

European International University, EIU-Paris

Doctoral Thesis

Title:

Certain investigation on crude oil price prediction using machine learning model.

Submitted by:

Mayur Malik

EIU0272

In Partial Fulfillment of the Requirements for the Degree of

DBA

Supervisor:

Mr Edward Roy Krishnan

Date:

Dec 2024

CHAPTER 1

INTRODUCTION

1.1 Introduction

A financial market is a marketplace where financial products, such as stocks, currencies, bonds, derivatives, commodities, and more, are exchanged and traded. It facilitates transactions between buyers and sellers, allowing them to trade their preferred financial assets at prices driven by market forces. The connection between the economy and the financial market greatly impacts the global landscape. The financial market is a vital factor in the growth of the economy, offering opportunities for individuals, businesses, and governments to invest, borrow, and raise capital. This leads to increased efficiency, the emergence of new businesses, and higher employment rates.

Financial markets serve as a crucial link between the investors and the borrowers, bridging the gap for mutual gain. Financial markets offer investors the option to invest their money in various securities or schemes for short- or long-term investing benefits. Figure 1.1. depicts the many forms of financial markets.

Capital market

A capital market is a platform where individuals and institutions engage in buying and selling financial securities like bonds and stocks. This market primarily deals with medium and long-term securities, and its scale is closely linked to the size of a nation’s economy. Hence, any fluctuations in this market can have significant

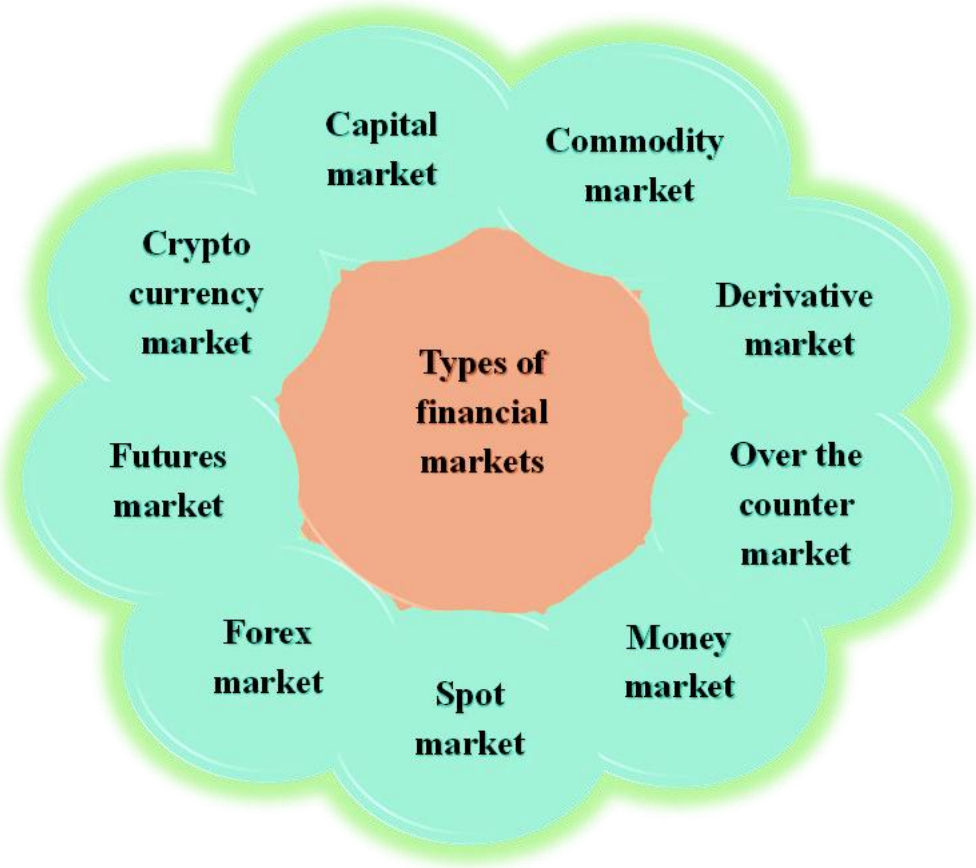


Figure 1.1 Types of financial markets

ripple in other areas. It consists of two types: (i) primary market and (ii) secondary market.

Primary market

Primary market refers to the market, where new shares or securities are offered for the first time. In this market, companies issue new securities and receive cash from investors or buyers in exchange for these securities. It involves the trade of newly issued stocks and other securities to interested investors.

Secondary market

Secondary market involves the trading of existing or previously-issued securities among investors. Once new securities are sold in the primary market, there needs to be an efficient mechanism for their resale. The secondary market provides investors with the opportunity to resell or trade their existing securities.

Commodity market

Investors and merchants buy and sell natural resources such as oil, corn, meat, and gold on the commodity market. These resources have a dedicated market due to their price volatility and unpredictability. In the future market for commodities, the price of goods to be delivered at a specific time is already established and fixed.

1.2 Background of the study

Crude oil is one of the most important commodities that its price and volatility has huge impact on people life across the globe. It comprises a blend of hydrocarbons found in liquid form within natural underground reservoirs. Crude oil maintains its liquid state under normal atmospheric pressure even after undergoing separation processes at the surface. Crude oil is the primary source of energy on a global scale, making it challenging to find alternative resources to replace it. Its impact extends to economic growth, spanning everyday goods to energy and military domains. Sudden fluctuations in oil prices influence economic stability for both producing and consuming nations, with heightened significance for import-dependent countries. Deng et al. (2021). Its pervasive presence in various aspects of human lives grants it a significant role in shaping nation's economy. Consequently, fluctuations in crude oil

prices are often indicators for changes in global market and overall economy. Research studies indicated that increased volatility and reduced predictability in oil prices can harm economic indicators, highlighting potential negative consequences Baumeister and Kilian (2014). However, accurately predicting oil prices is challenging across different time frames due to an array of factors that impact price changes.

For instance, the 2008 US recession caused a housing crisis that led to a drop in crude oil prices, rendering forecasting models ineffective. More recently, the onset of the COVID-19 pandemic in early 2020 drastically reduced crude oil demand. During this period, a conflict emerged between the Organization of the Petroleum Exporting Countries (OPEC) and non-OPEC nations, with Russia declining production cuts proposed by OPEC. This refusal triggered an oil price war, primarily involving Saudi Arabia, causing a sharp decline in crude oil prices Bollapragada et al. (2021).

There exist two primary grades of crude oil that function as reference points for other oil prices. These are West Texas Intermediate (WTI) located in Cushing and the North Sea Brent. WTI from Cushing, located in the U.S, serves as the benchmark for U.S. oil prices. On the other hand, North Sea Brent oil, sourced from Northwest Europe, acts as the benchmark for global oil prices. According to the Energy Information Administration's (EIA) short-term energy outlook published on November 2022, the price of Brent crude oil is predicted to average $95.33/b globally in 2023.In comparison, the average yearly amount in 2021 and 2022 was, $70.89 and $ 102.13, respectively. WTI averaged $87.55/b in October 2022. The EIA forecasts that WTI prices will average $87.33/b in 2023, down from $95.55/b in 2022. The EIA

also predicts that by 2025 Brent oil's nominal price will rise to $67/b. By 2030, world demand is seen driving Brent prices to $79/b. By 2040 and 2050, prices are projected to be $84/b and $90/b, respectively. WTI price is expected to rise to $65/b by 2025, increasing to $71/b by 2030, $81/b by 2040, and $87/b by 2050.

Recent studies highlight the ongoing significance of crude oil in the global economy, despite increasing efforts towards sustainability and energy transition. The IEA reports that oil remains a crucial commodity, accounting for approximately 3% of global GDP. Crude oil prices have a far-reaching impact, influencing inflation due to its integral role in numerous industries, including transportation, manufacturing, and petrochemicals. As of 2023, oil demand has surpassed pre-pandemic levels, driven by strong post-lockdown economic recovery and the resurgence of industrial activity, especially in China (IEA).

In 2023, global oil markets faced significant disruptions sue to geopolitical tensions such as the Russia-Ukraine conflict, which caused price volatility and reshaped trade flows. This period saw a strong supply response from non-OPEC+ producers like the U.S. and Brazil, while OPEC+ implemented cuts to stabilize prices. The oxford institute of energy studies notes that OPEC+ has taken a pre-emptive approach in managing oil supply to cope with uncertainties in demand.

The importance of oil in the global economy is highlighted by its impact on inflation and cost of living. Although energy transition efforts aim to reduce reliance of fossil fuels, current demand for crude oil remains robust, underscoring the commodity's enduring importance as both an economic driver and geopolitical tool.

These unpredictable actions and event can create substantial fluctuations in crude oil prices, with subsequent implications for the global economy and markets. Accurate forecasting of these price changes enables economic stakeholders, including businesses, to adopt to future market shifts and empowers decision-makers with precise insights for optimal choices. As a results, it is imperative to formulate a robust predictive model for anticipating oil prices and their subsequent variations.

1.3 Rationale for research

Crude oil is a crucial commodity that has a substantial impact on global economies. Accurate price predictions can help businesses, governments, and traders to make correction decisions about investments, consumption, and policy formation. Industries reliant on crude oil for their transportation and manufacturing, encounter substantial financial risks due to price fluctuations. Effective prediction models can aid in developing risk management strategies and hedging mechanism. Investors in the energy sector, as well as those in related markets like commodities and equities, can benefit from accurate price forecasts. These predictions provide valuable insights for portfolio diversification and investment allocation.

Predicting future crude oil prices presents a considerable challenge due to the highly non-linear, dynamic, non-stationary, and chaotic nature of the market, and perspectives of stakeholders. The price of crude oil experiences continual fluctuations, marked by millions of daily share transactions. Several factors, such as supply and demand, political events, reserves, OPEC’s actions, dollar value, and stakeholder perceptions, exert influence on the market value of crude oil. The intricate interplay among these factors and the crude oil price results in notable instability.

The research efforts in question hold substantial importance as they strive to establish a reliable and accurate framework for anticipating upcoming crude oil prices. Precise predictions in crude oil pricing can offer crucial support for decision-making in manufacturing, investments, corporate, and government domains, considering the essential role of crude oil as the predominant global energy source. The research intends to examine the power of machine learning models in predicting future crude oil prices. In addition to this, the research aims to boost the prediction accuracy by incorporating technical indicators and machine learning models.

1.4 Statement of the problem

Three major issues have been identified through investigations conducted on prior research.

- The previous predictions heavily relied on data from WTI or Brent crude oil prices, neglecting the consideration of other relevant market inputs. However, the crude oil price market is influenced by various contributing factors surrounding it, and overlooking these inputs undermines the predictive capabilities of the tool. A reliable prediction system should understand and correlate between various factors, providing accurate forecasts.
- Only a few studies implemented verification and validation techniques on the main fluctuating factors. While previous studies consider global crude oil prices, they often overlook other significant factors like demand and supply, which play vital roles in market volatility. Focusing solely on these observations is inadequate to fully capture the information offered by market

trends. Therefore, a comprehensive prediction requires embracing appropriate key factors and correlating them to achieve a thorough understanding.

- Time-series data remains the primary data type used for prediction. However, some previous research neglects essential data preprocessing and data representation procedures. These crucial steps are vital for cleaning and reducing errors and noise in the dataset while ensuring its uniformity. Implementing these processes will organize the prediction procedure more systematically, generating more stable and reliable results. Failing to include these processes compromises the reliability of the prediction tool.

The research investigates the correlation between historical data and global crude oil prices, aiming to build a model for predicting future crude oil prices. The research analyses the non-linear relationship between global crude oil prices and macroeconomic variable to create a precise model for predicting future crude oil prices. The research examines the relationship between technical indicators and global crude oil prices to design a more accurate model for forecasting future crude oil prices.

1.5 Research questions

Based on the identified research problems, the following research questions are formulated to be answered.

1. What type of relationship exists between the historical prices of WTI and Brent crude oil prices and future crude oil prices?
2. What sort of relationship exists between macroeconomic variables and WTI and Brent crude oil prices?

3. What is the dynamic relationship between technical indicators and WTI and Brent future crude oil prices?
4. Do the proposed machine learning models anticipate future crude oil prices?
5. Which set of input features can better describe the prediction performance?
6. What is the percentage of prediction accuracy for training data and testing data?

1.6 Research objectives

Crude oil is the most important energy source whose international price fluctuations affect all aspects of the global economy. Fluctuations in crude oil prices receive considerable attention from financial experts and individuals engaged in the market. However, forecasting crude oil prices is a highly complex and challenging task due to its irregular, non-linear, dynamic nature, coupled with high volatility. Consequently, preciously forecasting the time series of crude oil prices poses a major challenge, holding immense importance for energy economists aiming to make informed decisions across various managerial levels. As a result, numerous researchers have invested substantial resources into creating various models for forecasting. The prime target of this research is to develop crude oil prediction models based on machine learning. The main objectives are as follows:

- To conduct a comprehensive exploration of available techniques for crude oil price prediction, along with their advantages and disadvantages.
- To analyse the evaluation criteria used to evaluate the performance of the crude oil price prediction models.

- To build a Feed Forward Neural Network (FFNN) to predict future crude oil prices based on historical data
- To formulate a crude oil price prediction model using macroeconomic variables and Support Vector Machine (SVM), aiming to anticipate future crude oil prices
- To create a model for crude oil price prediction by utilizing using technical indicators and machine learning models
- To investigate how certain input variables may affect the prediction of future crude oil prices
- To validate the developed models using daily, weekly, and monthly data from WTI and Brent crude oil markets.
- To assess the applicability of designed models for predicting crude oil prices and conduct comparisons to identify the most effective model.
- To compare the effectiveness of the newly introduced models with previous methods to visually demonstrate their superior performance.

1.7 Organization of the thesis

This thesis demonstrates the development of models for crude oil price prediction using machine learning methodologies, with the intention of achieving a higher accuracy in predicting future crude oil prices. The designed crude oil prediction models have been implemented and validated using price data from WTI and Brent crude oil market. The efficacy of these models has been confirmed through evaluations utilizing performance metrics. The organization of this thesis is structured as outlined below:

Chapter 1 introduces the background of the research, rationale behind the research, and statement of the research. Research questions, objectives and organization of the thesis are given.

Chapter 2 discusses the various methods found in existing literature for addressing the challenge of anticipating crude oil prices. The chapter starts by giving an outline of the crude oil prediction methods. Subsequently, available strategies for predicting crude oil prices are deliberated.

Chapter 3 explains the research design. The chosen crude oil markets are described in depth. The operational principles of machine learning methods are discussed. Additionally, proposed crude oil prediction models are explained.

Chapter 4 examines the effectiveness of the developed crude oil prediction models by evaluating metrics such as root mean square error, mean absolute percentage error, directional statistic, and prediction accuracy. This chapter also includes a comparative assessment.

Chapter 5 summarizes the key achievements of this research. It identifies the main contribution of the models developed in this thesis. The chapter suggests some improvements that could be made to the existing models and research topics that might be interesting to cover in the future.

The works of several researchers are quoted and used as evidence to support the concepts explained in the thesis. All such evidence used is listed in the reference section, and a list of papers published based on this research work is given at the end of this research thesis report.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The onset of the industrial revolution marked the utilization of traditional energy sources such as primarily coal combustion and hydropower as the initial sources of energy. Subsequently, oil and nuclear energy gained prominence as significant contributors to the energy landscape Dunn (2002). Oil is a non-renewable resource and it is distributed randomly around the world. Tuo et al. (2011) pointed out that oil price analysis can be traced back to Harold Hotelling, a renowned mathematical economist in 1931, laid the foundation for market analysis using economic tools, exhaustible resources, and fundamental theoretical constructs in energy economics.

Oil plays a pivotal role in both national and international economies, serving as a crucial raw material for diverse industries. It serves as a crucial energy source and an indispensable component in numerous manufacturing and transportation processes. According to the Energy Information Administration (EIA) (https://www.eia.gov/outlooks), current global crude oil consumption stands at 100 million barrels daily. This consumption level makes oil the most significant contributor to global energy consumption compared to other sources Lin (2009). The volatility and fluctuations in oil prices are evident, with the commodity being highly active and traded in global markets. Sudden fluctuations in oil prices have an impact on world economy, politics, military, and society He et al. (2012), Gabralla and

Abraham (2013). As a result, governments, businesses, and investors have a growing interest in this matter.

Due to its fundamental role as an energy source and its substantial influence on a nation's economic growth, social stability, and security Zhang et al. (2015), accurately anticipating fluctuations in crude oil prices has remained a consistently hot topic of investigation. The international research on predicting fluctuations in crude oil prices is further complicated by a multitude of factors, such as supply and demand dynamics Kilian (2009), fluctuations in the value of the US dollar Lizardo and Mollick (2010), speculative trading activities Kilian and Murphy (2014), geopolitical tensions Sahir and Quershi (2007), and occurrences of natural disasters Ramsay (2011). These factors introduce considerable noise into the crude oil data, rendering the task of forecasting the complex and nonlinear volatility patterns of crude oil prices exceptionally challenging and yielding results with inherent uncertainty. As a results, the prediction of crude oil prices remains an immense challenge. To date, an extensive body of literature has been dedicated to exploring the prediction of crude oil prices.

Developing an accurate and high-performance model for predicting crude oil prices emerges as a highly valuable academic pursuit aimed at addressing the challenge of fluctuating oil prices. Hence, this chapter aims to offer comprehensive overview of methods for predicting crude oil prices.

2.2 Crude oil prediction models

The models devised for the purpose of forecasting crude oil prices can generally be categorized into three primary groups, as illustrated in Figure 2.1.

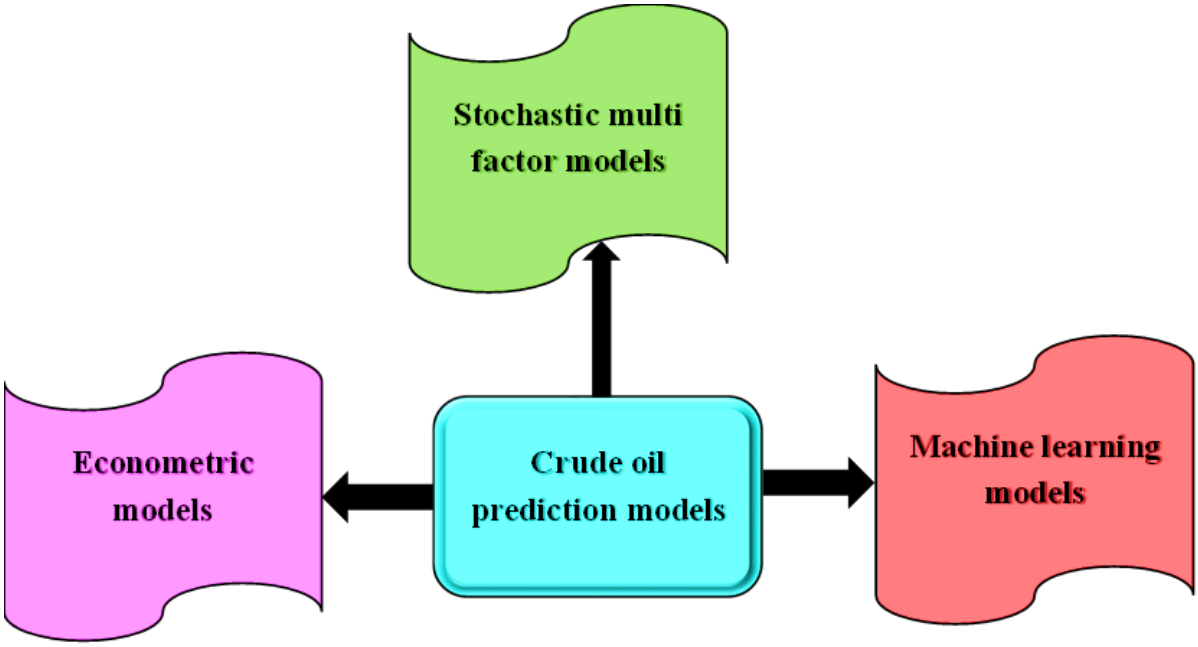


Figure 2.1 Classification of crude oil price prediction models

There exist two categories of econometric models: 1) structural models and 2) time series models. The structural models encompass the Error Correction Model (ECM) Lanza et al. (2005), the Vector Error Correction Model (VECM) Murat and Tokat (2009), and the Vector Auto-Regressive (VAR) Baumeister and Kilian (2012) model. These models employ linear regression techniques and rely on various data points such as demand and supply. Moreover, they incorporate additional explanatory variables in conjunction with past oil price data. On the other hand, the alternate approach involves time series models, which solely employ historical oil price data to make future predictions. However, these models require the data to be both stationary and linear, as they are capable of accurately handling the intricate complexities of crude oil prices alongside the methodologies of econometric analysis.

Stochastic multi-factor models, such as the one-factor as well as the two and three- factor models have been employed to price the crude oil market. Nonetheless, these approaches encounter limitations regarding the quantity and nature of pivotal factors used as inputs, demanding intricate parameter calibration. For instance, two-factor model, Shahmoradi (2016), lssaka (2020) necessitates the tuning of a minimum of seven parameters, often undertaken through time-consuming techniques like Particle Swarm Optimization (PSO).

Econometric models tend to lack accuracy in forecasting price time series, particularly when dealing with non-linear, dynamic, and non-stationary data. Consequently, Machine Learning (ML) models have gained popularity to tackle this challenge. These ML models, such as Artificial Neural Network (ANN) Gao and Lei (2017), Support Vector Machine (SVM) Altan and Karasu (2019), and Genetic Algorithm (GA) Tehrani and Khodayar (2011), Leonori et al. (2020), are frequently used in energy market contexts to overcome this limitation.

2.3 Review of papers on predicting crude oil price using econometric models

Econometric models have been widely used to capture the volatility and price dynamics of crude oil. These models often assume linearity and rely on time series data to make accurate predictions. Among these, Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models stand out their ability to model and anticipate volatility in crude oil prices.

2.3.1 Volatility modelling with GARCH and its variants

Numerous studies applied GARCH-family models to capture volatility in crude oil prices. A non-parametric GARCH model for forecasting international crude oil prices was presented by Hou and Suardi (2012). Focusing on Brent and WTI crude oil prices, the authors demonstrated that the non-parametric GARCH model provided superior test data volatility forecasts compared to a wide range of parametric GARCH models.

Ahmed and Shabri (2013) used crude oil spot prices to fit a GARCH model. This was done to demonstrate the benefits of nonlinear models. The findings of the study. The authors used three GARCH models: GARCH-N, GARCH-t and GARCH-G are applied to crude oil spot pricing. The GARCH-N model was shown to be the most effective. The GARCH-G model is used for forecasting Brent, while the most accurate at forecasting WTI crude oil spot prices.

Cheong et al. (2009) examined the volatility of the West Texas Intermediate (WTI) and Brent crude oil markets using a flexible Autoregressive Conditional Heteroskedasticity (ARCH) model. Results showed that WTI crude oil prices had greater long-term volatility and similar appreciation and depreciation stocks. Brent crude oil had a leverage effect. WTI models provided superior forecasts, while the simplest, most parsimonious GARCH offered the best forecasted evaluations.

Kang et al. (2009) examined the effectiveness of a volatility model for three crude oil markets: Brent, Dubai, and WTI. The authors concentrated on the model's ability to forecast and recognize volatility patterns, particularly in terms of persistence

or long memory. They evaluated volatility persistence in these crude oil prices using conditional volatility models. Among the models considered, Component-Generalized Autoregressive Conditional Heteroskedasticity (CGARCH) and Fractionally Integrated GARCH (FIGARCH) demonstrated better capability in capturing persistence compared to GARCH and IGARCH models. Moreover, CGARCH and FIGARCH models exhibited superior performance in predicting volatility beyond observed data.

Wei et al. (2010) extended the work of Kang et al. (2009), adopting nine linear and nonlinear Generalized Autoregressive Conditional Heteroskedasticity (GARCH)-type models. They accessed the accuracy of 1-, 5-, and 20-day out-of-sample volatility predictions using daily Brent and WTI crude oil prices. The time frame of the study spanned from January 1992 to December 2009. The outcomes of the test data forecast demonstrated that linear GARCH-type models performed better for short-term price predictions, while nonlinear GARCH models exhibited superior performance for long-term price predictions.

2.3.2 ARIMA-GARCH models

Xiang and Zhuang (2013) utilized ARIMA with an order of (1,1,1) to analyse the price changes of Brent oil. Their investigation spanned from November 2012 to April 2013.The findings indicated that ARIMA model exhibited a strong predictive capability, particularly for short-term forecasts of crude oil prices. This suggested that the ARIMA model, with its specific parameters, effectively captured the underlying patterns and dynamics of the Brent oil market during the specified time frame, enhancing the prospects for reliable price predictions.

Mohammadi and Su (2010) assessed the efficacy of various Autoregressive Integrated Moving Average (ARIMA), ARIMA-GARCH models for modelling and predicting the conditional mean and volatility of weekly crude oil prices on international markets from January 1997 to October 2009. They focused on four volatility models, such as GARCH, Exponential GARCH (EGARCH), Asymmetry Power GARCH (APARCH), and FIGARCH, to assess their test data forecasting performance from January 2009 to October 2009. Results revealed that the APARCH model showed superior predictive ability compared to the others. He et al. (2010) used a vector error correction mechanism and transfer function framework to examine the behaviour of the WTI crude oil price over a ten-year period. In terms of out-of-sample forecast accuracy, the results revealed that both models outperformed the naive random walk and univariate ARIMA models.

Abledu and Kobina (2012) attempted to estimate and analyse the macroeconomic impact of crude oil price fluctuations in Ghana using annual data from 2000 to 2011. The authors focused on forecasting feasibility using the nested conditional mean, ARIMA, and conditional variance, GARCH, families of models under such volatile market conditions. A regression-based forecast filtering simulation was presented and examined for any improvements in the predicted outcomes.

Xiang (2022) coupled ARIMA and GARCH for analysing and forecasting international crude oil prices. Simulation outcomes showed that the combination model of ARIMA (1,10) - GARCH (1,1) was more suitable for short-term forecasting of international crude oil prices with higher accuracy. Ng'ang'a and Oleche (2022) sought to provide an in-depth analysis of the forecasting abilities of various GARCH

models in order to identify the optimal GARCH model for crude oil price forecasting. The analysis examined Brent crude oil price data over a period of ten years, from 2011 to 2020. The study found that the Integrated GARCH (IGARCH) model was the best model out of five models for crude oil price prediction based on Root Mean Square Error (RMSE).

Most economic models like ARIMA, GARCH and their variants assume a linear relationship between past and future prices. However, crude oil prices exhibit nonlinearity due to chaotic and complex nature of oil market. This makes econometric models are less effective in capturing non-linear dynamics of price fluctuations. While GARCH models are widely used to predict volatility, they may struggle to capture sudden and extreme price changes caused by some factors like supply-demand imbalances. Furthermore, crude oil prices are influenced by multiple factors such as political instability, global economic conditions, and technological advancements. Econometric models rely on past prices and fail to incorporate external non-financial variables that can significantly affect crude oil prices. Advances econometric models like FIGARCH and EGARCG may provide more accuracy but often at the expense of overfitting where the model becomes overly complex and loses its predictive power on new data. This leads to model that may perform well in-sample but poorly when predicting future data. In summary, while econometric models have been beneficial for crude oil prediction, their limitations in handling non-linearity and non-stationarity have led to the increasing adoption of non-linear models.

2.4 Review of papers on predicting crude oil price using non-linear models

Although the statistical approaches improve the accuracy of forecasting crude oil prices to a certain degree, this improvement is restricted due to research studies demonstrating that the linearity assumption of crude oil prices is not fulfilled. This limitation hampers the overall accuracy of such forecasts. As a result, a range of Artificial Intelligence (AI)-based approaches has been put forward to address the complexities of nonlinearity and non-stationarity inherent in crude oil prices Wang et al. (2004), Baruník and Malinská (2016), Chen et al. (2017).

Computational techniques represent non-linear methodologies for forecasting that have gained increasing popularity. The proponents of computational techniques consider traditional methods of forecasting, such as time-series methods, unsuitable for highly nonlinear and chaotic time-series data, like oil prices. This is because conventional methods assume that the time-series data stem from linear processes. A primary computational method applied to oil price forecasting involves ANNs. Recent investigations into ANN reveal their capabilities in pattern classification and recognition problems. ANNs draw inspiration from human brain biology, enabling them to learn and generalize from experiences. Currently, ANNs find application in diverse fields across business, industry, and science.

AI models are designed to handle non-linear relationship in data. This makes them more effective in capturing complex and dynamic inherent in crude oil prices.AI models can easily incorporate a wide range of variables such as macroeconomic variables, market sentiments, and geopolitical factors, which play vital roles in crude oil price movements. AI models do not require strict assumptions about stationarity,

making them suited for dealing with the evolving and non-stationary nature of crude oil prices. In addition to this, AI models have the ability to learn and adopt from new data, which makes them more robust over time. This is crucial in the oil market, when new trends or policy changes can dramatically shift price patters.

2.4.1 Artificial neural network based models

ANNs mimic the human brain's learning ability which have been frequently employed to address nonlinearities in crude oil price data. Kulkarni and Haidar (2009) employed a Multilayer FFNN (MFFNN) to predict short-term trends in crude oil prices and explored the effectiveness of futures prices in predicting spot prices. The dataset comprises daily WTI crude oil spot prices spanning from 1996 to 2007, as well as prices 1-,2-,3-, and 4-month WTI futures contracts. The findings suggested that future prices offered novel insights into spot prices, particularly in the context of 1- and 2-month futures contracts.

Kadkhodaie-Ilkhchi et al. (2009) employed a Committee Machine with Training Algorithms (CMTA) that combined Levenberg-Marquardt (LM), Bayesian Regularisation (BR), Gradient Descent (GD), One-Step Secant (OSS), and Resilient Back-propagation (RP) algorithms. The results of this study revealed that the CMTA produced more reliable and acceptable outcomes than each of the individual neural networks with different training procedures.

Huang and Wang (2018) suggested a model to improve the prediction accuracy of crude oil price fluctuations. In this model, Wavelet Neural Network (WNN) and random-time effective functions were combined. WNN was a predictive system that

could perform strong nonlinear approximations. The random time effective function was used to formulate the variable impact of previous data on the current market, which endows historical data with time-variant weights to make them affect the training process of WNN. The empirical experiments were implemented in predicting crude oil prices as well as the moving average absolute return series of WTI and Brent. The empirical results showed that the proposed model is more accurate at predicting changes in the price of crude oil than the conventional BPNN, SVM, and WNN models and is advantageous in terms of enhancing prediction precision.

Jin and Xu (2024) presented a forecasting model via ANNs to anticipate feature crude oil prices. In this approach, four different models such as regression tree, SVR, autoregressive-GARCH, and BPNN were employed to predict future price of WTI and Brent. Daily prices from 2011 to 2021 were used of validation. Effectiveness of these models were assessed by computing RMSE. Experimental results demonstrated that BPNN model gave better performance than other models due to its ability to handle non-linear data.

2.4.2 Support vector machine based models

SVM has been applied to overcome limitations in traditional models, particularly in handling non-linear relationships. Xie et al. (2006) introduced a non-linear approach to forecasting crude oil prices using SVM. The method consisted of several steps: data sampling, sample preprocessing, training, and out-of-sample forecasting to build the SVM for time series prediction. The study evaluated the SVM's predictive capabilities by comparing it with ARIMA and Back Propagation Neural Network (BPNN) methods. Experimental results showed that SVM surpassed

both ARIMA and BPNN, indicating it is a strong candidate for accurate crude oil prediction.

Forecasting crude oil prices is gaining popularity around the world. This interest stems primarily from the economic value of the product. Ahmed and Shabri (2012) suggested an SVM-based technique for forecasting crude oil prices. In this approach, the data was standardized to a common range between 0 and 1. For its experimental purposes, the study used daily WTI crude oil prices from January 1986 to September 2006. The study used RMSE and Mean Absolute Error (MAE) metrics to assess the model's performance. In terms of error measurements, the model's effectiveness was compared to that of the ARIMA and GARCH models. Results showed that the method outperformed the other models.

Yu et al. (2017) designed a model to predict crude oil prices that is based on SVM. Monthly WTI crude oil data spanning from January 1990 to July 2008 was gathered and then normalized using the min-max technique. The normalized data was split into training and testing samples. Finally, SVM was trained and employed for forecasting the future prices of test data. The authors compared the model's performance with earlier models such as FFNN, ARIMA, and the Random Walk (RW) model. Results showed that the SVM model outperformed the other models in terms of a lower error rate.

2.4.2 Decomposition based models

Another strategy to anticipate crude oil prices is to decompose the price before building prediction models. Several studies have focused on combining decomposition

techniques such as Wavelet Analysis (WA) Jammazi and Aloui (2012), Reboredo et al. (2013), Shahbaz et al. (2015), Empirical Mode Decomposition (EMD) Zhang et al. (2008), Wang et al. (2014), Fan et al. (2016), Singular Spectral Analysis (SSA) Wen et al. (2014), and Variational Mode Decomposition (VAD) Dragomiretskiy and Zosso (2014) with ANN and other machine learning models to improve prediction rate.

Crude oil price prediction has usually proved to be an intractable task due to the intrinsic complexity of oil market mechanism. Jammazi and Aloui (2012) combined WA and ANN to forecast monthly crude oil prices of WTI from January 1988 to March 2010. Three different activation functions namely sigmoid, bipolar sigmoid, and hyperbolic tangent were used to test model's flexibility. Results proved that ANN with sigmoid activation function could significantly improve prediction accuracy compared to ANN with other activation functions.

Mingming and Jinliang (2012) formulated a new model known as Multiple Wavelet Recurrent Neural Network (MWRNN) to investigate the dynamics of crude oil prices. The utilization of WA was aimed at capturing the intricate multiscale attributes present within the data, while the RNN was adopted to replicate the behavior of crude oil prices via simulation. The outcomes demonstrated that the model had higher prediction accuracy.

In the year 2014, Shabri and Samsudin (2014) designed a model to forecast crude oil prices using WA and ANN. Authors applied Mallat Wavelet Transform (WT) to break down historical data into two main components: approximation and detailed series. Afterward, they aggregated the significant subseries and used them as inputs for the ANN to facilitate the prediction of crude oil prices. Performance of the WA-

ANN model was assessed by computing RMSE and MAPE. Results showed that the model exhibited a minimal level of prediction error.

Yu et al. (2008) utilized a multiscale neural network named Empirical Mode Decomposition-Feed Forward Neural Network-Adaptive Linear Neural Network (EMD-FFNN-ALNN) as an alternative to the single scale neural network for forecasting Brent and WTI crude oil prices. This model incorporated an EMD approach for breaking the original price into various intrinsic modes of different classes. By using a FFNN, they captured internal correlation patterns among different components. Eventually, selected vital subseries served as inputs for an ALNN to perform predictions. The study employed daily data from January 1986 to December 2006.Outcomes indicated that the EMD-FFNN-ALNN demonstrated superior performance compared to single scale neural network.

Yu et al. (2015) predicted the monthly Brent and WTI monthly data from January 1986 to July 2014, based on divide and conquer strategy. Initially, Ensemble Empirical Model Decomposition (EEMD) was applied to decompose the historical data into multiple series. Subsequently, component reconstruction based on the Data Characteristic Driven (DCD) modelling was introduced for capturing inner factors. In the next step, ANN, and Least Square Support Vector Regression (LSSVR) were employed to forecast the future values. Results demonstrated that the performance of the EEMD-DCD-ANN and EEMD-DCD-LSSVR models superior to all benchmark models, such as ANN and EEMD-ANN, in terms of accuracy and time.

Sun et al. (2018) introduced an Interval Decomposition Ensemble (IDE) learning model for forecasting interval-valued crude oil prices. The model had four

processes, such as complex-valued signal creation, decomposition, individual forecasting, and ensemble-based forecasting. Primarily, the original crude oil price was converted into a complex-valued signal. Subsequently, the complex-valued underwent decomposition, resulting in a set of complex-valued IMFs and a residual component, achieved using the Bivariate Empirical Mode Decomposition (BEMD) technique. Finally, Multilayer Perceptron with Inputs (MLPI) model was employed for forecasting oil prices. Performance of the model was tested using daily, weekly, and monthly WTI crude oil prices. Results showed improved performance compared with other methods, such as ANN, in terms of Average Relative Variance (ARV).

