Leveraging Deep Learning for Enhanced Forecasting

in

Downstream Oil and Gas Industry

Thesis submitted in partial fulfilment

for the Award of Degree of

Doctor of Philosophy

By

Praveen Kumar

RAJIV GANDHI INSTITUTE OF PETROLEUM TECHNOLOGY

JAIS, INDIA – 229304

PM1608 2025

CERTIFICATE

It is certified that the work contained in the thesis titled "**Leveraging Deep Learning for Enhanced Forecasting in Downstream Oil and Gas Industry**" by **Praveen Kumar** has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

It is further certified that the student has fulfilled all the requirements of Comprehensive, Candidacy and SOTA.

Dr. Debashish Jena

(Supervisor)

DECLARATION BY THE CANDIDATE

I, Praveen Kumar, certify that the work embodied in this thesis is my own bonafide work and carried out by me under the supervision of Dr. Debashish Jena from July 2016 to June 2024, at the "Department of Management Studies" (DoMS), Rajiv Gandhi Institute of Petroleum Technology, Jais (India).

The matter embodied in this thesis has not been submitted for the award of any other degree. I declare that I have faithfully acknowledged and given credits to the research workers wherever their works have been cited in my work in this thesis. I further declare that I have not willfully copied any other's work, paragraphs, text, data, results, etc., reported in journals, books, magazines, reports, dissertations, theses, etc., or available at websites and have not included them in this thesis and have not cited as my own work.

Date : 30th March 2025 **Praveen Kumar**

Place: RGIPT, Jais **PM-16-08**

CERTIFICATE BY THE SUPERVISOR

It is certified that the declaration made by Mr. Praveen Kumar regarding his thesis submission titled "**Leveraging Deep Learning for Enhanced Forecasting in Downstream Oil and Gas Industry**" for the award of PhD Degree in Operations Management from Department of Management Studies (DoMS), RGIPT, Jais, India is correct to the best of our knowledge.

Dr. Debashish Jena **Dr. Jaya Srivastava**

(Supervisor) (Head, DoMS, RGIPT, Jais)

CERTIFICATE

Certified that the work contained in the thesis titled "**Leveraging Deep Learning for Enhanced Forecasting in Downstream Oil and Gas Industry**" by Mr. Praveen Kumar has been carried out under my supervision. It is also certified that he fulfilled the mandatory requirement of paper publication that arose out of his work.

It is further certified that the publications (copies enclosed) of the aforesaid Mr. Praveen Kumar have been published in the journals by (a) SCOPUS (b) ABDC – B List.

Dr. Debashish Jena **Dr. Shrawan Trivedi**

(Supervisor) (Convener, DPGC, DoMS, RGIPT, Jais)

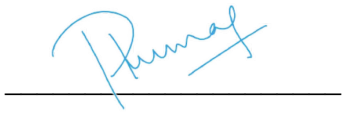
COPYRIGHT TRANSFER CERTIFICATE

Title of the Thesis: "**Leveraging Deep Learning for Enhanced Forecasting in Downstream Oil and Gas Industry**"

Name of the Student: **Praveen Kumar**

Copyright Transfer

The undersigned hereby assigns to the Rajiv Gandhi Institute of Petroleum Technology Jais all rights under copyright that may exist in and for the above thesis submitted for the award of the "Doctor of Philosophy ".



Date: 30th March 2025 **Praveen Kumar**

Place: RGIPT,Jais **Roll No: PM-1608**

Note: However, the author may reproduce or authorize others to reproduce material extracted verbatim from the thesis or derivative of the thesis for the author's personal use provided that the source and the Institute's copyright notice are indicated.

Acknowledgement

"The wise person acknowledges the transient nature of the material world and seeks the eternal truth within."

- Mundaka Upanishad

At the very outset, I would like to convey my gratitude and thanks to my beloved mother Late Smt. Malti Singh, my first teacher and father Shri Hausila Prasad Singh, a great visionary who embodied in us the quest and supremacy of knowledge and encouraged us to pursue the path of continuous learning at every stages of our lives. Being a veteran in Oil and Gas industry, his insights have been invaluable. Faith, Destiny, God or whatever name we may know thee, I bow before things that are unknown, uncontrollable by us mortals and they somehow always have a unique way of having their presence felt. In addition to the efforts that we may put in, I am sure that there is their role in completion of this thesis. I acknowledge thee.

I sincerely thank the Director, Rajiv Gandhi Institute of Petroleum Technology (RGIPT) for his encouragement and support to complete my research work.

I am really thankful to Dr. Jaya Srivastava, Head, Department of Management Studies for her constant support and guidance for timely complete my research work.

My sincerest thanks are due to Dr Debashish Jena, Department of Management Studies, RGIPT, my guide for his ever-supportive guidance and motivating me during the entire tenure of my research. I would not have been able to complete the work without his guidance, help, constant and generous support.

I sincerely thank Dr. Manoj K. Rajput, Dr. Rohit Bansal, Dr. Shrawan Trivedi for their input throughout my research study.

I am thankful to all faculty from the Department of Management Studies for their meaningful input at various stages of research at RGIPT. My sincere thanks to Dr. Ramakrushna Padhy, IIM Sambalpur for his valuable guidance during the research and review stage and Mr. Debasmit Mohanty, Founder, Stratlytics for insights into practical applications of forecasting.

I thank all my friends and colleagues who have supported me in various stages for their invaluable help. A special thanks to my research colleague Shri Abhijit Deb Roy, who has been my constant companion for his support during the research, Ms. Meenakshi Agarwal, Shri Dinker N. Jha, Ms Devisha Shruti, Shri Aman Mishra and Shri Jamyang Lotus, for their help in data processing and collation.

Last, but not the least, I would like to thank my family, especially my wife Preeti Kumar and sons Ankoor Kumar Kashyap and Prasoon Kumar Kashyap for their understanding, patience, and continuous support during the entire period of my research journey. Without their unstinting support and motivation, this would not have been possible. I would also like to thank my brothers Shri Ratin Kumar and Shri Himanshu Kumar for their continuous encouragement and motivation.

I am thankful to my organization Indian Oil Corporation Ltd. for permitting me undertake this arduous journey, seniors and colleagues for their understanding, support, encouragement and for pitching in for my absence and enabling me to continue this research.

Abstract

"*Prediction is very difficult, especially about the future."*

- Niels Bohr, Nobel laureate in Physics

Price fluctuations in the Crude oil and Gasoline markets are highly uncertain, driven by a host of factors such as supply and demand prospects, macroeconomic factors, geo-political condition, speculative trading etc. Further, there is an ongoing debate in the empirical literature on the causal link and asymmetric relationship between the crude oil and gasoline prices. In practice, petroleum refiners constantly strive to protect the erosion of refining margins or the price gap between the price of crude oil and that of the refined products, also known as *crack spread*. Integrated Oil Refiners and Marketers typically purchase crude oil from international markets, refine it, and then sell refined products in a competitive market; they are caught between two markets, exposing them to enormous risks that can severely impact crack spreads or profit margins. The ability to accurately predict crack spreads to protect margins and mitigate risks remains a major challenge. Furthermore, crack spreads can vary significantly by region and time period, depending on factors such as local market conditions, infrastructure and seasonal demand.

In contrast to extensively researched crude oil price forecasting, research on crack spread forecasting remains limited. To this end, in the first application of predictive analytics in this thesis, we propose a novel method for crack spread forecasting. This method is a hybrid of statistical forecasting techniques, specifically Auto Regression and Integrated Moving Average (ARIMA) for modeling linear patterns, and the Long Short-Term Memory (LSTM) deep learning model to capture the residual complexities present in the time series data of both crude and refined product prices. The crack spread model has been trained on Dubai Crude and Singapore Gasoline (refined product) data. The crack spread forecast thus obtained outperforms well-known standalone statistical or deep learning models on various performance metrices. This method not only yields more precise forecasts but is also conveniently deployable using widely available open-source tools. The improved crack spread forecasts so available is intended to assist procurement managers in selecting and configuring the most price efficient crude basket leading to increased profitability

As with crack spread forecasting, accurate LPG demand forecasts enable companies to optimize their supply chain operations, ensuring a balance between supply and demand. This requires addressing distinct challenges such as demand volatility, data availability and quality, seasonality, trends, customer behavior and preferences, geographic variability, and external influences like government policies (e.g., subsidy schemes, price regulations), energy market trends, competing fuel sources (e.g., natural gas, electricity), and environmental considerations (e.g., air quality regulations). In the second application, we demonstrate time series modeling techniques on real-world data, utilizing both statistical and AI/ML Deep Learning methods and subsequently comparing the results. Deep Learning standalone models based on Feed Forward Neural Networks (FFNN) outperformed standalone statistical models on the various performance metrices.

The predictive analytics framework developed in this thesis offers valuable insights to industry practitioners such as sourcing and procurement managers, demand planners and supply chain managers, augmenting their forecasting capabilities and contributing significantly towards improving profitability. The study also explores potential avenues for future research and the advancement of these techniques.

Table of Content

List of Figures

List of Tables

Chapter 1

Introduction

"Forecasting is not about predicting the future with certainty, but about preparing for the range of possibilities."

- Russell Cameron Thomas,

1.1 Motivation

In the dynamic landscape of petroleum refining and marketing, businesses navigate the dual challenge of safeguarding margins amidst volatile commodity (crude oil) and gasoline (or refined product) prices and their asynchronous movements (Ederington et al., 2018), while optimizing supply chain through accurate demand prediction and inventory management. Variable margins and unpredictable commodity price fluctuations compel petroleum refiners and marketers to seek innovative strategies to bolster their bottom line. Consequently, there is a growing emphasis on forecasting market trends and price variations, enabling refineries to make well-informed choices regarding inventory management, product-mix planning, and hedging strategies, which gives them a competitive advantage in the extremely unpredictable Oil and Gas business.

Downstream petroleum refining and marketing companies operate across the entire oil industry value chain, from crude procurement to product marketing. Due to the volatile prices of crude oil and refined products, these companies are always vulnerable to margin pressure (Clyde Russell, 2023, 2024). At a fundamental level, crude oil refining process involves 'cracking' crude oil into refined products, which creates a price differential known as the 'crack spread.' Among other standalone forecasts such as price of crude oil or refined products, of greatest significance are the crack-spreads, or differences, between the cost of crude oil and the prices of refined products (Apergis & Vouzavalis, 2018), such as gasoline and distillates (diesel and jet fuel). Accurate forecasting of crack spreads significantly impacts the profit margins of oil refining firms, especially during periods of stressed refinery margins (Steven et al., 2014a). It enables them to formulate effective pricing strategies, optimize production processes, manage inventory efficiently, and mitigate associated risks. Additionally, crack spread forecasting plays a pivotal role in optimizing procurement strategies, enabling refineries to secure crude oil at favourable prices and strategically plan production schedules to maximize profit margins (Dunis et al., 2016).

In contrast to extensively researched fields of crude oil price or refined product price forecasting, which often accounts for global trends and market conditions, crack spreads can vary significantly across different regions and time periods due to factors such as local market conditions, infrastructure constraints, and seasonal demand patterns. Research examining spatial and temporal variability in crack spreads, as well as the development of localized prediction models, could improve forecast accuracy for specific markets and timeframes.

Two distinct methodologies exist for forecasting the Crack spread. In the first approach, the predicted crack spread (notably, the gasoline crack in this context) can be determined by calculating the variance between the expected crude oil price and the anticipated gasoline price. In contrast, the second method entails calculating the crack spread as the difference between crude oil and gasoline prices, subsequently employing the resulting time series for forecasting purposes. The methodology of initially deriving the crack spread time series (the differential between the gasoline product price and the crude price) and subsequently forecasting the spread is widely acknowledged as the more effective alternative because of the following reasons:

i. Cointegration and Error Correction

   Since crude and gasoline prices are cointegrated, their difference (the crack spread) likely follows a stationary process or exhibits mean-reverting behavior. In a research on Co-integration and Error Correction (Engle & Granger, 1987) showed that if two non-stationary time series are cointegrated, their linear combination (such as the spread) is stationary or follows a well-defined long-run equilibrium. This makes direct forecasting of the spread statistically robust.

ii. Reduced Forecasting Error

   If we forecast crude and gasoline prices separately, each model will introduce independent forecast errors (Diebold, 2007). When the crack spread is derived as the difference of these two forecasts, the errors accumulate, leading to higher uncertainty in the final spread forecast . Forecasting the crack spread directly eliminates this issue by modeling the spread dynamics explicitly.

iii. Simpler and More Efficient Modeling

The crack spread is often stationary, or can be easily transformed into a stationary process (e.g., via first differences or seasonal adjustments). Stationary series are easier to model using time series techniques such as ARIMA, GARCH, or VAR (Box & Jenkins, 1976). On the other hand, crude and gasoline prices (individually) are likely non-stationary, requiring a more complex Vector Error Correction Model (VECM) if modeled jointly. Modeling the spread directly avoids the complexities of multivariate cointegration modeling.

iv. Market Dynamics and Structural Information

The crack spread reflects market fundamentals like refinery margins, seasonal effects, supply-demand imbalances, and regulatory changes. Directly modeling the spread ensures that these relationships are captured more effectively. If crude and gasoline prices are forecast separately, the spread forecast might fail to capture refinery-specific shocks (e.g., unexpected maintenance shutdowns or policy changes affecting gasoline demand) (Pindyck, 1999).

To summarize, it can therefore be said that forecasting the crack spread directly is superior because it:

i. Leverages cointegration, ensuring better statistical properties.
ii. Reduces accumulated forecast error, leading to more reliable predictions.
iii. Allows for simpler modeling, as the spread is likely stationary.
iv. Captures market dynamics better than separate crude and gasoline forecasts.

The only scenario where the first approach (forecasting crude and gasoline separately) might be preferable is when domain-specific factors require individual price modeling—such as when a firm wants separate price risk hedging strategies. However, for the most accurate and robust forecast of the crack spread, the second approach is typically superior.

1.2 Research Directions

There has been considerable interest in crack spreads from a variety of stakeholders, including traders, regulators, hedge funds, speculators, oil market participants, and academics (Fousekis & Grigoriadis, 2017a). Academic research has explored co-integration between petroleum product

and crude prices(Ederington et al., 2019a), the predictive power of crack spreads for oil price forecasting, and the efficacy of day trading strategies for crack spread modeling (Poitras & Teoh, 2003a). Regulators and international financial institutions monitor crack spreads to inform policy responses, facilitate trade, and regulate markets. Hedge funds use crack spreads for speculation, hedging against refinery stocks, or directional trades within energy portfolios (Murat & Tokat, 2009a). Investors may employ crack spread trading as a hedge against the stock value of refining companies. The low margins associated with crack spread trades, which generate substantial spread credit for margining purposes, make them an attractive option for professional traders considering directional trading strategies in their energy portfolios.

The majority of models now in use have mostly been developed for options trading and mitigating risk for key stakeholders, specifically traders and refiners. This indicates a paucity of models that are specifically tailored to meet the needs of the integrated Oil and Gas Refining and Marketing industry. Whereas there is abundant research in the field of Crude oil price forecasting over the decades, academic research on crack spread forecasting has been very limited (Baumeister et al., 2013). In the forecasting of crude oil prices, advanced analytical techniques such as machine learning and artificial intelligence have been more prevalent than in the prediction of *crack-spreads*. It follows that research opportunities exist for applying these advanced techniques to *crack spread* forecasting, which could improve prediction accuracy and reliability.

Together with crack spread forecasting, demand forecasting also plays an equally critical role, impacting supply chain and inventory management in the petroleum product and LPG business. Forecasting the demand for LPG (liquefied petroleum gas) cylinders involves distinct challenges, such as demand volatility, data availability and quality, seasonality, trends, customer behavior and preferences, geographic variability, and external influences like government policies (e.g., subsidy schemes, price regulations), energy market trends, competing fuel sources (e.g., natural gas, electricity), and environmental considerations (e.g., air quality regulations). Integrating these external factors into predictive models adds significant complexity that needs to be addressed. Accurate forecasts of demand for gasoline, diesel, jet fuel, and LPG allow companies to optimize their supply chain operations, ensuring the right balance between supply and demand. Predicting future consumption patterns enables refineries and distributors to plan production schedules, manage inventory levels, and coordinate logistics more effectively(Correia et al., 2017). This minimizes the risks of costly shortages and surpluses, reduces operational costs, and ensures a

reliable supply of products to meet market demand, thereby improving customer satisfaction and maintaining competitive advantage in the energy sector.

Against this backdrop, this study focuses on two specific challenges that Oil Refiners and Marketers face in regional contexts:

i. Development of robust predictive models for crack spread prediction using Machine Learning that account for these variations, offering more precise and region-specific forecasts.
ii. Developing LPG cylinder refill demand forecasting models

These predictive analytics framework and models developed for forecasting will significantly aid accurate predictions of crack spread thereby assisting in the selection / procurement of crude oil by refineries; and production, procurement and logistics planning of LPG by Oil marketeers. The framework will thereby enable managers and planners to make informed decisions, thereby contributing to improving Gross Refining Margin (GRM) and overall profitability.

1.3 Research Objectives

The overarching objective of this thesis can be stated as under:

- Analyze historical crack spread data and identify key factors influencing crack spread fluctuations.
- Investigate the applicability of various forecasting techniques, including statistical models, machine learning algorithms, and econometric approaches, in predicting crack spreads.
- To design and develop accurate and reliable predictive analytics framework for forecasting crack spread
- Forecasting model for LPG cylinder demand based on historical booking patterns, customer demographics, seasonal trends, and external factors influencing LPG consumption.
- Validate the predictive accuracy and robustness of the proposed models through empirical testing and comparison with established applicable performance metrices.

This thesis seeks to contribute by a) exploring how recent advancements in AI/ML can enhance the accuracy of forecasts for crack spreads and LPG demand, benefiting oil refiners and marketers,

and b) developing an accessible and cost-effective predictive analytics framework for forecasting, designed for easy adoption by industry practitioners.

1.4 Contributions and Outline

This thesis consists of six chapters including this Introduction chapter. The subsequent chapters are organized as follows:

Chapter 2: This Literature review chapter takes a historical view of crude price forecasting with a view to understand the evolution of techniques in this domain. A thematic evolution of Crude oil price and crack spread forecasting is presented from 2008, 2017 till 2022. The dominant themes identified with respect to forecasting in this domain are Time Series modelling and Neural networks. A chronological evolution of the forecasting techniques and their evolution is also presented. An attempt is also made to classify crude price forecasting techniques along techniques employed during the last 2 decades. The literature review of LPG Demand forecasting is also undertaken, and selection criteria identified for undertaking research in this domain.

