisman and Ziv Pollak (1995), The Perceptron, Ben Gurion University of the Negev.

55. *Overfitting vs Underfitting*. Docs.paperspace.com. (2019). Retrieved 6 October 2021, from https://docs.paperspace.com/machine-learning/wiki/overfitting-vs-underfitting.

56. P.M. Alderton (1995), Sea Transport (Operation & Economics) 4th Ed., T. Reed.

57. Patel, D. (2020). *Boosting your Model performance Using XGBoost*. Medium. Retrieved 6 October 2021, from https://medium.com/analytics-vidhya/boosting-your-model-performance-using-xgboost-80df2b642f60.

58. Petnehazi, G. (2019). Recurrent neural networks for time series forecasting.

59. Petropoulos, T. (2009). Petrofin.gr. Retrieved 4 October 2021, from https://www.petrofin.gr/wp-content/uploads/2018/04/CargoPlusAprilMay2009-1.pdf.

60. Psaraftis, Charilaos (2005), Course Notes: Economics of Maritime Transport 1, National Technical University of Athens, School of Marine Mechanical Engineering, Department of Ship Design and Maritime Transport.

61. Roman, V. (2020). *Recurrent Neural Networks*. Towards data science. Retrieved 6 October 2021, from https://towardsdatascience.com/recurrent-neural-networks-56e1ad215339.

62. Rosenblatt, F. (1957). The perceptron—a perceiving and recognizing automaton.

63. Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning representations by back-propagating errors.

64. Sahin, B., & Gurgen, S. (2018). Forecasting the Baltic Dry Index by using an artificial neural network approach. Turkish Journal Of Electrical Engineering & Computer Sciences. Retrieved 5 October 2021.

65. Sahiner, B., Chan, H., & Petrick, N. (1996). Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images (pp. 598-610).

66. Samuel, A. (1967). Some studies in machine learning using the game of checkers II. Stanford University.

67. Saurabh, K. (2012). Approximating Number of Hidden layer neurons in Multiple Hidden Layer BPNN Architecture. International Journal Of Engineering Trends And Technology-, 3(6). Retrieved 4 October 2021.

68. Simitis Konstantinos (2008), The Crisis, Polis Publications, Athens.

69. Singh, S. (2021). *Swish as an Activation Function in Neural Network*. Deep Learning University. Retrieved 6 October 2021, from https://deeplearninguniversity.com/swish-as-an-activation-function-in-neural-network/.

70. Soros George (2009), The 2008 Financial Crisis and its Importance, Livanis SA, Athens.

71. Stathakis, D. (2009). How many hidden layers and nodes. International Journal Of Remote Sensing, 30(8), 2133-2147. Retrieved 4 October 2021.

72. Stopford M. (1997), Maritime Economics 2nd edition, Routledge, London.

73. Stournaras Yannis, The International Crisis and the Greek Economy: Dilemmas and Proposals, presentation.

74. Stournaras, Y. (2019). *Lessons from the crisis and Greece's future economic prospects*. Retrieved 4 October 2021, from

https://www.bankofgreece.gr/enimerosi/grafeio-typoy/anazhthsh-enhmerwsewn/enhmerwseis?announcement=0b23cf67-8c2e-48f2-a6bc-0c8615eaddf5.

75. *Supervised Learning*. Tutorialspoint.com. (2021). Retrieved 6 October 2021, from https://www.tutorialspoint.com/artificial_neural_network/artificial_neural_network_supervised_learning.htm.

76. Team, K. (2021). Keras: the Python deep learning API. Keras.io. Retrieved 4 October 2021, from https://keras.io/.

77. Ted Petropoulos, Article: On hold until 2011, newspaper Cargo.

78. TensorFlow: Large-scale machine learning on heterogeneous systems. TensorFlow. (2021). Retrieved 4 October 2021, from https://www.tensorflow.org/.

79. Voudris, Athanasios V (2006), Analysis and forecast of the capesize bulk carriers shipping market using Artificial Neural Networks, Massachusetts Institute of Technology. Dept. of Mechanical Engineering.

80. Watkins, C. (1989). Learning from delayed rewards.

81. William Payne (barrister-at-law.), Edward Richard Hardy Ivamy, (1989), Payne & Ivamy's Carriage of Goods by Sea, Butterworth, Great Britain.

82. Wolfgang Banzhaf, Peter Nordin, Robert E. Keller, Frank D. Francone (1998) Genetic Programming – An Introduction, Morgan Kaufmann Publishers, San Francisco, California.

83. Zhang, X., Chen, M., Wang, M., Ge, Y., & Stanley, H. (2020). A novel hybrid approach to Baltic Dry Index forecasting based on a combined dynamic fluctuation network and artificial intelligence method. Retrieved 4 October 2021.

84. Zhang, X., Xue, T., & Eugene Stanley, H. (2019). Comparison of Econometric Models and Artificial Neural Networks Algorithms for the Prediction of Baltic Dry Index. IEEE Access, 7, 1647-1657. https://doi.org/10.1109/access.2018.2884877