Lahmiri (2016) presented a time series forecasting model which fuses VMD and General Recurrent Neural Network (GRNN). Authors compared the forecasting accuracy of crude oil prices using VMD-GRNN model against EMD based approaches, FFNN and ARIMA. Experiments with different statistical criteria clearly showed that VMD-GRNN model significantly attained the lowest forecasting error compared to EMD-FFNN and EMD-ARIMA.

Crude oil price forecasting has gained attentions due to its significant role in the global economy. Accurate crude oil price forecasts often lead to a rapid new production development with higher quality and less cost. Wang et al. (2018) created a semi-heterogeneous approach to combining crude oil price forecasts, which interacts a set of decomposition methods with a set of forecasting techniques. Original series was decomposed into multiple series using four decomposition methods, such as WT, SSA, EMD, and WMD. Then, four different forecasting techniques, such as Autoregressive models, ARIMA, ANN, and SVR were adopted to forecast the

components from each decomposition methods. Finally, the predicted price was reconstructed the price forecasts from the forecasted components. Semi-heterogeneous approach offered better performance than WT-ARIMA.

Zhang et al. (2021) constructed a hybrid model to predict crude oil prices by using the two-layer decomposition technique. In this approach, EEMD was used to break down the residual term, which contains intricate information after VMD. This was then fused with a Kernel Extreme Learning Machine (KELM), resulting in the VMD-EEMD-KELM. To assess the model's effectiveness, experiments were conducted on Brent and WTI crude oil prices for the period from May 1987 to December 2019. The empirical findings demonstrated that the model improved the accuracy of crude oil price forecasts.

2.4.3 Hybrid models

Azadeh et al. (2012) used a flexible approach based on ANN and Fuzzy Regression (FR) to anticipate oil prices in noisy, uncertain, and complex environments. It is concluded that the selected ANN models considerably outperformed the FR models in terms of Mean Absolute Percentage Error (MAPE).

Chiroma et al. (2016) conducted a comprehensive review of the existing literature pertaining to the prediction of crude oil prices. Their analysis revealed a growing enthusiasm within the research community for employing AI approaches to tackle the complexities linked to predicting crude oil prices. Fan et al. (2016) presented an approach for forecasting crude oil prices using Independent Component Analysis (ICA) and Support Vector Regression (SVR). The authors introduced ICA-SVR,

which involves decomposing the oil prices into three independent components using ICA. These components were then individually forecasted using SVR. Through a comparative study, the authors showed that the ICA-SVR model outperformed two other models effectively forecasting crude oil prices.

Mostafa and Masry (2016) attempted to forecast oil prices from January 1986 to June 2012 utilizing evolutionary techniques such as Gene Expression Programming (GEP) and ANN models. To test evolutionary model, the ARIMA model was employed to benchmark evolutionary models. The results revealed that the GEP technique exceeded ARIMA in terms of the Mean Square Error (MSE), RMSE, and MAE in predicting oil prices.

Deng et al. (2019) proposed a fusion model that combined Multiple Timeframes Dynamic Time Wrapping (MDTW) and GA to predict changes in both daily and weekly crude oil data. In this model, MDTW was used to effectively gather and match similar time series, whereas GA was used for optimization, ensuring that the model's parameters were fine-tuned to yield the most accurate predictions. The outcomes of these investigations showed notable advancements in predictive performance.

Abdollahi and Ebrahimi (2020) built a robust hybrid model that combines the Adaptive Neuro Fuzzy Inference System (ANFIS), Autoregressive Fractionally Integrated Moving Average (ARFIMA), and Markov-switching models. These models were chosen for their ability to capture specific features such as nonlinearity, lag, and market interrelationships present in oil price time series. The hybrid model assigned weights to each constituent model to enhance prediction accuracy, employing three

weighting scenarios: equal weights, error-value-based weights, and a genetic algorithm weighting function. The robustness and prediction quality of the hybrid model were compared to other models using the Diebold-Mariano (DM) test. The results demonstrated that the hybrid model weighted by a genetic algorithm generally outperformed the other approaches, including the constituent models and other hybrid weighting strategies.

2.4.4 Deep learning based models

A deep learning model, Deep Belief Network (DBN) and ARIMA were combined to capture unknown complex pattern of the crude oil prices by Chen et al. (2017). The combination involved using ARIMA used to capture linear data characteristics, while DBN was used to capture the non-linear data features. To evaluate the model, historical daily WTI crude oil prices spanning from July 2007 to February 2017 was used. Results indicated that the ARIMA-DBN approach exhibited a lower MSE compared to using the than ARIMA model. Nevertheless, it was observed that the performance of the model was sensitive to parameters.

To tackle the complexity of forecasting crude oil prices due to the multiple of influencing factors, Zhao et al. (2017) created a deep learning model. The model incorporated two main techniques: 1) Stacked Denoising Auto encoders (SDAE) to capture intricate relationship between oil prices and its influencing factors, and 2) bootstrap aggregation method to create multiple datasets for training a set of SADEs. In the empirical analysis, the method was applied to WTI crude oil prices along with 198 economic series used as exogenous variables.

Zhou et al. (2019) introduced a novel method known as Complete Ensemble Empirical Model Decomposition with Adaptive Noise and Extreme Gradient Boosting (CEEMDAN-XGBOOST). The method involved decomposing the complex and non-stationary crude oil price into Intrinsic Mode Functions (IMFs) and a residue using CEEMDAN. Subsequently, XGBOOST was employed to individually predict each IMF and residue individually. The residue and each IMF's forecast findings were aggregated to get the final prediction outcomes. WTI datasets were used to validate the model's performance.

In the year 2019, Cen and Wang (2019) established a model that employed Long-Short Term Memory (LSTM) for the prediction of Brent and WTI crude oil prices. They applied EEMD to divide time series into multiple IMFs, which were then utilized as training data for the LSTM model. Results showed improved accuracy. However, deep learning model requires a larger volume of data for effective training.

Wang and Wang (2021) created a hybrid model called WPD-SW-LSTM, which combines Wavelet Packet Decomposition (WPD), LSTM, and the Stochastic time-effective Weight (SW) function. The framework utilized WPD as a signal processing tool to decompose the original time series into subseries with distinct frequency components. The SW-LSTM model was designed and trained to perform prediction task. The performance of the model was compared with SVM, BPNN, WPD-BPNN, and WPD-LSTM in terms of error metrics.

Yan et al. (2021) established a model using ML to anticipate global crude oil prices with reduced dimension. The authors used several macroeconomic variables such as supply, demand, gold price, natural gas, silver price, and financial

factors like dollar exchange rate, and rubble exchange rate as input variables. They employed Principal Component Analysis (PCA), Multidimensional Scaling (MDS), and Locally Linear Embedding (LLE) techniques to reduce data's dimensionality. Prediction model was designed using LSTM. Through a comprehensive analysis and comparison outcomes.

LSTM and Bi-LSTM models were developed by Daneshvar et al. (2022) to predict Brent crude oil prices. Primarily, the database was generated using daily data from the financial markets and the price of Brent crude oil for the period January 2015 to March 2021. The performance of the models was analyzed by varying the number of layers and solvers. Results showed that the model with two layers of LSTM and SGDM had less error and better accuracy.

Jovanovic et al. (2022) adopted a decomposition-forecasting strategy to forecast the daily WTI crude oil price. The authors introduced a modified version of Salp Swam Optimization (SSO) to optimise the LSTM parameters. The proposed approach was assessed using real-world WTI price data via two sets of experiments: one utilising the original time series and the other applying VMD to the time series. Through comparative analysis, the study concluded that the model exhibited great promise in crude oil price forecasting.

Foroutan and Lahmiri (2024) implemented several deep learning models such as LSTM, BiLSTM, Convolutional Neural Network (CNN), and CNN-BiLSTM to predict future prices of WTI and Brent. These models were tested using daily prices of WTI and Brent from 2000 to 2022. Before applying the data to the models, the prices were normalized using min-max method to improve prediction

performance. While deep learning models reported high prediction performance, they require more training data and finely tuned hyperparameters for accurate predictions.

Tables 2.1 and 2.2 compile a summarized overview of the reviewed articles related to econometric and non-linear models, respectively. Table 2.3 provides a consolidation of the metrics employed to assess the efficiency of crude oil prediction models.

Crude oil price forecasting has evolved from econometric models to non-linear and hybrid models, driven by the need to address the complex, non-linear, and chaotic nature of oil prices. While traditional econometric models like ARIMA and GARCH remain useful. AI based models have significantly enhanced prediction accuracy. These advancements highlight the shift toward more sophisticated methods in response to the growing complexity of crude oil prices.

Table.2.1 A summary of recent studies utilizing econometric models to predict crude oil price

Researchers	Year	Predictor	Data		Data frequency			Period
			Brent	WTI	Daily	Weekly	Monthly	
Cheong	2009	ARCH	•	•	✓			Jan 1993-Dec 2008
Kang et al.	2009	CGARCH, FIGARCH	•	•	✓			Jan. 1996-Dec.2006
Wei et al.	2010	GARCH	•	•	✓			Jan.1992-Dec.2006
Mohammadi and Su	2010	GARCH, FIGARCH, APARCH	•			✓		Jan1997-Feb 2009
Hou and Suardi	2012	Non-parametric GARCH	•	•	✓			Jan 1992-Dec 2010

Xiang and Zhuang	2013	ARIMA	•		✓			Nov 2012-Apr 2013
Yu et al.	2015	Divide and conquer	•	•			✓	Jan 1986-Jul 2014
Wang et al.	2018	Semi-heterogeneous		•			✓	Mar1983-Dec 2016
Ng'ang'a and Oleche	2022	IGARCH	•		✓			Jan 2011-Dec 2020
Xiang	2022	ARIMA-GARCH		•	✓			Jul 2021-Dec 2021
Foroutan and Lahmiri	2024	CNN, LSTM, BiLSTM	•	•	✓			Jan 2000-Mar 2022

Table.2.2 A summary of recent studies utilizing non-linear models to predict crude oil price

Researchers	Year	Input	Predictor	Data		Data frequency			Period
				Brent	WTI	Daily	Weekly	Monthly	
Xie et al.	2006	-	SVM		•			✓	Jan.1997-Dec 2003
Yu et al.	2008	-	EMD-FFNN-ALNN	•	•	✓			Jan 1986- Dec 2006
Kulkarni and Harder	2009	7 previous days'	MFFNN		•	✓			Spe 1996-Aug 2007
Ahmed and Shabri	2014	5 previous days'	SVM	•	•	✓			Jan 1986-Sep 2006

Jammazi and Aloui	2012	-	WA-ANN		•			✓	Jan 1988-Dec 2010
Shabri and Samsudin	2014	6 previous days'	WA-ANN	•	•	✓			May1987-Sep 2006
Lahmiri	2016	-	VMD-GRNN		•			✓	Jan2008- Dec 2013
Mostafa and Masry	2016	-	GEP- ANN		•	✓			Jan1986-Jun 2012
Chen et al.	2017	6 previous days'	DBN		•	✓			Jul2007-Feb 2017
Yu et al.	2017	5 previous days'	SVM	•	•			✓	Jan1990-Jul 2008
Zhao et al.	2017	Macroeconomic variables	SADE		•			✓	Jan 1986-May 2016

Huang and Wang	2018	5 previous days'	WNN	•	•	✓			Jun 2008-Dec 2012
Sun et al.	2018		IDE	•	•	✓	✓	✓	Jan 1990-Jul 2017
Cen and Wang	2019	Open, low, high, close	LSTM	•	•	✓			Jan 2005-Oct 2017
Zhao et al.	2019	6 previous days'	CEEMDAN - XGBOOST		•	✓			Jan 1986-Mar 2018
Abdollahi and Ebrahimi	2020	Closing price	Hybrid	•		✓			Jan 2001-Oct 2019
Wang and Wang	2021	4 previous days'	WPD-SW-LSTM	•	•	✓			Jan 2009-Oct 2019
Yan et al.	2021	Supply, demand, gas,	LSTM		•	✓			Jan 1997-Feb 2019

		exchange rate, Gold, and Silver							
Daneshwar et al.	2022	Commodity and financial factors	LSTM, BiLSTM	•			✓		Jan 2015-Mar 2021
Jovanovic et al.	2022	6 previous days'	VMD-SSO-LSTM		•	✓			Jan 1986-Jul 2022
Foroutan and Lahmiri	2024	5 previous days'	CNN, LSTM, BiLSTM	•	•	✓			Jan 2000-Mar 2022

Table 2.3 Performance measuring metrics

Researchers	Year	Evaluation Metrics				
		MSE	MAE	RMSE	MAPE	D_{stat}
Xie et al.	2006			✓		✓
Yu et al.	2008			✓		✓
Cheong	2009	✓	✓			
Kang et al.	2009	✓	✓			
Kulkarni and Haider	2009	✓	✓	✓		
Wei et al.	2010	✓	✓			
Mohammadi and Su	2010	✓		✓		
Jammazi and Aloui	2012		✓	✓		
Hou and Suardi	2012	✓	✓			
Xiang and Zhuang	2013	✓	✓			
Ahmed and Shabri	2014	✓		✓		
Shabri and Samsudin	2014			✓	✓	
Yu et al.	2015			✓	✓	✓

Mostafa and Masry	2016			✓	✓	
Lahmiri	2016		✓	✓	✓	
Chen et al.	2017	✓				
Yu et al.	2017			✓		✓
Zhao et al.	2017			✓	✓	✓
Huang and Wang	2018	✓	✓	✓		
Sun et al.	2018	✓				
Wang et al.	2018		✓	✓	✓	
Cen and Wang	2019		✓	✓	✓	
Zhou et al.	2019		✓	✓		✓
Abdollahi and Ebrahimi	2020		✓	✓	✓	
Wang and Wang	2021		✓	✓	✓	
Yan et al.	2021		✓	✓		
Zhang et al.	2021	✓	✓	✓		
Daneshwar et al.	2022	✓		✓		
Jovanovic et al.	2022	✓		✓		

Ng'ang'a and Oleche	2022			✓		
Xiang	2022			✓	✓	
Foroutan and Lahmiri	2024		✓	✓	✓	

2.5 Research gaps identified

Crude oil prices play a vital role in the global economy and constitute are a crucial factor impacting government sectors and commercial sectors. Having proactive knowledge about the future changes in these prices can lead to improved decision-making across various managerial levels. However, predicting the movement of oil prices is a complex task due to their highly volatile and unpredictable dynamic patterns, characterized by non-linear and chaotic behavior. Despite use of various statistical and econometric models aimed at addressing this challenge, forecasting oil prices remains difficult due to the inherent difficulties associated with the dataset.

Reviewing the existing literature on crude oil forecasting reveals a shift towards the adoption of non-linear tools such as ML and neural network. It is noted that non-linear models outperform conventional methods in accuracy, as evidenced by the effectiveness demonstrated in relevant studies. Despite these advancements, there is still a lack of focus on choosing input variables and optimizing the construction of the neural network architecture in most major research efforts. The research gaps have been identified from the reviewed articles are as follows:

- ❖ Econometric models can forecast prices based on stationary and linear assumptions in the original data, but they are unable to capture the non-linearity and complex pattern in crude oil prices.
- ❖ Presently, no investigations into crude oil price prediction have delved into the potential of incorporating technical indicators to forecast future crude oil prices.
- ❖ A limited number of studies employing ML techniques for crude oil price prediction have demonstrated notably higher prediction accuracy compared to econometric models.
- ❖ While numerous models have been designed for crude oil prediction, majority of these methods have focused solely on utilizing historical data as an input feature.
- ❖ Existing predictive models have considered macroeconomic variables as input futures for crude oil price prediction. Nevertheless, these methods have yet to explore the impact of the most influential factor.
- ❖ Existing approaches have not maintained consistent predictive effectiveness across different market data.
- ❖ For validation, existing approaches have considered any of these data, whether they are daily, weekly, or monthly. As a result, these models are unsuitable for making long-term predictions.

This research seeks to address these gaps by implementing ML techniques to boost predictive accuracy and improve the prediction of crude oil prices. Through this research, the impact of influential factors, the potential of technical indicators, and the power of ML methods are all explored to contribute more robust prediction models.

2.6 Summary

This chapter has provided an extensive examination of crude oil price prediction models. An effort has been exerted to emphasize the features of various researchers engaged in the domain of crude oil price forecasting. The prevailing techniques have primarily relied on either historical data or macroeconomic variables to forecast crude oil prices. A limited number of researchers have used ML techniques to enhance prediction accuracy. Furthermore, specific gaps in existing research have been identified. The following chapter delves into the methodologies employed in this research.

CHAPTER 3

RESEARCH DESIGN

3.1 Conceptual framework

A research framework, often referred to as a conceptual framework, is a structured outline or a theoretical model that provides the foundational structure for a research study. In the context of research on crude oil price prediction, this conceptual framework plays a pivotal role in delineating step-by-step processes required to develop an effective predictive model. Figure 3.1 depicts the visual representation of this framework, encompasses eight major processes, each contributing indispensably to the goal of forecasting crude oil prices. These processes include:

1. Data collection
2. Variables selection
3. Preprocessing
4. Model selection
5. Model training
6. Testing
7. Post processing and
8. Evaluation

Data collection

Data collection process revolves around the gathering of historical data pertaining to selected crude oil prices. Additionally, macroeconomic variables and

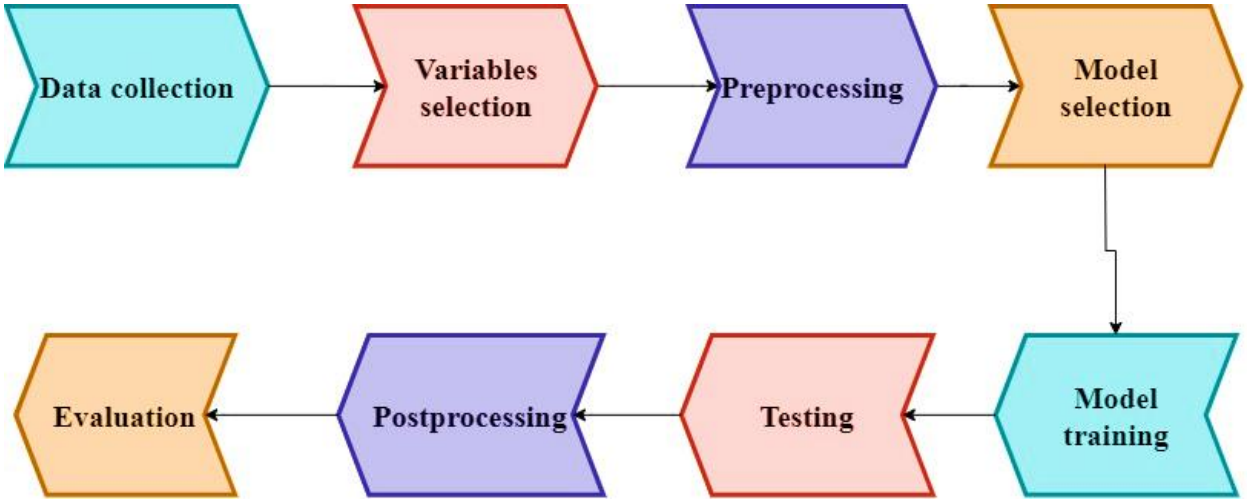


Figure 3.1 Conceptual framework

other variables directly linked to dynamic of the crude oil market. These data sources provide the raw materials essential for model development. In this research, Historical data and macroeconomic variables related to Brent and WTI are collected.

Variables selection

The variables selection unit is responsible for determining key input and output variables essential to the crude oil prediction model. This step involves a critical assessment of which factors are most pertinent to forecasting crude oil prices. In this research work, past crude oil prices, macroeconomic variables, and technical indicators are considered input variables and future crude oil prices are the output variables.

Preprocessing

The collected data often requires extensive preparation. The preprocessing unit is responsible for this task. Data undergoes a series of techniques. Cleaning processes

are implemented to address any inconsistencies or errors. Missing values are handled, ensuring data integrity is maintained. Additionally, normalization techniques may be applied to standardise variables, making them compatible for modelling. The min-max technique is used for normalization.

Model selection

xA fundamental aspect of predictive modelling is choosing the appropriate Machine Learning (ML) model. In this process, various ML algorithms are explored, seeking the most suitable one for the specific task of crude oil price prediction. This choice will greatly influence the model's accuracy and effectiveness. By reviewing previous papers, Feed Forward Neural Network (FFNN) and Support Vector Machine (SVM) are selected.

Model training

Once the model is selected, it enters the training phase. During this process, the model is exposed to the training data, where it learns to capture the intricate relationships between the chosen input and output variables. The training process is fundamental to building a predictive model that can generalize its insights to make future price predictions.

Testing

The efficacy of the trained model is assessed in this phase using separate test data. This step ensures that the model's performance extends beyond mere training and can generalize to make accurate predictions on new, unseen data.

Post processing

Post processing refers to the application of techniques to the predicted values to refine and align them with actual values. This step is vital for ensuring that the predictions are actionable and reliable for decision-making.

Evaluation

Finally, the performance of the predictive model is subjected to evaluation. Various performance metrics such as mean square error, prediction accuracy, and mean absolute error are employed to assess the predictive model's accuracy and reliability. This evaluation phase provides crucial insights into the model's strengths and areas for improvement.

Theoretical framework

Independent variables	**Dependent variables**
Historical stock price Macroeconomic variables Technical indicators	Future crude oil price

3.2 Research hypotheses

The following hypotheses have been formulated in response to the research questions outlined.

H0 -Null hypothesis

H1-Alternative hypothesis

The first research question aims to ascertain the relationship between past prices of Brent and WTI crude oil and future crude oil prices. This is addressed by hypothesis 1:

H0: There is no significant relationship between historical prices of Brent and WTI crude oil and future crude oil prices.

H1: There is a significant relationship between historical prices of Brent and WTI crude oil and future crude oil prices.

The second research question explores the relationship between macroeconomic variables and Brent and WTI crude oil market prices, as answered by hypothesis 2:

H0: Macroeconomic variables do not have a significant impact on Brent and WTI crude oil prices

H1: Macroeconomic variables have a significant impact on Brent and WTI crude oil prices

The third research question delves into the dynamic relationship between technical indicators and Brent and WTI crude oil prices, as addressed by the hypothesis 3:

H0: Technical indicators do not exhibit dynamic relationship with Brent and WTI future crude oil prices.

H1: Technical indicators exhibit a dynamic relationship with Brent and WTI future crude oil prices.

The fourth research question investigates whether the proposed ML models can accurately predict the future Brent and WTI crude oil prices, as answered by hypothesis 4:

H0: The proposed ML models cannot accurately predict future crude oil prices

H1: The proposed ML models can accurately predict future crude oil prices

The fifth research question examines the impact of different input features on future crude oil prices, as addressed by hypothesis 5:

H0: There is no significant difference in prediction performance between different sets of input features

H1: There is a significant difference in prediction performance between different sets of input features

The sixth research question investigates whether there is a difference between training and testing accuracy, as answered by hypothesis 6.

H0: The percentage of prediction accuracy for training data and testing data is not significantly different.

H1: The percentage of prediction accuracy for training data is significantly different form that of testing data.

3.3 Description of data sets

In this research work, the primary focus revolves around the analysis of two prominent crude oil prices, specifically Brent and WTI crude oil prices. WTI crude oil, a prevalent variety of crude oil found in North America, has gained prominence as the standard for global crude oil pricing due to its extensive utilization in the United States. Similarly, Brent crude oil, derived from the Brent oilfield, serves as the international benchmark for oil pricing and is utilized in setting prices for approximately two-thirds of the world's crude oil supplies. While other types of crude oil, like Dubai crude oil and Daqing crude oil, do exist, they tend to closely track the movements of WTI or Brent oil and are thus referenced to less frequently Huang and Wang (2018). The rationale behind the selection of these prices is rooted in their widespread recognition as benchmark prices within the area of crude oil pricing. Numerous studies, such as those conducted by Yu et al. (2008), Khashei and Bijari (2011), Hu et al. (2012), and Shabri and Samsusin (2014), have substantiated the significance of these indices by incorporating them as fundamental components in various crude oil price formulas.

To facilitate a comprehensive analysis, data spanning from January 1, 2001, to February 28, 2023, has been gathered for both Brent and WTI. This dataset encompasses daily, weekly, and monthly price records, excluding weekends and public holidays to ensure data integrity and consistency. The crude oil markets chosen for experimentation are listed in Table.3.1. To gain insight into the characteristics of the Brent and WTI crude oil prices, descriptive statistics are presented in Table 3.2. Notably, Table 3.2 highlights significant disparities between the minimum and maximum prices, revealing the inherent volatility in these markets. The mean values of both datasets exceed the medians, signifying a positive skew in the distribution. Both datasets exhibit a substantial Standard Deviation (SD), emphasizing the significant price fluctuations. The skewness of WTI on a daily, weekly, and monthly basis as well as Brent on a daily, weekly, and monthly basis is greater than zero, while the kurtosis is less than 3. These observations indicate a positively skewed or right-skewed distribution with a long right tail and platykurtic characteristics. Additionally, the Jarque-Bera (JB) statistic test yields a value larger than that of the normal distribution, confirming the unique nature of the crude oil market's dynamic. Overall, the chosen data points exhibit non-linear and non-stationary characteristics, which implies that any model construction should consider data preprocessing techniques.

Table 3.1 Selected crude oil markets

Crude oil market	Data frequency	Period
Brent	Daily	01/01/2001-28/02/2023

	Weekly	01/01/2001-28/02/2023
	Monthly	01/01/2001-28/02/2023
WTI	Daily	01/01/2001-28/02/2023
	Weekly	01/01/2001-28/02/2023
	Monthly	01/01/2001-28/02/2023

Table 3.2 Descriptive statistic of WTI and Brent crude oil prices

Parameters	**WTI crude oil price**			**Brent crude oil price**		
	Daily	**Weekly**	**Monthly**	**Daily**	**Weekly**	**Monthly**
No. of samples	5629	1156	266	5629	1156	266
Minimum	8.91	16.55	3.32	9.12	14.24	18.38
Maximum	145.31	133.88	142.52	143.95	141.07	132.72
Mean	64.14	63.61	63.95	66.97	67.06	66.68
Median	61.44	61.36	61.23	63.82	64.02	63.97
SD	25.71	25.74	25.71	29.13	29.22	29.23
Skewness	0.280	0.277	0.269	0.297	0.288	0.282

J-B statistic	186.26	37.65	8.358	265.54	54.85	12.81

For a more visually informative representation of these price trends, Figure 3.2 has been constructed, this graphical depiction offers a clear visualization of the daily, weekly, and monthly fluctuations in both Brent and WTI crude oil prices, thereby providing a valuable information for further analysis and insights in this research work.

3.4 Variables for the research

This research work employs a multifaceted strategy for predicting future crude oil prices. It considers three different sets of input variables for the development of three predictive models, namely:

1. Historical data
2. Macroeconomic variables and
3. Technical indicators

Each model brings a distinct viewpoint, offering insights into the complex dynamics that influence the trajectory of oil prices. This comprehensive analysis aims to provide valuable insights to enhance the field of oil price prediction.

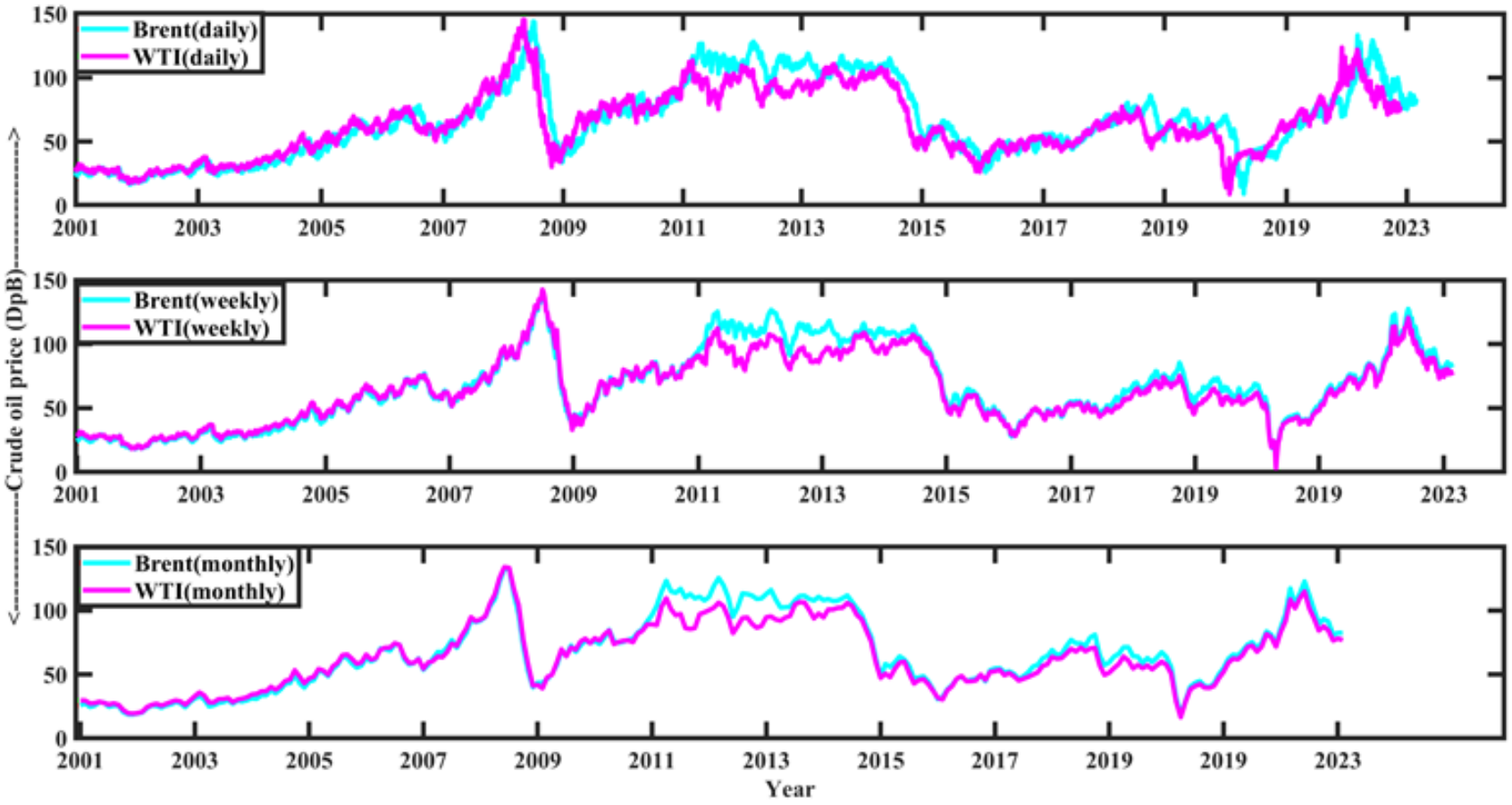


Figure 3.2 Crude oil prices from 2001-2023

Historical data

Historical data serves as the foundational input variable for the first predictive model. Specifically, this model relies on a lagged time series of past crude oil prices. The first model uses a lag of six previous points as input variables to anticipate oil prices one step ahead on daily, weekly, and monthly time frames for both Brent and WTI crude oil. The method leverages the principle that the past price trends can offer valuable insights into the potential trajectory of crude oil prices.

Macroeconomic variables

The second model adopts a macroeconomic perspective, incorporating seven distinct variables as input parameters. These variables encompass a wide range of economic indicators and market drivers. They include:

- **Gold Price (GP):** As a traditional safe-haven asset, the price of gold often influences investor sentiment and risk appetite, which can, in turn, affect oil prices.
- **Silver Price (SP):** Like gold, the price of silver can reflect market sentiment and economic conditions.
- **S&P 500 Index:** The performance of the S& P 500, a key equity market benchmark, can provide insights into broader economic trends and potential impacts on oil demand.
- **EUR/USD exchange rate:** Exchange rates can significantly influence the competitiveness of oil-exporting countries and consequently, oil prices.

- **Natural Gas Price (NGP):** Natural gas and crude oil are often interconnected in energy markets, with changes in one commodity's price affecting the other.
- **Dollar Index (DI):** The strength of U.S. dollar can have a substantial impact on oil prices, as crude oil is commonly priced in dollars.
- **GBP/USD exchange rate:** Fluctuations in the GBP/USD exchange rate can influence oil prices, especially for regions that trade oil in pounds.

These macroeconomic variables are employed to predict oil prices one step ahead on daily, weekly, and monthly timeframes for both Brent and WTI crude oil.

Technical indicators

The third predictive model adopts a more technical approach by incorporating 14 technical indicators as input variables. These indicators provide insights into price trends, momentum, and volatility. Some of the key technical indicators are given below:

- **Moving Average (MA):** A fundamental indicator, which smoothens price data to identify underlying trends.
- **Exponential MA (EMA):** Like MA bust assigns more weight to recent prices, making it sensitive to recent market movements.

- **Disparity and Exponential (ED):** Indicators that measure the deviation of the current price form a reference value, providing insights into market overbought or oversold conditions.
- **Price Oscillator (POSC) and Exponential POSC (EPOSC):** These indicators evaluate the price momentum and help identify potential turning points.

These technical indicators aim to predict oil prices one step ahead on daily, weekly, and monthly timeframes for both Brent and WTI crude oil.

3.5 Feed forward neural network

ANN is the widely used model for representing non-linear data. It is the subset of ML mimics the human brain's functionality by comprising numerous interconnected neurons. It is an emerging technology utilized in diverse fields such as forecasting weather conditions, anticipating stock prices, and predicting commodity trends, often achieving improved outcomes Cheng et al. (2022), Wang and Fang (2022). FFNN is a kind of feed forward, supervised network. Generally, FFNN consists of one input layer, one or more hidden layers, and a layer of computation output. The input layer is responsible for receiving input vectors, the hidden layers perform data processing, and the output layer generates the output values corresponding to the given input vectors. A general structure of the FFNN is illustrated in Figure 3.3. The input layer contains neurons equal in number to the independent variables used as attributes. The output layer serves as the network's output, representing the predicted value. Each subsequent layer employs the weights inherited from preceding layers and fine-tunes them to reduce the error between the actual and

predicted values for the dependent variable. Each neuron uses its respective input vectors and the associated weights to compute its output.