Chapter 3: This chapter delves into the domain of Artificial Intelligence and Machine Learning. The techniques identified as evolutionary themes in the previous chapter, are elaborated upon, including a characterization of top Machine Learning algorithms along with their advantages and disadvantages. The prevalent frameworks for forecasting using statistical and machine learning techniques are explained in the context of time series forecasting frameworks Concepts such as Sliding Window and Multistep Time Series forecasting with Walk-forward validation are also discussed. Subsequently, Deep Learning models are examined, with particular attention to issues like exploding gradients and vanishing gradients. A comparison between statistical machine learning techniques and ANN deep learning techniques is presented. Various Deep Learning architectures are compared, and the most suitable architecture for time series forecasting is identified. The chapter also addresses data preparation, Bayesian optimization, and Explainable AI. Metrics for measuring model accuracy and to determine the best performing model for the respective use cases are identified.

Chapter 4: In this chapter, the conceptual model for the crack spread forecasting use case is introduced, and a proposed three-stage methodology is detailed comprising of data selection, preprocessing, and the application of modeling techniques with hyperparameter tuning. The data

for the study has been sourced from Platts daily published prices for petroleum products. The selection of data for study and duration of data considered are explained. The chapter then goes on to compare the results of crack spread forecasting using statistical, deep learning, and hybrid models, ranking the best-performing models on metrices for measuring model accuracy, bias-variance trade-off, MAPE and R-MAPE. Additionally, Bayesian optimization is explained which plays a crucial role in enhancing hyperparameter tuning performance. The hybrid deep learning model emerges as the best performing model thereby opening possibilities for exploration of hybridization of Deep Learning techniques and other time series forecasting techniques for improvement of forecast accuracy.

Chapter 5: This chapter introduces the business context for LPG demand forecasting use case and explains the data collection and modelling methodology. Time series modelling technique is employed with both statistical and AI / ML Deep Learning techniques and results are compared. The models are ranked for efficacy on a relative score comprising of RMSE and MAPE metrices the best performing model identified. Feed Forward Neural Networks (FFNN) exhibited the best performance metrices.

Chapter 6: This chapter summarizes the findings of thesis and identifies the limitations. While the research is specific to Indian petroleum refiners, its techniques are adaptable globally. Recommendations are provided for exploring longer-term forecasts and integrating newer AI architectures for improved accuracy, promising advancements in petroleum product forecasting.

Chapter 2

Literature Review

"The real power of predictive analytics comes from marrying data sources to drive insights, intelligence, and actions."

- Vala Afshar, Chief Digital Evangelist at Salesforce

2.1 Asymmetry between crude oil and product prices

The literature on the relationship between crude oil prices and gasoline spot prices reveals a complex interplay of various factors. Research using statistical methods such as Vector Autoregressions (VAR) and impulse response functions has helped researchers better understand lead-lag dynamics, volatility spillovers, and causal effects between current and future crude oil prices. (Bhutto et al., 2023). Studies have revealed asymmetric responses in the gasoline market at both national and state levels, with differing adjustment speeds for regular and premium gasoline types (Kamyabi & Chidmi, 2023). There are several factors contributing to asymmetric pass-through of oil prices to gasoline prices, including city-level characteristics such as demand for gasoline and supply. Additionally, the study on the Brazilian market highlights the importance of internal price management structures in influencing the pass-through of oil prices to gasoline and diesel prices, with long-term equilibrium adjustments observed in fuel prices (Barbosa et al., 2024). Furthermore, research in Central Asian countries indicate that the effects of crude oil prices on retail gasoline and diesel prices vary between short and long terms, with asymmetric relationships observed where positive and negative shocks to oil prices have differing impacts on retail prices. In summary, these findings emphasize that city-level characteristics, internal price management structures, and timing of oil price shocks play a significant role in determining the asymmetric pass-through of oil prices to gasoline prices. (Chesnes, 2016). Additionally, studies have explored cointegration and causality relationships between gasoline and crude oil prices, which have revealed bi-directional short-term causality and uni-directional long-run causality from gasoline futures contracts to crude oil spot prices (Dunis et al., 2006a). It is crucial for stakeholders such as policy makers, refiners, marketers and investors, who are all involved in the energy market, to understand these dynamics in order to navigate the complex market effectively.

2.2 Crack spread and its significance for petroleum refiners

In the petroleum industry, Refiners' profit / margins are directly linked to the difference (spread) between the price of crude oil and the prices of refined products distillates (gasoline, diesel, jet fuel etc.) (Steven et al., 2014a). This spread is known as 'Crack spread' owing its origin to refining process that "cracks" crude oil into its major refined products.

'Cracks spread' is of interest to refiners, traders, regulators, hedge funds & speculators, oil market players and academicians (Fousekis & Grigoriadis, 2017). The perspectives and objectives with which they view 'Crack spread' and its asymmetry, however, varies widely. Academic researchers have explored co-integration between petroleum product and crude price, predictive power of crack spread for oil price forecasting, efficacy of day trading strategy for crack spread modeling etc (Poitras & Teoh, 2003). Regulators and International Financial institutions observe crack spread from the point of view of policy response to market inefficiencies, trade facilitation and regulation of the market. Hedge funds frequently use crack spreads to speculate in oil markets (Murat & Tokat, 2009a), and to hedge against refinery stocks or perform directional trade for energy portfolio. Further, the investor community may use crack spread trades as a hedge against a refining company's equity value. Other professional traders may consider using crack spread as a directional trade as part of their energy portfolio, with the added benefit of its low margins (the crack spread trade receives a substantial spread credit for margining purposes). Together with other indicators, such as crude oil inventories and refinery utilization rates, shifts in crack spreads or refining margins can help investors get a better sense of where some companies— and the oil market may be headed in the near term.

Refiners are exposed to volatility in both the markets in which they operate: crude oil price fluctuations on supply side and volatility in price of refined products in retailing market (Karathanasopoulos et al., 2016a). Because refiners are on both sides of the market at once (Mahringer & Prokopczuk, 2015), their exposure to market risk can be greater than that incurred by companies who simply sell crude oil, or sell products to the wholesale and retail markets (Steven et al., 2014a). The price of crude oil and its principal refined products are often independently subject to variables of supply, demand, production economics, environmental regulations and other factors (Ederington et al., 2019b).

Crude oil comprises about 85% (Robinson, 2007) of a typical refining operating cost. Refiners and non-integrated marketers are therefore at an enormous risk of crude oil price increase and decline of prices of the refined products on both sides of the market. Such a situation can severely narrow the crack spread. Consequently, refinery executives, are perpetually concerned about hedging their crack spread risk. With the ability to predict the movement of crack prices well in advance, the refiners can dynamically optimize the right crude-mix through multi-sourcing and development of a compatible hedging strategy for managing the price risk.

In addition to covering the operational and fixed costs of operating the refinery, refiners desire to achieve a rate of return on invested assets. Because refiners can reliably predict their costs, other than crude oil, an uncertain crack spread can considerably cloud understanding of their true financial exposure.

Oil refiners, therefore, in an effort to optimize the supply chain of oil products, model and predict cracks spread with the objectives (Dunis et al., 2006b) of (a) Profit maximization by deciding when to procure stocks of crude or processed products (b) Increasing GRM (Gross Refining Margin - The Gross Refining Margin (GRM) is the difference between the total value of petroleum products coming out of an oil refinery (output) and the price of the raw material, (input) which is crude oil.) by selectively choosing cracks to process, and (c) Implementing low risk trading / hedging against volatility in crude price fluctuations through futures contracts for protecting the margins. Cracks are thus very useful in implementing various procurement strategies and therefore oil refiners have substantial interest in protecting the crack spread (Haigh & Holt, 2002).

Indian Oil Companies are more integrated than general oil companies carrying out refining and marketing functions in tandem. Therefore, they are not only exposed to the volatility of Crude oil prices arising out of various reasons mentioned earlier but also to inelastic changes in product prices which in turn are impacted by variations in worldwide demand, regional factors, seasonality etc.

It has been studied that whereas the Futures of commodities are good forecasts for their spot prices, the same cannot be said for crude and petroleum Cracks prices (Reichsfeld & Roache, 2011). Therefore, good quality medium term forecasts for Crude and Cracks spot prices becomes the cornerstone of procurement decisions of Indian Oil Companies.

State run Indian oil companies have societal obligation to ensure undisrupted supply of fuels across the country. Even in the face of impending unfavorable marketing conditions owing to volatility in crude oil prices, the option to stall operations till favorable return is simply non-existent. Thus, the volatility of crude price and its impact has to be borne out.

With a very limited mandate to indulge in financial transactions such as hedging, state run oil companies in India are severely constrained in ensuring profitability and margin protection. In the absence of hedging as an option, Indian state-run oil companies need to resort to forecasted crack spread prices for their procurement planning decisions. The modelling and predicting crack spread is therefore of greater relevance to refiners for optimizing their supply chain operations and hedging against the volatility in crude oil prices.

Oil refiners, in an effort to optimize the supply chain of oil products (Lisitsa et al., 2019; Szucs & Hassen, 2012), model and predict crack spread with the following objectives (Dunis et al., 2006a)

i) Profit maximization by deciding when to procure stocks of crude or processed products
ii) Increasing GRM (Gross Refining Margin) by selectively choosing cracks to process
iii) Implementing low risk trading / hedging against volatility in crude price fluctuations through futures contracts for protecting the margins

Cracks are thus very useful in implementing various procurement strategies and therefore oil refiners have substantial interest in protecting the crack spread (Murat & Tokat, 2009b). Hedge funds also heavily use crack spread to speculate in oil markets.

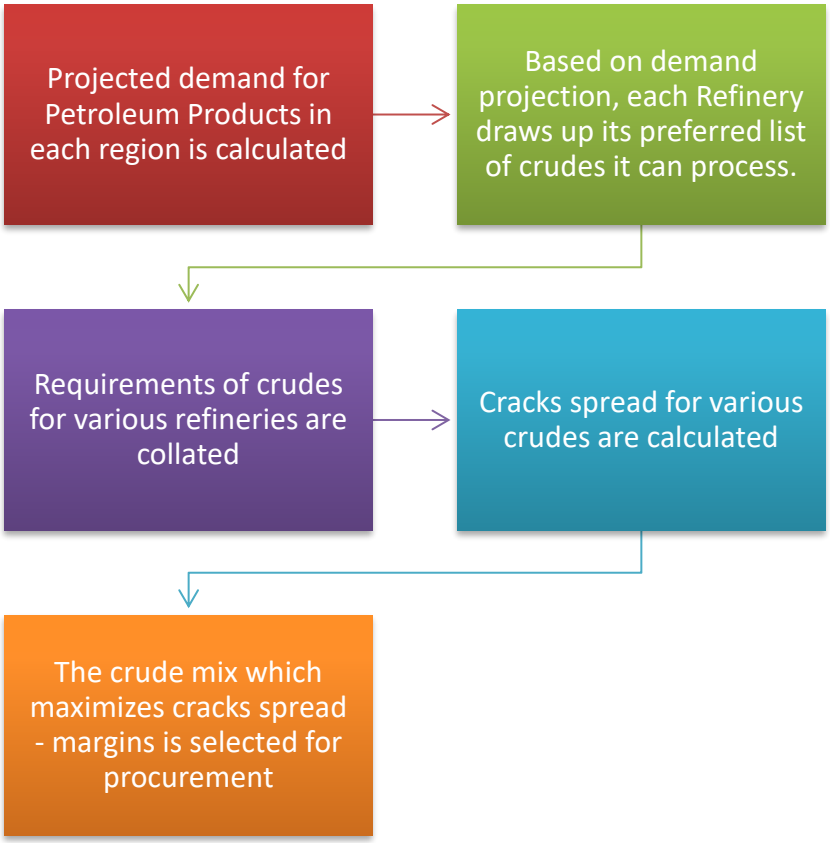


Figure 1 : Crude Selection Process based on crack spread

A recent paper study (Baumeister et al., 2018) has tried to study the hypothesis that whether there is a predictive power in the product spread (cracks) for forecasting crude oil prices. It was found that not all product spread (cracks) model are useful for forecasting. Cracks related to Gasoline and heating oil spot price spreads are good candidates and offer improved prediction accuracy

The crude oil roughly constitutes 90-95% of input cost for a refining company. The crude procurement strategy adopted by refiners is a mix of two components (i) Term contracts and (ii) Spot procurements as depicted in **Figure 2**.

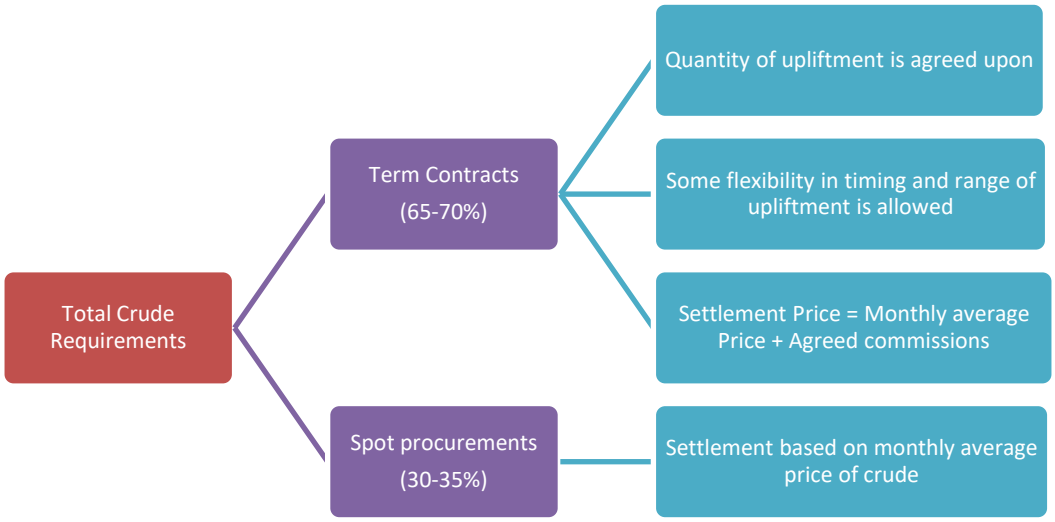


Figure 2 : Crude Procurement strategy adopted by refiners

This mix strategy is adopted with the twin objective of securing a steady supply of crude oil for the refineries which need to be continually operated and to hedge against the volatility in crude oil prices

A study conducted on the volatility of crude oil price from the period of 1991-2006 (Narayan & Narayan, 2007) arrived at the conclusion that the behavior of oil prices tends to change over short periods of time. Another study by (Haidar et al., 2008) attributed the volatility to three main factors: namely:

i) Supply / Demand imbalances, (Economic growth, behavior of oil producing countries)
ii) Exogenous events such as war, weather (storms, hurricane and earthquakes)
iii) Endogenous factors (speculation in the markets).

In addition to the above, the price of crude oil and its principal refined products are also independently subject to variables of environmental regulations, evolution of newer alternative technologies and other factors.

Forecasting of crude oil prices and its derivatives (futures), have evinced a lot of research due to the pivotal role played by crude oil in macroeconomic planning and meeting energy demand of countries since the turn of the last century. Forecasting crude oil prices or its derivatives have always been fraught with uncertainty owing to the highly volatile nature of crude oil prices.

From the above we can infer the following:

i. Crack spread model the profitability of downstream refining companies
ii. Crack spread are good predictor of spot crude oil prices.
iii. Modelling and predicting crack spread is of greater relevance to refiners for optimizing their supply chain operations and hedging against the volatility in crude oil prices.

2.3 Crude price forecasting

Crack spread is difference (spread) between the price of crude oil and the price of refined products distillates (gasoline, gas oil etc.). The principal component of the crack spread being crude oil, it has a direct bearing on the prices of cracks. In-fact the prices of cracks are correlated to prices of crude oil. Also, crude owing to its strategic significance has attracted a lot more interest over the years in terms of modelling and forecasting efforts. Therefore in order to get a comprehensive view of the forecasting landscape i.e. the evolution of the forecasting models and techniques over the years, it is important to first undertake the study of crude forecasting techniques and their evolution, which will provide us with greater insights to craft our modelling initiatives for Crack spreads. Accordingly, we have invested considerable effort in studying and understanding the evolution of Crude oil forecasting techniques and the same is presented in the next few sections.

Crude oil is the primary source for meeting energy demands of various countries. Fluctuations in the price of crude oil (volatility) puts a huge strain on the country's economy. Crude oil prices are very volatile as they are impacted by various natural and man-made factors. Forecasting the price of crude oil price is of immense significance to economists, policy makers and for supply chain planning.

Extensive research over forecasting crude oil price has been carried out in the last couple of decades. Several techniques have been developed, employed and evolved with respect to Crude Oil Price forecasting. With the evolution of new computing paradigms such as Artificial Neural Networks and Deep Learning Neural Networks, applicability of these new techniques for Crude Oil Price forecasting is being researched. Crack spread being a derivative of the crude and product prices and the cointegration between the crude and product prices, crack spread prices are amenable to similar kind of analysis. Therefore, to cover the research spectrum comprehensively,

the review of crude price forecasting and crack spread forecasting have been undertaken separately with a view that the methodologies learnt will help us arrive at a better forecasting model.

Figure 3: Crude prices for the last 4 decades

Figure 3 depicts the price of Crude oil over the last 4 decades. The volatility (rise and fall) and major shifts in price levels are clearly discernable and attributable to major events such as Gulf War (1990-91), The Y2K Bubble (1998-200), Sub-prime lending crisis in USA (2008), Excess shale gas production by US (2014-16) and the COVID-19 pandemic in 2020 to name some of the few major events. In order to undertake Literature Review of Crude oil forecasting techniques, the databases such as SCOPUS, Google Scholar, Research Gate etc. were searched.

Thematic Evolution of Crude Oil / Cracks Price Forecasting

For understanding the evolution of Crude oil / cracks price forecasting, a three-pronged search methodology was employed. The first stage involved searching documents in closed databases such as Scopus and in public domain through Google Scholar and Research Gate. The keywords on which the databases were searched are as under:

Keywords: (crude" OR "cracks") AND ("price" OR "prices") AND ("forecasting" OR "prediction" OR "predicting" OR "forecast"))

Initial search was performed in Scopus database on the keywords mentioned above. This scan yielded 828 research papers. The document attributes of qualifying documents were downloaded. This result set was used to analyze the thematic evolution of the Crude price forecasting over the period. Abstract of all these papers were eyeballed / screened to select only those papers which pertained to Crude Price Forecasting domain only. This resulted in the elimination of 225 papers from the set, leaving the result set of 603 papers. A thematic review using bibliometric tool "Biblimetrix" was carried on these 603 papers to understand the drift of research over this period and identify evolution of the techniques.