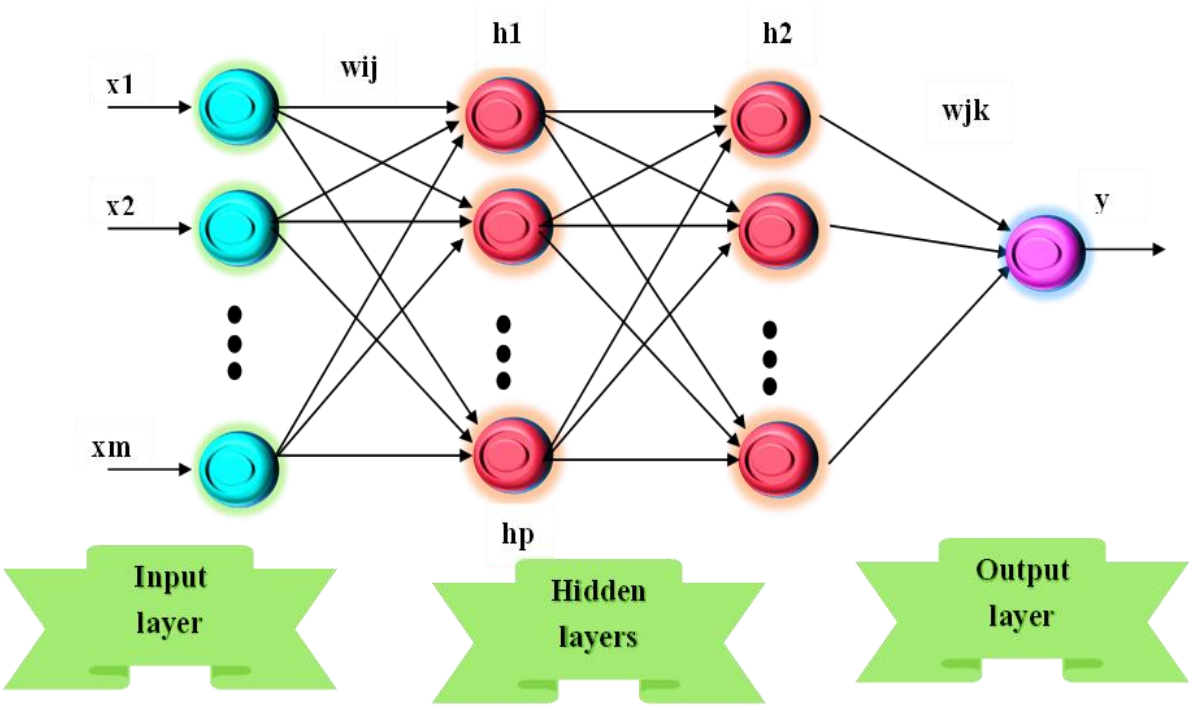


Figure 3.3 Structure of the FFNN

As in Figure 3.3, the input neurons, hidden, and output layers are denoted by the indices i, j, and k respectively. Each neuron in the input layer is connected to each neuron in the hidden layer via weight followed by the output layer. Each hidden neuron sums up the input vectors multiplied by the weight. Let's denote the following variables:

Input layer: X = [x1, x2, ...xi…xm]

Hidden layer: H = [h1, h2,…hj,…hp]

Output layer: Y= [y1, y2,..yk,…yr]

Where,

m- Number of input vectors

p-Number of hidden neurons in the hidden layer

r-Number of output neurons

Input to hidden layer output:

For each neuron hj in the hidden layer, the output can be computed as,

$$Net_{in} = \sum_{i=1}^{m} x_i w_{ij} + b_1 \quad (3.1)$$

$$h_j = f(Net_{in}) = f\left(\sum_{i=1}^{m} x_i w_{ij} + b_j\right) \quad (3.2)$$

Where,

Net_{in} – Weighted sum of input vectors

Xi-i^{th} input vector

W_{ij}-Weight between input and hidden layers

b-Bias for hidden neuron

f(.)-Activation function

hj- output of the j^{th} hidden neuron

The activation function, f(.) is required to activate the net and cause the j^{th} hidden neuron to produce an output. Sigmoidal activation is used in the hidden layer.

Hidden to output layer

For each neuron y_k in the output layer, the output can be computed as,

$$y_k = g\left(\sum_{j=1}^{p} w_{jk}\left(f(h_j)\right) + b\right) = g\left(\sum_{j=1}^{p} w_{jk}\left(f\left(\sum_{i=1}^{m} x_i w_{ij} + b_j\right)\right) + b\right) \quad (3.3)$$

Where,

W_{jk}-Weight between hidden and output layers

g(.) – Activation function

y_k- output of the k^{th} output neuron

b-Bias for output neuron

3.6 Support vector machine

SVM is a supervised machine learning model that can be used to solve both classification and regression problems. Vapnik (1995) introduced the concept of SVM. SVM operates based on the principle of Structured Risk Minimization (SRM), which involves reducing an upper limit of the generalization error, a departure from the approach used in other neural networks that typically focus on minimizing empirical error. It finds application in a range of tasks, with a focus on addressing classification and regression challenges. Notably, SVM exhibits its power in the area of time series forecasting, spanning form anticipating stock market prices to understanding chaotic systems Cao et al. (2005), Chen et al. (2006), Zhang et al. (2013), Ahmed et al. (2014). In time series analysis, SVM operates analogously to classification: it converts data into a higher-dimensional space and then establish a

maximum-margin hyperplane for separation. However, the key distinction lies in objective, which focus on identify a function capable of effectively predicting future prices.

Let's define a training dataset consisting of m points, $\{x_i, y_i\}_{i=1}^{m}$ with input data $x_i \in \Re^p$, where p represents the total number of data patterns and the output is $y_i \in \Re$. The concept of constructing a SVM to approximate a function entails the process of transforming the input data, x into a higher-dimensional features space via a nonlinear mapping and subsequently conducting linear regression within this feature space. The SVM regression function is formulated as follows:

$$\mathrm{y(x)} = \mathrm{w^T}\varphi(\mathrm{x}) + \mathrm{b} \tag{3.4}$$

Where,

$\varphi(x)$ – High-dimensional feature space

w and b – Coefficients

The coefficients, w and b are computed by minimizing the function given in Equation (3.7),

$$Min\left(\frac{1}{2}w^T w\right) \tag{3.5}$$

subject to the following constraints:

$$y_i - \mathrm{w^T}\varphi(x_i) - b \leq \varepsilon \tag{3.6}$$

$$\mathrm{w^T}\varphi(x_i) + b - y_i \leq \varepsilon \tag{3.7}$$

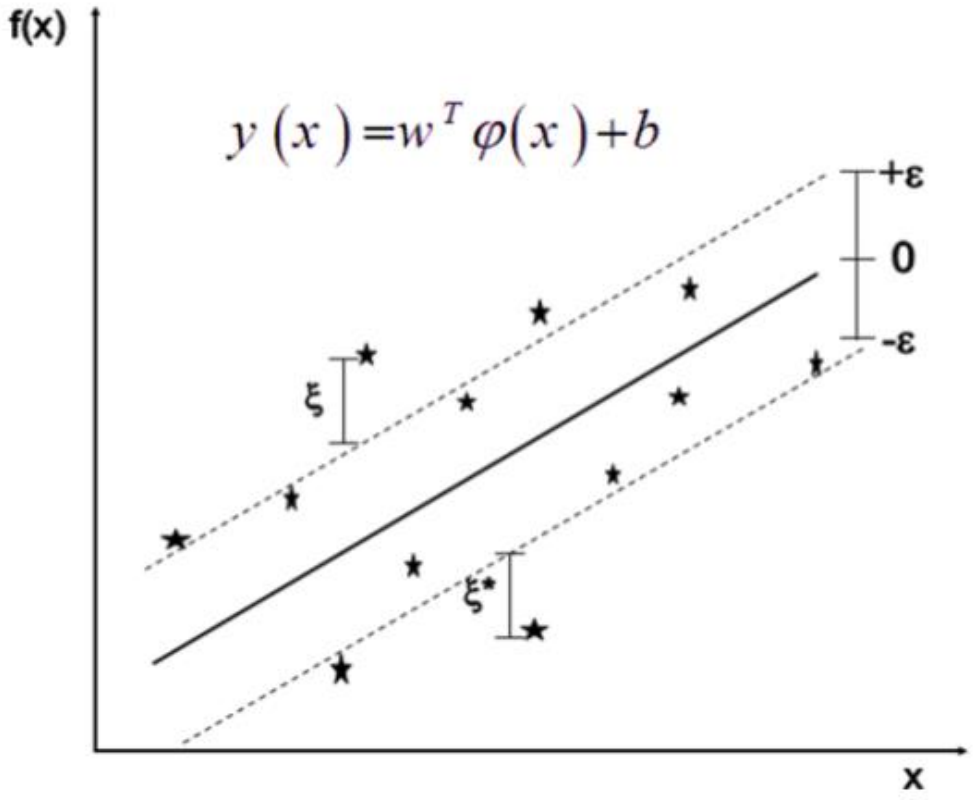


Figure 3.4 SVM hyperplane

This gives,

$$y_i(wx_i + b) - 1 \geq 0 \qquad i = 1,2, \ldots . m \qquad (3.8)$$

Figure 3.4 illustrate the SVM hyperplane. To calculate w and b, Equation (3.8) is converted to the prime function by introducing the positive slack variables ξ and ξ^* as follows:

Minimize

$$R(w, \xi, \xi^*) = \frac{1}{2} w^T w + C \sum_{i=1}^{N} (\xi_i + \xi_i^*) \qquad (3.9)$$

subject to the following constraints:

$$y_i - \mathrm{w}^{\mathrm{T}} x_i - b \leq \varepsilon + \xi_i \tag{3.10}$$

$$\mathrm{w}^{\mathrm{T}} x_i + b - y_i \leq \varepsilon + \xi_i^* \tag{3.11}$$

$$\xi_i, \xi_i^* \geq 0$$

Where,

$\mathrm{w}^{\mathrm{T}}\mathrm{w}$ – Weights vector norm

yi-Desired value

C-Regularization parameter

The term, ε is called the tube of size of SVM and is equivalent to the approximation accuracy placed on the training data points. Using Lagrange multipliers and exploiting the optimality constraints, the decision function represented by Equation (3.9) takes the following explicit form:

$$L = \frac{1}{2} w^T w + C \sum_{i=1}^{N} (\xi_i + \xi_i^*) - \sum_{i=1}^{N} \alpha_i (\varepsilon + \xi_i - y_i + w^T x_i + b) - \sum_{i=1}^{N} \alpha_i^* (\varepsilon + \xi_i^* - y_i + w^T x_i + b) - \sum_{i=1}^{N} (\eta_i \xi_i + \eta_i^* \xi_i^*) \tag{3.12}$$

with the following constrains:

$$\sum_{i=1}^{m} \alpha_i = \sum_{i=1}^{m} \alpha_i^*$$

$$0 \leq \alpha_i^* \leq C, \qquad i = 1,2, \ldots m$$

$$0 \leq \alpha_i \leq C, \qquad i = 1,2, \ldots m$$

This can be expressed as,

$$y(x) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*).K(x_i, x) + b \quad (3.13)$$

Where,

$K(x_i, x)$- Kernel function.

In this research work, SVM with RBF is used. The RBF can be expressed as,

$$K(x, x_i) = exp\left(-\frac{\|x-x_i\|^2}{2\gamma}\right),\ \gamma > 0 \quad (3.14)$$

Where, γ – Kernel parameter. Kernel parameter need to be chosen carefully, as they implicitly define the structure of the high dimensional feature space and thus controls the complexity of the final solution.

Predicting crude oil price is a very tough task due to the dynamic and volatile nature of the market, driven by multitude of factors including past price, geopolitical event, macroeconomic variables, and so on. In this research work, FFNN and SVM models are chosen due to their ability to handle such complex data. FFNN is capable of learning complex, non-linear relationships between input and output variables. Crude oil prices are influenced by non-linear interactions by many factors, making FFNN ideal for modeling these relationships. FFNN can approximate any complex functions using enough hidden layers and neurons. This makes highly flexible for anticipating crude oil prices, where the relationships between the input features may not be easily captured by simple linear models. FFNN can analyse historical crude oil price trends and identify patterns, enabling it to make predictions based on past market

behaviors. Furthermore, FFNN can be trained to continuously improve as more data is provided, which is important for adapting to changing market conditions or new trends in oil prices.

SVM is effective in handling high-dimensional data. Crude oil prices can exhibit sudden spikes due to unexpected events. SVM is robust to such outliers because it tries to find the hyperplane that best separates the data, focusing on the support vectors rather than entire dataset. Crude oil prices are inherently noisy and volatile. FFNN can filter noise via multiple layers of processing and SVM's robustness to outliers makes them well-suited to this challenge.

FFNN is selected for its ability to model complex, non-linear relationships and continuous data, which are essential for predicting actual crude oil prices based on a variety of input features. SVM is chosen for its strength in robustness to outliers, making it ideal for predicting crude oil prices.

3.7 Proposed methodologies for crude oil price prediction

The primary objective of this research is to build and implement crude oil prediction models using ML algorithms. Within the scope of this research, three distinct models have been developed for anticipating crude oil prices one step ahead, encompassing daily, weekly, and monthly timeframes. The models developed in this research are visually given in Figure 3.5. Further, elaboration on these models for crude oil price prediction is provided in the subsequent section. In all the three models, there are two phases: training phase and testing phase. During training phase, the selected ML models are trained to capture the relationship between independent and

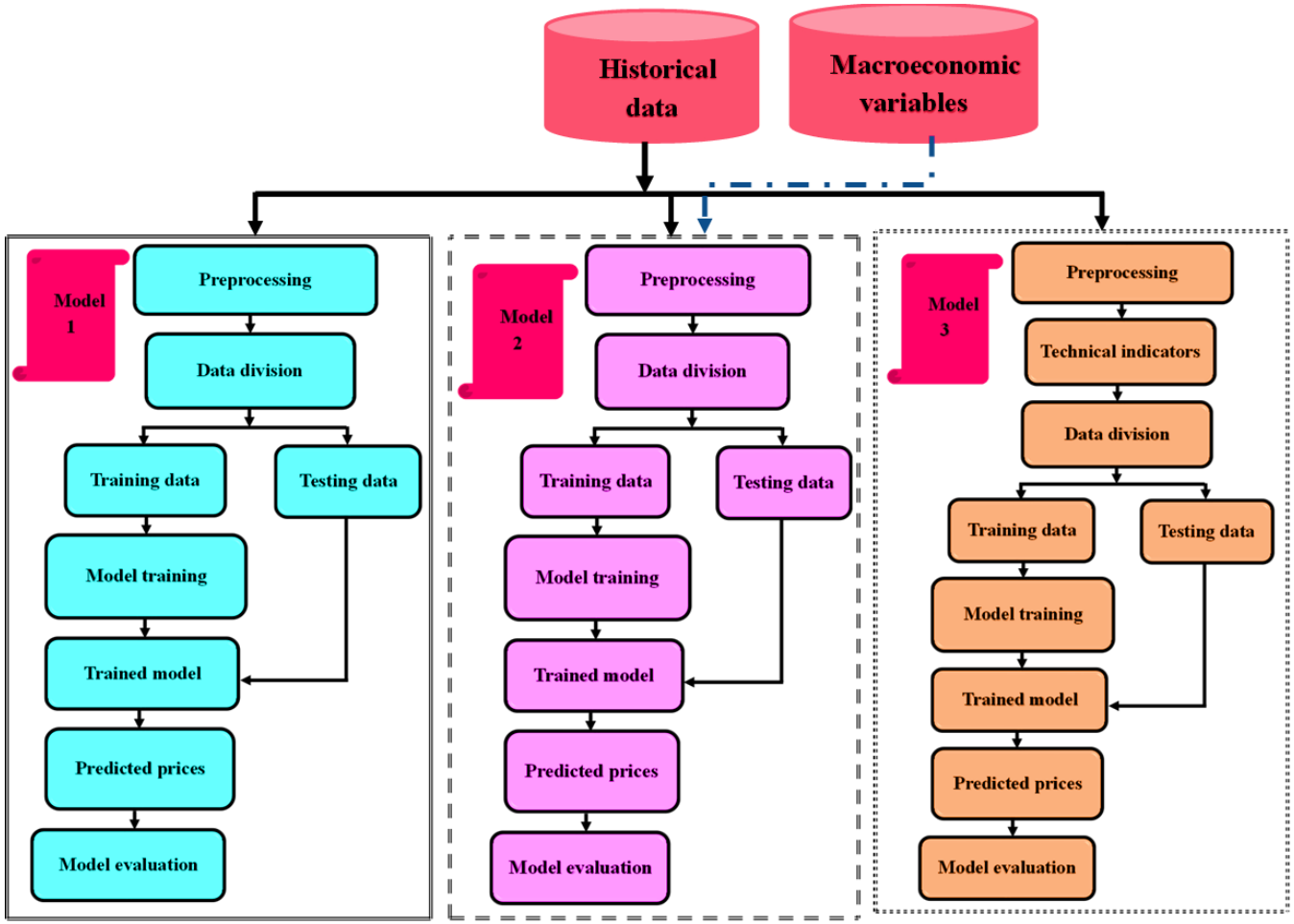


Figure 3.5 Overall structure of the developed crude oil prediction models

dependent variables. At the time of testing, trained network is used to predict the future prices of test data.

3.7.1 Crude oil price prediction model based on historical data and ML

A predictive model has been developed to anticipate future crude oil prices by using the power of ML algorithms. The main goal of this models is to provide valuable insights into the expected trends in crude oil prices, facilitating better decision-making in the energy and financial sectors. The overall structure of the introduced model is shown in Figure 3.6, offering a comprehensive view of its underlying processes and components. The key processes of the developed model are given below:

- Data preprocessing
- Data division
- Development of predictive models and training
- Model testing and
- Post processing

Data preprocessing

Data preprocessing is the one of the crucial steps in building the predictive model. The collected data may contain missing values, errors, and outliers. Hence, data preprocessing methods are adopted to clean and prepare the data for analysis. In this model, preprocessing is done via two distinct ways:

1. **Addressing missing values:** To handle missing values, a method involving the utilization of neigh boring data's mean is implemented.

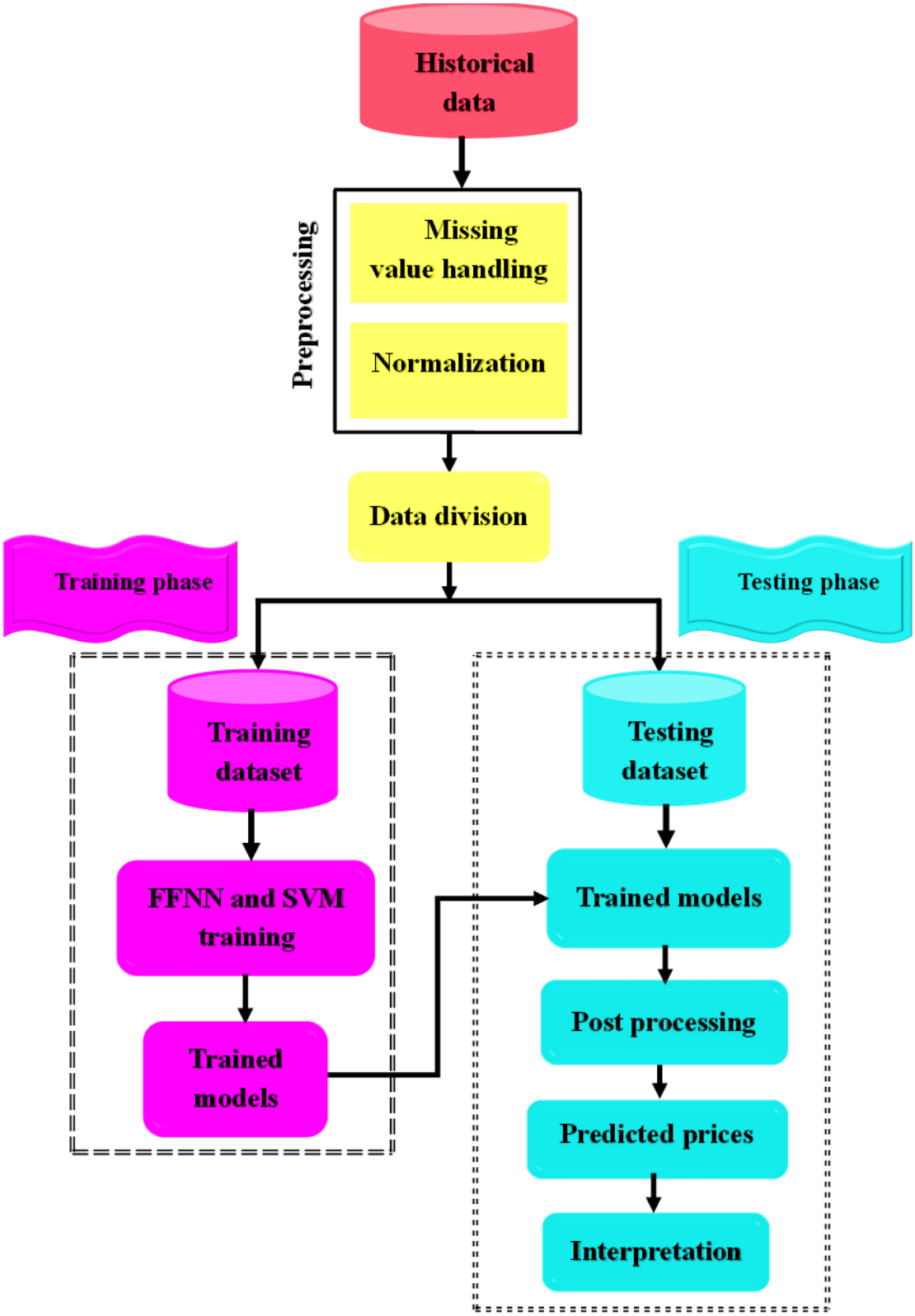


Figure 3.6 Pipeline of the developed crude based on historical data and ML

2. **Normalization:** For the purpose of data normalization, the min-max technique is used. This process can be expressed as

$$X_{\text{norm}} = \frac{\text{x} - x_{\text{min}}}{\text{x}_{\text{max}} - \text{x}_{\text{min}}} \quad (3.15)$$

Where,

x - Actual price

X_{norm} -Normalized price

x_{min} -Minimum price

x_{max}-Maximum price

Data division

After preprocessing, the dataset is typically split into two subsets: training set and testing set. This division is essential for evaluating model's performance. The training set is used to learn the model patterns and relationships within the data, whereas the testing data is reserved to assess how well the model generalizes to unseen data. For the training and testing processes, both the Brent and WTI datasets is partitioned into different ratios, including 60:40, 70:30, 80:20, and 90:10. In the 60:40 ratios, 60% of the data is allocated for training the model, while the remaining 40% is reserved for testing purposes. The effectiveness of the model is evaluated using various combination of training and test samples, leading to the selection of the top-performing model, which can provide predictions with a higher level of accuracy.

Development of predictive models and training

Two models, ANN and SVM, are used to develop predictive models for forecasting future crude oil prices.

Training

A FFNN is designed with four layers: input layer, two hidden layers with 10 hidden neurons in each hidden layer, and output layer. To generate the output signal at hidden and output layers, activation must be applied. Activation function allows for the introduction of non-linearity into the model, which enables the model to learn more complex decision boundaries. The sigmoidal activation function is used at the hidden layer to generate hidden layers' output. This function can be represented as,

$$f(x) = \frac{1}{1+e^{-x}} \tag{3.16}$$

Sigmoidal function maps any input value to a value between 0 and 1, making it useful for regression problems like crude oil price prediction. The main goal of regression problem is to predict a binary outcome given a set of independent variables. The sigmoid function output is between 0 and 1, can be interpreted as the probability of the sample belonging to positive class. Graphical representation of sigmoidal activation function is shown in Figure 3.7.

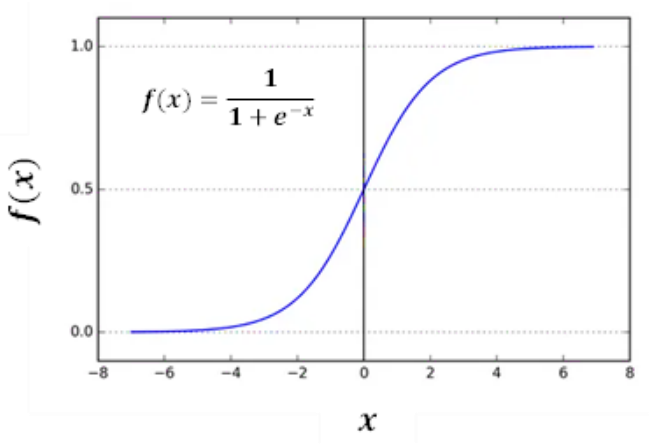


Figure 3.7 Sigmoidal activation function

To get final predicted value, the output layer is activated by using a linear activation function. Linear activation function can be expressed as,

$$g(x) = x \tag{3.17}$$

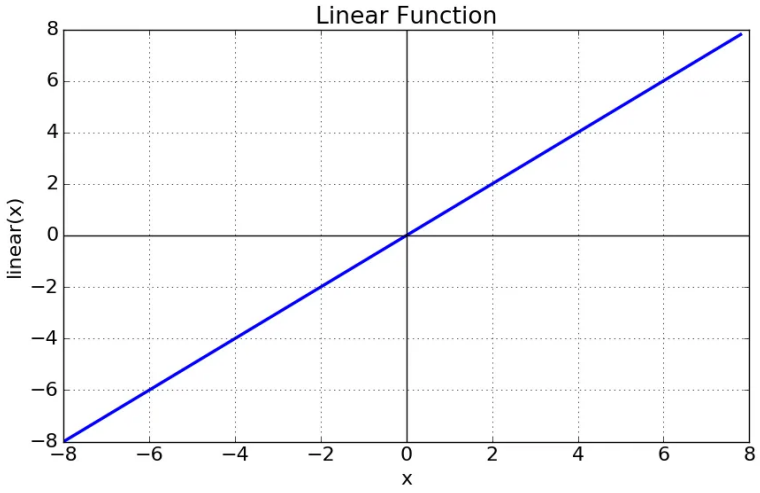


Figure 3.8 Linear activation function

Figure 3.8 shows the graphical depiction of linear activation function. Error, E is the difference between the target value, t and predicted value, y. Error on the presentation of nth training pattern can be defined as,

$$E(n) = t(n) - y(n) \tag{3.18}$$

If there are N patterns in the training data, average squared error energy is,

$$E_{Ave} = \frac{1}{M}\sum_{n=1}^{M} E(N) \tag{3.19}$$

The network is trained with Leven-Berg Marquart (LM) algorithm. To solve complex non-linear problems, LM algorithm is commonly used as a standard algorithm. This algorithm uses a combination of gradient descent and Gauss-Newton methods. If Back propagation (BP) is expressed as gradient descent, the algorithm becomes slow and does not give an optimal solution. On the other hand, if BP is expressed as Gauss-Newton, the algorithm has highest probability to give optimal solution. In LLM algorithm, the Hessian calculation approximation is shown in Equation (3.19) and the calculation of its gradient is expressed in Equation (3.20),

$$H = J^T J \tag{3.20}$$

$$g1 = J^T E \tag{3.21}$$

Where,

J – Jacobian matrix

E-Error

The LM algorithm behaves as Newton is expressed as by the following equation

$$\Delta w = w - [J^T J + \mu I]^{-1} J^T E \quad (3.22)$$

Where,

w – Old weight

Δw- New weight

μ-Mean

The SVM is developed with Radial Basis Function (RBF) and regularization parameter, 0.01. Both FFNN and SVM are trained using training data.

Model testing

The trained model is employed for predicting future value of test data.

Post processing

After the prediction process, the predicted prices are converted back to their original value, enabling a straightforward and intuitive comparison between the actual and predicted prices. The process of reverting to the original value can be expressed as,

$$x = y_p(x_{max} - x_{min}) + x_{min} \quad (3.23)$$

Subsequently, the de-normalized prices are compared with the actual prices to find the efficacy of the model.

3.7.2 Crude oil price prediction model based on macroeconomic variables and ML

The schematic representation of the designed crude oil prediction model is depicted in Figure 3.9. As illustrated in Figure 3.9, the model comprises three phases.

- Data collection
- Data preparation
- ML models training and
- Test data prediction

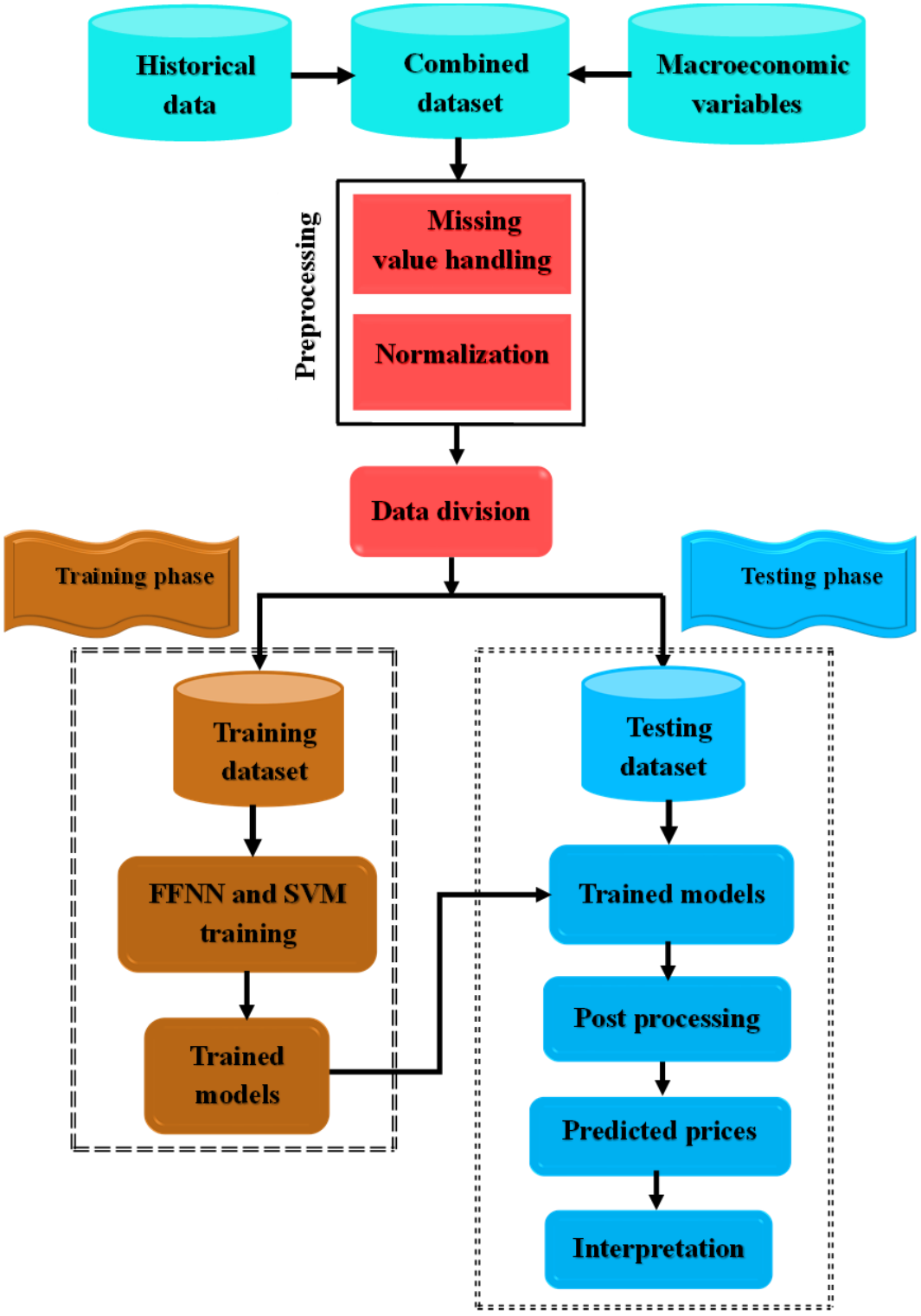


Figure 3.9 Framework of the crude oil prediction model based on macroeconomic variables and ML

Data preparation

The second phase involves data preparation, as essential step in the creation of an accurate crude oil prediction model. To build a crude oil prediction model using ML models, it is imperative to convert input data into a common rage between 0 and 1. Data preparation has two processes, such as handling missing value and data normalization. Mean of neighbor's method is used to handle missing values and then data is normalized into a common range between 0 and 1 by using Equation (3.15).

ML models training

The third phase of the prediction model revolves around FFNN and SVM training, a pivotal step in the development. This phase consists of three essential steps:

- **Determination of FFNN and SVM architecture:** Critical decisions regarding the architectural configuration are made. This includes selecting the appropriate architectural parameters for the FFNN and SVM. The FFNN is designed with an input layer, two hidden layers, and output layer. Sigmoidal and linear activation function is used at the hidden layer and output layer, respectively. The SVM is developed with Radial Basis Function (RBF) and regularization parameter, 0.01.
- **Data training:** This process involves the actual training of both FFNN and SVM models carefully prepared.
- **Validation:** Ensuring the FFNN and SVM models efficacy and fine-tuning their performance through validation procedures.

Test data prediction

The last phase in predictive model is the test data prediction phase. Once preceding processes are successfully executed. The trained ML models, FFNN and SVM can be used as a predictor for forecasting crude oil prices in test data scenario. The predicted prices are subjected to denormalization process using Equation (3.22) to obtain corresponding prices. Finally, the predicted values and observed values are compared to assess the efficacy of the models.

3.7.3 Crude oil price prediction model based on technical indicators and ML

The third predictive model brings forward a ground breaking contribution by introducing an entirely new framework designed to effectively capture the fluctuation patterns characterizing the Brent and WRI crude oil prices. This innovative model leverages a variety of technical indicators, providing a unified solution for comprehensively analysing and predicting these vital commodities. The complete workflow of the introduced model is depicted in Figure 3.10. Furthermore, the subsequent sections offer a detailed breakdown of the crucial components of the system, ensuring a through comprehension of their roles and functionalities.

- Data preprocessing
- Computation of technical indicators
- Model training and
- Model testing

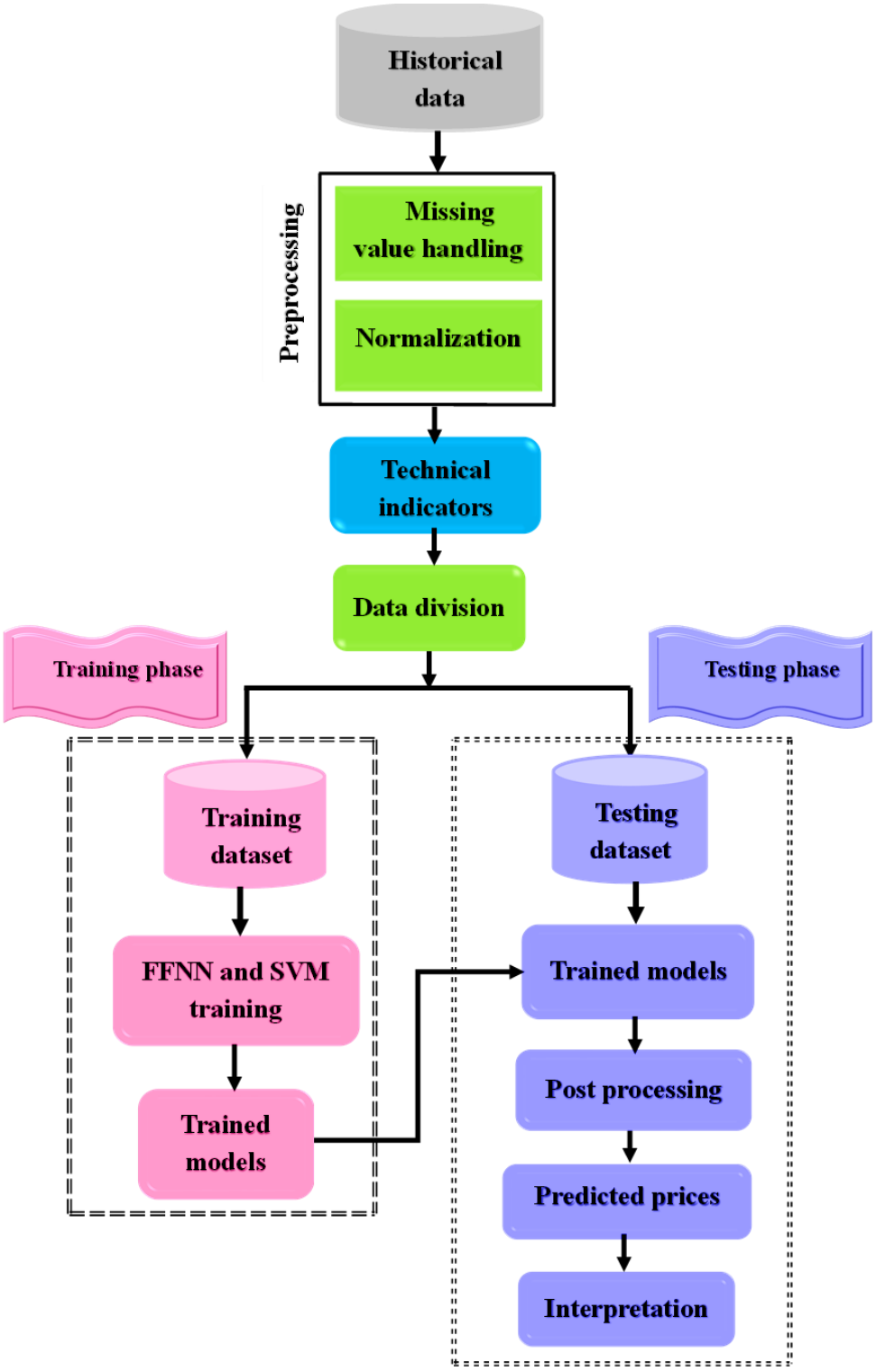


Figure 3.10 Developed crude oil prediction model based on technical indicators and ML

Data preprocessing

In order to prepare input data for further processing, two data processing techniques are applied: (i) addressing missing values, and (ii) normalizing the data using Equation (3.11). The procedure involves substituting missing values with the average of nearby data point and employing the min-max method to scale the data into a uniform range spanning from 0 to 1.

Computation of technical indicators

The development of an effective crude oil prediction model relies on the appropriate selection of input variables. While numerous existing models have traditionally relied on macroeconomic variables as their input variables, others have turned to past day's prices. However, a recent investigation by Manickavasagam et al. (2020), pursued a different approach by utilizing a set of technical indicators as input variables to forecast intraday prices of Brent and WTI crude oil. Their findings demonstrated remarkable performance when compared to other models. This model adopts a similar approach by incorporating 14 technical indicators for predicting future prices of Brent and WTI crude oil, spanning daily, weekly, and monthly timeframes. Technical indicators are used in this model is listed in Table 3.3.

Table.3.3 Selected technical indicators and their formulas

Indicators	Equation	Indicators used

Moving Average (MA)	$MA = \sum_{j=1}^{N} \frac{p_j}{N}$	MA5, MA10, MA20
Exponential Moving Average (EMA)	$EMA = p(t) \times K + EMA(y) \times (1 - K)$ $K = \frac{2}{N+1}$	EMA5, EMA10, EMA20
Disparity (D)	$D = \frac{p(t) - MA_t}{MA_t} \times 100$	D5, D10, D20
Exponential Disparity (ED)	$ED = \frac{p(t) - EMA_t}{EMA_t} \times 100$	ED5, ED10, ED20
Price Oscillator (POSC)	$POSC = \frac{MA_t - MA_t}{MA_t}$	POSC
Exponential Price Oscillator (EPOSC)	$EPOSC = \frac{EMA_t - EMA_t}{EMA_t}$	EPOSC
Where, P (t) – price at time t, MA_t-moving average of t days, EMA_t-moving average of t days		

Model training

After data collection and database creation, the amount of training, validation and testing samples are determined. The data are normalized and predictor variables are computed. To predict oil prices, FFFN is designed with an input layer with 14 neurons representing input variables, two hidden layers with 14 hidden neurons, and an output layer with a single neuron representing predicted price. The FFNN is trained with LM algorithm to capture the relationship between independent and dependent variables. SVM is designed with RBF and trained to predict future crude oil prices.

Model testing

The trained models are employed to predict future crude oil prices on test data. The predicted prices are de-normalized using Equation (3.22) to get corresponding values. The actual and predicted values are compared to assess the performance of the model.

3.7.4 Model validation

When evaluating the performance of machine learning models like FFNN and SVM for anticipating crude oil prices, choosing appropriate evaluation metrics is essential to ensure a comprehensive assessment of model accuracy and reliability. Data division approach is used to validate the model's performance. 70% of the entire data is used for training while remaining 30% of the data is utilized for validation. Four metrics such as Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Directional statistic (D_{stat}), and prediction accuracy are adopted to assess the effectiveness of the developed models. RMSE is one of widely used metrics for regression task. RMSE gives an absolute measure of fit in the same units as the variable being predicted prices. This helps stakeholders understand the magnitude of the error directly in terms of crude oil prices, making it more intuitive. MAPE measures the prediction error in percentage terms, which makes it easier to interpret in practical contexts. It is easy to interpret as it gives an error rate in percentage terms, which is more intuitive for non-technical stakeholders such as traders or analysts, who need to understand the model's performance. In the context of crude oil price prediction, MAPE is useful because the impact of errors may be more meaningful when compared relative to actual price values. It allows for easy comparison between

models, regardless of the magnitude of the actual price values. In many financial applications, correctly predicting the changes of crude oil prices is important, if not more so, than predicting the exact value. D_{stat} directly captures this. Traders and investors often care more about whether the price is trending upward or downward rather than the specific price itself. A model that correctly predicts the movement of prices can be highly useful in making trading decisions. Prediction accuracy is a fundamental metric that captures the overall performance of the model in a more simplified manner. It shows the percentage of correct prediction relative to all predictions, which is often an easy to understand measure.

3.8 Summary

In this Chapter, the conceptual framework and research hypothesis of the research work were explained. Detailed information about the selected crude oil markets was provided. An introduction to feed forward neural network and support vector machine was given. Furthermore, the functioning of the developed crude oil prediction models was provided.

CHAPTER 4

DATA ANALYSIS & RESULTS & FINDINGS

4.1 Simulation platform

The fundamental focus of this thesis is to utilize ML models for the prediction of future crude oil prices. Numerous approaches have been devoted for crude oil price prediction, aiming to improve prediction accuracy and reduce error, Huang and Wang (2018), Ng'ang'a et al. (2022). Most of the earlier approaches used macroeconomic variables as input vectors to forecast crude oil prices. Nevertheless, none of these methods consistently delivers precise predictions of crude oil price. Recent research investigations have proved that the ML models could predict crude oil prices. Thus, this research intends to construct crude oil prediction models employing ML algorithms with the intention of improving prediction accuracy and minimizing error rates.

As discussed in Chapter 3, the crude oil price is influenced by numerous factors. Relying solely on macroeconomic variables is insufficient for achieving precise predictions Sekati et al. (2020). Historical data and technical indicators have a substantial influence on crude oil prices. Consequently, this research work aims to use three different variables: historical data, macroeconomic variables, and technical indicators, as input vectors for the two ML models, namely, FFNN and SVM. Based on these three different input vectors, three different predictive models were formulated and trained to predict future crude oil prices. The trained models were employed to anticipate future prices of test data. The rationale behind utilizing

different input vectors and ML models was to pinpoint the most effective input vector and ML model that yielded improved predictions. The implementation and validation of these introduced frameworks were carried out using the MATLAB2022 platform on a system equipped with an Intel Core i5 (2.70 GHz) processor running a 64-bit Windows 11 operating system with 8 GB of RAM.

4.2 Data normalization

This research work focused on analysing two prominent benchmark crude oil markets: namely Brent and WTI crude oil prices, the specifics of which were thoroughly detailed in in Chapter 3. Acknowledging the inherent sensitivity of ML models to input parameter, it becomes imperative to standardize these parameters to a common range, typically within the interval of 0 and 1. This normalization process is essential to ensure uniformity and facilitate the effective operation of the ML models.

1. Predictive model based on Historical data

First predictive model is designed to predict future crude oil prices based on historical data. Hence, historical prices were normalized using min-max method, as illustrated in Figure 4.1.

2. Predictive model based on macroeconomic variables

Second predictive model focuses on predicting future crude oil prices using macroeconomic variables. Descriptive statistic of macroeconomic variables is given in Table. 4.1. It is noted that the values are in very different rages. Therefore, the

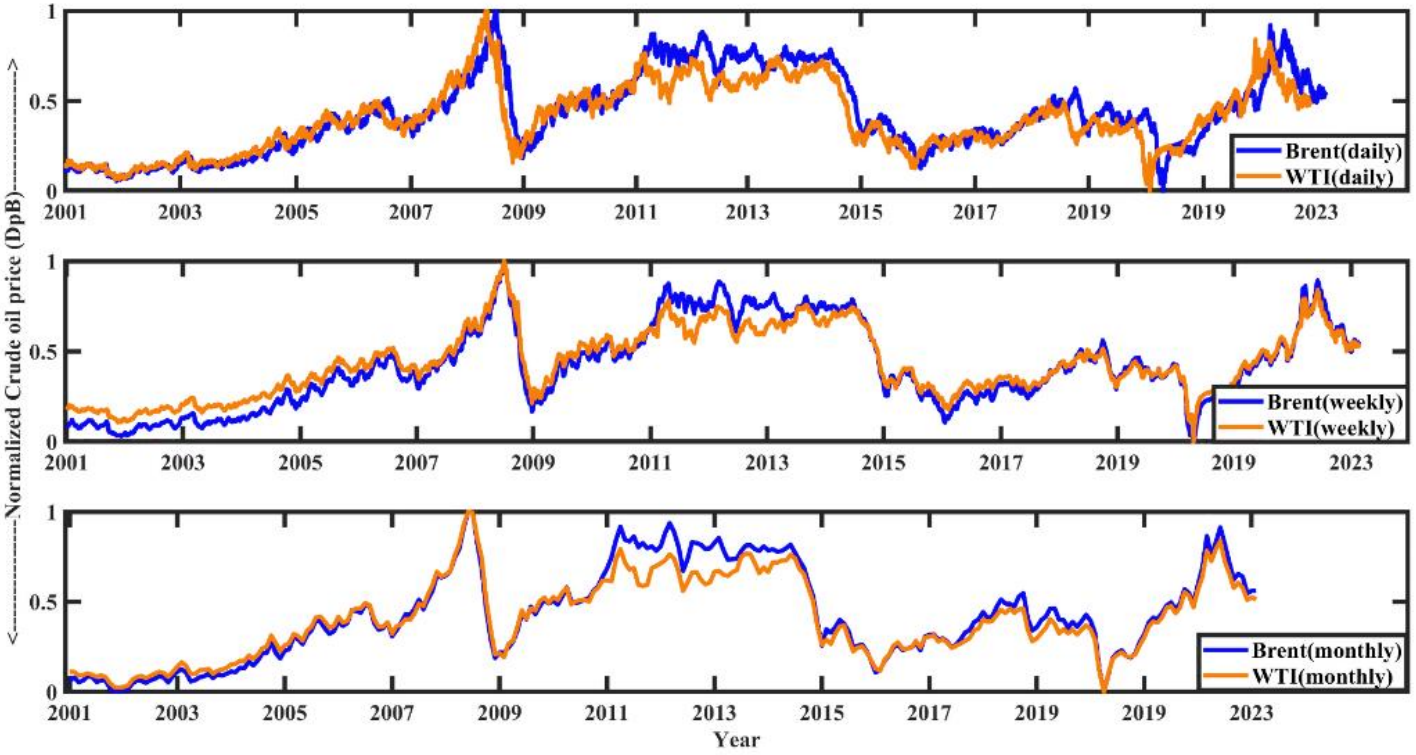


Figure 4.1 Normalized crude oil prices 2001-2023

macroeconomic variables were normalized into a common range [0,1] using min-max technique. The resultant normalized values are given in Table 4.2 for daily data, Table 4.3 for weekly data, and Table 4.4 for monthly data.

Table 4.1 Descriptive statistic of macroeconomic variables

Data frequency		Macroeconomic variables						
		GP	**SP**	**S& P 500**	**EUR/ USD**	**NGP**	**DI**	**GBP/ USD**
Daily data	**Minimum**	256.6	4.039	676.53	0.858	1.482	71.304	0.474
	Maximum	2069.4	48.599	4796.6	1.598	15.378	121.21	0.935
	Mean	1102.1	16.62	1908.9	1.218	4.522	90.534	0.657
	SD	504.53	8.274	990.1	0.144	2.246	10.602	0.094
Weekly data	**Minimum**	259.2	4.046	683.38	0.847	1.495	71.635	0.478
	Maximum	2018	48.59	4766.1	1.594	14.31	120.18	0.921
	Mean	1100	16.62	1932.8	1.21	4.552	90.87	0.659
	SD	504.26	8.239	1013.3	0.151	2.239	10.82	0.093
	Minimum	259.2	4.157	735.09	0.845	1.64	72.16	0.48

Monthly data	**Maximum**	1985.9	48.59	4766.2	1.577	13.92	120.59	0.895
	Mean	1093.4	16.76	1912	1.211	4.57	90.83	0.652
	SD	510.07	8.32	997.02	0.152	2.271	10.93	0.093

Table 4.2 Sample macroeconomic variables for daily data (Before and after normalization)

GP	**SP**	**S& P 500**	**EUR/USD**	**NGP**	**DI**	**GBP/USD**
Before normalization						
269.2	4.238	1197.7	0.8767	2.999	113.46	0.6956
270.3	4.265	1166.7	0.8801	3.094	113.5	0.6932
270.3	4.243	1173.6	0.883	3.468	113.56	0.6923
270.8	4.263	1150.5	0.8836	3.367	112.87	0.6881
269.9	4.26	1170.8	0.8909	3.303	113.54	0.6883
269.9	4.268	1142.6	0.8895	3.187	115.78	0.6895
270.7	4.308	1122.1	0.9036	3.166	115.68	0.6875
276.2	4.21	1117.6	0.8996	2.848	115.04	0.6915

276.8	4.248	1139.8	0.895	2.811	114.09	0.6953
278.6	4.288	1152.7	0.8935	2.706	114.64	0.6916
279.1	4.24	1182.2	0.8931	2.544	114.64	0.6947
278.4	4.24	1153.3	0.8914	2.415	111.86	0.6948
277.9	4.26	1148	0.8914	2.295	112.35	0.6956
282	4.26	1160.3	0.8926	2.395	112.08	0.6945
278.6	4.232	1145.9	0.8944	2.38	112.37	0.6901
277.4	4.24	1106.5	0.8834	2.359	113.63	0.6896
277.3	4.218	1103.2	0.8829	2.42	113.42	0.6943
277.7	4.254	1151.4	0.8806	2.44	112.7	0.6956
274.9	4.243	1128.4	0.8843	2.5	112.68	0.6967
274.9	4.249	1137.6	0.8838	2.392	113.5	0.6957
274.4	4.25	1168.4	0.8861	2.552	113.96	0.6966
275.8	4.243	1165.9	0.8779	2.369	113.37	0.6996
277.5	4.207	1183.5	0.8774	2.225	113.38	0.7022
276.5	4.187	1179.7	0.8661	2.102	113.67	0.7026

273.7	4.167	1191.8	0.8627	2.137	113.55	0.7095
After normalization						
0.0019	0.0045	0.1265	0.0246	0.1092	0.8553	0.4795
0.0025	0.0051	0.1190	0.0292	0.1160	0.8561	0.4743
0.0025	0.0046	0.1206	0.0331	0.1429	0.8574	0.4723
0.0028	0.0050	0.1150	0.0339	0.1357	0.8434	0.4632
0.0023	0.0050	0.1200	0.0438	0.1310	0.8570	0.4637
0.0023	0.0051	0.1131	0.0419	0.1227	0.9024	0.4663
0.0027	0.0060	0.1081	0.0609	0.1212	0.9004	0.4619
0.0058	0.0038	0.1071	0.0555	0.0983	0.8874	0.4706
0.0061	0.0047	0.1124	0.0493	0.0956	0.8681	0.4788
0.0071	0.0056	0.1156	0.0473	0.0881	0.8793	0.4708
0.0074	0.0045	0.1227	0.0467	0.0764	0.8793	0.4775
0.0070	0.0045	0.1157	0.0444	0.0671	0.8229	0.4778
0.0067	0.0050	0.1144	0.0444	0.0585	0.8328	0.4795
0.0090	0.0050	0.1174	0.0461	0.0657	0.8273	0.4771

0.0071	0.0043	0.1139	0.0485	0.0646	0.8332	0.4676
0.0064	0.0045	0.1044	0.0336	0.0631	0.8588	0.4665
0.0064	0.0040	0.1036	0.0330	0.0675	0.8545	0.4767
0.0066	0.0048	0.1153	0.0299	0.0689	0.8399	0.4795
0.0050	0.0046	0.1097	0.0349	0.0733	0.8395	0.4819
0.0050	0.0047	0.1119	0.0342	0.0655	0.8561	0.4797
0.0048	0.0047	0.1194	0.0373	0.0770	0.8655	0.4817
0.0055	0.0046	0.1188	0.0262	0.0638	0.8535	0.4882
0.0065	0.0038	0.1231	0.0255	0.0535	0.8537	0.4938
0.0059	0.0033	0.1221	0.0103	0.0446	0.8596	0.4947
0.0044	0.0029	0.1251	0.0057	0.0471	0.8572	0.5097

Table 4.3 Sample macroeconomic variables for weekly data (Before and after normalization)

GP	**SP**	**S& P 500**	**EUR/USD**	**NGP**	**DI**	**GBP/USD**
Before normalization						

265.4	4.46	1183.5	0.9247	6.21	111.54	0.6926
265.3	4.4	1242.98	0.9131	5.568	112.29	0.6923
266.9	4.36	1253.05	0.9176	5.131	112.07	0.688
268.9	4.337	1266.61	0.9347	5.27	111.2	0.6796
288.4	4.577	1245.67	0.9329	5.072	111.48	0.6802
278.7	4.583	1291.96	0.8976	5.035	115.24	0.6993
268	4.441	1277.89	0.89	5.273	116.12	0.7021
274.5	4.388	1260.67	0.8778	5.025	117.63	0.7063
272.3	4.418	1264.96	0.9042	5.388	114.9	0.6945
273.3	4.321	1214.36	0.8882	5.381	115.63	0.6949
272	4.327	1225.35	0.9022	5.128	114.48	0.6928
267.2	4.265	1224.38	0.8922	4.867	115.81	0.6942
268.3	4.308	1190.59	0.8927	4.49	115.09	0.6947
270.9	4.24	1215.68	0.8753	4.278	116.98	0.7047
268.2	4.218	1210.85	0.8814	4.291	116.46	0.6946
269.5	4.243	1205.82	0.8612	3.973	118.26	0.7029

275.4	4.16	1214.35	0.8477	3.93	118.99	0.7049
280.5	4.279	1190.16	0.8508	3.922	119.27	0.7253
273.5	4.199	1161.97	0.8615	3.979	118.28	0.7127
275.8	4.193	1184.93	0.8573	3.742	119.23	0.7077
274.5	4.171	1133.58	0.85	3.096	119.75	0.7063
291.6	4.318	1085.78	0.8475	3.218	120.15	0.7083
292.4	4.605	1092.54	0.8549	3.25	119.62	0.7122
294	4.675	965.8	0.8711	2.955	117.42	0.7004
292.5	4.685	1040.94	0.8777	3.167	117.14	0.7023
After normalization						
0.0035	0.0093	0.1225	0.1034	0.3679	0.8219	0.4843
0.0035	0.0079	0.1371	0.0879	0.3178	0.8373	0.4836
0.0044	0.0070	0.1395	0.0939	0.2837	0.8328	0.4739
0.0055	0.0065	0.1429	0.1168	0.2945	0.8149	0.4549
0.0166	0.0119	0.1377	0.1144	0.2791	0.8206	0.4563
0.0111	0.0121	0.1491	0.0671	0.2762	0.8982	0.4994

0.0050	0.0089	0.1456	0.0569	0.2948	0.9163	0.5058
0.0087	0.0077	0.1414	0.0406	0.2754	0.9474	0.5153
0.0074	0.0083	0.1424	0.0759	0.3037	0.8911	0.4886
0.0080	0.0062	0.1301	0.0545	0.3032	0.9062	0.4895
0.0073	0.0063	0.1327	0.0733	0.2835	0.8825	0.4847
0.0045	0.0049	0.1325	0.0599	0.2631	0.9099	0.4879
0.0052	0.0059	0.1242	0.0605	0.2337	0.8951	0.4890
0.0067	0.0044	0.1304	0.0372	0.2171	0.9340	0.5116
0.0051	0.0039	0.1292	0.0454	0.2181	0.9233	0.4888
0.0059	0.0044	0.1280	0.0183	0.1933	0.9604	0.5076
0.0092	0.0026	0.1301	0.0003	0.1900	0.9755	0.5121
0.0121	0.0052	0.1241	0.0044	0.1894	0.9812	0.5582
0.0081	0.0034	0.1172	0.0187	0.1938	0.9608	0.5297
0.0094	0.0033	0.1228	0.0131	0.1753	0.9804	0.5184
0.0087	0.0028	0.1103	0.0033	0.1249	0.9911	0.5153
0.0184	0.0061	0.0986	0.0000	0.1344	0.9994	0.5198

0.0189	0.0125	0.1002	0.0099	0.1369	0.9885	0.5286
0.0198	0.0141	0.0692	0.0316	0.1139	0.9431	0.5019
0.0189	0.0143	0.0876	0.0404	0.1305	0.9373	0.5062

Table 4.4 Sample macroeconomic variables for monthly data (Before and after normalization)

GP	SP	S& P 500	EUR/USD	NGP	DI	GBP/USD
Before normalization						
436.6	8.35	1220.33	1.2344	11.472	87.54	0.5544
472.3	8.89	1228.81	1.203	13.921	89.35	0.5669
468.8	9.915	1207.01	1.1988	12.205	89.93	0.5649
498.7	9.805	1249.48	1.1789	12.587	91.51	0.5782
521.2	11.56	1248.29	1.1843	11.225	90.96	0.5809
575.5	13.63	1280.08	1.2155	9.316	88.81	0.5622
566.5	12.505	1280.66	1.1921	6.714	90.08	0.5702
586.7	10.92	1294.83	1.2119	7.21	89.39	0.5753

657.7	11.39	1310.61	1.2637	6.555	85.88	0.5475
649	13.002	1270.09	1.2812	6.384	84.65	0.5347
619.3	11.54	1270.2	1.2789	6.104	84.9	0.5412
640.4	12.322	1276.66	1.2763	8.211	85.09	0.5354
631.2	14.052	1303.82	1.2807	6.048	85.02	0.5251
604.2	12.935	1335.85	1.2672	5.62	85.68	0.5342
609.8	13.632	1377.94	1.2763	7.534	85.11	0.5242
652.9	14.235	1400.63	1.3241	8.844	82.85	0.5088
641.1	13.51	1418.3	1.3199	6.299	83.43	0.5106
657.9	13.575	1438.24	1.3034	7.667	84.43	0.5091
675.6	13.538	1406.82	1.3235	7.3	83.5	0.5091
669	12.473	1420.86	1.3358	7.73	82.66	0.5081
686.7	13.072	1482.37	1.3647	7.863	81.3	0.5002
666.7	12.162	1530.62	1.3453	7.935	82.25	0.505
653.9	13.92	1503.35	1.3542	6.773	81.69	0.4975
673	14.494	1455.27	1.3673	6.191	80.66	0.4929

678.8	14.097	1473.99	1.363	5.468	80.745	0.4961
After normalization						
0.1027	0.0943	0.1204	0.5311	0.8006	0.3175	0.1782
0.1234	0.1065	0.1225	0.4882	1.0000	0.3549	0.2083
0.1214	0.1296	0.1171	0.4824	0.8603	0.3669	0.2035
0.1387	0.1271	0.1276	0.4552	0.8914	0.3995	0.2355
0.1517	0.1666	0.1273	0.4626	0.7805	0.3881	0.2420
0.1832	0.2132	0.1352	0.5053	0.6250	0.3437	0.1970
0.1780	0.1878	0.1353	0.4733	0.4132	0.3700	0.2162
0.1897	0.1522	0.1389	0.5003	0.4535	0.3557	0.2285
0.2308	0.1628	0.1428	0.5712	0.4002	0.2832	0.1616
0.2257	0.1990	0.1327	0.5951	0.3863	0.2578	0.1307
0.2085	0.1661	0.1327	0.5919	0.3635	0.2630	0.1464
0.2208	0.1837	0.1343	0.5884	0.5351	0.2669	0.1324
0.2154	0.2226	0.1411	0.5944	0.3589	0.2655	0.1076
0.1998	0.1975	0.1490	0.5759	0.3241	0.2791	0.1295

0.2030	0.2132	0.1595	0.5884	0.4799	0.2673	0.1055
0.2280	0.2268	0.1651	0.6537	0.5866	0.2207	0.0684
0.2212	0.2105	0.1695	0.6480	0.3794	0.2326	0.0727
0.2309	0.2119	0.1744	0.6254	0.4908	0.2533	0.0691
0.2412	0.2111	0.1666	0.6529	0.4609	0.2341	0.0691
0.2373	0.1871	0.1701	0.6697	0.4959	0.2167	0.0667
0.2476	0.2006	0.1854	0.7092	0.5067	0.1886	0.0477
0.2360	0.1801	0.1973	0.6827	0.5126	0.2083	0.0592
0.2286	0.2197	0.1906	0.6949	0.4180	0.1967	0.0412
0.2396	0.2326	0.1787	0.7128	0.3706	0.1754	0.0301
0.2430	0.2237	0.1833	0.7069	0.3117	0.1772	0.0378

3. Predictive model based on technical indicators

Third predictive model anticipates future crude oil prices based on technical indicators. 14 technical indicators were considered for both Brent and WTI crude oil markets. Descriptive statistic of these technical indicators for Brent and WTI crude oil is tabulated in Table 4.5 and Table 4.6, respectively. Like macroeconomic variables, the technical indicators exhibited a wide range of values. Therefore, to ensure uniformity, the min-max technique was applied for transforming these indicators into

the common range of [0,1]. The normalized values are given in Table 4.5 for Brent and Table 4.6 for WTI.

By undertaking these normalization procedures, not only brought the data into a standardized format suitable for ML training but also facilitated the fair comparison of predictive models across different input data types. This normalization approach empowers research to make accurate and meaningful predictions in the domain of crude oil price prediction.

Table 4.5 Descriptive statistic of technical indicators (Brent crude oil data)

Technical indicators	Daily data				Weekly data				Monthly data			
	Minimum	**Maximum**	**Mean**	**SD**	**Minimum**	**Maximum**	**Mean**	**SD**	**Minimum**	**Maximum**	**Mean**	**SD**
SMA5	14.147	141.52	66.980	29.109	17.732	137.08	67.114	29.211	19.556	125.53	66.846	28.656
SAM10	15.254	139.65	66.963	29.091	18.796	134.45	67.031	29.001	21.286	115.83	66.428	28.005
SMA20	17.229	138.17	66.928	29.054	19.271	127.91	66.867	28.734	22.991	113.31	65.490	26.822
EMA5	15.126	140.75	66.973	29.092	18.731	135.43	67.080	29.010	20.465	121.24	66.665	28.151
EMA10	16.330	138.87	66.947	29.054	19.366	130.87	66.951	28.757	21.922	114.58	66.014	27.102
EMA20	18.578	136.76	66.894	28.969	20.287	122.55	66.668	28.230	23.407	109.179	64.897	25.599
D5	-40.899	23.123	0.0283	2.0863	-33.894	26.808	0.143	4.324	-52.590	22.354	0.6017	9.123
D10	-49.186	33.332	0.0743	3.2631	-48.181	42.908	0.3737	7.2715	-62.447	31.724	1.5653	13.663
D20	-52.979	40.875	0.1699	4.8735	-62.324	31.894	0.8217	10.643	-67.321	49.661	3.5484	19.148
ED5	-42.475	22.684	0.0311	2.3158	-36.643	25.911	0.1528	5.0518	-54.547	22.046	0.6981	9.917
ED10	-49.973	32.877	0.0853	3.6829	-52.424	24.011	0.4147	8.0594	-63.30	36.855	1.9274	14.712
ED20	-58.131	35.632	0.1964	5.5774	-65.104	33.688	0.9845	11.474	-67.559	56.588	4.1988	20.518
POSC	-0.1741	0.1021	-0.00006	0.0168	-0.4092	0.1614	-0.0004	0.0435	-0.4597	0.1125	0.0001	0.0857
EPOSC	-0.1900	0.0094	-0.00004	0.0179	-0.4028	0.0793	-0.0006	0.0459	-0.2747	0.1082	0.0028	0.0742

Table 4.6 Descriptive statistic of technical indicators (WTI crude oil data)

Technical indicators	Daily data				Weekly data				Monthly data			
	Minimum	Maximum	Mean	SD	Minimum	Maximum	Mean	SD	Minimum	Maximum	Mean	SD
SMA5	13.494	142.95	64.158	25.680	14.847	138.95	63.994	25.622	20.14	127.28	63.752	25.106
SAM10	15.143	141.76	64.144	25.656	19.254	135.92	63.924	25.497	22.071	117.40	63.388	24.357
SMA20	17.809	140.25	64.115	25.613	19.865	129.45	63.783	25.198	24.080	103.74	62.547	23.076
EMA5	14.240	142.63	64.152	25.660	16.923	137.34	63.965	25.513	21.127	122.78	63.599	24.535
EMA10	16.176	141.18	64.130	25.616	19.907	132.53	63.853	25.228	22.985	111.24	63.038	23.393
EMA20	19.031	138.82	64.085	25.522	20.960	123.99	63.612	24.629	25.131	99.208	62.084	21.889
D5	-56.466	46.874	0.021	2.378	-78.639	38.135	0.110	4.920	-52.816	23.740	0.511	9.1133
D10	-57.728	55.810	0.061	3.562	-84.973	50.138	0.314	7.484	-62.594	27.895	1.373	13.622
D20	-59.028	60.801	0.143	5.013	-90.296	31.716	0.716	10.699	-67.059	44.895	3.154	18.805
ED5	-55.693	43.839	0.020	2.572	-81.060	30.368	0.108	5.4301	-54.694	19.570	0.544	9.887
ED10	-58.150	53.817	0.066	3.876	-87.474	26.769	0.332	8.1850	-63.262	33.137	1.585	14.52
ED20	-61.529	49.956	0.160	5.598	-90.969	30.316	0.813	11.474	-67.527	52.440	3.486	19.95
POSC	-0.207	0.127	-0.00009	0.017	-0.421	0.183	-0.0007	0.0424	-0.505	0.1125	0.001	0.086
EPOSC	-0.211	0.084	-0.0001	0.018	-0.512	0.080	-0.0011	0.0465	-0.301	0.1075	0.001	0.073

Table 4.7 Sample technical indicators for Brent crude oil data (Before and after normalization)

SMA5	SAM10	SMA20	EMA5	EMA10	EMA20	D5	D10	D20	ED5	ED10	ED20	POSC	EPOSC
Before normalization													
23.436	23.435	23.435	23.433	23.432	23.431	0.019	0.020	0.021	0.028	0.035	0.039	0.000	0.000
23.908	23.856	23.833	23.812	23.639	23.539	2.768	2.995	3.091	3.182	3.939	4.378	0.002	0.007
24.236	24.128	24.087	24.131	23.844	23.657	2.202	2.660	2.834	2.646	3.882	4.706	0.004	0.012
24.457	24.291	24.236	24.338	24.009	23.761	1.199	1.889	2.120	1.694	3.086	4.163	0.007	0.014
24.436	24.266	24.218	24.268	24.031	23.796	1.252	0.562	0.363	0.570	0.412	1.404	0.007	0.010
24.652	24.422	24.348	24.506	24.204	23.909	1.331	2.283	2.596	1.936	3.208	4.481	0.009	0.012
24.972	24.666	24.540	24.870	24.457	24.070	2.515	3.787	4.318	2.934	4.671	6.357	0.012	0.017

25.233	24.875	24.696	25.124	24.671	24.218	1.572	3.034	3.781	2.016	3.889	5.829	0.014	0.018
25.571	25.132	24.875	25.426	24.918	24.391	1.796	3.573	4.645	2.377	4.463	6.720	0.017	0.020
25.653	25.275	24.961	25.457	25.027	24.498	0.517	0.969	2.238	0.247	1.969	4.170	0.015	0.017
25.225	25.153	24.892	25.061	24.890	24.477	3.787	3.510	2.498	3.158	2.489	0.844	0.003	0.007
25.002	25.101	24.884	24.954	24.862	24.502	1.048	1.439	0.578	0.859	0.492	0.972	0.004	0.004
25.353	25.328	25.041	25.400	25.122	24.672	3.697	3.798	4.987	3.506	4.649	6.558	0.001	0.011
26.126	25.782	25.325	26.163	25.589	24.959	5.986	7.401	9.337	5.836	8.211	10.940	0.013	0.022
26.565	26.060	25.515	26.449	25.849	25.156	1.712	3.682	5.898	2.160	4.530	7.411	0.019	0.023
26.911	26.290	25.691	26.646	26.066	25.335	0.478	2.853	5.251	1.479	3.738	6.729	0.023	0.022
27.039	26.464	25.845	26.744	26.225	25.488	0.367	1.799	4.238	0.733	2.728	5.697	0.021	0.019

27.054	26.632	25.999	26.843	26.373	25.636	0.052	1.533	4.004	0.736	2.530	5.477	0.016	0.017
26.989	26.758	26.138	26.878	26.478	25.761	0.143	0.719	3.108	0.266	1.783	4.616	0.009	0.015
26.763	26.752	26.212	26.692	26.449	25.814	1.654	1.616	0.412	1.395	0.488	1.959	0.000	0.009
26.673	26.781	26.298	26.658	26.475	25.888	0.312	0.714	1.110	0.256	0.435	2.711	0.004	0.007
26.807	26.874	26.424	26.829	26.601	26.010	1.353	1.103	2.822	1.272	2.139	4.459	0.002	0.008
27.503	27.236	26.703	27.519	27.019	26.285	5.081	6.108	8.227	5.018	6.961	9.947	0.010	0.018
28.191	27.615	26.995	28.096	27.425	26.568	3.758	5.920	8.352	4.107	6.656	10.096	0.020	0.024
28.675	27.939	27.252	28.431	27.729	26.809	1.481	4.157	6.781	2.354	4.943	8.546	0.026	0.025
After normalization													
0.0503	0.0435	0.0412	0.0482	0.0429	0.0411	0.0014	0.0010	0.0009	0.0019	0.0018	0.0016	0.0001	0.0008

0.0541	0.0469	0.0446	0.0513	0.0446	0.0420	0.1997	0.1435	0.1345	0.2097	0.2022	0.1771	0.0232	0.0916
0.0567	0.0492	0.0467	0.0538	0.0464	0.0430	0.1588	0.1275	0.1233	0.1743	0.1993	0.1904	0.0470	0.1495
0.0585	0.0505	0.0479	0.0555	0.0477	0.0438	0.0865	0.0905	0.0923	0.1116	0.1584	0.1685	0.0712	0.1697
0.0583	0.0503	0.0478	0.0550	0.0479	0.0441	0.0903	0.0269	0.0158	0.0376	0.0211	0.0568	0.0731	0.1230
0.0601	0.0516	0.0488	0.0569	0.0493	0.0451	0.0960	0.1094	0.1130	0.1275	0.1647	0.1813	0.0980	0.1549
0.0626	0.0536	0.0505	0.0598	0.0514	0.0465	0.1814	0.1815	0.1879	0.1933	0.2398	0.2572	0.1290	0.2087
0.0647	0.0553	0.0518	0.0619	0.0532	0.0477	0.1134	0.1454	0.1645	0.1328	0.1996	0.2358	0.1494	0.2266
0.0675	0.0574	0.0533	0.0643	0.0552	0.0492	0.1296	0.1712	0.2021	0.1566	0.2291	0.2719	0.1806	0.2511
0.0681	0.0586	0.0540	0.0646	0.0562	0.0501	0.0373	0.0464	0.0974	0.0163	0.1011	0.1687	0.1550	0.2123
0.0647	0.0576	0.0534	0.0614	0.0550	0.0499	0.2732	0.1682	0.1087	0.2081	0.1278	0.0342	0.0302	0.0862