The time horizon between 2008 and 2021-22 has very discernible break around 2014-2016 during which the US had increased the shale gas production which caused a drastic drop in crude oil price (Figuerola-Ferretti et al., 2020). Also, rapid advances in computational techniques were taking place, more around ANN which was also leading to lot of research on their applicability in various domains. It was therefore felt that splitting of time horizon around this time will bring out some interesting insights.

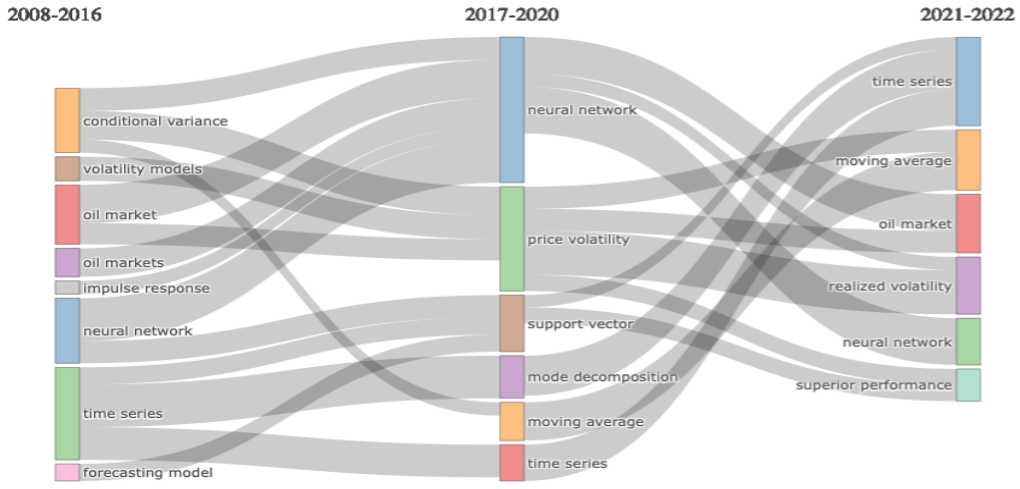


Figure 4 : Thematic evolution of crude forecasting

The **Figure 4** depicts the evolution of themes in research of Crude oil price forecasting. The themes that have consistently remained in focus over the period of time are Volatility, Time Series and Neural Networks. Other themes such as conditional variance, support vector, mode decomposition, moving average have evolved or morphed into one theme or the other.

The figures below provide a cross-sectional evolution of the themes in terms of Development Degree (Relevance) and Relevance Degree (Centrality) **Figure 5,6,7**. The four quadrants depict the position of themes around the Relevance and Centrality theme as below:

1. Emerging or declining (Low Relevance and Low Centrality)
2. Niche Theme (High Relevance and Low Centrality)
3. Basic Themes (Low on Relevance and High Centrality)
4. Motor Themes (High Relevance and High Centrality)

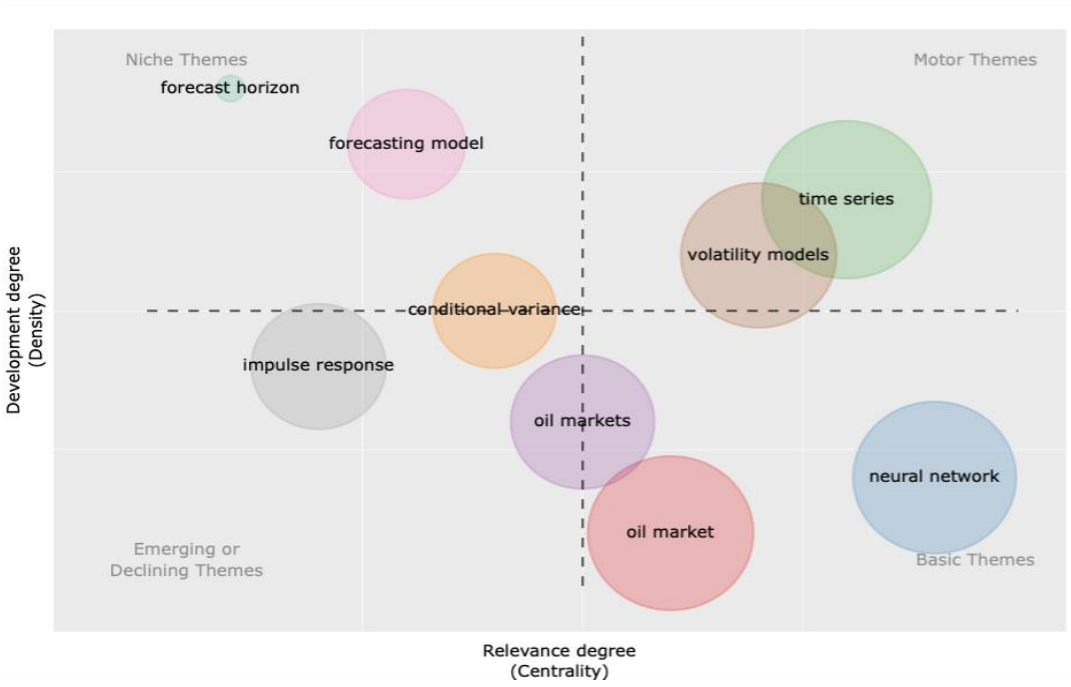


Figure 5: Thematic evolution of crude / crack price forecasting (2008)

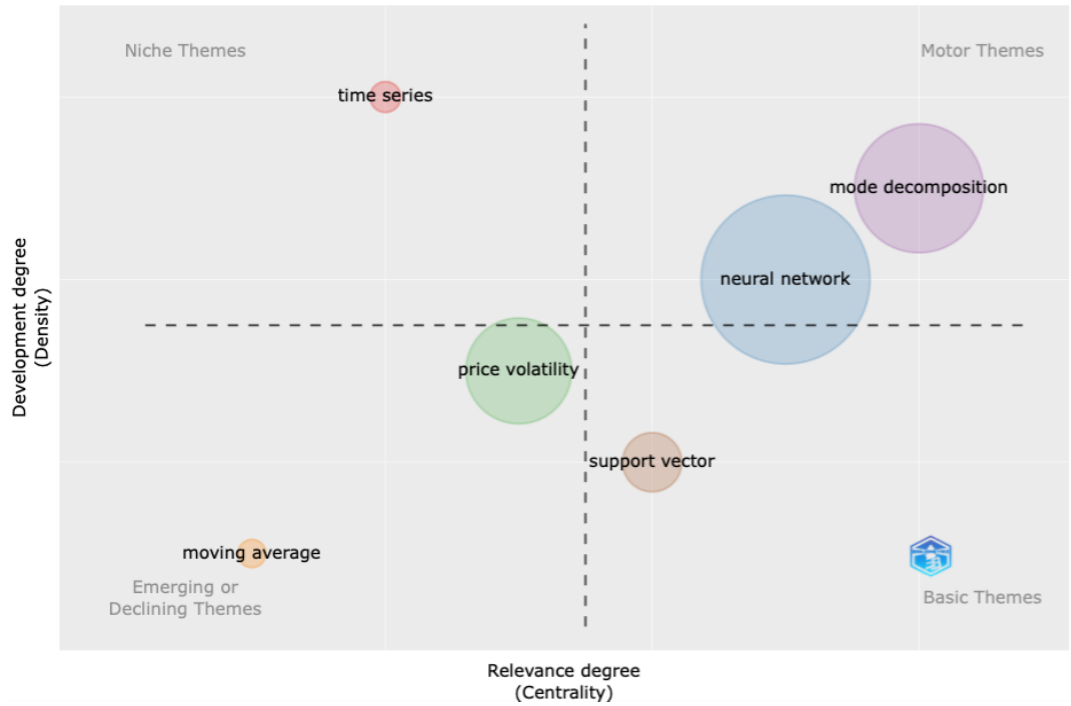


Figure 6: Thematic evolution of Crude / crack price forecasting (2016).

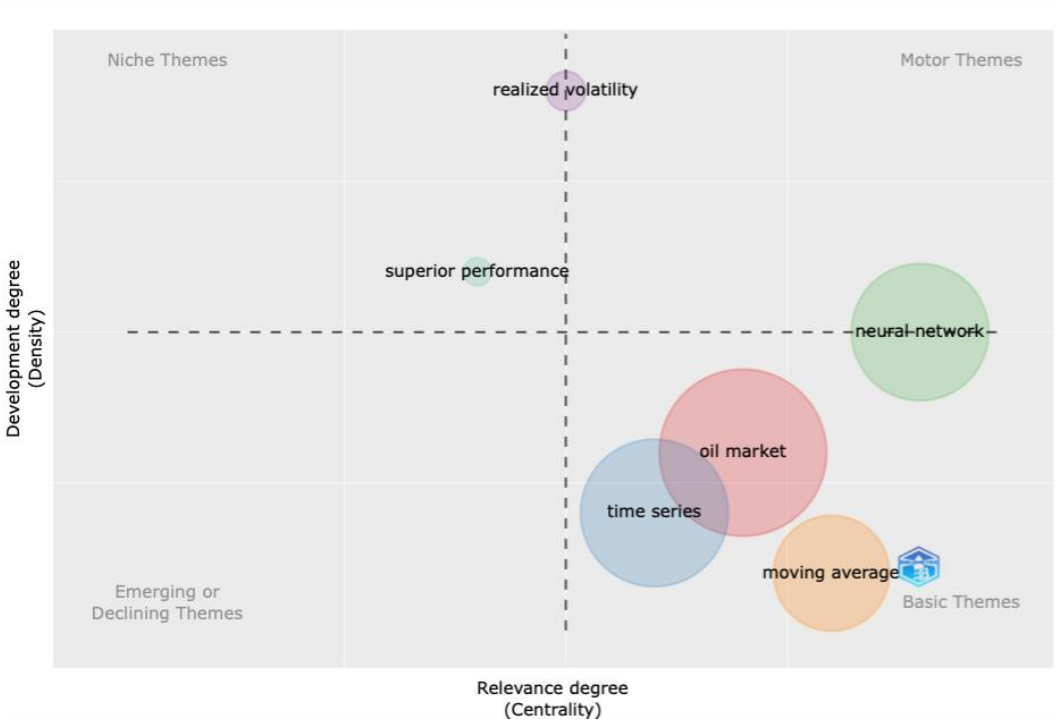


Figure 7 : Thematic evolution of Crude Price Forecasting (2022).

It is apparent from the above figures that Time Series sand Neural Network themes have prominently evolved and in the year 2022 being primarily researched.

Evolution of Techniques

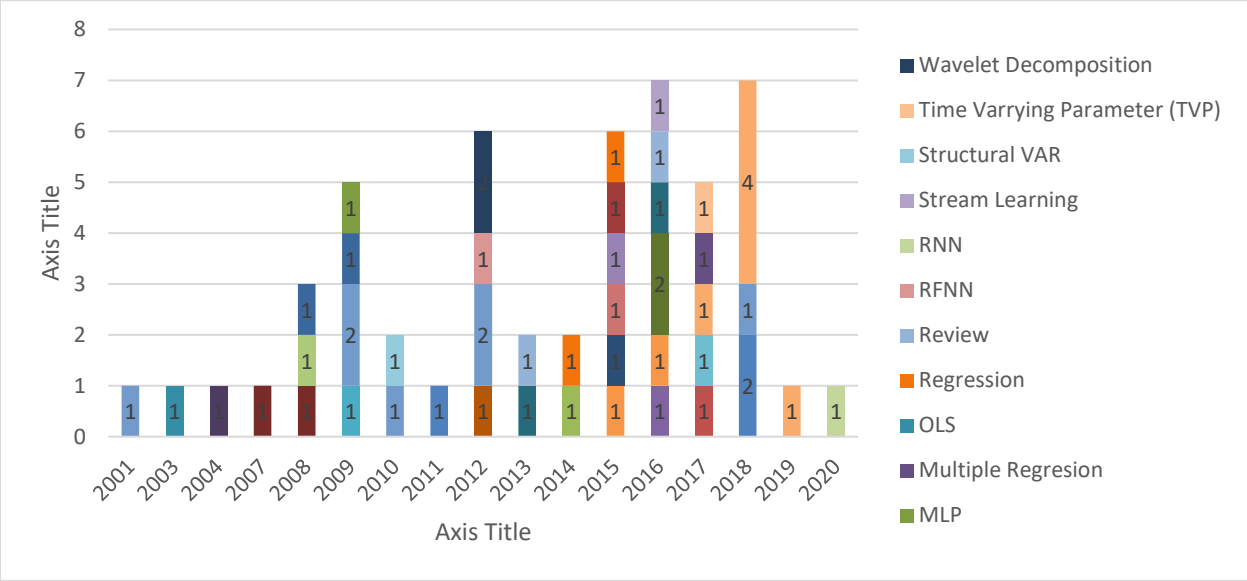


Figure 8 : Chronology of research in forecasting crude / crack price using various techniques.

Crude, the black gold has been engine of world's economic growth in the last century. Forecasting the price of crude, has always been in vogue (**Figure 8**). Application of Ensemble Empirical Mode Decomposition (EEMD) was a novel data analysis method (Zhang et al., 2008) in which the crude price time series is decomposed into a number of independent intrinsic modes from high frequency to low frequency. These modes are composed into a fluctuating process, a slowly varying part and a trend based on the fine-to-course reconstruction. Interval Forecasting of Crude Prices (Xu et al., 2008) was an attempt to forecast the price of crude oil spot prices by arriving at the relationship between commodity inventory levels and crude. The empirical study suggested that the interval method of ILS method produces results better than 95% confidence interval. An ensemble of Empirical Mode Decomposition and Neural network learning paradigm (Yu et al., 2008) was proposed for spot price forecasting of crude oil. A 3 layer Feed Forward Network modelled each of the small number of Intrinsic Mode Functions (IMFs) which were created for crude oil spot price series. The predictions of all the IMFs were combined by an adaptive linear neural network

(ALNN) to create an ensemble output for the original crude oil price series, which showed promising results

The volatility of the crude oil prices (Haidar et al., 2008) can be attributed to three main factors: namely i) Supply / Demand imbalances, (Economic growth, behavior of oil producing countries) ii) Exogenous events such as war, weather iii) Endogenous factors (speculation in the markets). For short term prediction of crude oil prices, they implemented a 3 layer Feed Forward based Artificial Neural Network (ANN) with back propagation algorithm. Two groups of inputs, crude oil futures and Stock market data were tested. Another attempt to forecast the price of crude oil in short term using ANN and commodity futures(Kulkarni & Haidar, 2009) modelled a multilayer feedforward neural network. The reported accuracy of the direction of the market was 76%, 66% and 53% for one, two and three days in future conclusively.

Refiners are major participants in oil markets and their substantial interest is in protecting the crack spread (Murat & Tokat, 2009b). Crack spread refers to the crude-product price relationship. Oil consumers are also an active participants is hedging and are exposed to crack spread. Also, hedge funds heavily use crack spreads to speculate in oil markets. Owing to these scenarios, the study tried to establish whether crack spread is a good predictor of oil price movements. The results indicated that (a) both the crack spread futures and the crude oil futures outperformed the RWM and (b) the crack spread futures are almost as good as the crude oil futures in predicting the movements in the spot oil markets. . The existence of this new relationship owes its origin to increasing uncertainty in oil market and therefore the increasing need for hedging particularly by the refiners. The neural networks have an advantage over conventional forecasting methods because they can learn more or less autonomously high dimensional and non-linear functions out of huge amount of data

An attempt to generate short, medium and long terms crude price forecast using ANN(Lackes et al., 2009) ended up with mixed results. The study used 3 layered Feedforward Multi Layered Perceptron, WTI crude prices with 2000 data points over a period from 1999 to 2006. The results indicated that the implemented neural networks achieved good results for mid to long term predictions but not for short term. To model the volatility in crude oil prices using GARCH type models on time series data of WTI future contracts(Agnolucci, 2009), nothing conclusive could be established.

A comparative study to assess the application of Neural Networks for the study of crude oil futures (Hu et al., 2012) arrived at the following conclusions: (a) Learning performance can be improved by increasing the training time (b) Recurrent Fuzzy Neural Network (RFNN) has the best predictive power (c) MLP has the worst performance. A hybrid technique for employing Multilayer Backpropagation neural network and wavelet decomposition (HTW-MPNN) (Jammazi & Aloui, 2012) was implemented to achieve better prediction of crude oil price. The findings indicated that this model performed better than conventional BPNN. Crude prices are notoriously volatile. A study to forecast energy market volatility using both Univariate and Multivariate GARCH class models (Y. Wang & Wu, 2012a) established that univariate models allowing for asymmetric effects exhibit greater accuracy.

The review paper on Crude Oil price forecasting (Gabralla & Abraham, 2013) lists out various efforts in terms of modelling crude price forecasting over the last 2 decades. An alternative method for Crude oil price forecasting explored the use of Genetic Algorithm and Neural Network. The studies indicated that the performance of the GA-NN model was better than the baseline algorithms in terms of prediction accuracy and efficiency.

In order to increase the forecasting accuracy of predictions of crude oil price and its fluctuations (J. Wang & Wang, 2016), a combination of Multilayer Perceptron and ERNN with stochastic time effective function was implemented. The stochastic time effective function implies that the recent information will have a greater effect in the investors than the old information. The proposed model was able to perform better and faster than the other comparative models. The application of new machine learning paradigm called the Stream Learning (Gao & Lei, 2017) has been proposed which capitalizes the fact that machine learning models will be able to capture new changing oil price data as soon as they are available with very small constant overhead. In order to predict complicated time series with high volatility and irregularity such as crude oil price, a novell approach (Yu et al., 2016) adopted decomposition and ensemble learning paradigm. This integrated the esemble empirical mode decomposition (EEMD) and extended extreme learning machine (EELM) for crude oil price forecasting. Empirically, the model was able to perform better at predicting spot prices of WTI crude compared to benchmark models in prediction accuracy and effectiveness.

A review of recent AI methodologies for forecasting crude oil price (Chiroma et al., 2016) provides a compilation of various AI techniques that have been put to use for forecasting crude oil price. A hybrid model incorporating statistical modelling technique Exponential Trend and Seasonality (ETS) and ANN (Panigrahi & Behera, 2017) has been proposed. For comparative analysis, sixteen time series datasets have been used. The technique has been compared with other standalone techniques ARIMA, ETS and hybrid ARIMA + ANN models. Experimental results have shown that the proposed model delivers better results for the datasets used.