0.0629	0.0572	0.0533	0.0605	0.0548	0.0501	0.0756	0.0689	0.0252	0.0566	0.0253	0.0393	0.0417	0.0463
0.0657	0.0590	0.0546	0.0641	0.0569	0.0516	0.2667	0.1820	0.2170	0.2310	0.2387	0.2653	0.0102	0.1374
0.0719	0.0628	0.0570	0.0703	0.0608	0.0540	0.4319	0.3547	0.4062	0.3845	0.4215	0.4427	0.1387	0.2759
0.0755	0.0651	0.0586	0.0726	0.0630	0.0557	0.1235	0.1765	0.2566	0.1423	0.2325	0.2999	0.2001	0.2850
0.0782	0.0670	0.0601	0.0742	0.0648	0.0572	0.0345	0.1367	0.2285	0.0975	0.1919	0.2723	0.2431	0.2737
0.0793	0.0684	0.0614	0.0750	0.0661	0.0585	0.0265	0.0862	0.1844	0.0483	0.1400	0.2305	0.2240	0.2441
0.0794	0.0698	0.0626	0.0758	0.0673	0.0597	0.0037	0.0735	0.1742	0.0485	0.1299	0.2216	0.1643	0.2200
0.0789	0.0708	0.0638	0.0761	0.0682	0.0608	0.0103	0.0345	0.1352	0.0176	0.0916	0.1868	0.0901	0.1874
0.0770	0.0708	0.0644	0.0746	0.0679	0.0612	0.1193	0.0774	0.0179	0.0919	0.0251	0.0793	0.0041	0.1145
0.0763	0.0710	0.0651	0.0743	0.0682	0.0618	0.0225	0.0342	0.0483	0.0168	0.0224	0.1097	0.0426	0.0865

0.0774	0.0718	0.0662	0.0757	0.0692	0.0629	0.0976	0.0529	0.1228	0.0838	0.1098	0.1804	0.0260	0.1067
0.0830	0.0748	0.0685	0.0813	0.0727	0.0652	0.3665	0.2927	0.3580	0.3306	0.3574	0.4025	0.1019	0.2284
0.0885	0.0779	0.0710	0.0860	0.0760	0.0676	0.2711	0.2837	0.3634	0.2706	0.3417	0.4085	0.2149	0.3004
0.0924	0.0805	0.0731	0.0887	0.0786	0.0696	0.1068	0.1992	0.2950	0.1551	0.2538	0.3458	0.2705	0.3102

Table 4.8 Sample technical indicators for WTI crude oil data (Before and after normalization)

SMA5	SAM10	SMA20	EMA5	EMA10	EMA20	D5	D10	D20	ED5	ED10	ED20	POSC	EPOSC
Before normalization													
27.646	27.627	27.618	27.503	27.406	27.351	1.029	1.097	1.129	1.551	1.911	2.117	0.001	0.004
27.778	27.748	27.735	27.652	27.505	27.408	0.618	0.729	0.776	1.077	1.617	1.978	0.001	0.005
27.876	27.830	27.812	27.775	27.599	27.466	0.515	0.683	0.746	0.883	1.526	2.016	0.002	0.006
27.752	27.736	27.730	27.663	27.570	27.464	1.124	1.066	1.047	0.807	0.471	0.087	0.001	0.003

27.750	27.733	27.729	27.682	27.597	27.488	0.108	0.048	0.031	0.137	0.445	0.843	0.001	0.003
28.286	28.079	28.013	28.261	27.929	27.672	4.009	4.777	5.023	4.099	5.340	6.316	0.007	0.012
28.723	28.352	28.229	28.648	28.200	27.839	2.428	3.769	4.220	2.696	4.327	5.680	0.013	0.016
29.278	28.700	28.489	29.122	28.540	28.051	2.705	4.774	5.550	3.256	5.362	7.197	0.020	0.020
29.737	29.026	28.716	29.478	28.840	28.255	1.524	4.010	5.134	2.416	4.682	6.849	0.024	0.022
29.872	29.249	28.854	29.575	29.009	28.399	0.341	1.782	3.175	0.659	2.623	4.827	0.021	0.019
30.087	29.545	29.046	29.857	29.266	28.592	1.106	2.963	4.729	1.886	3.945	6.395	0.018	0.020
30.803	30.104	29.400	30.611	29.785	28.928	4.277	6.696	9.251	4.929	7.841	11.036	0.023	0.027
31.368	30.604	29.721	31.144	30.226	29.240	2.684	5.246	8.375	3.422	6.566	10.156	0.024	0.029
31.607	30.928	29.951	31.316	30.486	29.471	0.167	2.366	5.705	1.098	3.850	7.429	0.021	0.026
31.682	31.139	30.140	31.364	30.663	29.660	0.701	1.030	4.379	0.306	2.598	6.068	0.017	0.022
31.694	31.309	30.329	31.446	30.835	29.846	0.265	0.960	4.223	0.521	2.512	5.911	0.012	0.019

31.020	31.109	30.330	30.894	30.645	29.841	3.965	4.240	1.780	3.574	2.791	0.169	0.003	0.008
30.261	30.771	30.264	30.286	30.359	29.767	3.937	5.527	3.945	4.015	4.246	2.342	0.017	0.002
29.729	30.460	30.210	29.897	30.134	29.706	2.047	4.399	3.607	2.600	3.364	1.971	0.025	0.008
29.199	30.077	30.113	29.472	29.858	29.602	1.982	4.846	4.958	2.889	4.148	3.318	0.030	0.013
29.278	29.945	30.130	29.608	29.862	29.629	2.056	0.217	0.830	0.920	0.059	0.849	0.023	0.009
29.936	30.075	30.270	30.162	30.118	29.785	4.456	3.972	3.302	3.674	3.824	4.986	0.005	0.001
30.255	30.090	30.326	30.291	30.197	29.858	0.974	1.529	0.738	0.854	1.170	2.318	0.005	0.003
30.383	30.084	30.343	30.284	30.210	29.897	0.371	0.618	0.241	0.047	0.198	1.247	0.010	0.002
30.767	30.285	30.442	30.613	30.403	30.028	1.636	3.252	2.720	2.147	2.852	4.137	0.016	0.007
After normalization													
0.0755	0.0729	0.0724	0.0731	0.0711	0.0686	0.0716	0.0560	0.0443	0.1088	0.0902	0.0716	0.0053	0.0324
0.0766	0.0739	0.0733	0.0743	0.0719	0.0691	0.0430	0.0372	0.0304	0.0755	0.0763	0.0669	0.0086	0.0488

0.0774	0.0746	0.0740	0.0752	0.0726	0.0696	0.0358	0.0348	0.0293	0.0619	0.0720	0.0682	0.0131	0.0582
0.0764	0.0738	0.0733	0.0744	0.0724	0.0696	0.0782	0.0544	0.0411	0.0566	0.0222	0.0029	0.0046	0.0309
0.0763	0.0738	0.0733	0.0745	0.0726	0.0698	0.0075	0.0025	0.0012	0.0096	0.0210	0.0285	0.0047	0.0282
0.0806	0.0766	0.0756	0.0792	0.0753	0.0713	0.2789	0.2439	0.1971	0.2876	0.2520	0.2135	0.0575	0.1081
0.0841	0.0788	0.0774	0.0823	0.0775	0.0727	0.1689	0.1924	0.1656	0.1891	0.2042	0.1920	0.1014	0.1435
0.0886	0.0817	0.0796	0.0861	0.0803	0.0745	0.1882	0.2437	0.2178	0.2284	0.2530	0.2433	0.1549	0.1834
0.0923	0.0843	0.0814	0.0890	0.0828	0.0762	0.1061	0.2047	0.2014	0.1695	0.2209	0.2315	0.1875	0.1986
0.0934	0.0861	0.0826	0.0897	0.0842	0.0774	0.0238	0.0910	0.1246	0.0462	0.1238	0.1632	0.1637	0.1757
0.0951	0.0885	0.0841	0.0920	0.0863	0.0790	0.0769	0.1513	0.1856	0.1323	0.1862	0.2162	0.1415	0.1817
0.1008	0.0931	0.0871	0.0981	0.0905	0.0818	0.2975	0.3418	0.3630	0.3458	0.3701	0.3731	0.1779	0.2478
0.1054	0.0971	0.0897	0.1024	0.0941	0.0844	0.1867	0.2678	0.3286	0.2401	0.3099	0.3433	0.1910	0.2707
0.1073	0.0998	0.0916	0.1038	0.0962	0.0863	0.0116	0.1208	0.2238	0.0770	0.1817	0.2511	0.1685	0.2431

0.1079	0.1015	0.0932	0.1041	0.0977	0.0879	0.0488	0.0526	0.1718	0.0215	0.1226	0.2051	0.1344	0.2050
0.1080	0.1029	0.0947	0.1048	0.0991	0.0895	0.0184	0.0490	0.1657	0.0366	0.1185	0.1998	0.0952	0.1782
0.1026	0.1012	0.0947	0.1004	0.0975	0.0894	0.2759	0.2164	0.0698	0.2507	0.1317	0.0057	0.0225	0.0739
0.0965	0.0985	0.0942	0.0955	0.0952	0.0888	0.2739	0.2821	0.1548	0.2817	0.2004	0.0792	0.1321	0.0221
0.0922	0.0960	0.0937	0.0923	0.0933	0.0883	0.1424	0.2245	0.1415	0.1824	0.1588	0.0666	0.1930	0.0725
0.0880	0.0929	0.0929	0.0889	0.0911	0.0874	0.1379	0.2473	0.1945	0.2027	0.1957	0.1122	0.2361	0.1205
0.0886	0.0918	0.0931	0.0900	0.0911	0.0877	0.1431	0.0111	0.0326	0.0645	0.0028	0.0287	0.1788	0.0789
0.0939	0.0928	0.0942	0.0945	0.0932	0.0890	0.3100	0.2028	0.1296	0.2578	0.1805	0.1686	0.0365	0.0132
0.0964	0.0930	0.0947	0.0955	0.0939	0.0896	0.0678	0.0780	0.0290	0.0599	0.0552	0.0784	0.0429	0.0286
0.0975	0.0929	0.0948	0.0955	0.0940	0.0899	0.0258	0.0316	0.0095	0.0033	0.0094	0.0422	0.0771	0.0224
0.1005	0.0945	0.0956	0.0981	0.0955	0.0910	0.1138	0.1660	0.1067	0.1506	0.1346	0.1398	0.1228	0.0629

4.3 Evaluation criteria

Three models for anticipating crude oil prices were developed and applied to predict the future crude oil prices of Brent and WTI markets. Numerous experiments were conducted to assess the efficacy of the introduced models. The performance of these frameworks was evaluated using historical data, macroeconomic variables, and technical indicators as input variables. In this research, various commonly used metrics, as listed in Table 4.9, were computed to evaluate the predictive performance of the developed models.

Table 4.9 Evaluation metrics

Metrics	Equation
Root Mean Square Error (RMSE)	$RMSE = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\left(A_j - P_j\right)^2}$
Mean Absolute Percent Error (MAPE)	$MAPE = \frac{1}{M}\sum_{j=1}^{M}\left\|\frac{A_j - P_j}{A_j}\right\|$
Directional statistic (D_{stat})	$Dstat = \frac{1}{M}\sum_{j=1}^{M} d_j$ $d_j = \begin{cases} 1, & \left(A_{j+1} - A_j\right)\left(P_{j+1} - P_j\right) \geq 0 \\ 0, & otherwise \end{cases}$

Prediction Accuracy (PA)	PA $= \frac{\text{Number of correctly predicted value}}{\text{Total number of samples}} X100$
Where, M-Number of samples, A – Actual price, and P-Predicted price	

Lower values of RMSE and MAPE are indicative of a better predictive performance. These metrics essentially measure the extent to which the predictions align with the actual prices. A lower RMSE value suggests that the predicted prices are closer to the actual prices on average, while a lower MAPE value indicates that the prediction have a small percentage error compared to the actual prices. Additionally, a higher value of PA signifies a more accurate prediction. PA quantifies how well the model's predictions match the actual prices. When D_{stat} is close to 1, it indicates that the predictions are very close to the actual prices, suggesting a high level of accuracy. However, if D_{stat} deviates significantly from 1, it represents reduced prediction accuracy.

4.4 Experimental results of the predictive model based on historical data and ML

A crude oil prediction model using ML is designed to anticipate future crude oil prices of Brent and WTI markets. Historical data was sourced from publicly accessible sources, specifically (http://www.eia.gov). The collected data was normalized to a common range utilizing min-max method. Subsequently, normalized data was split into training and testing samples. Two distinct ML models, namely

FFNN and SVM were constructed and trained for the purpose of predicting future crude oil prices. For the sake of conducting experiments, both Brent and WTI data were divided into training and testing datasets at different ratios, 60:40,70:30,80:20, and 90:10. detailed in Table 4.10.

Table 4.10 Details of data division

Data frequency	**Training period**	**No. of training samples**	**Testing period**	**No. of testing samples**	**Total No. of samples**
Daily	02/01/2001-22/04/2014	3376	23/04/2014-28/02/2023	2253	5629
	02/01/2001-07/07/2016	3939	08/07/2016-28/02/2023	1690	5629
	02/01/2001-19/09/2018	4502	20/09/2018-28/02/2023	1127	5629
	02/01/2001-04/12/2020	5065	07/12/2020-28/02/2023	564	5629
Weekly	05/01/2001-11/04/2014	693	18/04/2014-24/02/2023	463	1156
	05/01/2001-24/06/2016	808	01/07/2016-24/02/2023	348	1156
	05/01/2001-14/09/2018	924	21/09/2018-24/02/2023	232	1156

	05/01/2001- 27/11/2020	1039	04/12/2020- 24/02/2023	117	1156
Monthly	01/2001-03/2014	159	04/2014- 02/2023	105	266
	01/2001-05/2016	185	06/2016- 02/2023	81	266
	01/2001-08/2018	212	09/2018- 02/2023	54	266
	01/2001-10/2020	238	11/2020- 02/2023	28	266

A FFNN neural network was built with n input neurons corresponding to number of input neurons, m hidden layers, and a single output neuron representing predicted prices. Several experiments were conducted to explore the most effective architecture and parameter settings for achieving the best performance of the FFNN. These efforts led to improved results and an assessment of the FFNN's overall efficacy.

Number of neurons for hidden layer

The main aim of this test to explore how different quantities of neurons in hidden layer affects the performance of the FFNN. The selection of the best configuration is based on achieving the highest PA. The FFNN is constructed with an input layer comprising 5 input neurons, a single hidden layer, and an output layer with one neuron. To identify the ideal number of neurons for the hidden layer, the neuron

is varied from 8 to 12, and the performance of the FFNN was assessed. Figure 4.2 presents a visual representation of the prediction accuracy of the designed FFNN for different numbers of hidden neurons in the hidden layer. The highest PA attained by this network is 99.25% when utilizing 10 neurons. As a results, the optimal number of hidden neurons is set to be 10.

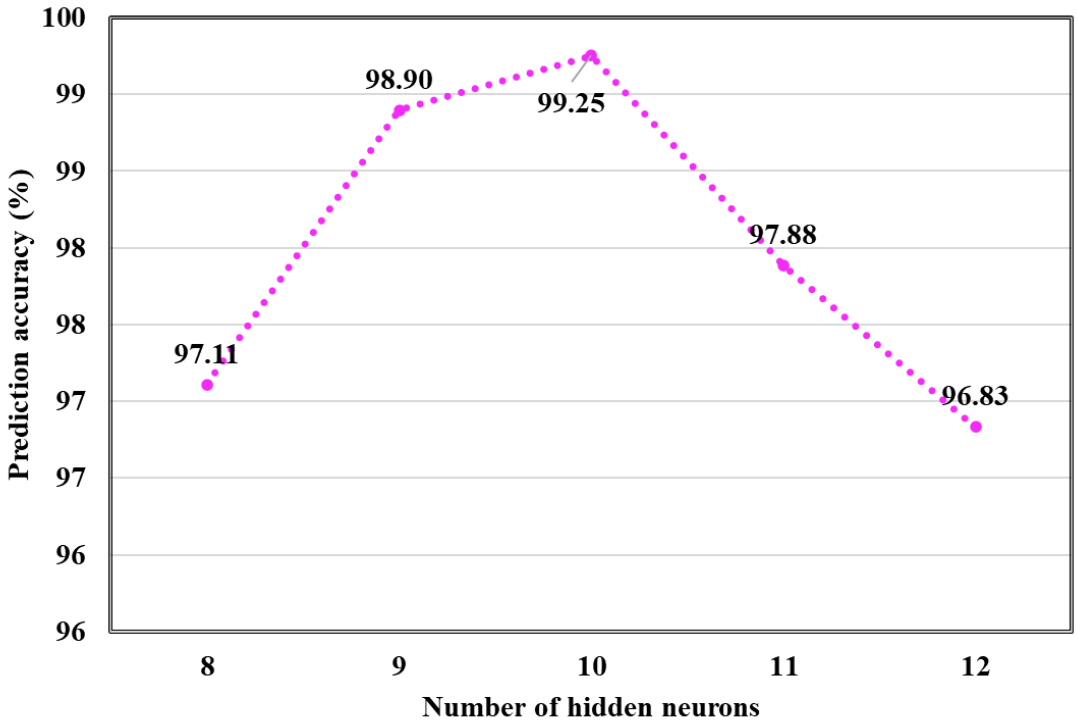


Figure 4.2 Efficacy of the FFNN for varying number of neurons in hidden layer

Number of hidden layers

The prime target of this test is to investigate the influence of varying the number of hidden layers in the FFNN on its performance. The purpose of the experiment is to determine the configuration that produces the highest PA. The experiment involves varying the number of hidden layers, ranging from 1 to 5, to

observe their impact on prediction outcomes. As illustrated in Figure 4.3, the FFNN's PA is plotted against the different number of hidden layers. The introduced network achieved its highest PA of 99.25% when utilizing two hidden layers, each consisting of 10 neurons. As a result, it can be concluded that the optimal number of hidden layers for this problem is two.

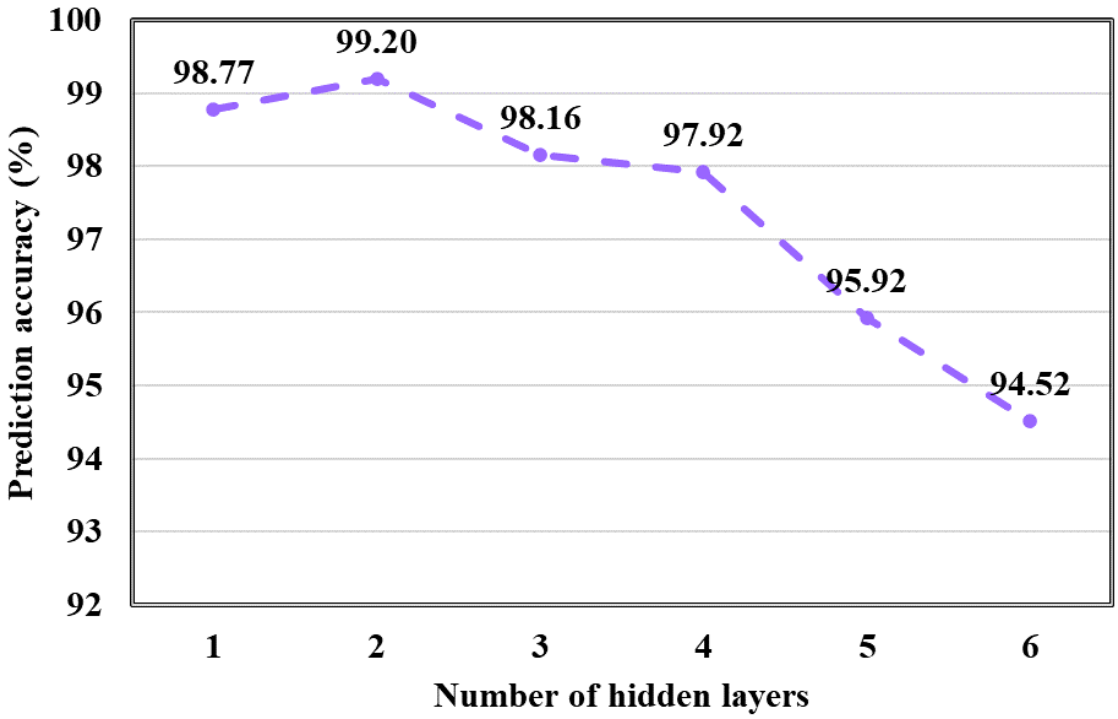


Figure 4.3 Efficacy of the FFNN for varying number of hidden layers

Results on Brent crude oil price prediction

For the purpose of experimentation, Brent crude oil prices spanning from January 1, 2001, to February 28, 2023, has been utilized. Three different frequencies of data were chosen: daily prices, weekly prices, and monthly prices. To assess the generalizability of the FFNN and SVM models, the data is split into training and

testing samples using various ratios, including 60:40, 70:30, 80:20, and 90:10. Table 4.11 and Table 4.12 presents the performance models in predicting daily, weekly, and monthly Brent crude oil prices for different testing sample sizes. The outcomes given in Table 4.11 and Table 4.12 reveals that both the FFNN and SVM models performs superior predictions for daily prices compared to weekly and monthly prices. For instance, when considering weekly and monthly prices, the FFNN achieved RMSE of 0.028 and 0.068, MAPE of 0.055 and 0.137, D_{stat} of 0.998 and 0.991, and PA of 97.99% and 94.91%, respectively. In contrast, for daily prices, the FFNN model attained the lower error values with RMSE of 0.014, MAPE of 0.023, a higher D_{stat} of 0.985, and a higher PA for 99.10%.

Regarding PA specifically, the FFNN exhibited outstanding performance at the 70:30 ratio for daily, weekly, and monthly prices when compared to other ratios, including 60:40, 80:20, and 90:10. For instance, when considering daily prices, the FFNN model achieved a remarkable PA of 99.10% for 70:30 ratio, compared to 99.05% for 60:40, , 98.90% for 80:20, and 98.03% for 90:10. When analyzing weekly prices, the FFNN yielded PA values of 97.94%, 97.94%, 97.62% , and 95.67% for 40%, 30%, 20%, and 10% testing samples, respectively. In case of monthly prices, the FFNN attained PA of 94.91% for 60:40, 94.97% for 70:30, 93.85% for 80:20, and 91.27% for 90:10. This analysis provides compelling evidence of the FFNN's strong performance when employing the 70:30 dataset ratio. In addition to this, these results demonstrate that the daily observations provide more accurate and reasonable predictions from the FFNN compared to the weekly and monthly observations.

Table 4.11 Efficacy of the FFNN for Brent crude oil price prediction

Data frequency	Testing period	RMSE	MAPE	D_{stat}	PA (%)
Daily	23/04/2014-28/02/2023	0.014	0.024	0.992	99.05
	08/07/2016-28/02/2023	0.014	0.023	0.993	99.10
	20/09/2018-28/02/2023	0.016	0.028	0.989	98.90
	07/12/2020-28/02/2023	0.029	0.074	0.985	98.03
Weekly	18/04/2014-28/02/2023	0.031	0.056	0.997	97.94
	01/07/2016-28/02/2023	0.028	0.055	0.998	97.99
	21/09/2018-28/02/2023	0.034	0.064	0.996	97.62
	04/12/2020-28/02/2023	0.061	0.148	0.991	95.67
Monthly	04/2014-02/2023	0.068	0.137	0.991	94.91
	06/2016-02/2023	0.068	0.125	0.988	94.97
	09/2018-02/2023	0.085	0.154	0.981	93.85
	11/2020-02/2023	0.119	0.198	0.963	91.27

Table 4.12 Efficacy of the SVM for Brent crude oil price prediction

Data frequency	Testing period	RMSE	MAPE	D_{stat}	PA (%)
Daily	23/04/2014-28/02/2023	0.014	0.019	0.985	99.04
	08/07/2016-28/02/2023	0.013	0.018	0.988	99.10
	20/09/2018-28/02/2023	0.016	0.022	0.992	98.89
	07/12/2020-28/02/2023	0.028	0.018	0.992	98.03
Weekly	18/04/2014-28/02/2023	0.028	0.036	0.997	97.97
	01/07/2016-28/02/2023	0.027	0.035	0.998	98.02
	21/09/2018-28/02/2023	0.033	0.043	0.995	97.63
	04/12/2020-28/02/2023	0.060	0.036	0.991	95.83
Monthly	04/2014-02/2023	0.075	0.098	0.990	94.54
	06/2016-02/2023	0.074	0.090	0.987	94.74
	09/2018-02/2023	0.087	0.117	0.981	93.74
	11/2020-02/2023	0.126	0.068	0.962	91.31

As given in Table 4.12, the SVM model exhibited lower error values and a higher PA for the daily prices when compared to weekly and monthly prices. Furthermore, the SVM model provided excellent performance specifically with the 70:30 ratio, surpassing other ratios such as 60:40, 80:20, and 90:10. To illustrate this, in terms of error metrics, the model achieved lower error values, including RMSE of 0.013, MAPE of 0.018, a higher D_{stat} of 0.988, and a PA of 99.10% at the 70:30 ratio for daily prices. Similar trends were observed for both weekly and monthly prices.

When examining weekly prices, the SVM model attained PA of 97.97%, 98.02%, 97.63%, and 95.83% for 60:40, 70:30, 80:20, and 90:10 ratios, respectively. Likewise, for monthly prices, the model achieved PA values of 94.54% for 60:40, 94.74% for 70:30, 93.72% for 80:20, and 91.31% for 90:10 ratios. In addition to the tabulated outcomes, the graphical depiction of actual versus predicted prices for daily, weekly, and monthly predictions of Brent prices are shown in Figure 4.4, Figure 4.5, and Figure 4.6, respectively, further showing the proposed models predictive performance. These findings reveal that the SVM model exceptionally performed well in predicting crude oil prices when utilizing 70:30 ratio. Overall, the findings highlight the superiority of the SVM model over the FFNN.

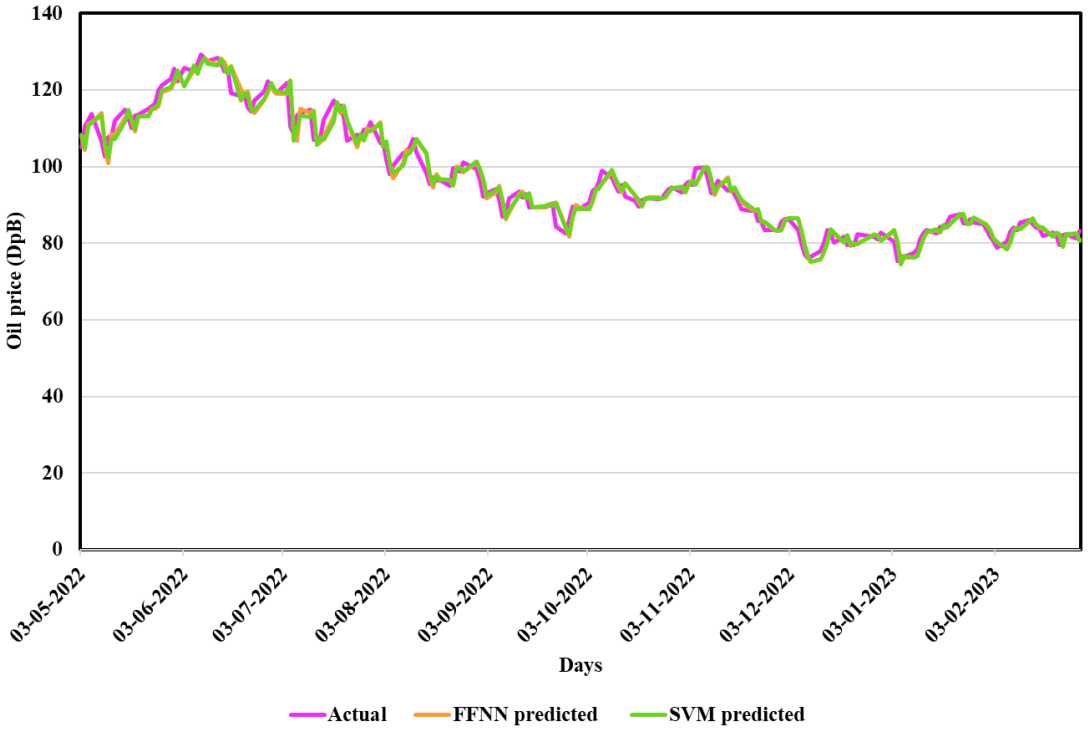


Figure 4.4 Actual Vs. predicted values of daily Brent crude oil prices

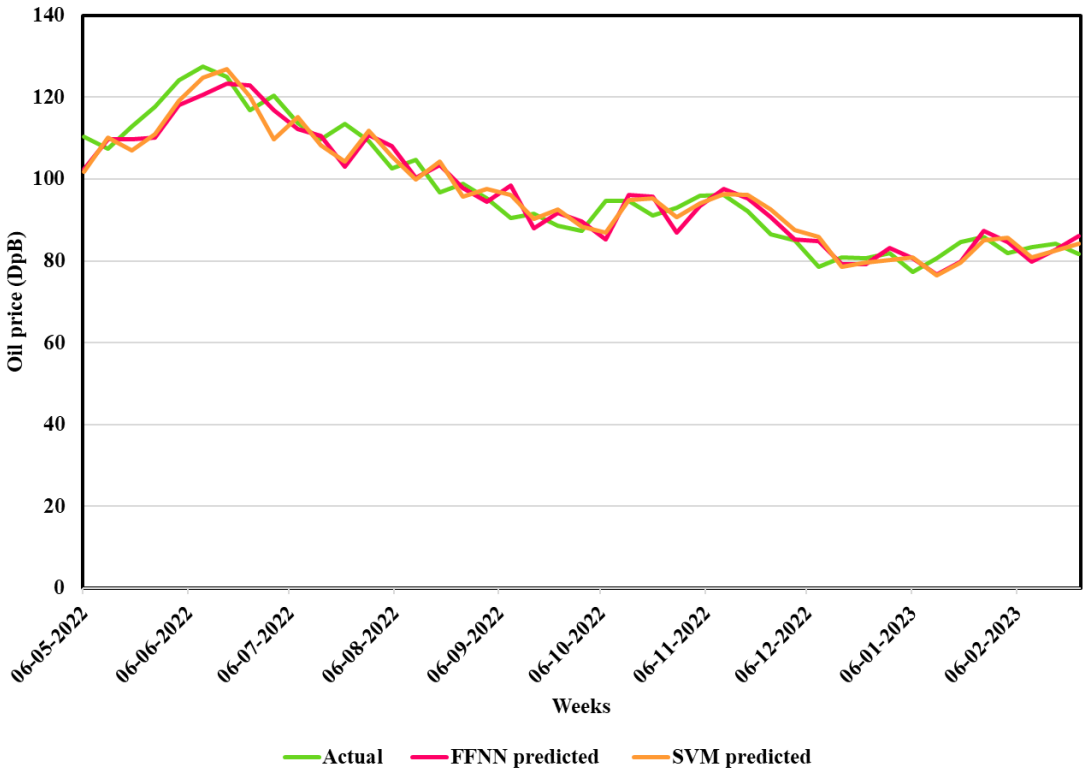


Figure 4.5 Actual Vs. predicted values of weekly Brent crude oil prices

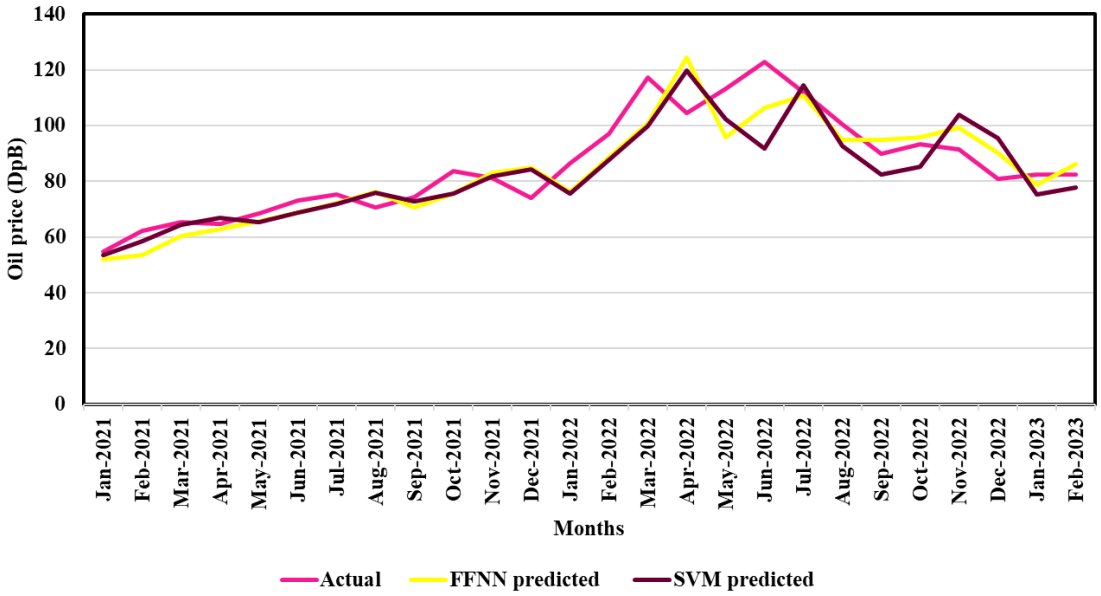


Figure 4.6 Actual Vs. predicted values of monthly Brent crude oil prices

Results on WTI crude oil price prediction

For this case, WTI oil prices from 1 January 2001 to 28 February 2023 has been considered. Three sets of data are chosen, i.e., daily prices, weekly prices, and monthly prices. From these prices, the data is randomly spilt into training and testing samples at several ratios, 60:40, 70:30, 80:20, and 90:10, to validate the effectiveness of the developed model. The daily, weekly, and monthly WTI crude oil price prediction performance for the developed model for various testing samples is summarized in Table 4.13 for FFNN and Table 4.14 for SVM models. Notably, it is evident from both Table 4.13 and Table 4.14 that both the FFNN and SVM models achieved lower prediction error and higher D_{stat} and PA when predicting daily price prices in comparison to their predictions for weekly and monthly prices. For daily prices using the 70:30 ratios, the FFNN exhibited a significantly superior performance. It achieved RMSE of 0.016, indicating a very small prediction error. The MAPE was 0.026, emphasizing the model's accuracy in its prediction. Furthermore, the D_{stat} value

reached 0.993 and PA stood at 98.99%, signifying that nearly 99% of the predictions were accurate.

Conversely, when dealing with weekly and monthly prices, the FFNN yielded slightly less favourable outcomes. The RMSE values for weekly and monthly prices were 0.026 and 0.059, respectively. Although higher that the daily price RMSE, these values still indicate relatively accurate predictions. The MAPE for weekly and monthly prices was 0.041 and 0.111, respectively, suggesting a reasonable level of PA. The D_{stat} value for both weekly and monthly prices were quite high at 0.998, showing a strong model fit. However, it is worth noting that the PA values for weekly and monthly prices, which were 98.75% and 95.32% respectively, were slightly lower than the PA achieved for daily prices. This highlights the model's greater PA in predicting daily prices compared to its predictions for weekly and monthly price fluctuations.

Table 4.13 Efficacy of the FFNN for WTI crude oil price prediction

Data frequency	Testing period	RMSE	MAPE	D_{stat}	PA (%)
Daily	23/04/2014-28/02/2023	0.017	0.027	0.991	98.70
	08/07/2016-28/02/2023	0.016	0.026	0.993	98.99
	20/09/2018-28/02/2023	0.019	0.032	0.991	98.82
	07/12/2020-28/02/2023	0.019	0.032	0.991	98.82

Weekly	18/04/2014-28/02/2023	0.027	0.042	0.997	98.17
	01/07/2016-28/02/2023	0.026	0.041	0.998	98.75
	21/09/2018-28/02/2023	0.034	0.049	0.996	97.77
	04/12/2020-28/02/2023	0.061	0.112	0.991	95.81
Monthly	04/2014-02/2023	0.106	0.163	0.991	92.87
	06/2016-02/2023	0.059	0.111	0.988	95.32
	09/2018-02/2023	0.068	0.141	0.981	94.87
	11/2020-02/2023	0.093	0.164	0.963	92.62

Table 4.14 Efficacy of the SVM for WTI crude oil price prediction

Data frequency	**Testing period**	**RMSE**	**MAPE**	**Dstat**	**PA (%)**
Daily	23/04/2014-28/02/2023	0.015	0.021	0.992	99.01
	08/07/2016-28/02/2023	0.013	0.020	0.997	99.97
	20/09/2018-28/02/2023	0.019	0.026	0.993	98.78
	07/12/2020-28/02/2023	0.032	0.021	0.996	97.74
Weekly	18/04/2014-28/02/2023	0.029	0.051	0.997	98.01

	01/07/2016-28/02/2023	0.026	0.050	0.997	98.24
	21/09/2018-28/02/2023	0.035	0.065	0.995	97.77
	04/12/2020-28/02/2023	0.062	0.039	0.991	95.59
Monthly	04/2014-02/2023	0.081	0.118	0.990	93.93
	06/2016-02/2023	0.069	0.092	0.987	94.98
	09/2018-02/2023	0.105	0.124	0.981	92.54
	11/2020-02/2023	0.115	0.071	0.962	91.28

As detailed in Table 4.14, the SVM model displayed superior performance metrics, including lower error values and a higher D_{stat} and PA, particularly in the context of daily prices, when compared to its performance on weekly and monthly prices. This superiority was most pronounced when the model was assessed under the 70:30 ratio, outperforming alternative ratios, including 60:40, 80:20, and 90:10. For daily prices, the SVM model showed better results, achieving a low RMSE of 0.013, signifying a highly accurate prediction. The MAPE stood at 0.020, showing the model's precision. Moreover, the D_{stat} was 0.997, reflecting the model's ability to capture daily fluctuations. The PA for this case reached 99.97%, illustrating that the SVM model's predictions closely aligned with actual crude oil daily prices.