With the evaluation of Deep Learning ANN models, recent efforts are directed towards studying the applicability of these models in forecasting crude oil price (Chen et al., 2017). Crude oil price movement may have a both linear and nonlinear component to it. Deep Learning networks are better suited to deal with non-linear data than the traditional data. They are also better suited for feature extraction through its multiple layers. Deep Belief Networks (DBNs) are better at dealing with issues related to back propagation, including long learning times, need for large labelled data sets inadequate parameter selection technique. Long Shot Term Memory (LSTM) networks are good at modelling long sequences and issues related to back propagation.

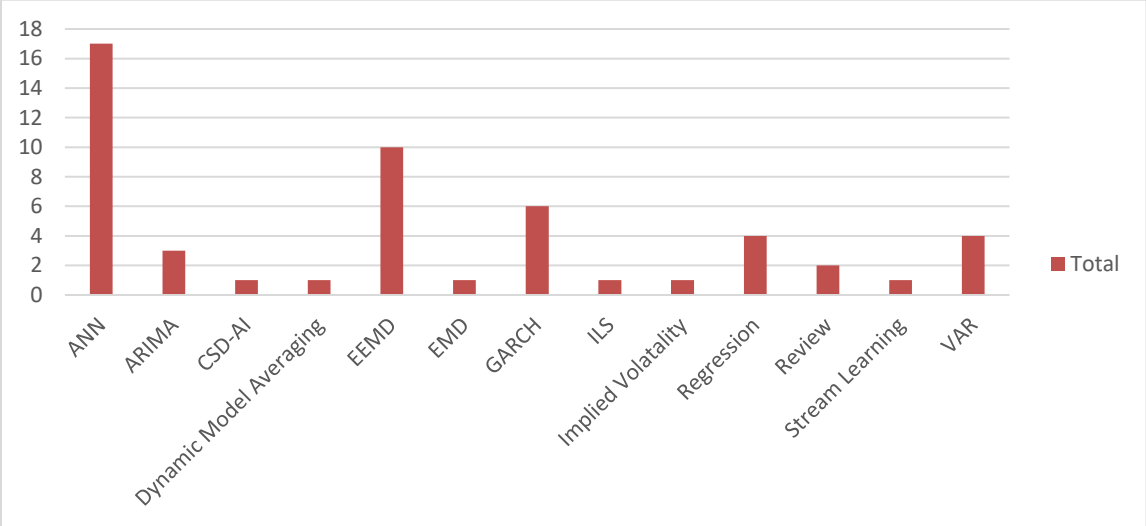


Figure 9 : Research papers published during on a particular technique 2001-2021.

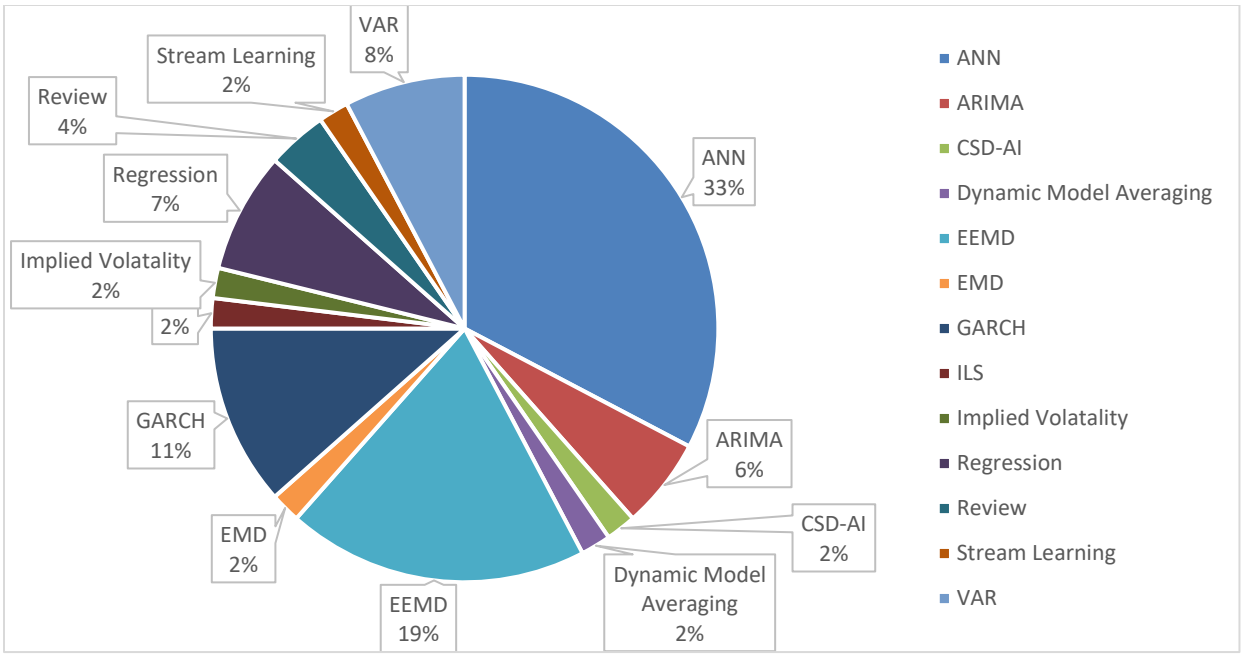


Figure 10 : Techniques employed in research papers published during 2001-2021.

A standalone LSTM model (Gupta & Pandey, 2018) was proposed for crude price forecasting and various combinations of LSTM models were tried out. A comparision of LSTM models and ARIMA models for forecasting petroleum production (Sagheer & Kotb, 2019) employed historical and new observations to improve model performance. Hyper parameter tuning was done using Grid search. Another hybrid model (M. Wang et al., 2018) used Data Fluctuation Network (DFN) and several AI algorithms named DFN-AI. In this method a complex network time series analysis technique is performed as a pre-step on the original data to extract the fluctuation features and reconstruct the original data, and then AI is employed to model the reconstructed data and predict future values. The results indicate that this model performs better than any single AI model. Concerns and way forward with respect to statistical and machine learning methods (Makridakis et al., 2018) have been suitably discussed. A comparision between ARIMA, LSTM and GRU for forecasting time series data (Yamak et al., 2019) on an individual basis lists ARIMA as the best option. A comparision of ARIMA, LSTM and BiLSTM (Siami-Namini et al., 2019) reports that the BiLSTM outperforms regular LSTM models and ARIMA models. It was also observed that BiLSTM models reach the equilibrium much slower than LSTM based models. Forecasting of crude price using a hybrid of EEMD, LSTM and enhancing data transfer rate (Cen & Wang, 2019) yielded better results. For predicting the price of Brent crude price, a comparision of LSTM and ARIMA / GARCH (Salvi, 2019) used Dropout layers to enhance the learning rate.

A model for forecasting WTI futures (Matsuoka & Hamori, 2021) has employed a Recurrent Neural Network in a standalone configuration. Hybridization of RNN in various configuration tends to improve accuracy.

Classification of Forecasting Techniques

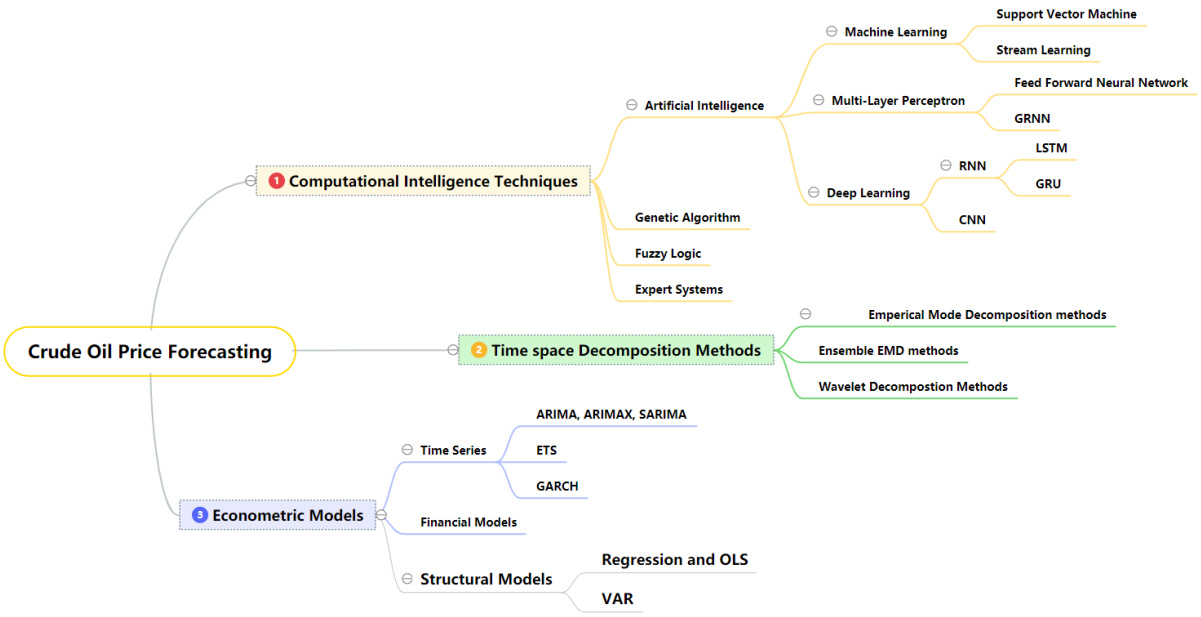


Figure 11 : Classification of crude oil price forecasting techniques published during 2001-2021.

The Crude Oil Price forecasting techniques that have been employed over the period of time can be broadly classified under the following heads:

1. Computational Intelligence Technique

   These techniques are algorithms that are modelled based on some object / concept and implemented using programs.

2. Time Space Decomposition Methods

   These technique attempts to decompose the Crude oil price time series data into various constituents waves and then tries to model each constituent wave for generating forecasts.

3. Econometric Models

   These models are built on well researched and established econometric techniques / body of knowledge.

Summary of Literature Review of Crude oil price forecasting

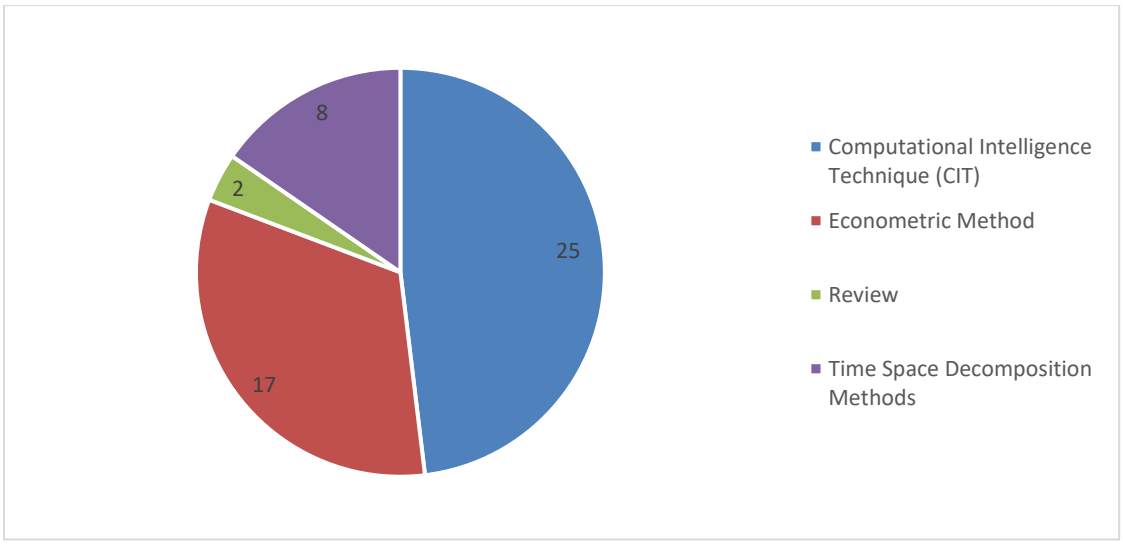


Figure 12 : Types of models used for crude / crack price forecasting (2001-2021).

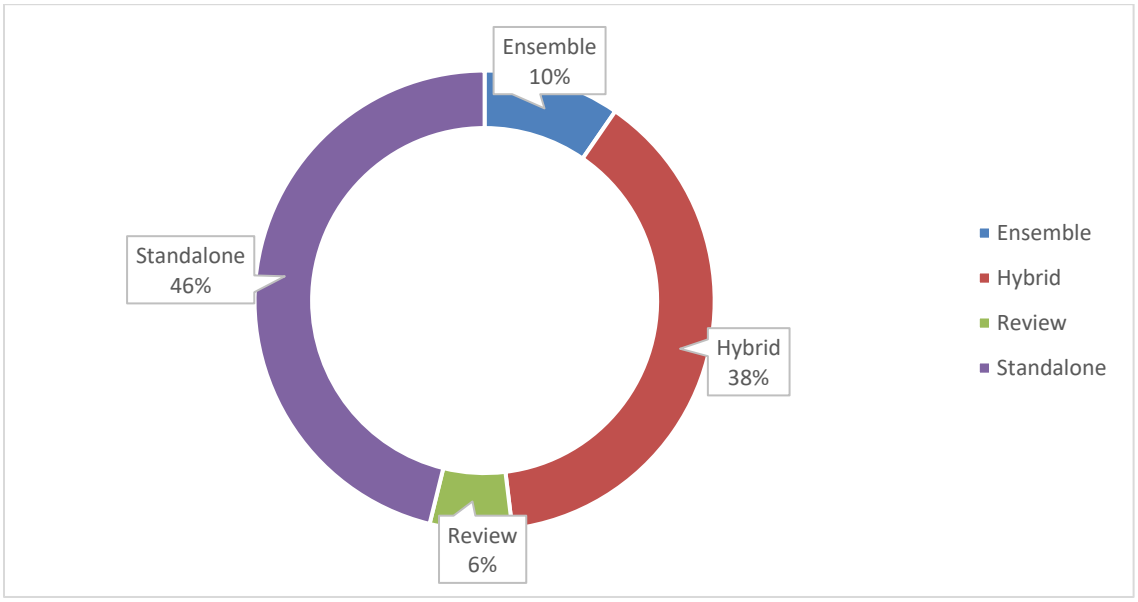


Figure 13 : Types of models used for crude / crack price forecasting (2001-2021).

2.4 Crack Spread forecasting and its evolution

A non-linear approach to modelling and tracking of the crack spread (Karathanasopoulos et al., 2016a) combined radial basis function (RBF) neural network and particle swarm optimizer (PSO) and was trained using a sliding window approach. The model when trained over 300-400 days sliding window and used to generate forecasts for 5 days gave superior performance and statistical accuracy when compared with a multi-layer perceptron (MLP) neural network. The forecast time horizon of 5 days though is appropriate for use in trading decision making but is not suitable for Refiners whose turn-around time for procurement-refining-sale of products is 3 months.

Another research on modelling and trading of Gasoline crack spread (Dunis et al., 2006a) employed High Order Neural Networks (HONN) on its time series data to investigate co-integration between WTI crude and Gasoline prices. The technique was tested to predict the change in the spread from one closing price to the next. The study established that there was a non-linear relationship between WTI crude price and Gasoline prices. The horizon of predicted using the HONN is too short to be considered a valid fit for time horizon for 3 months and hence of little use to refiners.

For modelling and forecasting crack spreads (Y. Wang & Wu, 2012b) have studied ARFIMA (Auto-Regressive Fractionally Integrated Moving Average) models applied to single product cracks. The objective was to verify that crack spread and their volatilities display long range autocorrelation. Also the study attempted to ascertain that whether crack spreads exhibit multi-fractionality wherein small and large fluctuations exhibit different scaling behaviors. It concluded that Crack-spreads exhibited weak mean reverting behavior for time horizon of days and weeks and exhibited strong mean reverting behavior in weeks and months' time horizon. On comparison it was found that ARFIMA could not outperform RWM out of sample forecasting and it could not capture the complexity of time series leading to larger prediction bias. ARFIMA could however capture long range autocorrelation but could not capture scaling. Also, it was able to capture multi-fractionality but not small-scale fluctuations.

Multiple regression analysis model for modelling for Naptha cracks (Sung et al., 2012) studied major parameters affecting naphtha prices and tried to identify the most dominating factors. Actual and predicted variations of naphtha crack were also analyzed. It concluded that naphtha price

depended mainly on Asia supply and demand of naphtha as well as naphtha substitute and margin. A statistical model for the same was prepared.

A more recent study (Ewing & Thompson, 2018) to model the response of Gasoline-Crude Oil Price Crack Spread to macroeconomic shocks such as economic growth, inflation, corporate default risk and monetary policy shocks has been attempted. This approach employs Vector Auto-Regressive approach with generalized impulse response analysis. The findings though were localized (i.e., relevant to the local context of U.S.A.) assessed the impact of shocks to these variables and consequent impact on gasoline crack spread.

In a study by (Mahringer & Prokopczuk, 2015) a comparison of forecast generated using the crack spread option prices using 2 different approaches were compared. Option 1 involved modelling the crack spread option using the two-factor mean reverting model in a continuous time framework with constant volatility and option 2 involved developing a bivariate GARCH model with co-integrated underlying futures forming the crack spreads between crude and heating oil and crude oil and gasoline, respectively. The empirical results showed that, when including maturity effects, the univariate and the more simplistic approach yielded a better pricing performance, mainly because of lower unsystematic pricing error.