Similar trends were observed when analyzing the model's performance for both weekly and monthly prices. For weekly prices, the SVM model achieved PAs of 98.01%, 98.24%, 97.77%, and 95.59% for the 60:40, 70:30, 80:20, and 90:10 dataset ratios, respectively. This consistency in performance across various ratios proves the SVM model's reliability in predicting weekly prices. Similarly, for monthly prices, the SVM model demonstrated PA values of 93.93 % for 60:40, 94.98% for 70:30, 92.54% for 80:20, and 91.28% for 90:10, confirming its reliability in predicting monthly prices. These outcomes indicate that the SVM model maintains a high level of PA in predicting monthly prices across different dataset split ratios. To provide visual representation of the FFNN and SVM performance, Figure 4.7, Figure 4.8, and Figure 4.9 display the comparison between actual and predicted prices for daily, weekly, and monthly predictions for WTI crude oil prices, respectively, further emphasizing the SVM model's prowess in accurately predicting oil prices across different timeframes.

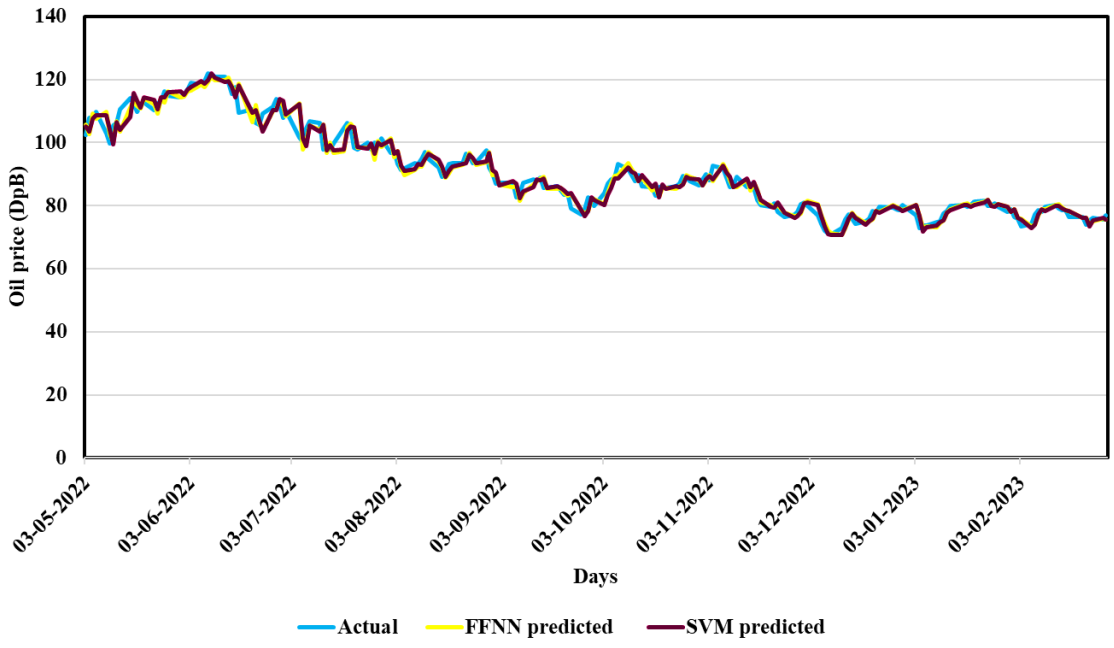


Figure 4.7 Actual Vs. predicted values of daily WTI crude oil prices

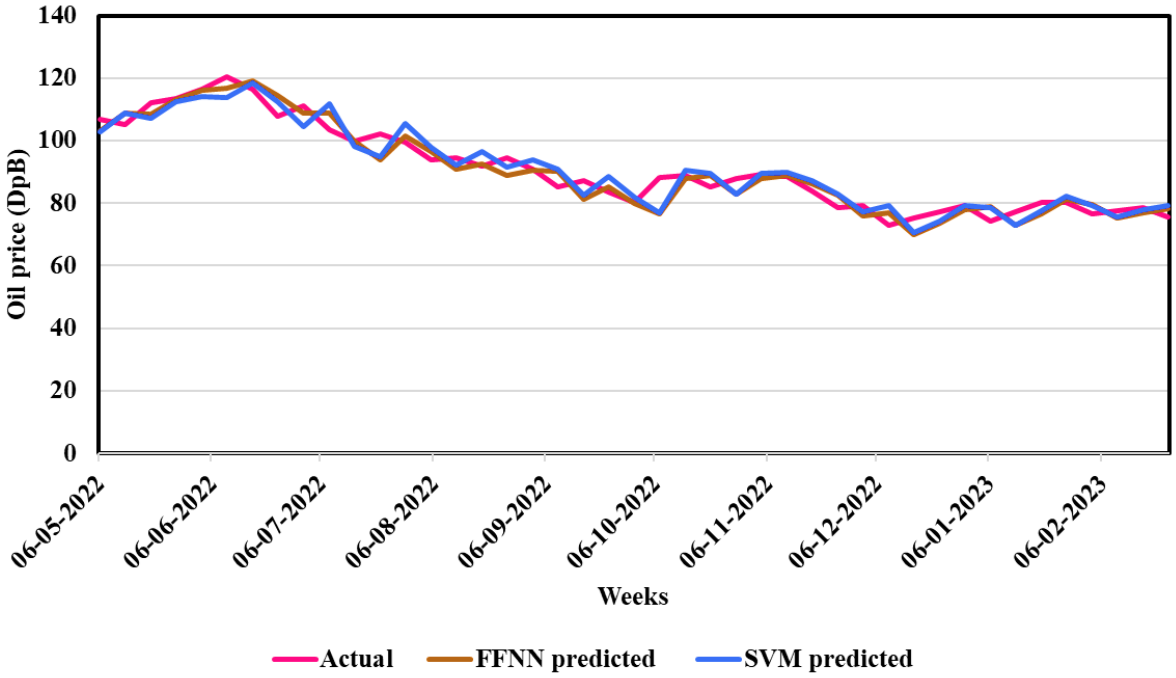


Figure 4. 8Actual Vs. predicted values of weekly WTI crude oil prices

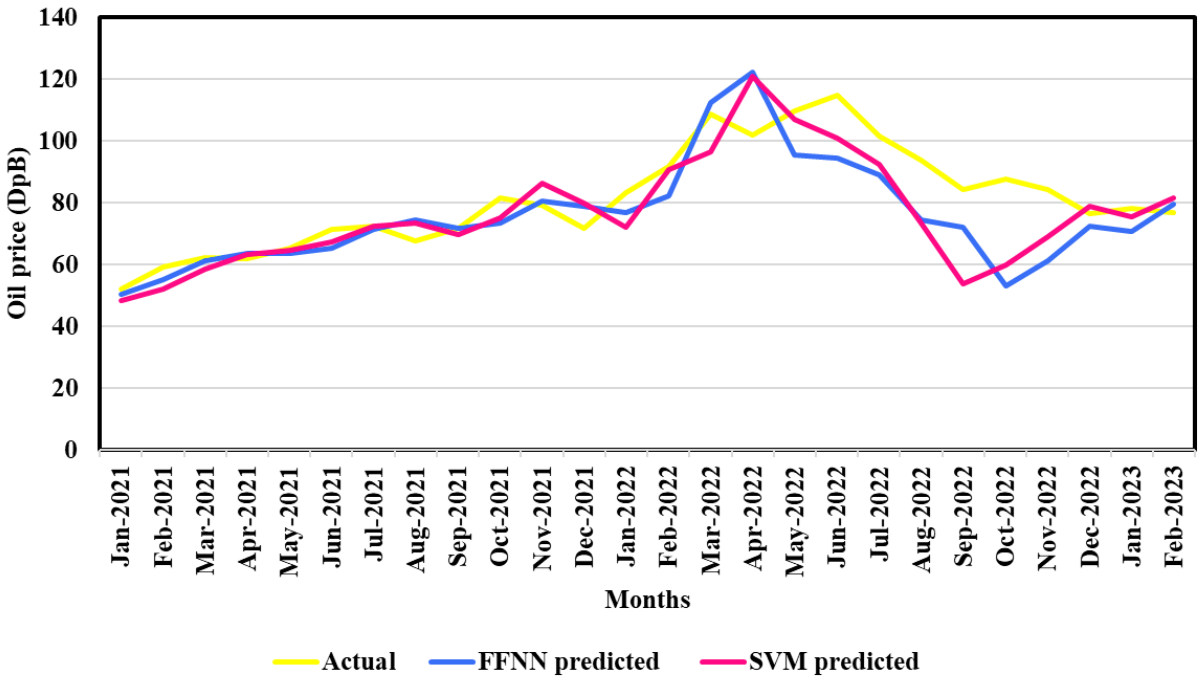


Figure 4.9 Actual Vs. predicted values of monthly WTI crude oil prices

Comparison with earlier methods

The prediction performance of the developed model is compared with the earlier models, including regression tree, Chen et al. (2019), RF Costa et al. (2021), and cat boost, Jabeur et al. (2022). The evaluation is based on RMSE for daily, weekly, and monthly predictions. For both of selected datasets, the proposed model consistently demonstrated superior performance across all prediction horizons- daily, weekly, and monthly. Figure 4.10 and Figure 4.11 presents visual comparisons of the introduced model's performance in relation to the previous models concerning Brent and WTI price predictions over various timeframes, respectively.

For daily Brent crude oil price predictions, the RMSE values of the models are given below:

- Regression tree = 0.029
- RF =0.025
- Cat boost = 0.023
- Introduced model with FFNN= 0.014 and
- Introduced model with SVM = 0.013

It is evident from the comparison results that the developed model with SVM outperformed all the other models for daily price predictions, with an impressively low RMSE of 0.013. This demonstrates the effectiveness of the proposed model in capturing the fluctuations of daily price, which is crucial in financial prediction. Furthermore, when considering, weekly and monthly price predictions, the introduced model maintained a consistent level of performance. This consistency is evidence to its robustness and adaptability in various prediction scenarios.

For WTI crude oil price predictions, the RMSE values of the models are as follows:

- Regression tree = 0.032
- RF =0.028
- Cat boost = 0.027
- Introduced model with FFNN= 0.016 and
- Introduced model with SVM = 0.013

For the WTI crude oil data, the introduced model with SVM achieved a remarkably low RMSE of 0.013 for daily price predictions, which is significantly lower than the RMSE values of the compared models. This result shows the exceptional predictive power of the developed model. Consistent with the Brent oil data, the model also exhibited competitive performance for weekly and monthly price predictions for the WTI dataset, further confirming its potential as a reliable tool for anticipating various time horizons.

The outcomes of this comparative analysis establish the superiority of the developed model in predicting daily, weekly, and monthly oil prices when compared to the earlier models in the literature. Its consistently lower RMSE values across different datasets and prediction horizons highlight its effectiveness and versatility in the field of financial prediction, particularly in the volatile and dynamic energy market. These findings reinforce the practical utility of the introduced model and its power for real-world applications in financial decision-making and risk-management.

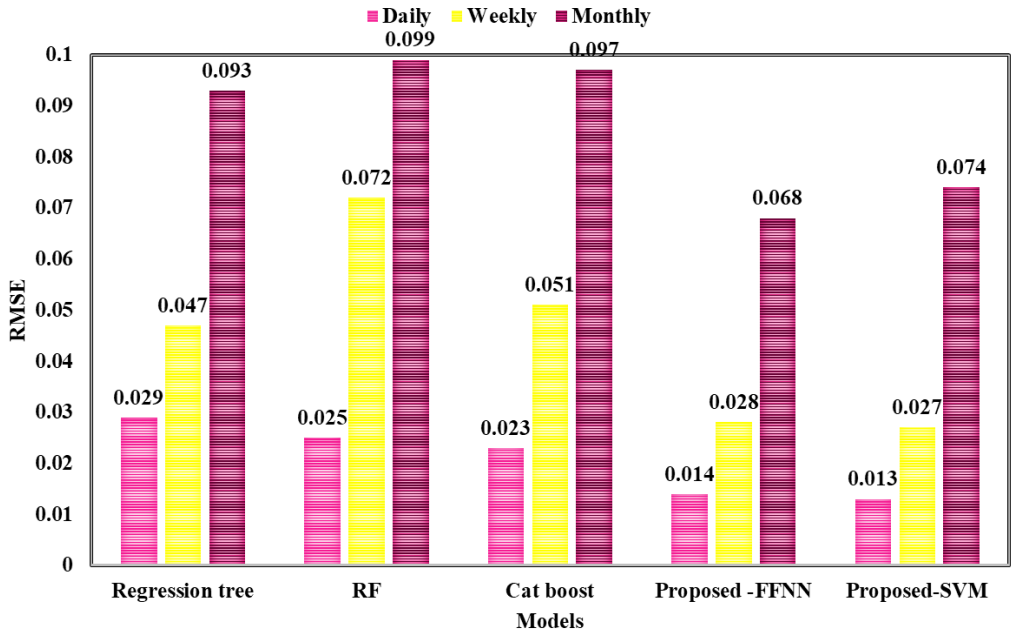


Figure 4.10 Performance comparison with earlier models for Brent crude oil price prediction

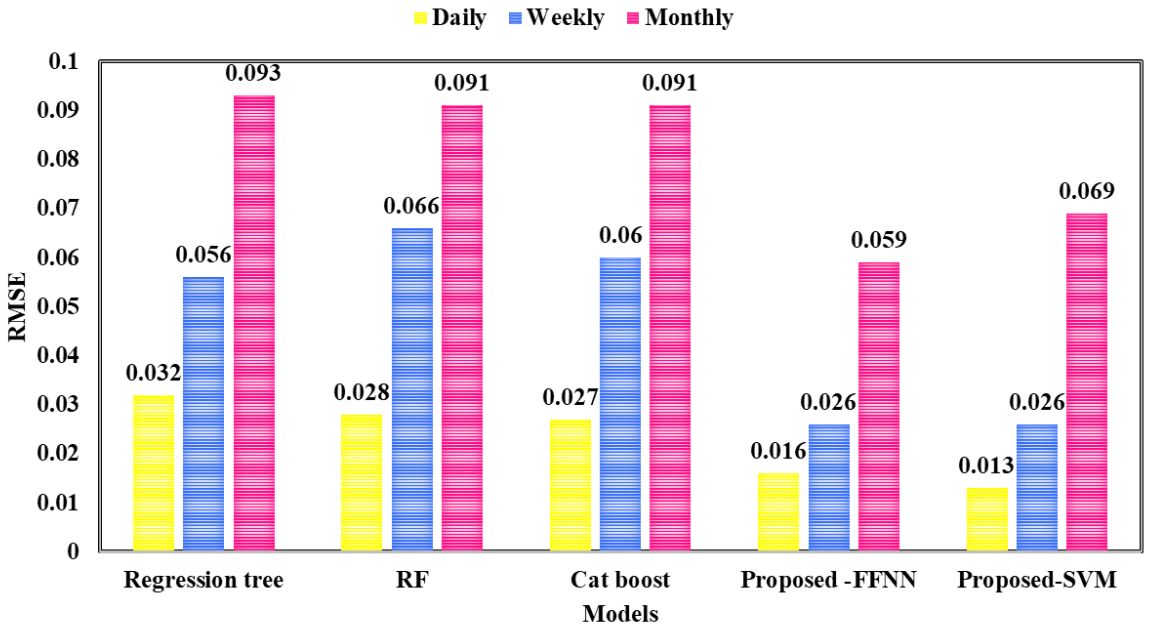


Figure 4.11 Performance comparison with earlier models for WTI crude oil price prediction

4.5 Experimental results of the predictive model based on macroeconomic variables and ML

The main objective of this model is to anticipate future crude oil prices leveraging the power of macroeconomic variables in conjunction with ML models, FFNN and SVM. The model was designed to predict two benchmark crude oil prices namely Brent and WTI, which play pivotal roles in the global energy market. To build a predictive framework, a set of macroeconomic variables based on the exhaustive analysis of the determinants of crude oil prices were chosen. These selected factors encompassed WTI and Brent oil prices, GP, SP, NGP, EUR/USD, S&P 500, DI, and GBP/USD. These variables were collected from January 1 2001, to February 28,2023, providing a substantial dataset for analysis. Before subjecting the data to prediction framework, a crucial preprocessing step was adopted to normalize the data, scaling it within a common range of 0 to 1. This normalization process ensures that all variables contribute equally to the prediction process, preventing any undue bias based on their original values.

To evaluate the efficacy of the prediction model, 10-fold cross validation was used. This involved partitioning the normalized data into ten equally sized subsets, as depicted in Figure 4.12. During each iteration, one subset was used as the test set, while the remaining nine subsets were formed as the training set. This process was repeated ten times, with each subset having the opportunity to serve as the test set once. The final assessment of the model's performance was computed by taking the mean of the results obtained in these individual iterations.

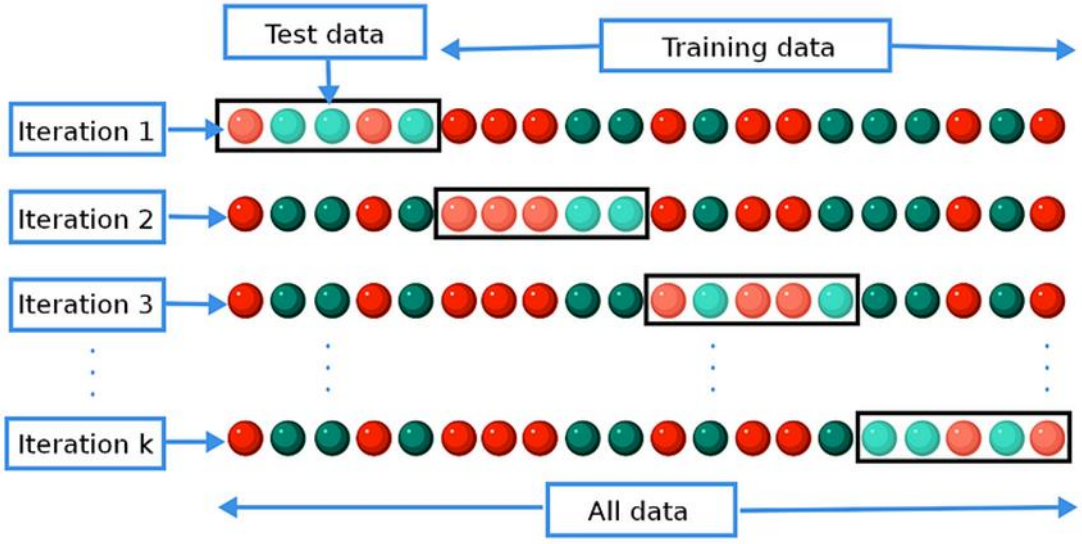


Figure 4.12 K-fold cross-fold validation

This approach to evaluation provides a comprehensive assessment of the proposed model's predictive capabilities, accounting for variations in data distribution and ensuring its generalizability to unseen data. Through this methodology, this research intends to provide valuable insights into the dynamic interplay between the macroeconomic variables and ML, with implications for decision-making in the energy and financial sectors.

Results on Brent crude oil price prediction

In order to evaluate the performance of the developed framework, various combinations of macroeconomic factors were used. These combinations are detailed in Table 4.15. Effectiveness of the introduced prediction model in predicting Brent crude oil prices at different time horizons, including daily, weekly, and monthly, utilizing these diverse combinations of macroeconomic factors, as tabulated in Table 4.16.

Table 4.15 Various combinations of macroeconomic variables

Number of influential factors	Data selection	Exceptional influential factor
1	GP	SP, S&P 500, EUR/USD, NGP, DI, GBP/USD
2	GP, SP	S&P 500, EUR/USD, NGP, DI, GBP/USD
3	GP, SP, S&P 500	EUR/USD, NGP, DI, GBP/USD
4	GP, SP, S&P 500, EUR/USD	NGP, DI, GBP/USD
5	GP, SP, S&P 500, EUR/USD, NGP	DI, GBP/USD
6	GP, SP, S&P 500, EUR/USD, NGP, DI	GBP/USD
7	GP, SP, S&P 500, EUR/USD, NGP, DI, GBP/USD	N/A

Table 4.16 Brent crude oil price prediction results

Data frequency	Number of influential factors	FFNN			SVM		
		RMSE	MAPE	PA (%)	RMSE	MAPE	PA (%)

Daily	1	0.389	0.232	65.52	0.388	0.231	65.74
	2	0.334	0.193	71.89	0.324	0.189	72.92
	3	0.183	0.079	84.82	0.179	0.077	85.23
	4	0.180	0.062	85.01	0.178	0.059	85.12
	5	0.111	0.040	90.63	0.098	0.041	91.88
	6	0.081	0.035	93.51	0.071	0.035	94.41
	7	0.076	0.034	93.92	0.071	0.034	94.42
Weekly	1	0.411	0.232	65.02	0.395	0.227	65.74
	2	0.388	0.201	67.20	0.384	0.198	67.67
	3	0.243	0.125	80.97	0.237	0.125	81.87
	4	0.233	0.079	79.95	0.214	0.080	82.31
	5	0.103	0.049	91.75	0.101	0.050	92.03
	6	0.105	0.445	91.67	0.089	0.044	92.95
	7	0.085	0.044	92.99	0.081	0.040	93.56
Monthly	1	0.447	0.246	60.82	0.440	0.227	61.90
	2	0.301	0.163	76.24	0.291	0.170	77.47

	3	0.186	0.115	85.23	0.169	0.111	86.65
	4	0.141	0.069	89.56	0.130	0.067	90.46
	5	0.134	0.064	90.30	0.133	0.062	90.74
	6	0.115	0.061	92.17	0.115	0.060	92.40
	7	0.104	0.059	92.71	0.103	0.057	93.10

Upon analyzing the empirical outcomes derived from daily, weekly, and monthly data, the following findings emerged:

In analysis of daily price predictions, it became evident that the prediction performance of both the FFNN and SVM models excels, when using all seven macroeconomic variables, surpassing the performance of the models that rely on different input combinations. It is worth noting that actual crude oil prices exhibit a high degree of complexity, leading to significant disparities between the predicted prices of models relying on a single input and the actual prices. This phenomenon results in limited PA. For instance, the RMSE values for the FFNN and SVM models incorporating input variables 1 through 7, were as follows: 0.389 and 0.388, 0.334 and 0.324, 0.183 and 0.179,0.180 and 0.178, 0.111 and 0.098, 0.081 and 0.071, 0.076 and 0.071, respectively. These statistics clearly indicate that both models achieved lower prediction errors when all seven inputs were considered. Remarkably, the SVM model achieved its lowest PA of 65.74% when using only the GP input, whereas it attained

the highest PA of 94.4% when all seven input variables were incorporated. Likewise, the FFNN model reported its lowest PA of 65.52% when using only the GP input but yielded the highest PA of 93.92% when using all seven input variables. When compared to models utilizing alternative input combinations, both the FFNN and SVM with all seven inputs consistently delivers exceptional results.

These trends were also observed in the weekly and monthly prediction data. Both the FFNN and SVM employing all seven input variables consistently outperformed models using fewer input variables. The FFNN attained higher PA, with values of 92.99% and 92.71% for weekly and monthly data, respectively. The SVM model demonstrated superior performance, yielding PA of 93.95% for weekly and 93.10% for monthly data. These outcomes reveal the significance of selecting appropriate macroeconomic variables, which effectively mitigates the inherent complexity of crude oil prices, boosting the PA of the model.

To provide a more visual representation of findings, Figure 4.13, Figure 4.13, and Figure 4.15. Depicts comparisons between the actual and predicted prices of Brent crude oil for daily, weekly, and monthly predictions, respectively. These visualizations offer a clear way to grasp the model's performance across various time horizons and input combinations, reinforcing the significance of research in the field of crude oil price prediction.

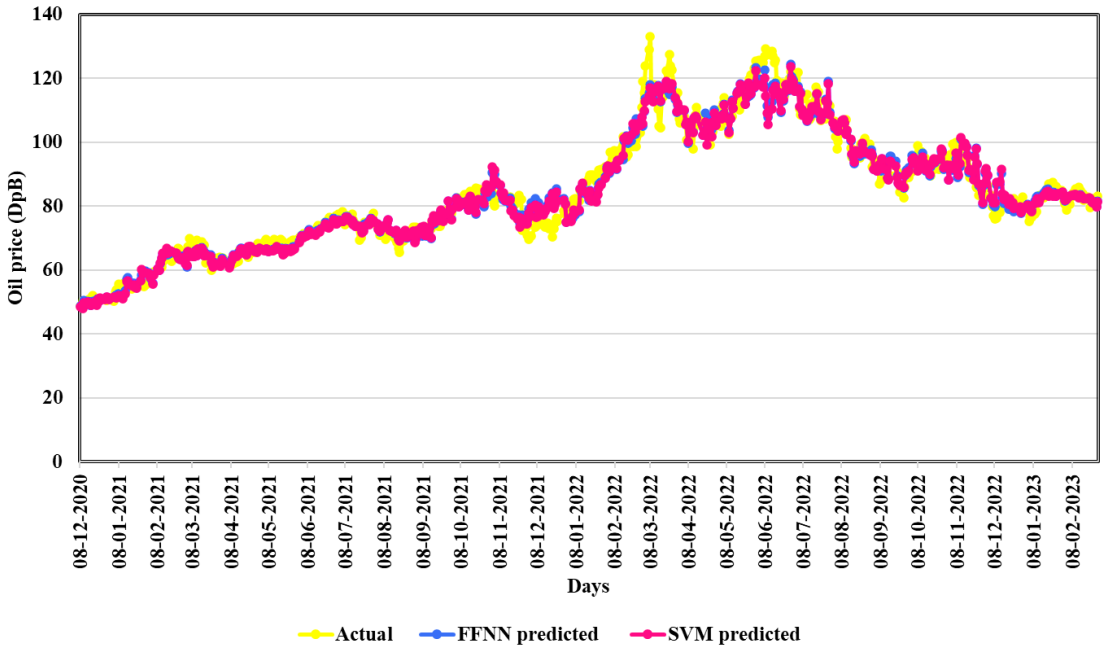


Figure 4.13 Comparison between actual and predicted daily Brent crude oil prices

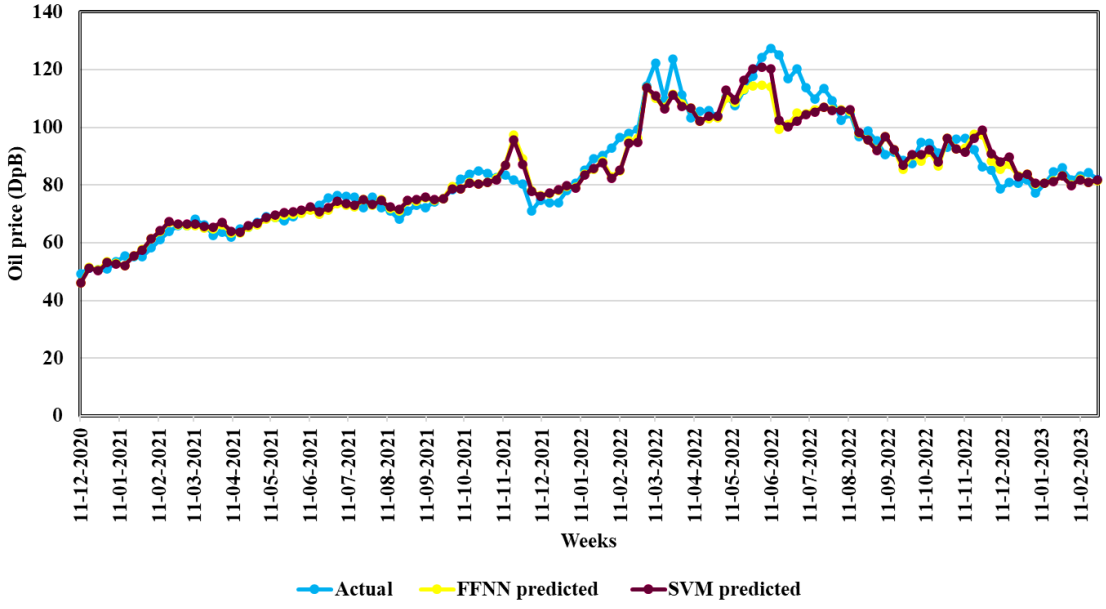


Figure 4.14 Comparison between actual and predicted weekly Brent crude oil prices

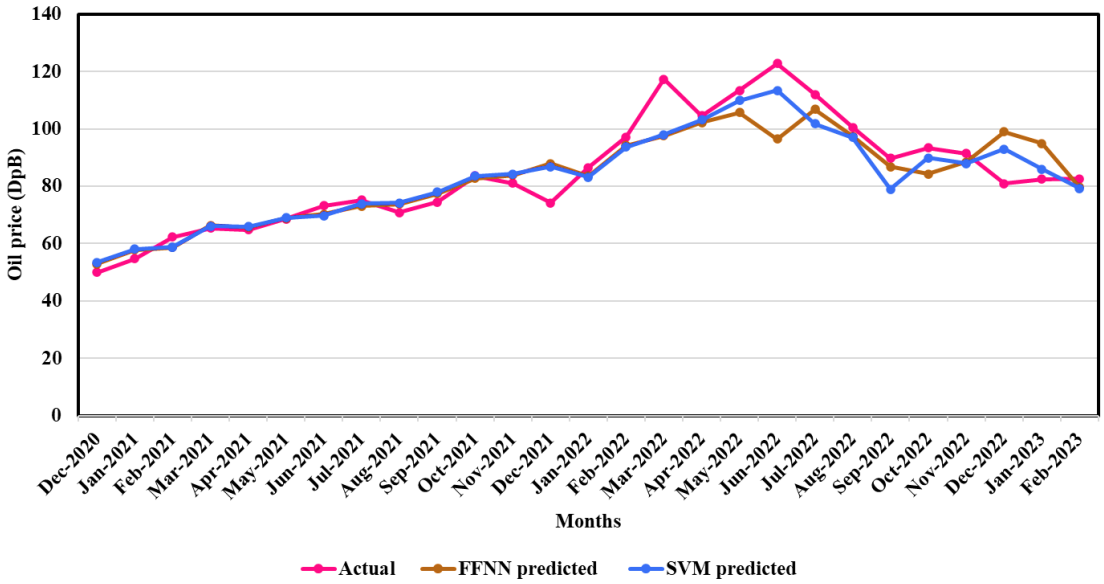


Figure 4.15 Comparison between actual and predicted monthly Brent crude oil prices

Results on WTI crude oil price prediction

Table 4.17 summarizes the effectiveness of designed prediction models in the task of predicting daily, weekly, and monthly crude oil prices. This assessment encompassed a range of combinations of macroeconomic variables which are listed in Table 4.15. The analysis of daily, weekly, and monthly price predictions yielded noteworthy observations. Specifically, it is observed that the models provided outstanding performance when all seven macroeconomic variables were integrating into the modelling frameworks, as compared to models utilizing alternative input combinations.

Table 4.17 WTI crude oil price prediction results

Data frequency	Number of influential factors	FFNN			SVM		
		RMSE	MAPE	PA (%)	RMSE	MAPE	PA (%)
Daily	1	0.415	0.201	62.85	0.411	0.201	63.32
	2	0.363	0.184	68.97	0.355	0.181	69.82
	3	0.179	0.075	85.07	0.177	0.076	85.31
	4	0.192	0.060	83.57	0.189	0.059	83.90
	5	0.111	0.040	90.45	0.101	0.040	91.55
	6	0.089	0.038	92.64	0.078	0.036	93.63
	7	0.080	0.036	93.50	0.073	0.035	94.23
Weekly	1	0.410	0.195	64.03	0.398	0.194	65.23
	2	0.404	0.198	65.09	0.403	0.188	65.68
	3	0.244	0.127	80.91	0.242	0.127	81.07
	4	0.223	0.082	81.82	0.209	0.080	83.19
	5	0.104	0.050	91.93	0.092	0.050	93.10
	6	0.088	0.049	92.52	0.080	0.045	93.22

	7	0.079	0.046	93.92	0.078	0.040	94.66
Monthly	1	0.457	0.198	59.84	0.455	0.197	60.19
	2	0.281	0.151	76.56	0.226	0.142	81.50
	3	0.185	0.115	85.41	0.167	0.109	86.98
	4	0.135	0.068	89.99	0.125	0.067	90.80
	5	0.133	0.060	90.93	0.124	0.060	91.52
	6	0.125	0.061	91.33	0.110	0.059	92.76
	7	0.103	0.057	93.39	0.094	0.053	93.85

For daily price predictions, the FFNN model with input combinations 1,2,3,4,5,6, and 7 produced impressive outcomes, as evidenced by RMSE, MAPE, and PA values of (0.415,0.201, and 62.85%), (0.363,0.184, and 68.97%), (0.179,0.075, and 85.07%), (0.192,0.060, and 83.57%), (0.111,0.040, and 90.45%), (0.089,0.038, and 92.64%), and (0.080,0.036, and 93.50%), respectively. Likewise, the SVM model employing various input combinations 1 through 7, yielded RMSE, MAPE, and PA values of (0.411,0.201, and 63.32%), (0.355,0.181, and 69.82%), (0.177,0.076, and 85.31%), (0.189,0.059, and (83.90%), (0.101,0.040, and 91.55%), (0.078,0.036, and 93.63%), and (0.073,0.036, and 94.23%), respectively.

It is quite clear from these outcomes that the models achieved significantly lower prediction errors and higher PA when all seven input variables were used. Moreover, the FFNN model, when equipped with all macroeconomic variables, attained the highest PA of 93.92% for weekly data and 93.39% for monthly data. Similarly, the SVM model, employing all seven input variables, exhibited a superior PA of 94.66%, and 93.85% for weekly and monthly data, respectively. These findings highlight the importance of selecting suitable macroeconomic variables, as it effectively mitigates the complexity of crude oil prices, consequently enhancing the model's PA. To provide a visual depiction of these outcomes, Figure 4.16, Figure 4.17, and Figure 4.18 offer a graphical representation of the actual and predicted prices for daily, weekly, and monthly predictions of WTI crude oil prices, respectively.

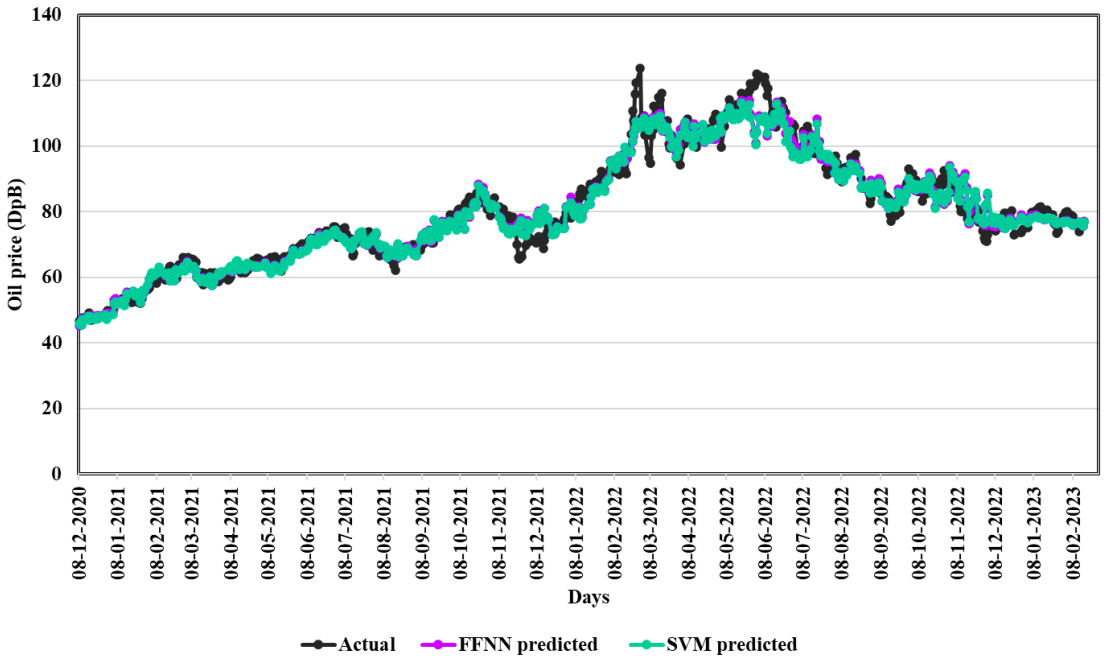


Figure 4.16 Comparison between actual and predicted daily WTI crude oil prices

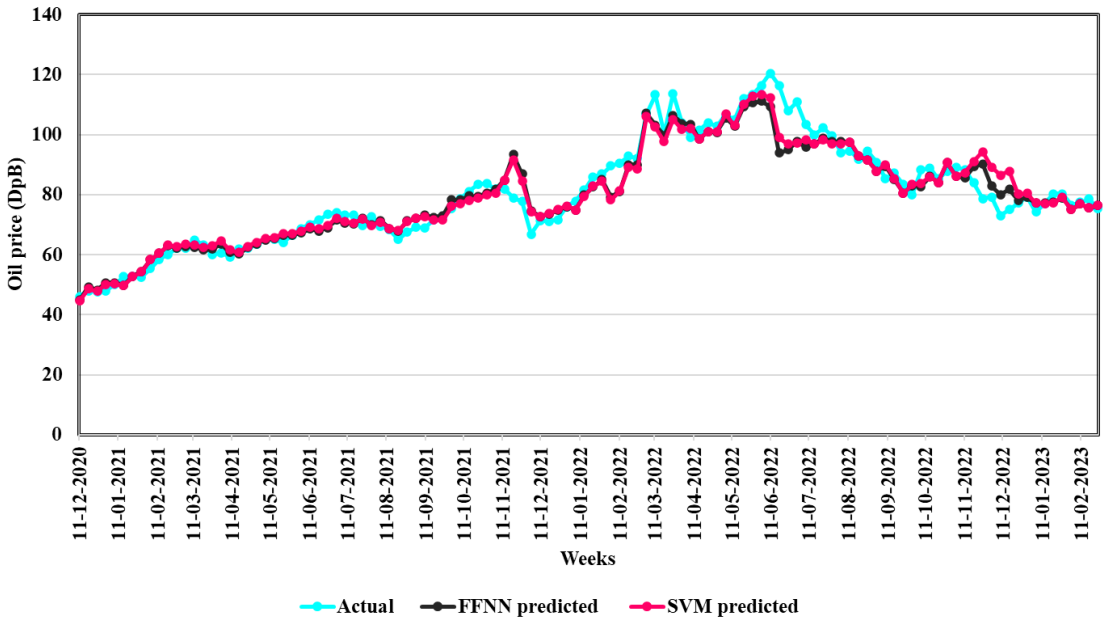


Figure 4.17 Comparison between actual and predicted weekly WTI crude oil prices

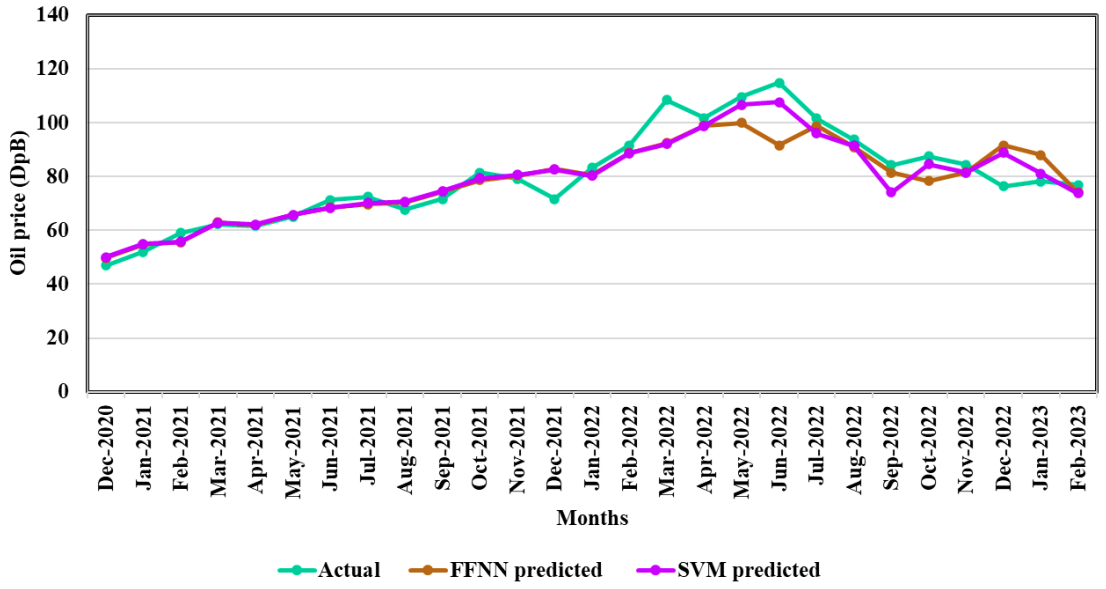


Figure 4.18 Comparison between actual and predicted monthly WTI crude oil prices

Comparison with the well-known methods

To further demonstrate the predictive capabilities of the developed models, a comparative analysis was conducted against earlier approaches, including Hybrid model, Abdollahi and Ebrahimi (2020) and LSTM model, Yan et al. (2021). Figure 4.19 presents a comparison of developed model's performance against earlier models on Brent oil data. For Brent oil prices, the hybrid model achieved RMSE values of 0.311 for daily, 0.325 for weekly, and 0.364 for monthly predictions. In contrast, the LSTM model attained RMSE values of 0.111, 0.152, and 0.183 for daily, weekly, and monthly predictions, respectively. Remarkably, both the FFNN and SVM models demonstrated exceptional performance, significantly lower RMSE values of 0.076 and 0.071 for daily predictions, 0.085 and 0.081 for weekly projections, and 0.104 and 0.103 for monthly price predictions, respectively.

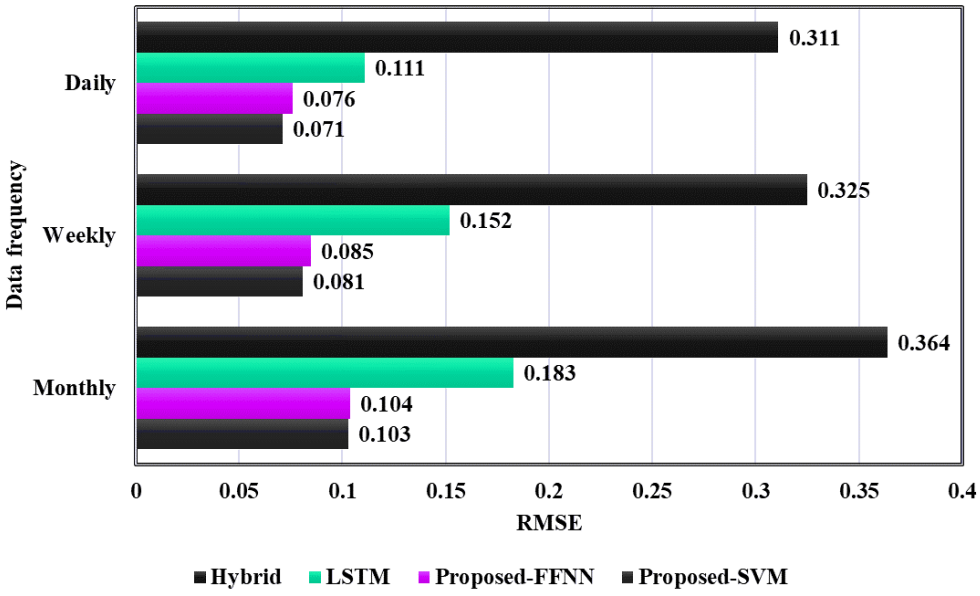


Figure 4.19 Prediction outcomes comparison between the designed models and earlier models (Brent oil data)

Figure 4.20 compares the efficacy of the developed models with the past approaches, including hybrid and LSTM. For WTI crude oil data, both the hybrid and LSTM models attained RMSE values of 0.32 and 0.124 for daily data, 0.331 and 0.147 for weekly data, and 0.336 and 0.172 for monthly data, respectively. On the contrary, the FFNN and SVM models showed better results with RMSE values of 0.08 and 0.073, 0.079 and 0.078, and 0.103 and 0.094 for daily, weekly, and monthly predictions, respectively. These findings confirm the potential of the developed models, demonstrating its superior PA when compared to the hybrid and LSTM models in the task of crude oil price prediction.

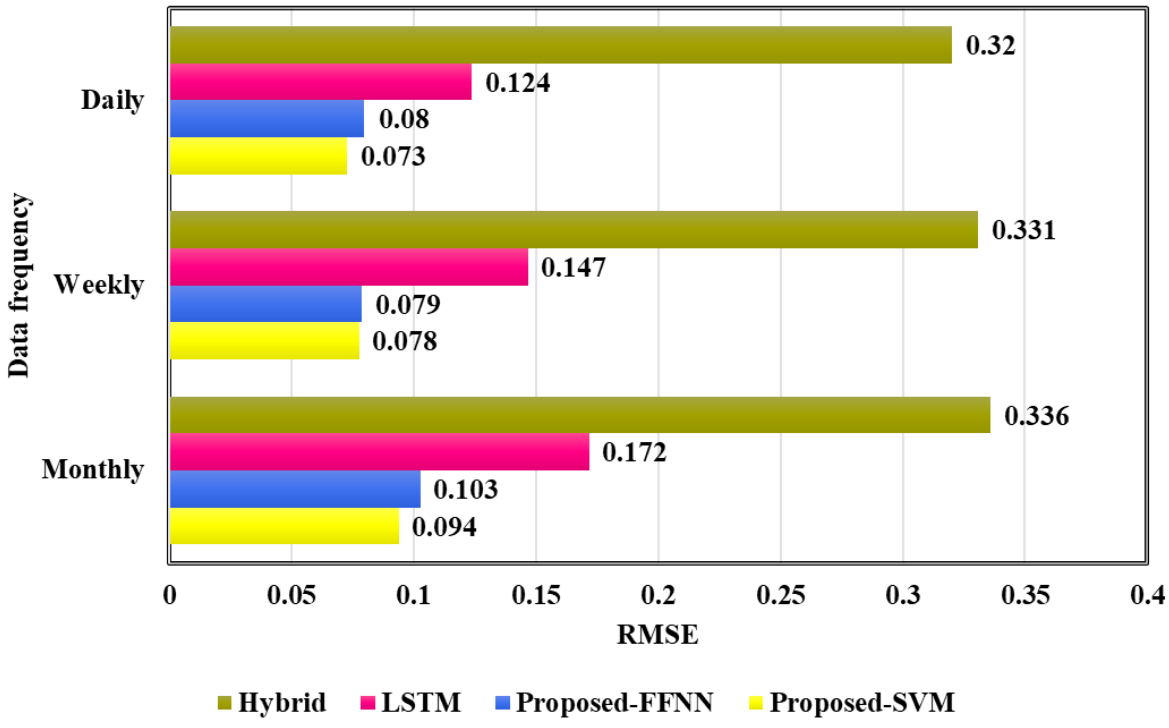


Figure 4.20 Prediction outcomes comparison between the designed models and earlier models (WTI oil data)

4.6 Experimental results of the predictive model based on technical indicators and ML

The fundamental focus of this model is to predict future crude oil price with the help of technical indicators and ML models such as FFNN and SVM. Specifically, the models were constructed to predict two key benchmark crude oil prices: Brent and WTI, both of which hold significant importance in the global energy market.

To establish a reliable predictive model, a total of 14 technical indicators were selected. These chosen technical indicators included SMA5, SMA 10, SMA 20, EMA 5, EMA 10, EMA 20, D5, D10, D20, ED5, ED10, Ed20, POSC, and EPOSC. Data was gathered from January 1, 2001, to February 28,2023. Primarily, historical data was collected and then normalized to a common range of 0 to 1. Subsequently, 14 technical indicators were derived and used as input for the predictors such as FFNN and SVM.

A FFNN model was designed with 14 input neurons representing 14 these technical indicators, along with a hidden layer and an output layer with single neuron representing the predicted value. Sigmoidal and linear activation functions were applied to the hidden layer and output layer, respectively. The determination of the number of hidden layers and hidden neurons typically involved a trail-and-error approach. Hence, multiple experimental scenarios were carried out to investigate the most optimal topology and parameters for the best FFNN, resulting in superior outcomes and an evaluation of the FFNN's overall performance.

Number of hidden neurons

In this scenario, the main goal is to investigate the impact of different quantities of neurons in hidden layer on the overall performance the FFNN. It is essential to identify the appropriate number of hidden neurons to achieve the highest PA. To pinpoint the optimal number of neurons for the hidden layer, experiments were conducted by varying the number of neurons within the range of 10 to 19. The performance of the FFNN was assessed under each of these configurations.

Figure 4.21 presents a pictorial depiction of the PA achieved by the developed FFNN across the spectrum of different hidden neurons in the hidden layer. It is noteworthy that the FFNN reached its peak PA of 97.9% precisely when using 14 neurons in the hidden layer. This choice ensures that the FFNN is finely tuned to achieve the best results in crude oil prediction. Therefore, optimal number of neurons was set to 14.

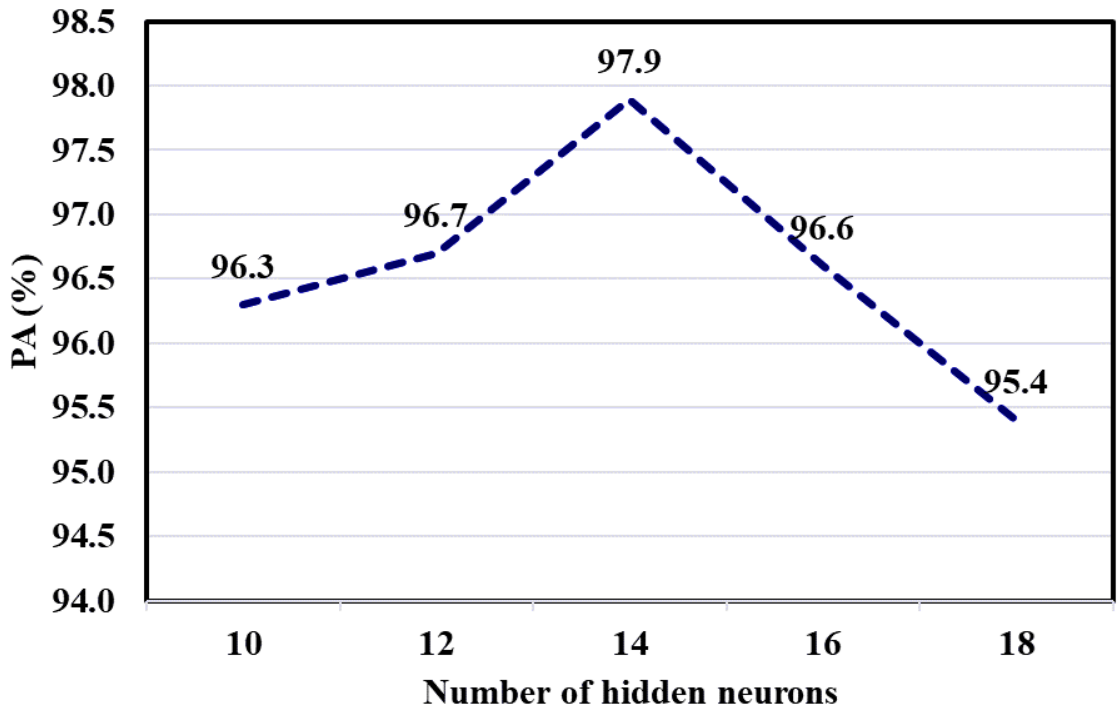


Figure 4.21 FFNN performance for varying number of hidden neurons

Number of hidden layers of the best FFNN performance

This test represents a comprehensive examination of the influence exerted by alterations in the number hidden layers in the FFNN on its overall performance. The primary aim of this test is to pinpoint the specific configuration that yields the highest degree of PA, which is of paramount importance in predictive modeling task. FFNN typically starts with an input layer, followed by one or more hidden layers, and an output layer responsible for producing prediction. In this experiment, the number of hidden layers were varied from 1 to 5, with the intent of observing how these variations impact the accuracy of predictions.

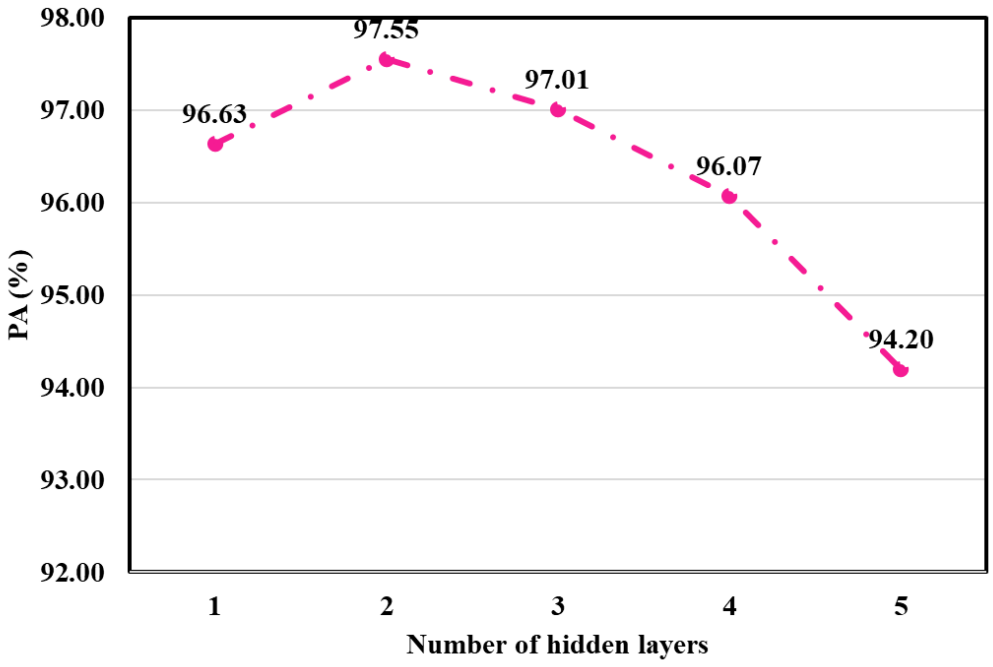


Figure 4.22 FFNN performance for varying number of hidden layers

The outcomes of this exploration are visually given in Figure 4.22, which graphically portrays the PS across a spectrum of different hidden layers. This graphical depiction provides an insightful and easily comprehensible overview of the experiment's results. Remarkably, the FFNN performed exceptionally well when configured with two hidden layers, each comprising 14 neurons. Under this specific configuration, the FFNN achieved its peak PA of 97.55%. Hence, the number of hidden layers was set to two.

Results on Brent crude oil price prediction

After conducting the two tests mentioned above, the FFNN topology was fixed as 14-14-14-1. To assess the robustness and generalization capabilities of the both FFNN and SVM models, Brent crude oil data divided into training and testing sets at varying ratios: 60:40, 70:30, 80:20, and 90:10. The main goal was to assess how effectively FFNN and SVM could adapt to these different data split ratios.

Table 4.18 provides a detailed insight into the prediction performance of the FFNN and SVM models in the task of Brent crude oil price predictions. The analysis encompassed daily, weekly, and monthly prediction frequencies, using data generated from varying testing ratios. A remarkable observation emerged from these outcomes. Both the models consistently exhibited superior performance when subjected to a 70:30 data split ratio in comparison to the other ratios across all data frequencies. This highlights the model's ability to generalize effectively at this specific ratio.

For daily price predictions, wherein 30% of the data was allocated for testing, the FFNN and SVM models attained exceptional performance metrics, including

RMSE of 0.028 and 0.025, MAPE 0f 0.057 and 0.049, D_{stat} of 0.992 and 0.993, and PA of 97.55% and 98.01%, respectively. Similarly, when dealing with weekly price predictions, the models, FFNN and SVM demonstrated its proficiency with an RMSE, MAPE, D_{stat}, and PA values are 0.042 and 0.039,0.066 and 0.054, 0.991 and 0.992, and 96.82% and 97.37%, respectively. These findings show the model's PA and reliability in the weekly prediction.

For monthly data prediction, the FFNN model achieved RMSE of 0.133, MAPE of 0.210, Dstat of 0.982, and PA of 88.68%. The SVM model reported RMSE of 0.131, MAPE of 0.200, D_{stat} of 0.983, and PA of 89.99%. While the accuracy slightly decreased compared to daily and weekly data, it remains highly reliable. To further visualize the model's predictive power, Figure 4.23, Figure 4.24, and Figure 4.25 present a graphical representation of the actual versus predicted prices for daily, weekly, and monthly predictions for Brent crude oil prices, respectively. These graphical representations provide demonstration of the FFNN's and SVM's ability to accurately predict crude oil prices across various timeframes, further proving the superiority of the 70:30 training-testing data split in achieving consistently accurate and reasonable predictions for daily, weekly, and monthly data frequencies.

Table 4.18 Brent crude oil price prediction outcomes

Data frequency	Testing period	FFNN				SVM			
		RMSE	MAPE	D_{stat}	PA (%)	RMSE	MAPE	D_{stat}	PA (%)
Daily	23/04/2014-28/02/2023	0.030	0.059	0.988	97.40	0.027	0.051	0.989	97.76
	08/07/2016-28/02/2023	0.028	0.057	0.992	97.55	0.025	0.049	0.993	98.01
	20/09/2018-28/02/2023	0.030	0.058	0.992	97.42	0.029	0.053	0.992	97.60
	07/12/2020-28/02/2023	0.067	0.060	0.981	94.15	0.041	0.057	0.991	97.27
Weekly	18/04/2014-28/02/2023	0.042	0.069	0.991	96.80	0.044	0.065	0.990	96.86
	01/07/2016-28/02/2023	0.042	0.066	0.991	96.82	0.039	0.054	0.992	97.37
	21/09/2018-28/02/2023	0.046	0.078	0.990	96.33	0.049	0.072	0.994	96.63
	04/12/2020-28/02/2023	0.141	0.096	0.991	89.00	0.150	0.093	0.991	89.60
Monthly	04/2014-02/2023	0.189	0.261	0.980	84.23	0.167	0.252	0.991	85.66
	06/2016-02/2023	0.133	0.210	0.982	88.68	0.131	0.200	0.983	89.99
	09/2018-02/2023	0.139	0.244	0.981	87.57	0.149	0.218	0.981	87.02
	11/2020-02/2023	0.230	0.153	0.961	80.50	0.243	0.140	0.962	80.99

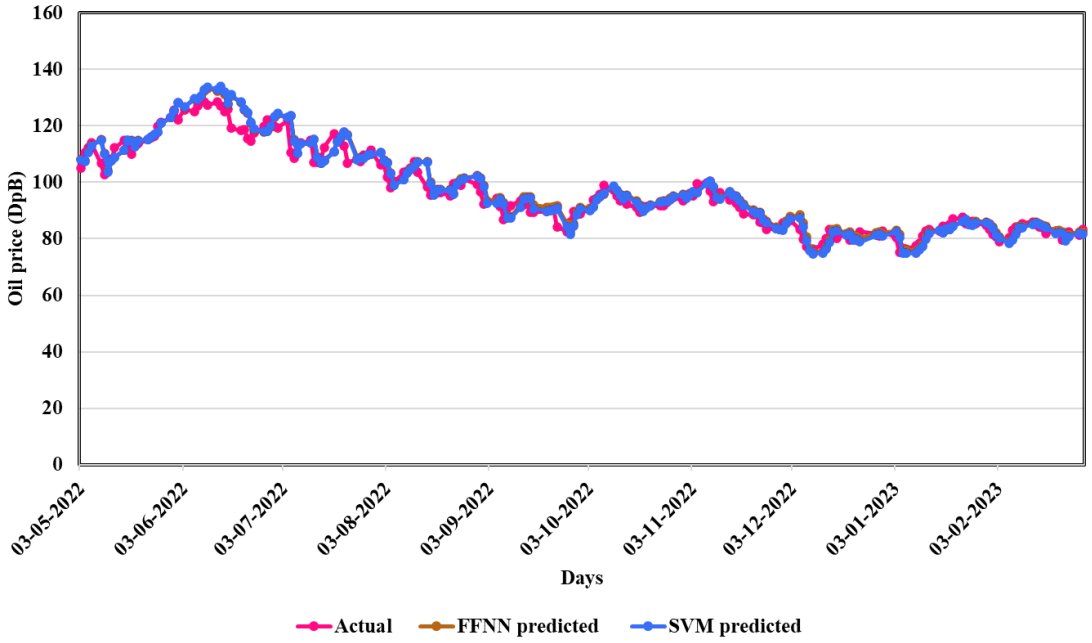


Figure 4.23 Evaluation of actual and predicted daily Brent crude oil prices

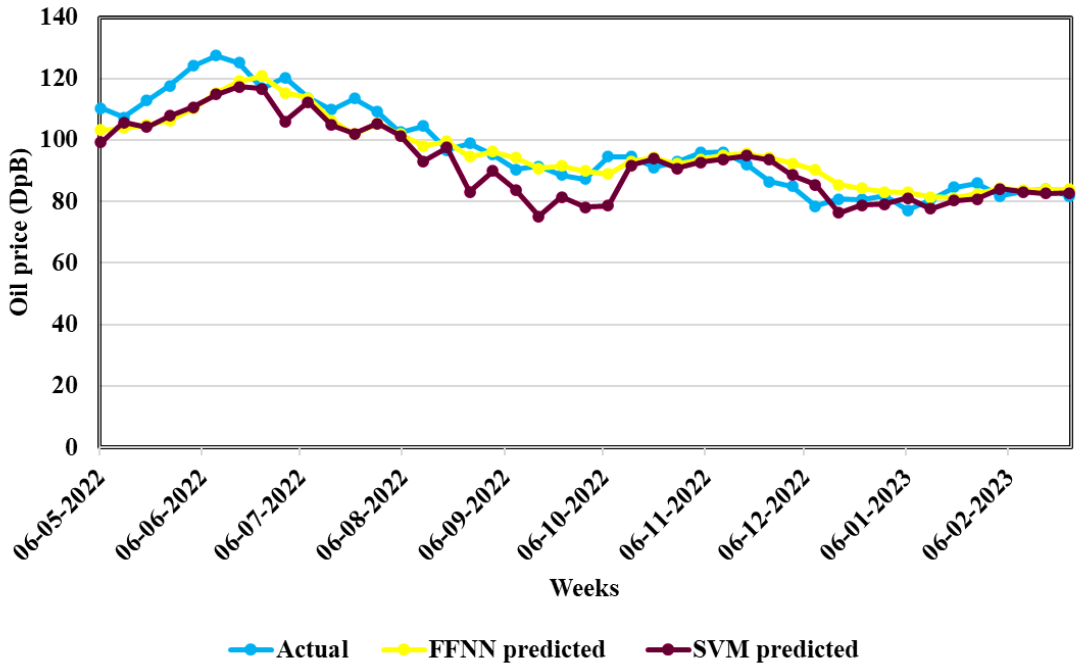


Figure 4.24 Evaluation of actual and predicted weekly Brent crude oil prices

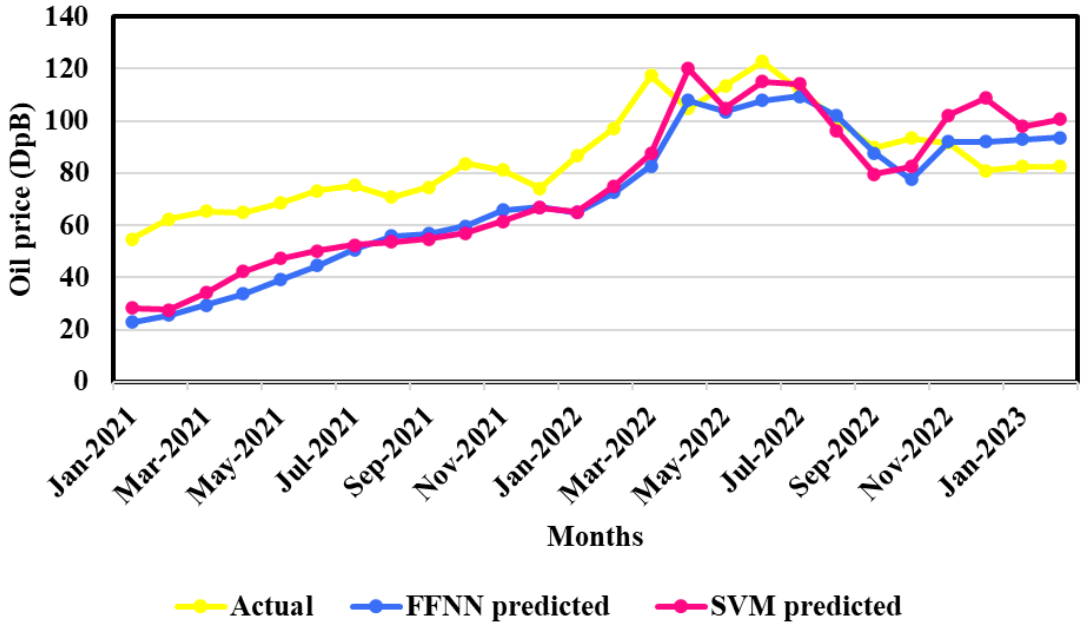


Figure 4.25 Evaluation of actual and predicted monthly Brent crude oil prices

Results on WTI crude oil price prediction

In order to evaluate the generalization capabilities of the developed models, WTI crude oil data was partitioned into training and testing samples at varying ratios, specifically, 60:40, 70:30, 80:20, and 90:10. Prediction performance of the models for WTI crude oil prices on a daily, weekly, and monthly basis, using different test samples is reported in Table 4.19. It is observed that the models consistently deliver better results when the training-testing data ratio was set at 70:30, regardless of the data frequency.

For daily data, with 70% of the data allocated for training and 30% of the data allocated for testing, the models provided outstanding performance. The FFNN model reported RMSE of 0.017, MAPE of 0.030, D_{stat} of 0.992, and PA of 98.59%. The SVM model reached RMSE of 0.016, AMPE of 0.219, D_{stat} of 0.986, and PA of 98.66%.

These metrics underscore the FFNN and SVM model's ability to provide highly accurate daily predictions of WTI crude oil prices. For weekly data, both the FFNN and SVM maintained its excellence, yielding RMSE of 0.051 and0.024, MAPE of 0.100 and 0.045, D_{stat} of 0.995 and 0.997, and PA of 95.60% and 97.93, respectively. These results further attest to the robustness of the model's predictions across varying data frequencies.

Likewise, for monthly data, the FFNN and SVM models proves their power, with RMSE of 0.147 and0.145, MAPE of 0.227 and 0.225, D_{stat} of 0.987 and 0.988, and PA of 87.03% and 88.87%, respectively. While the PA slightly diminished compared to daily and weekly data, the model's predictions remained reliable. The figures accompanying these analysis, Figure 4.26 for daily predictions, Figure 4.27 for weekly predictions, and Figure 4.28 for monthly predictions, visually illustrate the alignment between actual and predicted WTI crude oil prices, further underscoring the model's ability to consistently deliver accurate forecasts across different timeframes.