In analyzing the price risk to crack spread due to volatility dynamics and seasonality(Suenaga & Smith, 2011), had modelled using econometric techniques the shocks to prices of crude oil Gasoline and heating oil. This study attempted to factor in the impact inventory held (Geman & Smith, 2013) and the maturity of futures contract. The estimated model revealed highly nonlinear volatility dynamics of the three commodity prices that are consistent with the observed seasonality in demand and storage. For all three commodities, volatility exhibits both time-to- delivery effects

Table 1: Summary of Literature Review

Year	Citations	Objective	Technique	Findings
2022	(H. Guliyev & E. Mustafayev, 2022)	To predict West Texas Intermediate (WTI) oil price dynamics from dataset of 1991 to 2021	Machine Learning Models like Logistic Regression, Decision Tree, Random Forest, AdaBoost, and XgBoost and compare the performance of best machine learning models with DeLong statistical test procedures and SHAP (SHapley Additive exPlanations) values to support model evaluation and interpretability	Explainable Machine Learning (XML) model along with the ML techniques can improve the prediction of WTI crude oil prices and severely impact policy decisions across the world economies.
2022	(Y. Bai et al, 2022)	Price forecasting of crude oil using news Text	Using two novel indicators of topic and sentiment analysis for short and	As an improvement of text indicators from news headlines and exogenous financial parameters in

Year	Citations	Objective	Technique	Findings
			sparse text data, AdaBoost.RT with the proposed text indicators outperforms the other benchmarks in crude oil price prediction.	forecasting model, sparse and short news headlines are used in modelling a framework for crude oil price forecasting and is more suitable for price forecasting with historical data.
2022	(Dmitry Vedenov & Gabriel J. Power, 2022)	To develop heuristics for Hedging effectiveness	Comparing two hedging approaches namely minimizing downside risk and minimization of variance in a market exhibiting strong momentum using Copulas based approach	Downside risk minimization hedging technique improves the decision-maker's expected utility in nearly all (99.6%) cases. This technique also appears to be most advantageous when the spot and futures prices exhibit strong upward or downward momentum.
2021	(He H, Sun M, Li X et al, 2021)	To Propose a Hybrid forecasting model for crude oil price trend prediction using Machine Learning algorithm.	Crude oil price series is decomposed by variational mode decomposition algorithm and multi-modal data features are extracted based on the decomposed modes. Using time series analysis, volatility of crude oil prices is simultaneously converted into trend symbols. Machine learning multi-classifier are then trained with multi-modal data features and historical volatility as input and trend symbols as output.	The hybrid model (ML classification prediction method based on multi-modal data features) uses the Variational mode decomposition (VMD) algorithm to obtain the low frequency and high frequency Intrinsic Mode Functions (IMF) of crude oil prices, and then extracts multi-modal data features based on the IMFs.The accuracy of the proposed model for predicting high volatility of crude oil prices is evidenced to be better than that of low volatility.
2021	(Busari G & Lim D, 2021)	Study compares the AdaBoost-LSTM and AdaBoost-GRU models for improving forecasting performance of Crude oil price prediction.	Hybrid approach for comparing AdaBoost algorithm and Long Short-Term Memory (LSTM) network and AdaBoost algorithm and Gated Recurrent Unit (GRU) network, with an aim for forecasting the crude oil price resulting in improved forecasting accuracy.	AdaBoost-LSTM and AdaBoost-GRU models along with the two benchmarked models (single LSTM and single GRU) were used on the crude oil prices data to test the forecasting measures. From five different metrics used in the paper, it is found that the AdaBoost-GRU model outperforms all other models, making it a more realistic and powerful tool for the prediction of crude oil price.
2020	(Abdollahi H & Ebrahimi S, 2020)	Hybrid model for forecasting Brent crude oil price	For Brent Crude Oil price forecasting, a hybrid model like the Adaptive Neuro Fuzzy Inference System (ANFIS), Autoregressive Fractionally Integrated Moving Average (ARFIMA), and Markov-switching models are employed where features like nonlinearity, lag, and market interrelationships existing in oil price time series are captured.	The complexity of Brent crude oil prices pertinent to time series comprises lag, nonlinearity, and interrelationships with different markets and to optimize these features a hybrid model with genetic algorithm weighting produces quality forecasting values was presented to forecast daily prices.
2018	(Ewing & Thompson, 2018)	Identify the impact of macroeconomic factors such as real output growth, inflation, corporate default risk and monetary policy stance on single product gasoline crack spread	Econometric techniques namely Vector Auto-Regression (VAR) and Generalized Impulse Response analysis	Upstream fundamentals are more sensitive to changes in economic output than downstream (i.e., refined petroleum products). Small and unanticipated changes in aggregate demand can lead to relatively substantial changes in the crack spread. Effect of inflation lasts from 1-5 months. Corporate Default Risk has no impact. Monetary policy shocks are short lived (up to 2 months)
2016	(Karathanasopoulos et al., 2016b)	The modeling and the tracking of the crack spread with a non-linear approach.	Combining radial basis function (RBF) neural network and particle swarm optimizer (PSO) trained using a sliding window approach	5 days forecasts with 80-85% accuracy

Year	Citations	Objective	Technique	Findings
2016	(Dunis et al., 2016)	Predict WTI-GAS crack spread time-series using non-linear models	The best trading model of the spread is the higher order neural network (HONN) with the threshold filter	Relationship between WTI and GAS is non-linear. HONN outperformed the MLP out of sample, despite shorter computational times and limited variables
2015	(Mahringer & Prokopczuk, 2015)	Comparing pricing of crack spread options using GARCH volatility model or crack spread model.	Time series data for options prices are modelled using bivariate GARCH model and univariate crack spread model	The more simplistic univariate approach to option pricing was found to be better than bivariate GARCH model
2014	(Kallestrup et al, 2014)	Decision support systems based on Hierarchical planning structures for crude oil procurement planning in the Oil Refining industry.	Implicit Anticipation model is used to develop a DSS system building on the case study for procurement planning	Case study approach resulted in significant savings in planning efforts and procurement costs thereby improving the procurement planning for Oil Refining industries.
2013	(Hong, Z., & Lee, C., 2013)	The paper proposes a novel decision support framework to help the buyer make optimal and robust procurement decision including supplier selection and order allocation among multiple supplier sources in the existence of correlated demand, yield and spot price uncertainties.	Monte Carlo simulation algorithm termed as Expected Profit–Supply at Risk (A-EPSaR) is used to quantify each supplier's risk so as to let decision maker realize the trade-off between profit and risk. The goal programming model helps to allocate orders among the supplier pool and the contract-spot allocation model can assign orders between the spot market and the supplier pool, respectively.	Procurement Risk Management (PRM) framework is designed to support the DSS to identify the risks in the procurement model and build profit model of their owm.
2013	(Shin, H. et al, 2013)	Forecast the upward and downward movement of oil prices	Machine learning algorithm known as semi-supervised learning (SSL) is used to create the network representation of entities and the explicitness of inference which is expressed through relations between different entities.	A novel method for oil price prediction using the SSL algorithm is proposed that modifies the existing SSL algorithm for application to time-series prediction, including. measuring the similarity between different sets of time-series data and the labels of pricing ups and downs, and the advanced techniques using TI transformation and feature extraction.
2013	(Suenaga & Smith, n.d.)	To develop two copula-based crack spread option models for pricing of a heating oil–crude oil "crack" spread option	Following two models developed (i) a Geometric Brownian motion crack spread option model (GBM-CSOM) and (ii) a mean reversion crack spread option model (MR-CSOM)	Copula approach is methodologically more robust technique in capturing non-linear dependencies and the unique features of energy commodities namely mean reversion and seasonality.
2010	(C.-F. Tsai & Y.-C. Hsiao, 2010)	Predict stock price for investors using prediction model for stock prediction used for future investment decisions	Data mining techniques such as Principal Component Analysis (PCA), Genetic Algorithms (GA) and decision trees (CART) are used to model the prediction of stock price	Comparing the values from Principal Component Analysis (PCA), Genetic Algorithms (GA), and decision trees (CART) and based on union, intersection, and multi-intersection approaches, a superior prediction accuracy and errors are achieved.

and substantial seasonal variations, yet these volatility patterns differ substantially across the three commodities and by delivery month of the contract.

A survey on factors affecting the petroleum product prices (Ederington et al., 2019b) identified fundamental factors such as refinery outages and the weather on product prices, the way that price discovery occurs for petroleum products, and the predictive accuracy of petroleum product futures prices for future spot prices. As per the empirical evidence, it concluded that speculation does not

increase volatility in petroleum product prices or excess returns, compared to risk-free returns on such products during the period of the study. Petroleum product futures prices (gasoline and heating oil) are generally found to be unbiased predictors of future spot prices three months out, but not for the six- and 12-month contract horizons. In examinations of predictive ability, gasoline futures prices perform better (i.e., generate fewer mean-squared errors) than a simple random walk, and perform better than the forecast ability of oil and heating oil futures.

To develop heuristics for Crack spread hedgers, the study by (Vedenov & Power, 2022) compared two hedging strategies namely the minimizing the downside risk criterion and minimization of variance measure on Expected Utility. It concluded that Downside risk minimization hedging technique improves the decision-maker's expected utility in nearly all (99.6%) cases. This technique also appears to be most advantageous when the spot and futures prices exhibit strong upward or downward momentum.

For enterprises involved in downstream oil refining and marketing, the crack spread functions as a pivotal indicator of profitability. This measure is more effective than merely monitoring absolute prices of crude or refined products for inventory management, production strategies, pricing mechanisms, and hedging tactics for several salient reasons:

a. Inventory Planning: Efficient Management of Stock Levels

   Variations in crack spread are reflective of refining margins and serve as indicators regarding the advisability of increasing stock levels of either crude or refined products. When the crack spread is elevated, refiners may opt to augment crude oil acquisitions and enhance the output of refined products to capitalize on the more favorable margins. Conversely, if the spread is contracting or is in a negative state, this may signal an oversupply or diminished demand for products, thereby prompting refiners to adjust inventory levels strategically to mitigate potential losses. This approach enables improved decision-making concerning crude procurement and the storage of refined products, thereby enhancing overall profitability. For instance, a refiner might defer crude oil purchases in response to a decreasing crack spread, which may indicate a downturn in gasoline demand (Geman, 2005).

b. Production Planning: Enhancing Refinery Operations

Crack spreads serve as a critical mechanism for refiners to modulate refinery throughput and enhance product yields. Variations in crack spreads (e.g., 3:2:1 for gasoline-centric refiners, 2:1:1 for diesel-centric refiners) are instrumental in determining the prioritization of product output. Seasonal influences have a significant impact on spreads: During the summer months, elevated gasoline crack spreads result in an augmentation of gasoline production. Conversely, in the winter season, increased distillate spreads (including heating oil and diesel) necessitate a transition towards distillate production. The optimization of refinery operations is pivotal for maximizing profitability in response to prevailing market signals. Should gasoline crack spreads experience an uptick prior to the summer season, refiners are likely to adjust their production focus towards gasoline while reducing output of diesel (Pirrong, 2014).

c. Pricing Strategy: Competitive and Profitable Pricing

Crack spreads are inherently correlated with the profitability of refining processes, establishing them as a standard for the determination of product pricing. Retail fuel pricing is influenced not solely by the fluctuations in crude oil prices but also by the prevailing refining margins, thereby rendering crack spreads as a pivotal element. An elevated crack spread indicates that refiners are positioned to enhance product prices, whereas a diminished spread may necessitate price modifications to sustain market competitiveness.

More calculated pricing strategies are essential to achieve a harmonious balance between profitability and market competitiveness.

Example: In the event that diesel crack spreads experience a significant increase attributable to supply limitations, refiners may opt to raise wholesale diesel prices to correspond with the augmented refining margins (Alquist & Kilian, 2010).

d. Hedging Strategy: Mitigating Margin Volatility

Given that the prices of crude oil and refined products do not consistently exhibit synchronous movement, reliance solely on crude oil price hedges may render refiners vulnerable to various risks. Crack spreads can be effectively hedged through the utilization of futures and options contracts, thereby enabling refiners to secure their margins irrespective of fluctuations in the market. Crack spread swaps, such as the acquisition of gasoline futures coupled with the sale of crude futures, serve to safeguard against the volatility inherent in refining margins. Robust risk management is achieved through the establishment of stable refining margins. A refiner anticipating a contraction in the crack spread may engage in hedging by selling gasoline futures while concurrently purchasing crude futures to mitigate potential financial detriment (Ederington, 1979).

Table 2: Review summary of Crack spread related papers

Title	Description	Technique	Limitations	Region of Study / Indian Implications	Reference
An empirical model comparison for valuing crack spread options	The paper investigates the pricing of crack spread options, comparing univariate and bivariate modelling approaches. It finds that the univariate model, which directly models the crack spread, outperforms the bivariate GARCH model in terms of pricing accuracy and computational efficiency. This conclusion is supported by empirical analysis of options traded on the New York Mercantile Exchange, revealing lower pricing errors for the univariate approach, especially when including maturity effects. The study suggests future research could explore the hedging performance of these models and extend findings to other spread options with potentially cointegrated underlying.	The paper investigates the pricing of crack spread options using a bivariate GARCH volatility model for cointegrated underlyings . - It contrasts this with a simpler univariate modelling approach, found superior for option pricing performance . - The estimation approach employs implicit estimation under a risk-neutral pricing measure, minimizing discrepancies between observed and model prices . - Numerical optimization procedures are used to solve the minimization problem, representing an in-sample fitting exercise .	- Univariate modelling approaches neglect the correlation between underlying assets, affecting spread distribution . - The univariate model cannot capture negative spreads, limiting its applicability . - The B-NGC-M model may be biased due to unaccounted factors, leading to higher pricing errors . - High parameterization in the B-NGC-M model induces estimation errors, impacting its reliability . - Univariate modelling results in a loss of informational content regarding Greeks, affecting risk management. - The assumption of log-normal distribution may not hold for all types of spreads.	US. Indian Downstream industries use a different basket for computation of profitability.	(Mahringer et al, 2015)

Title	Description	Technique	Limitations	Region of Study / Indian Implications	Reference
What can we learn from the history of gasoline crack spreads?	- The paper investigates the long-range auto-correlations of gasoline crack spreads using a nonparametric method called detrended moving average (MF-DMA). - It finds that these auto-correlations exhibit multiscaling behaviors dominated by anti-persistence, indicating mean-reversion in the long-term. - The study reveals that extreme events can increase anti-persistence and multifractality, negatively impacting market efficiency. - The results suggest that while ARFIMA-GARCH models capture large fluctuations, they are misspecified for small fluctuations, leading to low predictability. - The findings highlight the inefficiency of the crude oil market, particularly in periods of high variability.	- The paper employs a nonparametric method called detrended moving average (MF-DMA) to investigate long-range auto-correlations of crack spreads . - It utilizes a technique of rolling windows to analyze extreme events affecting anti-persistence and multifractality . - The study also mentions the use of ARFIMA-GARCH models to capture dynamics of large fluctuations in crack spreads .	- The ARFIMA model cannot outperform the random walk model in out-of-sample prediction despite capturing long-range auto-correlations . - The complexity of the time series leads to larger prediction bias in the ARFIMA model . - ARFIMA-GARCH models are misspecified for small fluctuations, limiting their effectiveness . - The strong long-range auto-correlated behaviors do not guarantee better predictive performance of the ARFIMA model . - Market complexity contributes to the low predictability of the ARFIMA model .	US. The models studied ARFIMA-GARCH is ineffective in periods of high volatility.	(Wang et al, 2012)

Title	Description	Technique	Limitations	Region of Study / Indian Implications	Reference
Forecasting Naphtha Price Crack Using Multiple Regression Analysis	- The paper focuses on forecasting naphtha price crack, which is the price difference between naphtha and crude oil, using multiple regression analysis. - It identifies three major factors influencing naphtha prices: supply and demand, margin, and naphtha substitute. - The proposed model explains 65.1% of the variations in naphtha crack, highlighting the significant impact of supply and demand. - The study aims to provide more accurate predictions to minimize losses from price fluctuations, especially amid increasing uncertainty due to geopolitical factors and rising demand.	- The technique employed in this study is multiple regression analysis, which is utilized to forecast naphtha crack, defined as the price difference between naphtha and crude oil. - The model incorporates three major factors: supply and demand of naphtha, margin, and naphtha substitute, which are identified as having significant correlations with naphtha crack. - The regression coefficients are derived from unstandardized coefficients, allowing for the analysis of actual and predicted variations in naphtha crack over time. - The model aims to enhance prediction accuracy and minimize losses associated with naphtha price fluctuations.	The provided contexts do not mention any specific limitations of the study. Therefore, there is no information available regarding the limitations discussed in the paper.	- The study focuses on the naphtha market, particularly in Asia, as it emphasizes the supply and demand dynamics of naphtha in that region. - The fluctuations in naphtha prices are linked to political unrest in Middle East Asia and the growing demand in developing countries. - The proposed model for forecasting naphtha crack can be extended to other downstream chemicals, indicating a broader relevance to the Asian market.	(Sung et al, 2012)

Title	Description	Technique	Limitations	Region of Study / Indian Implications	Reference
Volatility Dynamics and Seasonality in Energy Prices - Implications for Crack-Spread Price Risk	- The paper investigates the dynamics of commodity prices, specifically focusing on heating oil, using a multivariate GARCH(1,1) model with dynamic conditional correlation (DCC) to analyze price volatility. - It highlights the impact of seasonal demand and supply shocks on prices, particularly during the winter peak-demand season when inventory is low. - The study also discusses the regulatory requirements affecting commodity specifications and their influence on production and storage decisions by refiners. - Overall, the findings emphasize the importance of short-term shocks and inventory management in understanding price fluctuations in heating oil markets.	- The POTS model is a factor model used for commodity price dynamics, similar to Schwartz-type models. - A continuous-time process for the spot price is specified, incorporating the absence of arbitrage. - The model captures features of petroleum commodity markets, focusing on volatility and co-variability. - The analysis includes crack-spread volatility, indicating potential daily effectiveness in hedging price risk. - The model reveals substantial price risk at a one-year horizon for crack-spread positions.	- The estimated mean parameters were very small and had no effect on the results presented in the paper . - Firms cannot borrow inventory from the future during market shortages, creating a limitation in inter-temporal price linkage . - Refiners have limited flexibility in product extraction due to the inability to expand coker or downstream processing capacity . Notably, these limitations impact the analysis of seasonality in petroleum product demand and its implications for volatility dynamics.	US. The study is centered around heating oil and its nuances, moreover from a hedging perspective.	(Suenaga, 2011)

Title	Description	Technique	Limitations	Region of Study / Indian Implications	Reference
Modelling and Trading the Gasoline Crack Spread – A Non Linear Story	- The paper investigates the gasoline crack spread, focusing on its non-linear characteristics and the implications for market participants. - It employs non-linear cointegration methods and various neural network architectures to forecast the spread, comparing them against a fair value model. - Results indicate that the spread exhibits asymmetric adjustment, with larger movements on the downside. - The higher order neural network with a threshold filter emerged as the most effective trading model, offering superior risk-return profiles. - The findings suggest that consumers have been receiving a fair deal over the analyzed period.	- The research employs non-linear cointegration techniques developed by Enders and Granger to analyze the gasoline crack spread time series. This method allows for the testing of asymmetric adjustments in the spread, which is crucial for understanding market dynamics. The study utilizes various neural network architectures, including multilayer perceptron, recurrent neural networks, and higher order neural networks, to forecast the spread. These models are benchmarked against a fair value non-linear cointegration model to evaluate their predictive performance. The results indicate significant asymmetry in the spread's movements, with larger adjustments occurring on the downside.	- Recurrent Neural Networks (RNNs) require more connections and memory, leading to increased computational time compared to standard backpropagation networks . - The implementation complexity of RNNs is noted as a minor disadvantage .	US. The downward and upward price movement of cointegrated products crude and gasoline have little implications in Indian context where the price of Gasoline and Gas Oil is artificially regulated.	(Dunis et al, 2006)

2.5 Research Gap

The summary of related research is tabulated in depicted in **Figure 14** below:

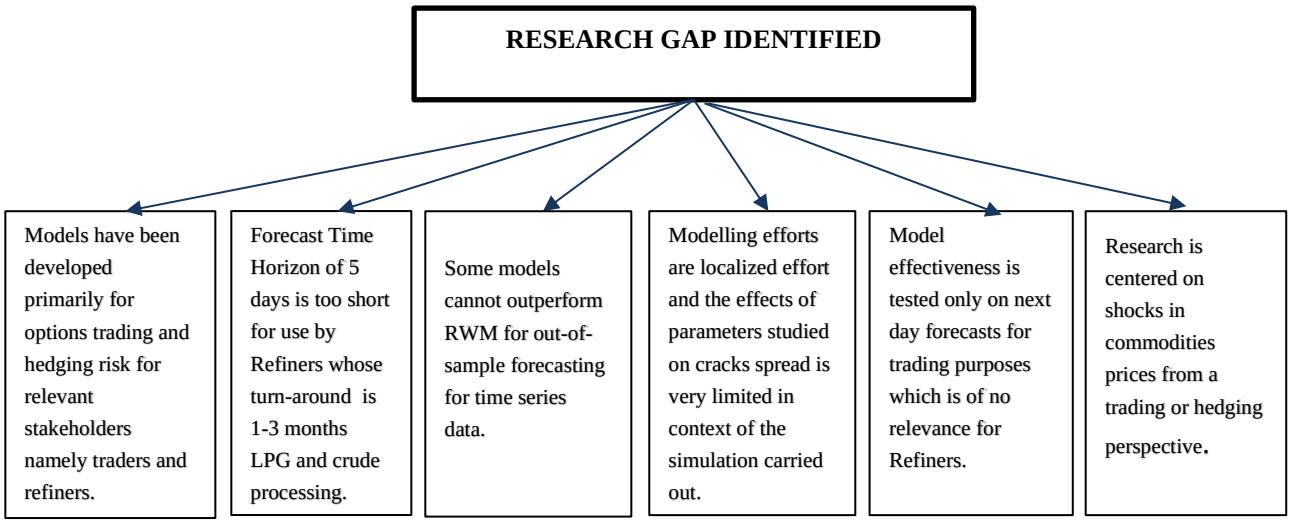


Figure 14: Research Gap Analysis

The research gaps identified in the earlier sections can therefore be enumerated as below:

i. Whereas, crude price forecasting is a variedly studied area, very few models exist for crack spread forecasting
ii. Crack spread forecast models that exist have been developed for a very short time horizon (5 days) which are of little relevance to oil
iii. The models that exist have been developed are in the context traders and are of little use to refiners and that too for trading and hedging purposes.

The objective of this research is therefore to address this gap by

i. Developing a model for forecasting price of single product crack spread using time series data and recent advances in AI Deep Learning techniques
ii. Developing models with forecast horizon of 2-3 months using crude and product benchmarks used by Indian refiners to be of use relevance to the Indian context.

2.6 LPG Demand Forecasting

Consumption of LPG in India has been on a steady rise, with 27.6 million metric tonnes consumed in the fiscal year 2020-21, marking a 5% increase from the prior year. This upward trend is projected to continue, thanks to supportive government initiatives. As per the study done by (PPAC, 2022), the number of active LPG domestic customers in India has more than doubled from 14.9 crores in 2015 to 30.5 crores as of April 2022. The number of LPG distributors has also significantly increased, with 25269 in April 2022, compared to 9686 in 2010 and 15930 in 2015. Moreover, there are now 202 LPG bottling plants in India, leading to a 99.8% LPG coverage in 2021, compared to just 56.2% in 2015. The demand for clean cooking fuels is driving India's LPG market, especially among rural and urban households. The Pradhan Mantri Ujjwala Yojana (PMUY) beneficiaries scheme implemented by the government is expected to remain a key driver of LPG market growth in India. In conclusion, the increase in LPG demand, coupled with government initiatives, is forecasted to fuel market growth and lead to significant developments in the foreseeable future.

In urban and semi-urban areas of India, LPG is the most commonly used fuel for cooking. However, the demand for LPG exceeds what domestic refiners produce. To meet this demand, Oil Marketing Companies (OMCs) import bulk LPG and distribute it across the country through their network of LPG bottling plants and distributors. The amount of LPG imported is determined based on projected demands, which can be cyclical. Accurate LPG demand forecasts are therefore crucial for effective supply chain planning, especially given the long lead times involved (Agarwal et al., 2021).

OMCs in India source bulk LPG from their refineries or other OMCs for bottling at their plants. If neither of these options is available, OMCs obtain bottled LPG directly from other OMCs. LPG is packaged into 14.2 Kg cylinders for domestic use and 19 Kg cylinders for commercial use. OMCs use a secondary distribution model by supplying domestic-use cylinders to distributors responsible for selling and distributing LPG to households. Commercial-use LPG accounts for less than 8% of total LPG sales by OMCs (Petroleum Planning and Analysis Cell (PPAC), MoPNG, 2021).

Due to insufficient domestic production, India relies heavily on LPG imports to meet demand. To make effective import decisions, it is essential to accurately estimate the demand shortfall. This

forecasted demand can then be fed to an optimisation model which can optimally plan the import requirement and schedule them over a period. This model can take the predicted demands and plan out the supply of LPG to the vast distributor network.

Studies in the past have attempted to forecast the Demand for LPG using Time Series modelling (Lopes et al., 2017). A recent study in Indonesia employed Auto-Regressive Forecasting Technique for estimating the monthly household LPG demand in Indonesia. (Nurochman & Moeis, 2021). In case of a recent study in Cameroon to determine the household LPG demand, it was found that price, income and urbanization are the important factors determining the demand (Sapnken et al., 2023) .

This use case for LPG demand forecasting attempts to develop, compare and study a forecasting model which can be of assistance to demand planners and supply chain managers.

When it comes to forecasting future demand, there are two primary methods to consider: Quantitative and Qualitative (Caniato et al., 2011). Quantitative methods rely on historical data to make predictions, and while they tend to be more precise, they can also be more challenging to utilize. Some examples of quantitative methods include trend projection, time series analysis, and econometric models. In trend projection, it is assumed that past patterns will persist into the future. While this method is straightforward, it may not be reliable if there are sudden shifts in demand. Another option is time series analysis, which divides historical data into trend, seasonal, and random components. This approach is more sophisticated than trend projection and is particularly valuable for estimating demand for products with seasonal fluctuations. The most advanced method for forecasting demand is econometric modelling, which employs statistical methods to link demand to economic factors, prices, and competitive activity. However, this approach is also the most complex and time-consuming to develop and put into practice.

Forecasting future demand using Qualitative methods, on the other hand, can be achieved through three common methods: seeking expert opinion, conducting market research, and using the Delphi method (Ebrahimi Meimand et al., 2012). While expert opinion is a quick and simple approach, it may not always be entirely objective or trustworthy. Market research involves surveying customers or potential customers, and while it can be more reliable than expert opinion, it can be time-consuming and costly. The Delphi method, on the other hand, involves a group of experts who refine their forecasts based on feedback from each other, making it a more dependable method

than expert opinion. However, it can also be a time-consuming and expensive process. While they may be less precise, they are typically faster and easier to use.

Newer techniques like machine learning and anticipatory forecasting while offering greater accuracy, can also be more complicated and expensive. An effective approach for forecasting future demand is utilizing machine learning algorithms that analyse historical data. However, this method can be intricate and require a significant investment of time. An alternative technique is anticipatory forecasting that employs real-time data to predict demand. Although more costly than conventional methods, it can offer exceptional accuracy. The selection of a forecasting technique depends on various factors namely the context of the forecast, the relevance and availability of historical data, the degree of accuracy desired (Moroff et al., 2021; Niknam et al., 2022; Thalor et al., 2021; Thomasson, 2017) by the researcher undertaking the study.

Chapter 3

Framework for Predictive Analytics and Modelling

"Just as electricity transformed almost everything 100 years ago, today I actually have a hard time thinking of an industry that I don't think AI will transform in the next several years."

- Andrew Ng, Computer Scientist

3.1 Artificial Intelligence and Machine Learning

Artificial intelligence is a branch of computer science which studies ways to build intelligent programs and machines that can creatively solve problems, which humans can solve innately.

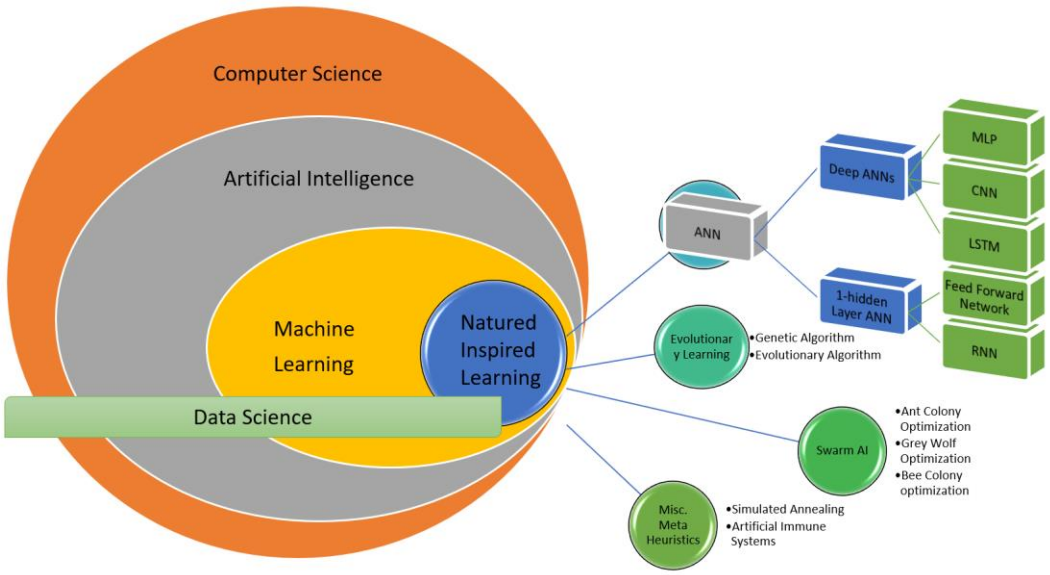


Figure 15 : Visualization of relationship between AI, ML and Deep Learning ANNs

Machine learning is a subset of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. In ML, there are different algorithms (e.g. neural networks) that help to solve problems. There are various

types of machine learning namely 1) Supervised Learning 2) Semi Supervised Learning 3) Re-enforcement Learning and 4) Unsupervised Learning (**Figure 15**).

Deep learning, or Deep Neural Learning (Deep ANN), is a subset of machine learning, which uses the deep embedded neural networks to analyze different factors with a structure that is similar to the human neural system. Some of the applications of Deep Learning are in Image Recognition, Natural Language processing, Portfolio management and Prediction of Stock and price movements, Drug discovery, Speech recognition, Robots and self-driving cars.

The characterization of Top machine learning algorithms is provided in **Table *3***.

Table 3 : Characterization of Top Machine Learning Algorithms

Machine Learning Method	**Description**	**Use Cases**	**Pros**	**Cons**
Linear Regression (Fahrmeir et al, 2021)	A statistical method to model the relationship between a dependent variable and one or more independent variables using a linear equation.	Predicting Sales estimates, Assessment of risk in financial services and insurance, evaluation of trends.	Simple to implement, interpretable, efficient for small datasets.	Assumes linearity, sensitive to outliers, not suitable for complex relationships.
Logistic Regression (Bertsimas et al, 2017)	It is a supervised machine learning algorithm which estimated the probability of an event occurring based on dependent variables.	Predicting the risk of developing a disease, predicting the failure of a process, system or product, predicting customer's propensity to purchase a product or discontinuation of subscription.	Easy and simple to implement, unlikely to overfit, model flexibility, fast.	Assumption of linearity, can only be used to predict discrete functions, cannot be used for solving complex relationship problems
Decision Tree (Rokach et al, 2010)	It is a supervised machine learning algorithm for classification and	In finance for forecasting future outcomes, Product	Easy to implement and interpret, can work with	It tends to overfit, inadequate for regression, takes

Machine Learning Method	Description	Use Cases	Pros	Cons
	prediction. A decision tree is a tree, where each internal node represents a test on a feature, each branch represents an outcome of each test, and each leaf node represents the class label.	planning, Loan approval, Customer willingness to buy a product, customer segmentation and fraud detection.	numerical and categorical features, requires little data preprocessing, feature selection happens automatically.	higher time to train the model.
Support Vector Machine (Cortes et al, 2011)	It is a supervised learning algorithm which is used for classification as well as regression problems. The SVM create the best decision boundary that can separate n-dimensional space into different classes.	Text categorization, classification of images, classification of satellite data, hand-written character recognition, biological science for protein classification.	Effective in high dimensional spaces, memory efficient, works with clear margin of separation.	It doesn't perform well when dataset is large, Prone to noise.
Naïve Bayes (Vikramkumar, 2014)	It is a supervised machine learning algorithm which is based on Bayes theorem and used for solving classification problems.	Spam filtration, Sentimental analysis, classifying articles, collaborative filtering and naïve bayes is used for building recommendation engine.	Fast, can solve multi-class prediction problems, performs exceptionally well with categorical input variables.	Zero frequency problem (if test data has a categorical variable of a category not present in training data set it will assign it zero probability), assumption of all features as independent.
KNN (Kramer, 2013)	It is a supervised learning algorithm used for classification and regression problems. It works by finding the distance between a query point and all the points in the data selecting the	Forecasting stock market, bank customer profiling, identify the risk factors for diseases, recommendation systems.	Simple to implement, can learn non-linear decision boundaries, constantly evolve with new data, single hyperparameter.	High time complexity for large datasets, assumes equal importance to all features, sensitive to outliers.

Machine Learning Method	Description	Use Cases	Pros	Cons
	specified number(K) closest to the query, then votes for the most frequent label (in the case of classification) or averages the labels (in the case of regression).			
KMeans (Ahmed et al, 2020)	It is an unsupervised learning algorithm that is used to solve the clustering problems. It groups the unlabeled dataset into pre-defined K clusters based on intra cluster distance and inter-cluster distance. It is a centroid-based algorithm.	Document classification, delivery optimization, identifying crime localities, customer segmentation, insurance fraud detection, call record detail analysis.	Simple to implement, scales to large data sets, guarantee convergence, generalizes to clusters of different sizes and shapes	Choosing k manually, dependent on initial values, clustering outliers, scaling with number of dimensions
Random Forest (Liu et al, 2012)	It is a supervised learning algorithm used for classification and regression problems. It builds decision trees on different samples and take the majority votes for classification and average vote in case of regression.	Fraud detection, detect customers more likely to repay their debt on time, predicting stock's future behavior, identifying diseases in patient, predicting customer behavior.	Solves the problem of overfitting by taking majority vote, robust to outliers, immune to curse of dimensionality, better accuracy.	Highly complex, high time complexity

3.2 Framework for forecasting using Statistical and Machine Learning

The law of large numbers is a theorem from probability and statistics that suggests that the average result from repeating an experiment multiple times will better approximate the true or expected underlying result. The field of machine learning is concerned with the question of how to construct computer programs that automatically improve with experience.

The 5 basic steps (**Figure 16**) in a forecasting task are summarized by Hyndman and Athanasopoulos in their book Forecasting: principles and practice. These steps are:

1. Problem Definition.

The careful consideration of who requires the forecast and how the forecast will be used. This is described as the most difficult part of the process, most likely because it is entirely problem specific and subjective.

2. Gathering Information.

The collection of historical data to analyze and model. This also includes getting access to domain experts and gathering information that can help to best interpret the historical information, and ultimately the forecasts that will be made.

3. Preliminary Exploratory Analysis.

The use of simple tools, like graphing and summary statistics, to better understand the data. Review plots and summarize and note obvious temporal structures, like trends seasonality, anomalies like missing data, corruption, and outliers, and any other structures that may impact forecasting.

4. Choosing and Fitting Models.

Evaluate two, three, or a suite of models of varying types on the problem. Models may be chosen for evaluation based on the assumptions they make.

5. Using and Evaluating a Forecasting Model.

The model is used to make forecasts and the performance of those forecasts is evaluated and skill of the models estimated. This may involve back-testing with historical data or waiting for new observations to become available for comparison.

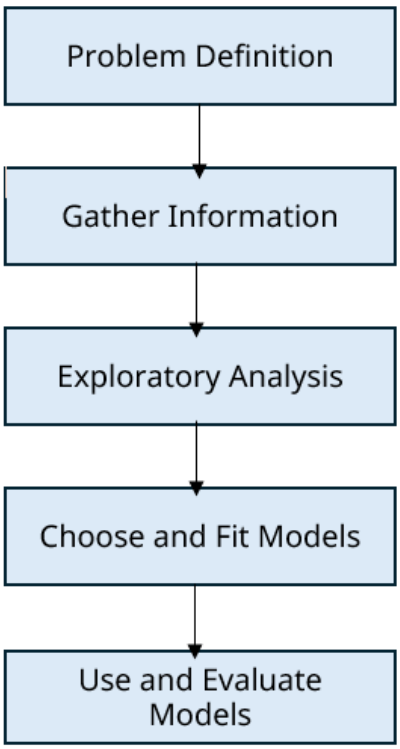


Figure 16 : Forecasting methodology using machine learning.

3.3 Framework for forecasting using Time Series Data

A time series data is one that has data which has time only dependent as the independent factor. A time series is usually modelled using a stochastic process Y(t) , which is a sequence of random variables. At time t, we are interested in forecasting Y(t+h), using information available at time 't' (Burba, 2019).