Table 4.19 WTI crude oil price prediction outcomes

Data frequency	Testing period	FFNN				SVM			
		RMSE	MAPE	D_{stat}	PA (%)	RMSE	MAPE	D_{stat}	PA (%)
Daily	23/04/2014-28/02/2023	0.018	0.031	0.992	98.58	0.019	0.030	0.995	98.52
	08/07/2016-28/02/2023	0.017	0.030	0.992	98.59	0.016	0.029	0.996	98.66
	20/09/2018-28/02/2023	0.022	0.035	0.993	98.46	0.040	0.055	0.995	97.37
	07/12/2020-28/02/2023	0.053	0.039	0.996	95.93	0.044	0.065	0.995	96.76
Weekly	18/04/2014-28/02/2023	0.054	0.103	0.993	95.39	0.025	0.046	0.992	97.91
	01/07/2016-28/02/2023	0.051	0.100	0.995	95.60	0.024	0.045	0.997	97.93
	21/09/2018-28/02/2023	0.054	0.104	0.995	95.53	0.027	0.045	0.994	97.90
	04/12/2020-28/02/2023	0.115	0.077	0.991	91.39	0.074	0.045	0.996	95.23
Monthly	04/2014-02/2023	0.215	0.299	0.990	81.49	0.157	0.234	0.981	86.16
	06/2016-02/2023	0.147	0.227	0.987	87.03	0.145	0.225	0.988	88.87
	09/2018-02/2023	0.191	0.224	0.981	85.10	0.174	0.270	0.987	84.30
	11/2020-02/2023	0.267	0.152	0.961	79.34	0.204	0.307	0.991	81.66

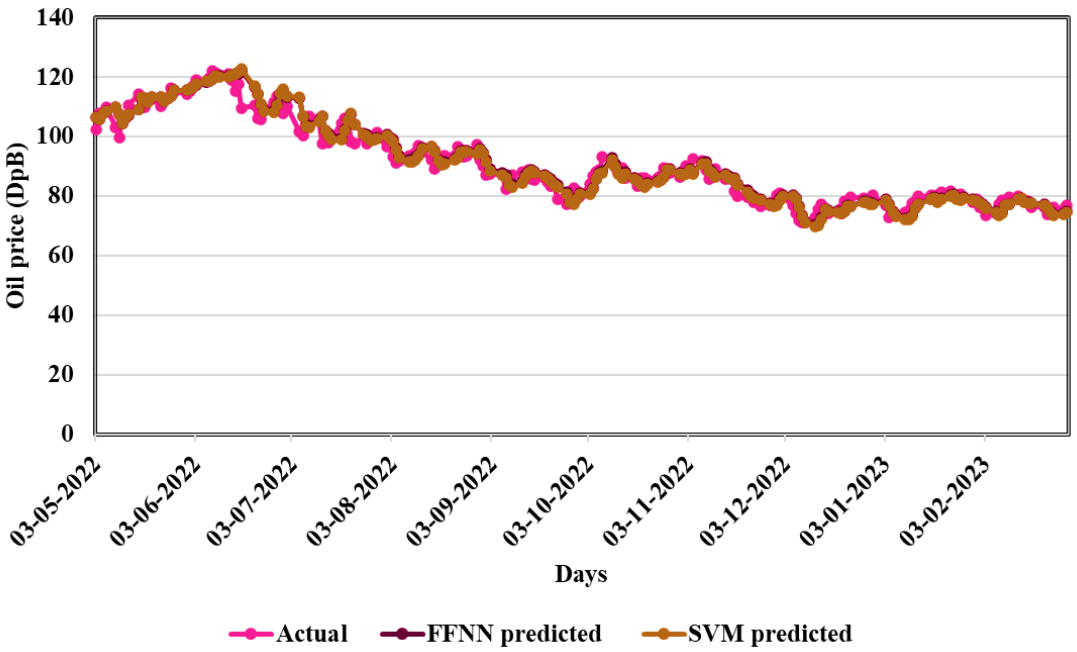


Figure 4.26 Evaluation of actual and predicted daily WTI crude oil prices

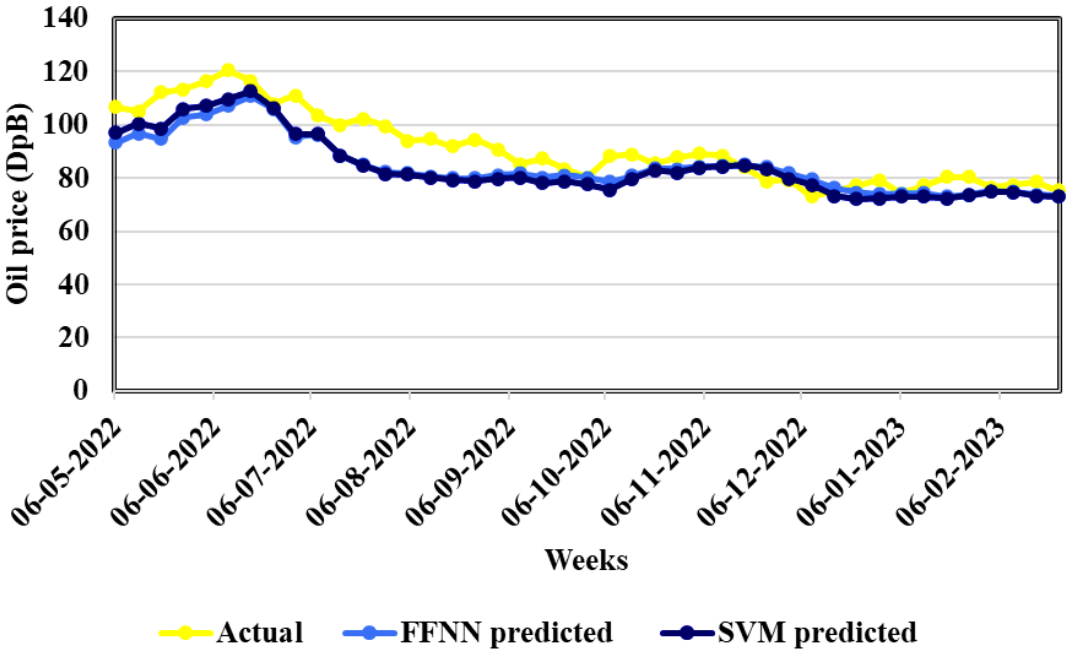


Figure 4.27 Evaluation of actual and predicted weekly WTI crude oil prices

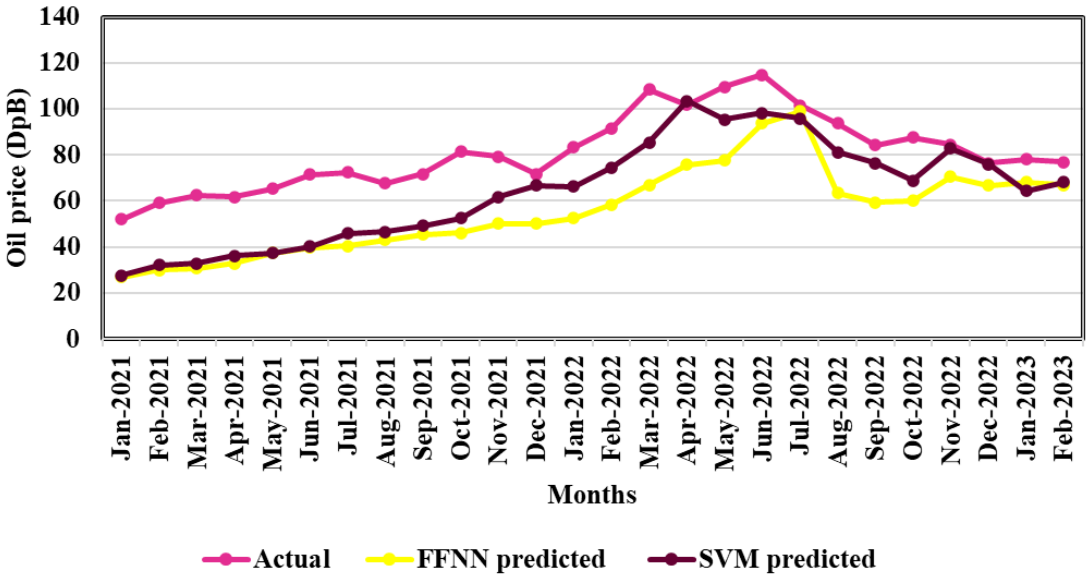


Figure 4.28 Evaluation of actual and predicted monthly WTI crude oil prices

4.7 Hypothesis summary

The simulation results provide compelling evidence supporting the efficacy of the three predictive models introduced in this research. Each model offers unique insights into the factor influencing crude oil prices, ultimately rejecting the H0 in favour of H1.

The Diebold-Mariano (DM) test is designed to assess and compare the relative predicting accuracy of competing predictive models. This test is particularly useful for comparing the forecasting performance of different models by analysing the differences in their predictive errors. Specifically, the DM test evaluates whether these differences are statistically significant with the assumption that the predictive error differences are normally distributed under large sample sizes, as dictated by the central limit theorem.

In this research work, DM test is conducted with a significance level (α) of 0.05 to measure the effectiveness of the developed models. The results of the DM statistic tests are reflected in Table 4.20 for Brent and in Table 4.21 for WTI. As can be seen in Table 4.20 and Table 4.21, significant differences are found between the prediction performance of the developed model and those of the models compared. These tests showed that the proposed models performed significantly better for both Brent and WTI forecasts. The results indicate a clear advantage of the developed models, regardless of the forecast horizon, demonstrating their robustness in anticipating crude oil prices across varying granularities.

Table 4.20 DM statistical test result for the Brent crude oil prices

Models	Daily	Weekly	Monthly
Regression tree	0.996	1.555	1.123
RF	0.011	0.488	0.058
Cat boost	0.761	1.328	0.871
LSTM	0.437	1.289	1.039
Hybrid	0.358	1.032	1.104

Table 4.21 DM statistical test result for the WTI crude oil prices

Models	Daily	Weekly	Monthly

Regression tree	3.182	1.512	1.115
RF	1.158	0.486	0.069
Cat boost	2.896	1.414	1.015
LSTM	2.003	0.955	1.029
Hybrid	2.721	339	1.045

From the simulation outcomes, it is evident that the first model, which is based on ML and historical data, rejected the H0 and accepted the H1. This outcome signifies a significant relationship between historical prices of Brent and WTI crude oil and future crude oil prices. The use of ML techniques with historical price data proved to be effective for predictive accuracy. The second predictive model, incorporating ML and macroeconomic variables, showed better performance in predicting crude oil prices for both Brent and WTI. This model also rejected the H0 and accepted the H1, representing that the macroeconomic variables play a vital role in influencing the prices of Brent and WTI crude oil. Likewise, the third model based on ML and technical indicators, exhibits superior accuracy in predicting crude oil prices for both Brent and WTI. This model also rejected the H0 and accepted the H1, which indicates that technical indicators exhibit a dynamic relationship with Brent and WTI future crude oil prices.

Based on the results of the experimental analysis, it is proved that the ML models can predict future crude oil prices with greater precision. It is important to note that the accuracy of these models is not uniform and varies depending on the input

features employed. This variance clearly highlights a significant difference in prediction performance when different sets of input features are used. Different sets of features whether historical data, macroeconomic factors, technical indicators result in varying levels of predictive accuracy. This observation highlights the need for careful feature selection based on the specific context of the prediction task.

4.8 Summary

This chapter discussed the simulation software employed for the experimentation purpose. A detailed description of the performance measuring metrics used for evaluation was included. Experimental results of the developed crude oil prediction models were presented in detail. A through comparison was done between these models and the earlier approaches employed in similar studies. This comparison confirmed the advancements and improvements achieved through the approaches taken in this research, providing valuable insights into the effectiveness of the developed models.

CHAPTER 5

CONCLUSION AND RECOMMENTATIONS FOR FEATURE RESEARCH

5.1 Summary of the thesis

Crude oil is considered as an important energy source on a global scale. It serves as a key indicator of worldwide economic trends, with its price fluctuations exerting a substantial influence on economic growth. The sales and profits within the oil industry are intricately tied to the oscillations in crude oil prices, significantly affecting decisions related to capital budgeting. The price shifts in crude oil serve as a clear signal of whether an economy is experiencing a recession or a period of prosperity, prompting central banks to adjust their monetary policies correspondingly. Crude oil prices are characterized by high uncertainty, influenced by many factors, including natural disasters, government interventions, and fluctuations in demand and supply. Additionally, various factors, such as the new oil resources, weather conditions, and geopolitical conflicts, further contribute to the dynamics of crude oil prices. Over the past decades, crude oil prices have exhibited characteristics of non-stationarity, non-linearity, and considerable volatility. Both investors and researchers are actively engaged in designing predictive models to increase their returns on investments in the crude oil market. The intricate and multifaceted nature of these price dynamics presents a challenge in accurately forecasting oil prices. Therefore, the development of robust predictive models for predicting crude oil prices is imperative.

The main objective of this research is to employ non-linear models, specifically FFNN and SVM, to predict the fluctuations in crude oil prices. Numerous methods have been developed to predict crude oil prices. Many of these earlier models were in the assumption that the data underlying crude oil prices adhered to linearity and stationarity. However, extensive research investigations have revealed a different reality. Crude oil prices exhibit characteristics of nonlinearity, dynamics, and non-stationarity, deviating significantly from the assumptions of linearity and stationarity. These distinctive features of crude oil price data present a challenge, restricting the accuracy of conventional methods. To deal with the complex nature of crude oil prices, this research work aimed to develop crude oil prediction models using the power of machine learning. By embracing non-linear models, this research seeks to harness the potential of these ML models to deliver more precise and insightful predictions in the volatile world of crude oil markets. In this research work, three different prediction models have been built to predict future crude oil prices across a range of timeframes.

In the previous chapters of this thesis, the research work has been devoted to designing machine learning models for predicting future crude oil prices across different time horizons. The concluding chapter encapsulates the following:

- Research contributions
- Key research findings
- Identified limitations and
- Suggestions for potential extensions in future research

5.2 Research contributions

The fundamental focus of the research work is to establish reliable predictive models for anticipating crude oil prices using ML models. The scientific contributions of the research are outlined as follows:

- ✓ Formulated predictive models for crude oil price predictions by utilizing historical data, FFNN, and SVM, facilitating predictions for both Brent and WTI crude oil prices across various time horizons.
- ✓ Devised predictive models for crude oil prices, integrating macroeconomic factors with ML methods, thereby enabling predictions for both Brent and WTI crude oil across diverse timeframes.
- ✓ Constructed predictive models for crude oil price predictions by using technical indicators in conjunction with ML models. These models enable the prediction of crude oil prices for both Brent and WTI crude oil, over various time periods.

5.3 Key research findings

This research work developed three different predictive models based on machine learning to accurately anticipate future crude oil prices.

In the first phase, the research involved the construction of predictive models employing FFNN and SVM. These models were designed to predict future crude oil prices across different time horizons. To facilitate this, historical data was collected and normalized to a common range between 0 and 1. The normalized data was then split into training and testing data. During the training phase, both the FFNN and SVM

models were designed and fine-tuned using the training data. In the testing phase, these trained FFNN and SVM models were employed to predict future crude oil prices using the testing data. The predictive models used a lag of six preceding prices as input variables to anticipate oil prices one step ahead within daily, weekly, and monthly time frames. The assessment of the model's predictive capabilities involved the computation of various metrics, including RMSE, MAPE, D_{stat}, and PA. Effectiveness of these models was analyzed with respect to the daily, weekly, and monthly prices of both WTI and Brent crude oil. Additionally, performance of the models was validated for varying number of training and testing data. A comparative analysis was conducted to show the superiority of the models. It is observed from the simulation results that the SVM model outperformed the FFNN and earlier models, delivering notable better performance for both Brent and WTI crude oil data.

In the second phase, predictive models were formulated using FFNN and SVM to anticipate the future crude oil prices across varying timeframes. Historical data and macroeconomic variables were gathered and normalized using min-max technique, bringing them into a consistent range between 0 and 1. During the training phase, the ML models were designed and fine-tuned using the training data. At the time of testing, the trained FFNN and SVM models were applied to predict future crude oil prices in the testing data. The efficacy of these models was verified by using the daily, weekly, and monthly prices of WTI and Brent oil. Performance of these models was evaluated by measuring various metrics, including RMSE, MAPE, D_{stat}, and PA. Furthermore, the performance of these models was evaluated by incorporating various combinations of macroeconomic variables to determine the most influential factor. Experimental

results demonstrated that the SVM model achieved excellent performance compared to the FFNN and earlier methods, and the results are significantly superior to those of benchmark model.

In the third phase, predictive models were built using FFNN and SVM to predict crude oil prices across different timeframes. Historical data was collected, and a set of fourteen technical indicators was derived. These technical indicators were subsequently normalized with min-max method. The normalized data was then divided into training and testing subsets. During the training phase, the ML models were designed and tuned using the training data. At the time of testing, the trained FFNN and SVM models were applied to predict future crude oil prices in the testing data. The evaluation of the model's predictive capabilities involved the computation of metrics such as RMSE, MAPE, D_{stat}, and PA. Effectiveness of these models was analyzed with respect to the daily, weekly, and monthly prices of both WTI and Brent crude oil. A comparative analysis was conducted to show the superiority of the models. It is observed from the simulation results that the SVM model exhibited outstanding performance compared to the FFNN and earlier models.

5.4 Identified limitations

The aim of this research was to develop predictive models for predicting crude oil prices across different timeframes using ML methods. While the predictive models presented in this research yielded outstanding outcomes, it is important to acknowledge certain limitations, which are listed below:

- **Macroeconomic variables:** The predictive model developed in this research relied on a set of seven macroeconomic variables as input features. While these factors were effective in generating promising predictions, there exists broader landscape of other macroeconomic factors that could further enhance prediction potential. Future research should explore the inclusion of additional macroeconomic variables to evaluate their impact on prediction accuracy and accuracy.
- **Technical indicators:** The research limited its scope by considering only fourteen technical indicators. While this selection is valuable, the potential for prediction improvements could be further explored by evaluating additional technical indicators. A more comprehensive and diverse set of technical indicators might provide a richer and more accurate understanding of the complex dynamics influencing crude oil prices, thus enhancing the predictive power of the model.
- **Memory:** One significant constraint encountered during the research was related to the MATLAB software. Given that MATLAB is a interpreted environment, it imposes limitations on the speed and memory resources available for working with extensive crude oil data. This limitation could potentially hinder the scalability of the model when dealing with large datasets.

5.5 Recommendation for future research

The possible extensions of the present work are as follows:

- Future research should explore the inclusion of other macroeconomic variables to assess how they influence the accuracy of predictions.

- Analyse the performance of the developed models by using news and twitter data
- Hybrid models can be designed to further improve prediction accuracy.

REFERENCES

Abdollahi, H., & Ebrahimi, S.B. (2020). A new hybrid model for forecasting Brent crude oil price. *Energy, 200*,117520.

Abledu, G.K., & Agbodah, K. (2012). Stochastic forecasting and modelling of volatility of oil prices in Ghana using ARIMA time series model. *European Journal of Business and Management, 4(16)*, 122-131.

Ahmed, R. A., & Shabri, A.B. (2013). Fitting GARCH models to crude oil spot price data. *Life Science Journal, 10(4)*, 654-661.

Ahmed, R.A, and Shabri, A. B. (2014). Daily crude oil price forecasting using ARIMA, generalized autoregressive conditional heteroscedastic and support vector machines. *American Journal of Applied Sciences, 11(3)*,425-432.

Altan, A., & Karasu, S. (2019). The effect of kernel values in support vector machine to forecasting performance of financial time series. *The Journal of Cognitive Systems, 4 (1)*, 17–21.

Azadeh, A., Moghaddam, M., Khakzad, M., & Ebrahimipour, V. (2012). A flexible neural network-fuzzy mathematical programming algorithm for improvement of oil price estimation and forecasting. *Computers and Industrial Engineering, 62(2)*, 421–430.

Barsky, R.B. & Kilian, L. (2004). Oil and the macroeconomy since the 1970s. *Journal of Economic Perspectives, 18(4)*, 115-134.

Baruník. J., & Malinská, B. (2016). Forecasting the term structure of crude oil futures prices with neural networks. *Applied Energy,164*, 366–379.

Baumeister, C., & Kilian, L. (2012). Real-time forecasts of the real price of oil. *Journal of Business & Economic Statistics*, *30(2)*, 326-336

Baumeister, C., & Kilian, L. (2014). Real-time analysis of oil price risks using forecast scenarios. *IMF Economic Review ,62*, 119-145.

Bollapragada, R., Mankude,A., & Udayabhanu, V. (2021). Forecasting the price of crude oil. *Decision, 48(2)*,207-231.

Boubaker, S., Liu, Z., & Zhang, Y. (2022). Forecasting oil commodity spot price in a data-rich environment. *Annals of Operations Research*,1-18.

Bristone, M., Prasad, R., & Abubakar, A.A. (2020). CPPCNDL: Crude oil price prediction using complex network and deep learning algorithms. *Petroleum, 6(4)*, 353–361.

Cao, D-Z., Pang, S.L. & Bai, Y-H. (2005). Forecasting exchange rates using support vector Machines. *In. Fourth International Conference on Machine Learning and Cybernetics, Guangzhou,* 18-21.

Cen, Z., & Wang, J. (2019). Crude oil price prediction model with long short-term memory deep learning based on prior knowledge data transfer. *Energy, 169*, 160–171.

Chen, E., & He, X.J. (2019). Crude oil price prediction with decision tree-based regression approach. *Journal of International Technology and Information Management, 27(4)*, 2–16.

Chen, H., Liu, L., Wang, Y., & Zhu, Y. (2016). Oil price shocks and U.S. dollar exchange rates. *Energy, 112*,1036-1048.

Chen, W-H., & Shih, J.Y. (2006). Comparison of support vector machines and back propagation neural networks in forecasting the six major Asian stock markets. *International Journals Electronics Finance, 1*, 49-67.

Chen, Y., He, K., & Tso, G.K.F. (2017). Forecasting crude oil prices-a deep learning based model. *Procedia Computer Science*,*122*,300–307.

Cheng, Y., Jiayu, Y., Yang, X., Lai, K.K., & Seco, L. (2022). A CEEMD-ARIMA-SVM model with structural breaks to forecast the crude oil prices linked with extreme events, *Soft Computing,26*,8537-8551.

Cheong, C.W. (2009). Modeling and forecasting crude oil markets using ARCH-type models. *Energy, 37(6)*, 2346-2355

Chiroma, H., Abdulkareem, S., & Herawan, T. (2015). Evolutionary neural network model for west Texas intermediate crude oil price prediction. *Applied Energy, 142*, 266-273.

Chiroma, S. A.K., & Noor, A.S.M. (2016). A review on artificial intelligence methodologies for the forecasting of crude oil price. *Intelligent Automation and Soft Computing, 22(3)*,449–462.

Costa, A. B. R., Ferreira, P. C. G., Gaglianone, W. P., Guillén, O. T. C., Issler, J. V., & Lin, Y. (2021). Machine learning and oil price point and density forecasting. *Energy Economics, 102*, 105494.

Daneshvar, A., Ebrahimi, M., Salahi, F., Rahmaty, M., & Homayounfar, H. (2022). Brent crude oil price forecast utilizing deep neural network architectures, *Computational Intelligence and Neuroscience,2022*,6140796

Deng, C., Ma, L., & Zeng, T. (2021). Crude oil price forecast based on deep transfer learning: shanghai crude oil as an example. *Sustainability, 13*, 13770.

Deng, S., Xiang, Y., Fu, Z., Wang, M., & Wang, Y. (2019). A hybrid method for crude oil price direction forecasting using multiple timeframes dynamic time wrapping and genetic algorithm. *Application of Soft Computing Journal, 82*,105566.

Drachal, K. (2016). Forecasting spot oil price in a dynamic model averaging framework -Have the determinants changed over time? *Energy Economics, 60*, 35-46.

Dragomiretskiy, K., & Zosso, D. (2014). Variational mode decomposition. *IEEE Transactions on Signal Processing,62* (3), 531-544.

Dunn, S. (2002). Hydrogen futures: toward a sustainable energy system. *International Journal of Hydrogen Energy*, *27(3)*, 235-264.

Energy Information Administration, Annual energy outlook 2019 with projections to 2050. Technical report, US Department of Energy, Washington, DC, PP. 32.

Energy Information Administration, Annual energy outlook 2023 with projections to 2050. Technical report, US Department of Energy, Washington, DC, 1-25.

Fan, G.-F., Peng, L.-L., Hong, W.-C., & Sun, F. (2016) Electric load forecasting by the SVR model with differential empirical mode decomposition and auto regression. *Neurocomputing, 173*, 958-970.

Fan, L., Pan, S., Li, Z., & Li, H. (2016). An ICA-based support vector regression scheme for forecasting crude oil prices. *Technological Forecasting and Social Change, 112*, 245–253.

Foroutan,P., & Lahmiri,S. (2024). Deep learning systems for forecasting the price of crude oils and precious metals. *Financial Innovation* , 10,111

Gabralla, L.A., & Abraham, A. (2013). Computational modelling of crude oil price forecasting: A review of two decades of research. *International Journal of Computer Information Systems and Industrial Management Applications, 5*,729-740.

Gao, S., & Lei, Y. (2017). A new approach for crude oil price prediction based on stream learning. *Geosciences Frontiers, 8(1)*, 183–187.

He, A., Kwok, J., & Wan, T. (2010). An empirical model of daily highs and lows of West Texas Intermediate crude oil prices. *Energy Economics, 32*, 1499–1506.

He, K., Yu, L., & Lai, K.K. (2012). Crude oil price analysis and forecasting using wavelet decomposed ensemble model. *Energy, 46*, 564-574.

Herrera, G.P., Constantino, M., Tabak, B.M., Pistori, H., Su, J.J., & Naranpanawa, A. (2019). Long-term forecast of energy commodities price using machine learning. *Energy ,179*, 214-221.

Hou, A., & Suardi, S. (2012). A nonparametric GARCH model of crude oil price return volatility. *Energy Economics, 34(2)*, 618-626.

http://fxtop.com

http://www.eia.gov

https://www.investing.com/commodities/crude-oil-historicaldata

Huang, H. & Tian, Y. (2013). A novel visual modeling system for time series forecast: application to the domain of hydrology. *Journal of Hydro informatics, 15(1)*, 21-37.

Huang, L., & Wang, J. (2018). Global crude oil price prediction and synchronization-based accuracy evaluation using random wavelet neural network. *Energy,151*,1-10.

Issaka, A. (2020). Variance swaps, volatility swaps, hedging and bounds under multifactor Heston stochastic volatility model. *Stochastic Analysis and Applications*, 1–19.

Jabeur, S. B., Mefteh-Wali, S., & Viviani, J. L. (2022). Forecasting gold price with the XGBoost algorithm and SHAP interaction values. *Annals of Operations Research*, 1–21.

Jammazi, R. & Aloui, C. (2012). Crude oil price forecasting: experimental evidence from wavelet decomposition and neural network modelling. *Energy Economics, 34(3)*, 828–841.

Jin,B., & Xu,X. (2024).Price forecasting through neural networks for curde oil, heating oil, and natural gas. *Measurement: Energy* , 1,1-15.

Jovanivic, L., Jovanivic, D., Bacanin, N., & Stakic, A.J. (2022). Multi step crude oil price prediction based on LSTM approach tuned by salp swarm algorithm with disputation operator. *Sustainability, 14*, 14616.

Kadkhodaie- Ilkhchi, A., Rezaee, M., & Rahimpour-Bonab, H.2009. A committee neural network for prediction of normalized oil content from well log data: an example from South Pars Gas Field, Persian Gulf. *Journal of Petroleum Science and Engineering, 65(1)*, 23–32.

Kang, S. H., Kang, S. M., & Yoon, S. M. (2009). Forecasting volatility of crude oil markets. *Energy Economics, 31(1)*,119-125.

Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review, 99(3)*, 1053-1069.

Kilian, L., & Murphy, D. P. (2014). The role of inventories and speculative trading in the global market for crude oil. *Journal of Applied Econometrics, 29(3)*, 454-478.

Kulkarni, S., & Haidar, I. (2009). Forecasting model for crude oil price using artificial neural networks commodity futures prices. *International Journal of Computer Science and Information Security, 2(1)*,1-9.

Lahmiri, S. (2016). A variational mode decomposition approach for analysis and forecasting of economic and financial time series. *Expert Systems with Applications, 55*,268-273.

Lanza, A., Manera, M., & Giovannini, M. (2005). Modeling and forecasting cointegrated relationships among heavy oil and product prices. *Energy Economics, 27*, 831-848.

Leonori, S., Paschero, M., Mascioli, F. M. F., & Rizzi, A. (2020). Optimization strategies for microgrid energy management systems by genetic algorithms. *Applied Soft Computing, 86*,105903.

Li, J., Zhu, S., & Wu, Q. (2019). Monthly crude oil spot price forecasting using variational mode decomposition. *Energy Economics,83*, 240–253.

Lin, A. (2009). Prediction of international crude oil futures price based on GM (1,1). *In. IEEE International Conference on Grey Systems and Intelligent Services (GSIS),* 692-696.

Liu, X.T. (2010). Study on data normalization in BP neural network. *Mechanical Engineering & Automation,3,* 122-123.

Lizardo, R. A., & Mollick, A. V. (2010). Oil price fluctuations and US dollar exchange rates. *Energy Economics, 32(2)*, 399-408.

Makridakis, S. (1993). Accuracy measures: theoretical and practical concerns. *International Journal of Forecasting, 9(4)*, 527-529.

Manickavasagam, J., Visalakshmi, S., & Apergis, N. (2020). A novel hybrid approach to forecast crude oil futures using intraday data. *Technological Forecasting & Social Change,2020*,120126.

Mingming, T., & Jinliang, Z. (2012). A multiple adaptive wavelet recurrent neural network model to analyze crude oil prices. *Journal of Economics and Business, 64(4)*,275–286.

Mohammadi, H., & Su. L. (2010). International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models. *Energy Economics, 32(5)*, 1001-1008.

Morana, C. (2013). The oil price-macroeconomy relationship since the mid-1980s: a global perspective, *Energy Journal,4(3)*, 1-6.

Moshiri, S., & Foroutan, F. (2006). Forecasting nonlinear crude oil futures prices. *Energy Journal, 27(4)*,81-95.

Mostafa, M.M., & El-Masry, A.A. (2016). Oil price forecasting using gene expression programming and artificial neural networks. *Economic Modelling, 54*, 40-53.

Mozetic, I., Torgo, L., Cerqueira, V., & Smailovic, J. (2018). How to evaluate sentiment classifiers for twitter time-oriented data. *Plos One, 13(3)*,1-9.

Murat, A., & Tokat, E. (2009). Forecasting oil price movements with crack spread futures. *Energy Economics, 31(1)*, 85-90.

Ng'ang'a, F. W., & Oleche, M. (2022). Modelling and forecasting of crude oil price volatility comparative analysis of volatility models. *Journal of Financial Risk Management, 11*, 154-187.

Picciolo, F., Papandreoce, A., Hubacek, K., & Ruzzenenti, F. (2017). How crude oil prices shape the globe division of labor. *Applied Energy, 189*, 753-761.

Ramsay, K. W. (2011). Revisiting the resource curse: natural disasters, the price of oil, and democracy. *International Organization, 65(3)*, 507-529.

Reboredo, J.C., & Rivera-Castro, M.A. (2013). A wavelet decomposition approach to crude oil price and exchange rate dependence. *Economic Modelling, 32 (32)*,42-57.

Sahir, M. H., & Qureshi A H. (2007). Specific concerns of Pakistan in the context of energy security issues and geopolitics of the region. *Energy Policy, 35(4)*, 2031-2037.

Sekati, B.N.Y., Tsoku, J.T., & Metsileng, L.D. (2020). Modelling the oil price volatility and macroeconomic variables in South Africa using the symmetric and asymmetric GARCH models. *Cogent Economics & Finance*, *8(1)*, 1792153.

Shabri, A., & Samsusin, R. (2014.) Daily crude oil price forecasting using hybridized wavelet and artificial neural network model. *Mathematical Problems in Engineering,2014*,1-10.

Shahbaz, M., Tiwari, A.K., & Tahir., M.I. (2015). Analysing time-frequency relationship between oil price and exchange rate in Pakistan through wavelets. *Journal of Applied Statistics,42(4)*,690-704.

Shahmoradi, A. (2016). Valuation of crude oil futures, options, and variance swaps. *Doctoral dissertation*, University of Calgary.

Smales, L. A. (2021). Geopolitical risk and volatility spillovers in oil and stock markets. *The Quarterly Review of Economics and Finance,80*,358-366.

Sun, S., Sun, Y., Wang, S., & Wei, Y. (2018). Interval decomposition ensemble approach for crude oil price forecasting. *Energy Economics, 76*, 274–287.

Tehrani, R., & Khodayar, F. (2011). A hybrid optimized artificial intelligent model to forecast crude oil using genetic algorithm. *African Journal of Business Management,5(34)*, 13130–13135.

Tuo, J., & Yanbing. (2011). Summary of world oil price forecasting model. *In. 2011 Fourth International Symposium on IEEE Knowledge Acquisition and Modeling (KAM)*, 327-330.

Vapnik, V. N., Golowich, S.E., & Smola, A. J. (1996). Support vector method for function approximation, regression estimation, and signal processing. *Advances in Neural Information Processing Systems, 9*, 281-287.

Vapnik, V.N. (1995). The Nature of Statistical Learning Theory. 1st Edition., Springer, New York.

Wan, J.Y., & Kao, C.W. (2015). Interactions between oil and financial markets do conditions of financial stress matter. *Energy Economics,52*, 160-175.

Wang, D., & Fang, T. (2022). Forecasting crude oil prices with a WT-FNN model, *Energies,16*,1955.

Wang, J., & Wang, J. (2021). A new hybrid forecasting model based on SW-LSTM and wavelet packet decomposition: A case study of oil futures prices. *Computational Intelligence and Neuroscience., 2021*,7653091.

Wang, J., Li, X., Hong, T., & Wang, S. (2018). A semi-heterogeneous approach to combining crude oil price forecasts. *Information Sciences, 460*, 279–292.

Wang, J., Zhang, W., Li, Y., Wang, J., & Dang, Z. (2014). Forecasting wind speed using empirical mode decomposition and Elman neural network. *Applied Soft Computing, 23*, 452-459.

Wang, S., Yu, L., & Lai, K.K. (2004). A novel hybrid AI system framework for crude oil price forecasting. In. *Data Mining and Knowledge Management*, Y. Shi, W. Xu, and Z. Chen, Eds., 3327 of *Lecture Notes in Computer Science*,233-242, Springer, Berlin, Germany.

Wang. Y., Liu, L., & Wu, C. (2017). Forecasting the real prices of crude oil using forecast combinations over time-varying parameter models. *Energy Economics, 66*,337–348.

Wei, Y., Wang, Y., & Huang, D. (2010). Forecasting crude oil market volatility: further evidence using GARCH-class models. *Energy Economics, 32*,1477–1484.

Wen, F., Xiao, J., He, Z., &Gong, X. (2014) Stock price prediction based on SSA and SVM. *Procedia Computer Science, 31,* 625-631.

Xiang, Y., & Zhuang, X. H. (2013). Application of ARIMA model in short-term prediction of international crude oil price. *Advanced Materials Research, 798*, 979-982.

Xiang, Y., (2022). Using ARIMA-GARCH model to analyze fluctuation law of international oil price.*Big Data Modelling of Engineering and Managements*,1-10.

Xie, W., Yu, L., Xu, S., & Wang, S. (2006). A new method for crude oil price forecasting based on support vector machines. *In: Alexandrov, V.N., van Albada,*

G.D., Sloot, P.M.A., Dongarra, J. (eds) Computational Science – ICCS 2006.Lecture Notes in Computer Science,3994. Springer, Berlin, Heidelberg.

Yan, L., Zhu, Y., & Wang, H. (2021). Selection of machine learning models for oil price forecasting: Based on dual attributes of oil. *Discrete Dynamics in Nature and Society, 2021*,1566093.

Yang, Y., Guo, J.E., Sun, S., & Li, Y. (2021). Forecasting crude oil price with a new hybrid approach and multi-source data. *Engineering Applications of Artificial Intelligence, 101*, 104217.

Yu, L., Wang, S., & Lai, K. K. (2008). Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy Economics, 30*,2623-2635.

Yu, L., Wang, Z., & Tang, L. (2015). A decomposition-ensemble model with data-characteristic-driven reconstruction for crude oil price forecasting. *Applied Energy,156*, 251–267.

Yu, L., Zhang, X., & Wang, S. (2017). Assessing potentiality of support vector machine method in crude oil price forecasting. *EURASIA Journal of Mathematics. Science and Technology Education.13*, 7893–7904.

Yu, L., Zhao, Y., & Tang, L. (2014). A compressed sensing-based AI learning paradigm for crude oil price forecasting. *Energy Economics, 46*, 236-245.

Zhang, J. L., Zhang, Y. J., & Zhang, L. (2015). A novel hybrid method for crude oil price forecasting. *Energy Economics, 49*, 649-659.

Zhang, L., Zhou, W-D., Yang, J-W. & Li, F-Z. (2013). Iterated time series prediction with multiple support vector regression models. *Journal Neurocomputing, 99*, 411-422.

Zhang, T., Tang, Z., Wu, J., Du, X., & Chen, K. (2021). Multi-step-ahead crude oil price forecasting based on two-layer decomposition technique and extreme learning machine optimized by the particle swarm optimization algorithm. *Energy, 229*,120797.

Zhang, X., Lai, K.K., Lai, & Wang, S.-Y. (2008). A new approach for crude oil price analysis based on Empirical Mode Decomposition. *Energy Economics, 30*, 905-918.

Zhang, Y., Ma, F., & Wang, Y. (2019). Forecasting crude oil prices with a large set of predictors: can LASSO select powerful predictors? *Journal of Empirical Finance, 54,* 97–117.

Zhang, Y.J., & Wei, Y.M. (2011). The dynamic influence of advanced stock market risk on international crude oil returns: an empirical analysis. *Quantitative Finance 11(7)*, 967–978.

Zhao, C.L., & Wang, B. (2014). Forecasting crude oil price with an autoregressive integrated moving average (ARIMA) model. *In. Proceedings of the Fuzzy Information and Engineering and Operations Research and Management. Springer, Berlin, Heidelberg, 275–286*

Zhao, L., Cheng, L., Wan, Y., Zhang, H., & Zhang, Z. (2015). A VAR-SVM model for crude oil price forecasting. *International Journal of Global Energy Issues ,38(1)*, 126–144.

Zhao, Y., Jianping, L., & Yu, L. (2017). A deep learning ensemble approach for crude oil price forecasting, *Energy Economics, 66*, 9-16,

Zhao, Y., Zhang, W., Gong, X., & Wang, C. (2021). A novel method for online real-time forecasting of crude oil price. *Applied Energy, 303,* 117588

Zhou, Y., Li, T., Shi, J., & Qian, Z. (2019). A CEEMDAN and XGBOOST-Based Approach to Forecast Crude Oil Prices. *Complexity, 2019*, 4392785.

Zhu, B., Han, D., Wang, P., Wu, Z., Zhang, T., & Wei, Y.M. (2017). Forecasting carbon price using empirical mode decomposition and evolutionary least squares support vector regression. *Appl Energy,191*, 521–530.