In order to make forecasts, it is important to ensure that the time series is stationary (Dickey & Fuller, 1979). A stationary time series is one whose properties do not depend on time at which the time series is observed (Zhang et al., 2008). A time series is said to be stationary if:

i. Mean (u) is constant over time
ii. Variance (U) is constant over time
iii. The covariance between two time periods (Yt) and (Yt+k) depends on the lag k and not on time t.

The conceptual framework for Time Series Forecasting is described as below:

1. Data Collection:

Involves gathering of historical time-ordered data, ensuring a consistent and regular time interval between observations.

2. Data Preprocessing:

Involves cleaning the data by handling missing values, outliers, and inconsistencies. It also involves converting time-related information into a datetime format.

3. Exploratory Data Analysis (EDA):

This step attempts to visualize the time series data to understand patterns, trends, and seasonality. Data is identified for any outliers or anomalies that may need special attention.

4. Decomposition:

This step aims to decompose the time series into its components, including trend, seasonality, and residual. This step helps in understanding the underlying patterns in the data.

5. Stationarity Check:

Involves ensuring that the time series is stationary, meaning its statistical properties do not change over time. Techniques such as differencing may be applied to achieve stationarity.

6. Model Selection:

This step involves choosing an appropriate Time Series model based on the characteristics observed during EDA. Common models include ARIMA (AutoRegressive Integrated Moving Average), SARIMA (Seasonal ARIMA), and more advanced models like Prophet or deep learning models.

7. Parameter Estimation:

Estimation of the parameters of the chosen model. This may involve selecting the order of autoregressive (AR) and moving average (MA) components.

8. Model Training:

Train the Time Series model using the historical data, considering a certain portion of the data for training and the rest for validation.

9. Model Validation:

Validate the trained model using the validation set to assess its accuracy and performance.

Metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), or Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) can be used.

10. Forecasting:

In this step, the trained and validated model is used to make future predictions or forecasts. Visualize the forecast along with the actual values for evaluation.

11. Model Fine-Tuning:

If necessary, model parameters are fine tuned based on the performance during validation. Additional features are considered and / or adjustments made to improve accuracy.

12. Scenario Analysis:

Perform scenario analysis to understand the impact of potential external factors on future forecasts. Adjust the model or forecasting approach accordingly.

13. Deployment:

Deploying the final model for making real-time predictions or incorporate it into decision-making processes.

3.4 Types of Time Series Model

Various types of Time Series Forecasting techniques are:

1. **The Naive Model**

The Naive Model is the simplest model to think of to make your predictions and is used as a baseline to compare all your other models. Usually, it means using the last observation as your forecast for the next period.". If dealing with seasonal data, one might predict this year's December sales by looking at last year's December sales. This approach is useful for evaluating whether the forecasting process adds value or if simple naive predictions are sufficient.

2. Seasonal Decomposition (+any model)

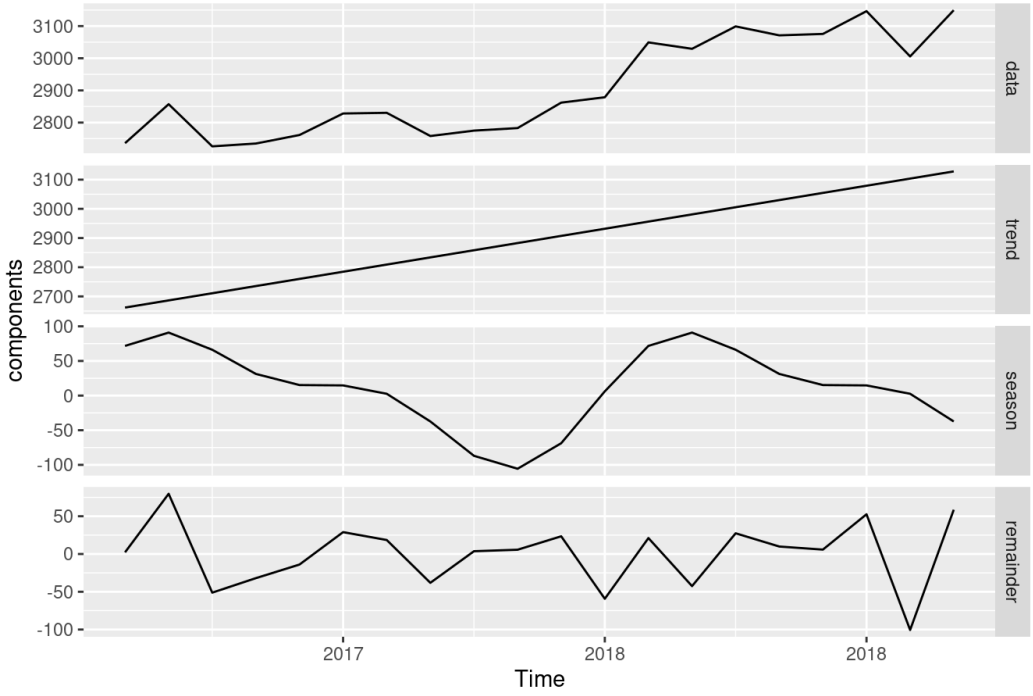


Figure 17: Decomposition of Time Series data into Trend, Seasonality and Residuals (Courtesy: Rob J Hyndman)

If there is some seasonality present in the data (daily, weekly, quarterly, yearly), the original time series data can be decomposed into the sum of three components :

$Y(t) = S(t) + T(t) + R(t)$

here

S(t) = Seasonal component

T(t) = Trend cycle component

R(t) = Remainder component

Classical decomposition technique consists of

i. Estimating trend T(t) through a rolling mean
ii. Computing S(t) as the average detrended series Y(t) - T(t) for each season
iii. Computing the remainder as R(t) = Y(t) -T(t) - S(t)t

Extensions to the classical decomposition technique allows us to:

i. handle non constant seasonality
ii. compute initial and last values of the decomposition
iii. avoid over smoothing effects

3. ARIMA, SARIMA

ARIMA stands for Auto-Regressive Integrated Moving Average. In an Auto-Regressive (AR) model, the forecast is a linear combination of past values of the variable. In a Moving Average (MA) model, the forecast is a linear combination of past forecast errors.

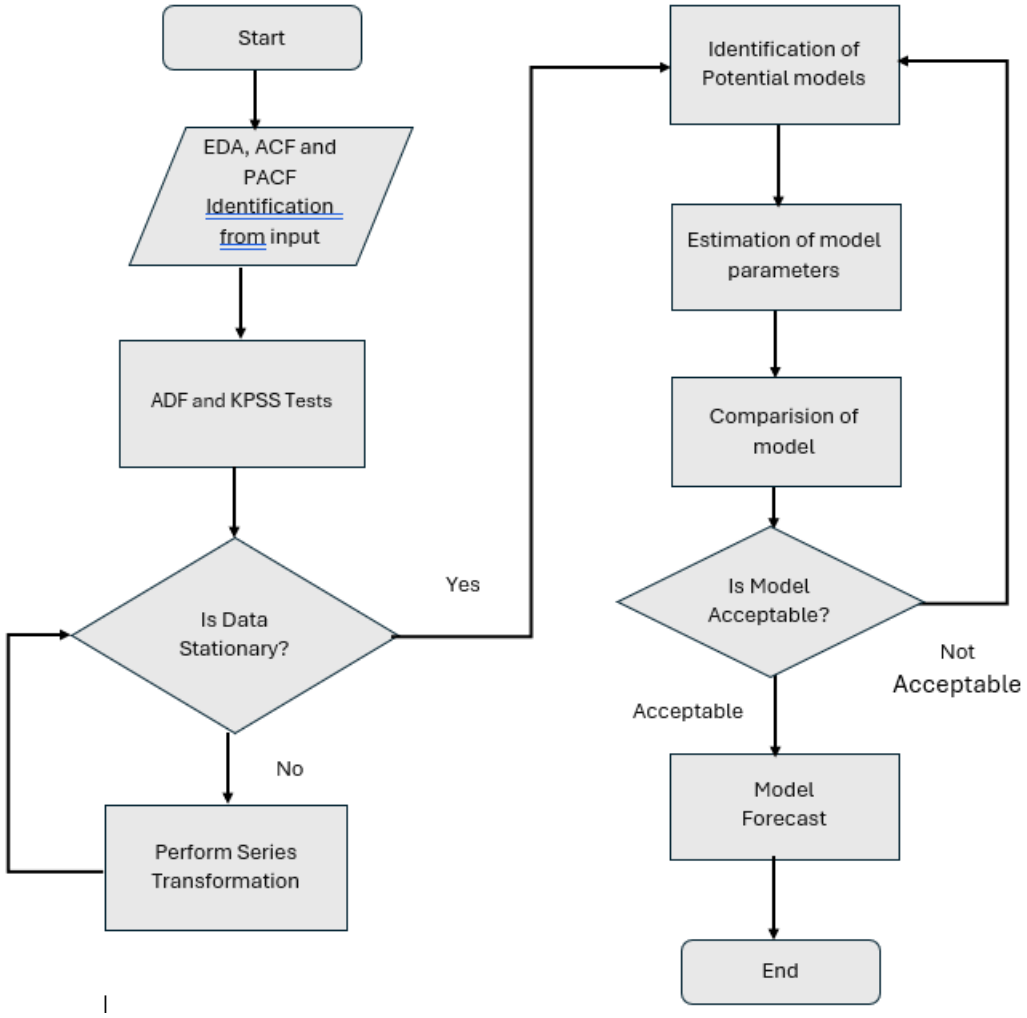


Figure 18: Forecasting using ARIMA

The ARIMA model is a combination of the two approaches. A prerequisite for forecasting time series data is that it should be stationary. An extension of the ARIMA model is the SARIMA model which extends the ARIMA model by adding a linear combination of seasonal past values.

Differencing is a technique where we consider the time series of differences instead of the original one.

Auto Regressive Integrated Moving Average (ARIMA) statistical models have been used for their reliability in forecasting and analysis (Ho, S.L. and Xiw, 1998). The time series technique is a forecasting technique that makes no assumptions about the data and is very flexible. The Box – Jenkins methodology (Box and Jenkins, 1971) summarized below is employed for modelling and generating forecasts based on ARIMA.

1. Removing Heteroscadacity in time series data (e.g. By applying box cox transformation, differencing or taking logarithm).
2. Ensuring the stationarity of Time series data (Mushtaq, 2011)
3. Identification – identify the ARIMA model using ACF and PACF plots. This would give us the p,d,q values for the forecasting model.
4. Estimation – Estimate the model parameters (using maximum likelihood)
5. Diagnostics – Check the residual for any issue such as not providing White Noise.

An improvement over the ARIMA model is the SARIMA model which includes more realistic dynamics of data specifically non-stationary in mean and seasonal behaviors(Gerolimetto, 2010). Since time series forecasting analysis can be conducted only on stationary data, differencing is carried out to convert non-stationary data to stationary data. If the series is still non-stationary data is subjected to order of differencing (d=0,1,2) to make it stationary (Solo, 1984). Heteroscedasticity refers to the spread of the data along the mean. Taking the logarithm of the data or differencing reduces the Heteroscedasticity. The traditional ARIMA model does not incorporate impact of external independent variables. Exogenous variables can be incorporated in the ARIMA and SARIMA models which will incorporate effects of any external variables which have exhibit regressive relationships to the base model (Zhu & Laptev, 2017). Auto Regressive Integrated Moving Average (ARIMA) models has been employed for reliability.

For selecting the best ARIMA model for generating forecasts, each of the candidate model needs to be evaluated using AIC and BIC scores (Kuha, 2004) which are explained below:

1. Akaike Information Criteria (AIC) score

   AIC = -2LL + 2m : where m is the number of variables estimated in the model

2. Bayesian Information Criteria (BIC) values

   BIC = -2LL +m ln(n) : where m is the number of variables estimated in the model and n

Before running any models, we should plot and analyze your time series data. If there is an evident trend or seasonality, exponential smoothing methods (ETS) are suitable because they explicitly model these components. In case we see autocorrelation in the data, i.e. the past explains the present, we go ahead for the ARIMA methodology. Plotting the partial autocorrelation function can help us with this.

4. ETS Models

Exponential Trend and Seasonality (ETS) smoothing was introduced towards the end of 1950s (Brown, 1959; Holt, 1957; Winters, 1960), and has been one of the most successful forecasting methods. This technique uses weighted averages of past observations, where weights decrease exponentially for older data, giving more importance to recent observations. This approach enables quick and reliable forecasts for various time series and is widely used in the industry.

In the ETS method, which is also known as the Holt-Winters seasonal method, there is the forecast equation and three smoothing equations — one for the level ℓt, one for the trend bt, and one for the seasonal component st, with corresponding smoothing parameters α, $\beta*$ and γ. We use m to denote the frequency of the seasonality, i.e., the number of seasons in a year. For example, for quarterly data m=4m, and for monthly data m=12.

There are two variations of this method that differ based on the seasonal component's nature. The additive method is used when seasonal variations are consistent throughout the

series, while the multiplicative method is preferred when these variations change proportionally with the series level. In the additive method, the seasonal component is expressed in absolute terms and subtracted from the series, summing to approximately zero annually. In the multiplicative method, the seasonal component is expressed as a percentage and divided by the series, summing to approximately one annually.

It's a common misconception that ARIMA models are more general than exponential smoothing models. While linear exponential smoothing models are indeed special cases of ARIMA models, non-linear exponential smoothing models do not have equivalent ARIMA counterparts. Conversely, many ARIMA models do not correspond to exponential smoothing models. Notably, all ETS models are non-stationary, whereas some ARIMA models are stationary.

ETS models featuring seasonality, a non-damped trend, or both require two levels of differencing to become stationary, indicating they have two unit roots. Other ETS models have one unit root and thus need only one level of differencing to achieve stationarity.

5. **GARCH**

Genaralised Autoregressive Conditional Heteroscadacity (GARCH) models owe their origin to the earlier assumption that the in a stochastic processes generating the time series is a homoscedastic process.

The GARCH process assumes that the variance of the error terms in a timeseries follows an auto-regressive ARMA process, therefore allowing it to change in time. Thus these models are very useful in scenarios where volatility is very high.

3.5 Sliding Window

The sliding window protocol (Zhan & Kim, 2024)employed in the domain of time-series forecasting—particularly in conjunction with deep learning architectures such as Long Short-Term Memory (LSTM) networks—represents a methodological framework for converting sequential data into a format amenable to effective model training. Rather than presenting the LSTM with the entirety of the time series in a singular instance, the series is partitioned into overlapping segments or "windows" of predetermined dimensions. Each window encapsulates a sequence of consecutive

time steps from the series, which are subsequently utilized as input to forecast the subsequent value(s) that directly follow the given window.

Let us consider a time series denoted as $[x_1, x_2, x_3, \ldots, x_t]$. Within the context of the sliding window protocol, one might select a window size of, for instance, 5. This specification implies that for the initial training sample, the input could be $[x_1, x_2, x_3, x_4, x_5]$ with the corresponding target being x_6. Following this, the window is advanced by one time step: thus, the subsequent input evolves to $[x_2, x_3, x_4, x_5, x_6]$ with the target being x_7, and this iterative process continues. This methodology effectively reframes the forecasting dilemma into a supervised learning challenge, whereby each sample grants the model insight into the recent historical context to project the imminent future.

This approach is characterized by several notable advantages:

i. **Local Context Capture:** Through the implementation of a fixed-size window, the model concentrates on a confined, recent segment of data, which frequently retains the most pertinent information for short-term predictions. This characteristic is especially advantageous for identifying local trends or seasonal variations.
ii. **Data Augmentation:** The utilization of overlapping windows facilitates the creation of a substantial number of training samples derived from a singular time series, which is particularly advantageous in scenarios where data availability is constrained.
iii. **Simplified Learning Task:** The transformation of the time series into input-output pairs is congruent with the supervised learning paradigm. This alignment permits models such as LSTMs to accommodate variable-length sequences (by preserving a fixed input size) and to discern how prior observations impact forthcoming outcomes.

However, the efficacy of the sliding window protocol is contingent upon the judicious selection of the window size. An inadequately small window may overlook critical long-term dependencies, whereas an excessively large window could introduce extraneous noise and result in overfitting. Typically, the optimal window size is ascertained through domain-specific expertise or via a process of hyperparameter optimization.

This methodology has demonstrated substantial utility across a multitude of applications, including stock market forecasting, meteorological predictions, and energy consumption modeling. In conjunction with LSTM networks, which are particularly adept at capturing temporal

dependencies, the sliding window protocol guarantees that the network processes data in manageable, focused segments, thereby enhancing the efficiency of the learning process and tailoring it to the specific task.

Furthermore, once the model has been trained utilizing these sequential windows, it can be employed with a similar sliding technique during the forecasting phase—continuously refreshing the input window with the most recent predictions to generate multi-step forecasts. This recursive application further capitalizes on the model's acquired relationships from historical segments to anticipate future observations.

3.6 Multistep Time Series Forecasting

In normal cases, the time series forecasting techniques are able to predict observations upto one step head. This methodology is called one-step forecasting.

However, in many scenarios there is a need to predict the values of a time series more than one step ahead. These types of problems are called as multi step time series forecasting scenarios.

The four commons strategies for performing multi-step forecasts are as below:

1. Direct Multi-step Forecast strategy
2. Recursive multi-step forecast strategy
3. Direct-recursive hybrid multi step forecast strategies
4. Multi step output forecast strategy

In the direct multi step forecast technique, a seperate model is generated for each forecast time. In the second method, the recursive multi step forecast method, one step method is recursive called upon to predict the forecast for the 2nd period using the forecast of the 1st period. One of the drawbacks of this method that the error of one stage gets accentuated at each step.

In the Direct recursive hybrid strategy, a seperate model is constructed for each step of the prediction and each models predictions is used as an input for the second model.

In the multiple output strategy model there is one model that is capable of making predictions for the entire time series in one go. These models are obviously more complex than the previous

models as they can learn the dependence structure between inputs and outputs and other outputs as well.

3.7 Walk-forward validation

Walk-forward validation, occasionally referred to as rolling forecast origin, represents an evaluative methodology specifically designed for time-series data, ensuring that the temporal integrity is upheld throughout the processes of model training and testing. This strategy is particularly vital for advanced deep learning methodologies such as Long Short-Term Memory networks (LSTMs) (Nguyen, n.d.) which necessitate datasets that retain the inherent sequential order of occurrences.

Steps in implementing walk-forward validation:

i. Initial Training: Commence with an initial dataset segment (for instance, the initial 70% of your time series) to facilitate the training of your model.
ii. Step-by-Step Forecasting: Employ the trained model to predict the subsequent time period (or a series of time steps). For example, in the context of daily forecasts, one might endeavor to predict the following day's values.
iii. Update the Training Set: Subsequent to the forecasting activity, integrate the actual observations from the predicted period into your training dataset. This newly augmented dataset now serves as the foundation for the ensuing training iteration.
iv. Slide Forward: Persist in this methodology by "walking" forward through the time series: training, forecasting, updating, and retraining. The model consistently predicts the next unobserved period based on all prior observed data.

Walk forward is particularly effective because of the following reasons:

i. Respects time order: In contrast to arbitrary partitions utilized in conventional k-fold cross-validation, walk-forward validation preserves the chronological order, thereby ensuring that the training data consistently precedes the testing (forecasting) data.
ii. Realistic Performance Evaluation: This approach emulates authentic forecasting situations where predictions are generated exclusively from historical data, subsequently refined as new information becomes accessible. It facilitates a more authentic evaluation of the model’s potential real-world performance.

iii. Adaptability: Through the continuous updating of the training dataset, the model becomes responsive to alterations in data patterns, including trends or shifts in seasonality. This adaptive characteristic is particularly advantageous when the foundational data evolves over time.

iv. Robust Error Analysis: Walk-forward validation offers the capability to monitor model performance across multiple forecasting intervals. This enables the identification of periods where the model may encounter difficulties, such as during abrupt changes or shifts within the dataset.

Walk-forward validation is especially advantageous in time series contexts as it not only assesses the forecasting capabilities of your model across diverse time segments but also simulates the conditions encountered during live forecasting tasks. This continual process of learning and updating mirrors the operational deployment of models in environments where data is received in a sequential manner.

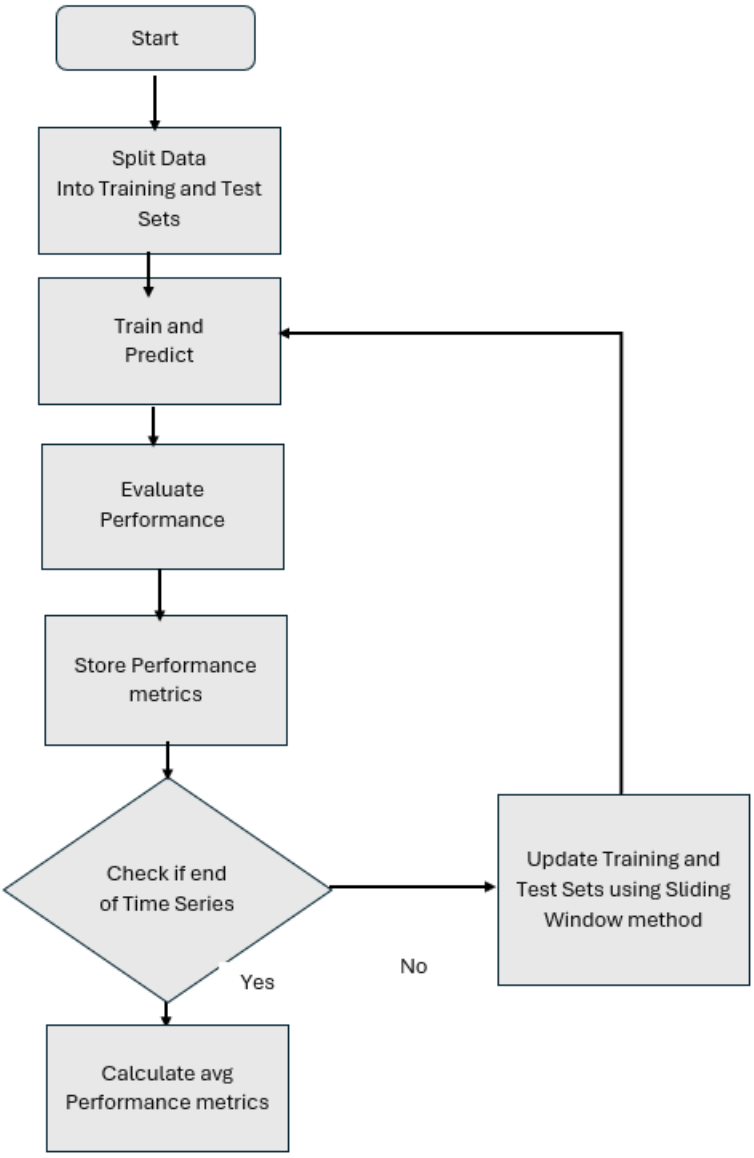


Figure 19: ARIMA forecasting using Walk-forward validation

A comparative study of various time series forecasting techniques is depicted in **Figure 20**.

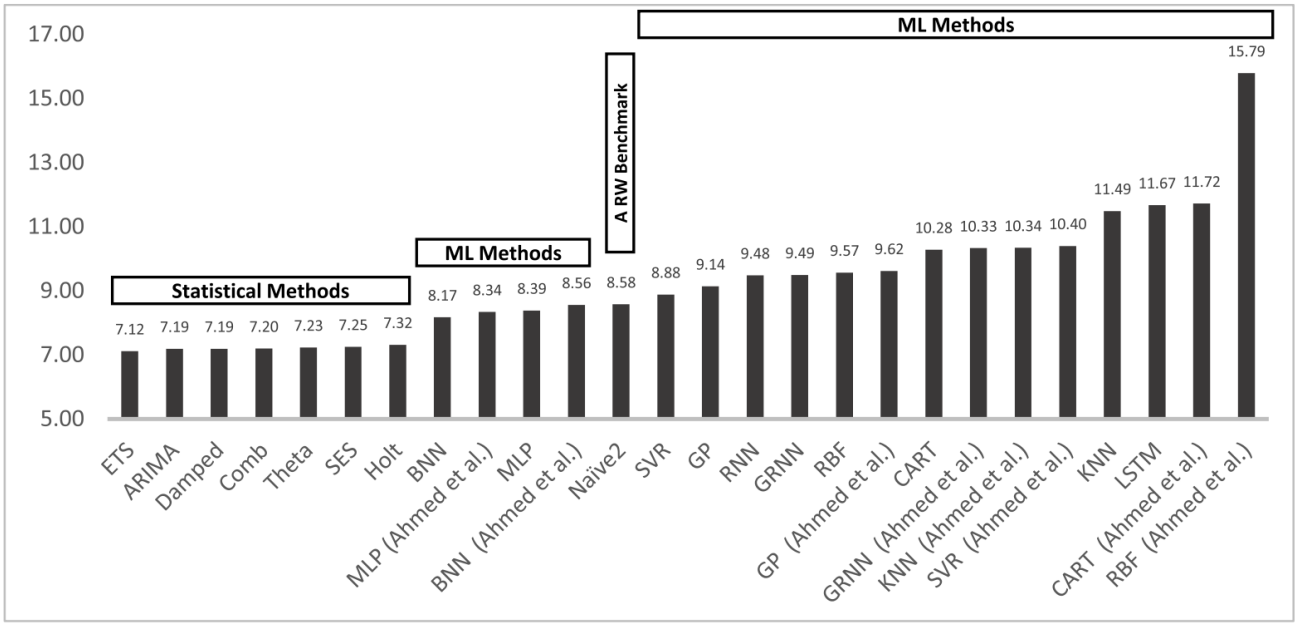


Figure 20 : Statistical Machine Learning forecasting methods: Concerns and way forward (Courtesy: Makridakis et al)

3.8 Test for Stationarity and Augmented Dickey Fuller (ADF) Test

Statistical tests depend on certain assumptions about the data and are used to determine whether to reject or retain a null hypothesis. Proper interpretation of the results is essential for their applicability to a specific problem. These tests offer quick verification and confirm he stationarity or non-stationarity of a time series. A unit root test, in particular, assesses the strength of a trend within a time series, providing valuable insight into its behavior.

Among the various tests for unit root, the Augmented Dickey-Fuller test is the most commonly used test. This test employs an autoregressive model and searches for the optimal lag values by optimizing an information criterion.

The null hypothesis of the ADF test states that the time series can be represented by a unit root, indicating non-stationarity (i.e., presence of time-dependent structure). Conversely, the alternative hypothesis (rejection of the null hypothesis) suggests that the time series is indeed stationary.

Null Hypothesis (H0): If the null hypothesis is not rejected, it implies that the time series does contain a unit root, indicating its non-stationarity and the presence of time-dependent structure.

Alternate Hypothesis (H1): The rejection of the null hypothesis indicates that the time series does not have a unit root, suggesting its stationarity and absence of time-dependent structure.

We derive the interpretation of this outcome based on the p-value obtained from the test. A p-value below a specified threshold (e.g., 5% or 1%) indicates rejection of the null hypothesis (stationarity), while a p-value exceeding the threshold suggests a failure to reject the null hypothesis (non-stationarity).

- **p-value > 0.05:** Failre to reject the null hypothesis (H0), the data has a unit root and is non-stationary.
- **p-value <= 0.05:** Able to reject the null hypothesis (H0), the data does not have a unit root and is stationary.

3.9 ANN and Deep Learning Models

Artificial Neural Network (ANN) are models of computational systems that is architected on the model of biological brains and neurons. They mimic the structures and functions of biological neural networks. A single unit representing a neuron is called as ‘Perceptron’ (**Figure 21**). These systems "learn" to perform tasks by considering examples, generally without being programmed with task-specific rules.

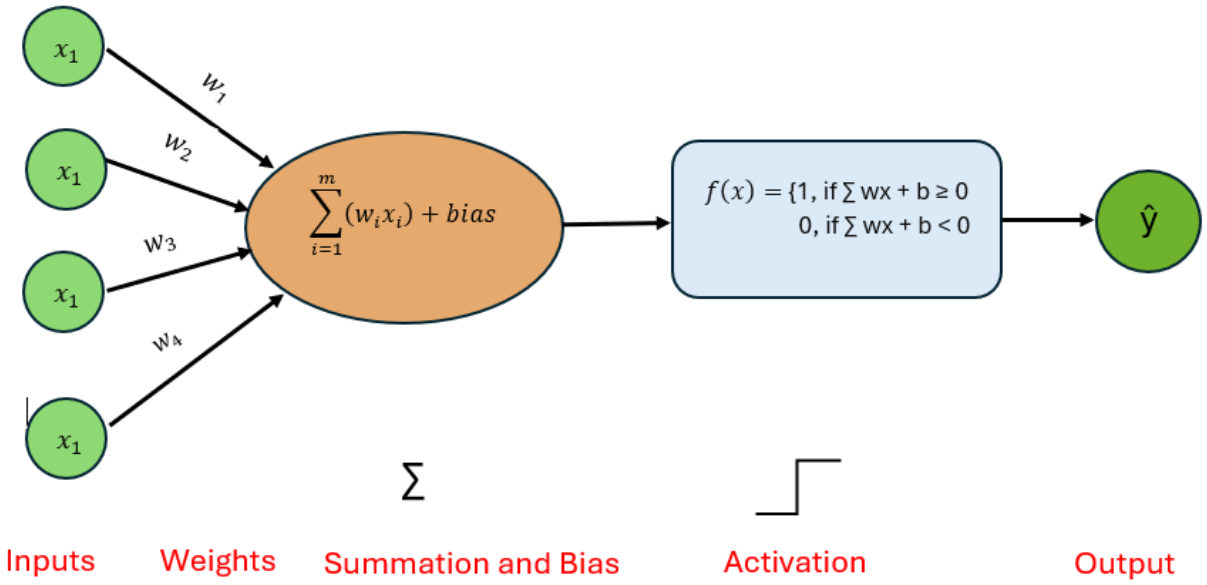


Figure 21 : Structure of a Perceptron (Courtesy: İlyurek Kılıç)

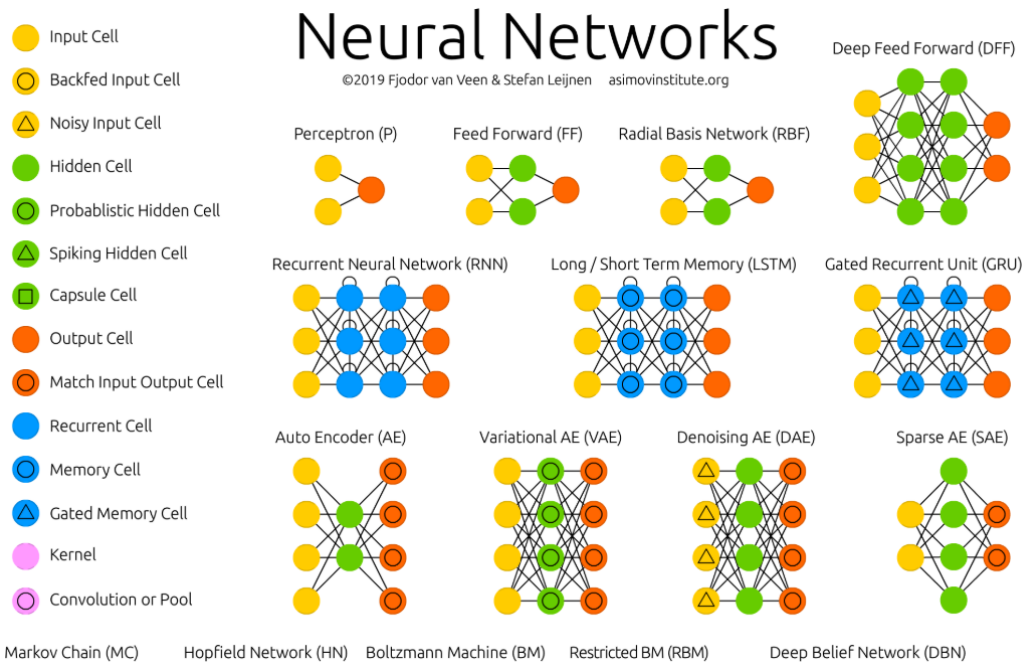


Figure 22 : Various types of Neural Networks (Courtesy: Asimov Institute)

There are various types of Neural Networks (**Figure 22**). The Deep Learning Artificial Neural Networks (ANN) can be broadly classified into three types:

1. **Multi-Layer Perceptron (MLP)**

   They are comprised of one or more layers of neurons. Data is fed to the input layer. The hidden layer which can range from 1 to many, provide levels of abstraction and predictions. The output is made available on the output layer, which is a visible layer. They are suitable for classification / prediction problems where inputs are assigned a label / class viz image data, text data, time series data etc.

2. **Convolutional Neural Network (CNN)**

   They are designed to map image data to an output variable. The CNN was developed for reading two-dimensional spatial-temporal data, but can be adapted for use with univariate time series forecasting

3. **Recurrent Neural Network (RNN)**

   They are designed to work with a sequence detection problems. The Long Short Term memory (LSTM) is the most successful RNN because it was able to overcomes the problem of training a RNN as in turn has been used on a wide range of applications such as Speech data recognition, classification and prediction problems, regression prediction problems and generative models.

As mentioned above, each type of network is particularly a special type of problems. MLP are suited for classification tasks. CNN are better suited towards image recognition and RNN are suitable for speech recognition and NLP. Long Short Term Memory is a type of RNN which is suitable for pattern recognition, both long term and short term

3.10 The Exploding Gradient Problem

The exploding gradient problem is a numerical instability issue that can occur during the training of artificial neural networks (ANNs), especially deep neural networks. It is a counterpart to the vanishing gradient problem and occurs when the gradients of the loss function with respect to the

network's parameters become extremely large. This can lead to numerical overflow, making the training process difficult or even impossible.

The primary cause of the exploding gradient problem is the chain of derivatives involved in backpropagation, the algorithm used to update the network's weights during training. In a deep network with many layers, the product of these derivatives can become very large, leading to exploding gradients. When this happens, weight updates become excessively large, causing the model's parameters to diverge instead of converging to optimal values.

The exploding gradient problem can manifest in various ways:

1. Numerical Overflow:

The weights and gradients become so large that they exceed the limits of numerical precision, causing overflow issues in the computer's representation of numbers.

2. Unstable Training:

The training process becomes highly unstable, and the model fails to converge to a solution. This makes it challenging to train deep networks effectively.

3. Divergent Behaviour:

The model’s parameters diverge rather than converge during training, leading to an unsuccessful learning process.

Mitigating the Exploding Gradient Problem:

1. Weight Initialization:

Properly initializing the weights can help mitigate the exploding gradient problem. Techniques such as Xavier/Glorot initialization are designed to maintain signal magnitudes during forward and backward passes.

2. Batch Normalization:

Batch normalization normalizes the inputs of each layer, which can help stabilize and reduce the risk of exploding gradients.

3. Gradient Clipping:

Gradient clipping involves setting a threshold for the gradient values during training. If the gradients exceed this threshold, they are scaled down to prevent them from becoming too large.

4. Learning Rate Adjustment:

Adjusting the learning rate can also be effective. Using adaptive learning rate methods or reducing the learning rate over time can help control the update step sizes.

5. Proper Activation Functions:

Choosing appropriate activation functions, such as Rectified Linear Unit (ReLU) or variants, can also contribute to mitigating the exploding gradient problem.

6. Normalization Techniques:

Techniques like layer normalization and weight normalization can be employed to stabilize the learning process and prevent gradient explosion.

3.11 The Vanishing Gradient Problem

The vanishing gradient problem is a challenge that can occur during the training of artificial neural networks (ANNs), especially deep neural networks. It is characterized by the gradients of the loss function with respect to the network's parameters becoming extremely small. As a result, the weights of the neural network fail to update effectively during the training process, hindering the learning of meaningful representations and making it difficult for the network to converge to an optimal solution.

The vanishing gradient problem is particularly pronounced in deep networks with many layers. During backpropagation, which is the process of updating the network's weights based on the calculated gradients, the gradients are successively multiplied as they are propagated backward through the layers. If these gradients are small, the product of these small values can become vanishingly tiny, approaching zero. Consequently, the weights of the early layers in the network receive minimal updates, and the network fails to learn useful features from the input data.

Key Characteristics and Effects:

i. Stalled Learning:

The weights of early layers in the network may stop learning altogether or learn at an extremely slow rate. This can result in slow convergence or even prevent convergence during training.

ii. Diminished Representational Power:

Layers that are too deep in the network may not effectively capture meaningful features or representations from the input data, leading to a degradation of the network's ability to generalize.

iii. Network Degradation:

The deeper layers may become practically inactive, essentially reducing the network to a shallower architecture. This limits the capacity of the network to model complex relationships in the data.

Mitigating the Vanishing Gradient Problem:

i. Weight Initialization:

Proper weight initialization can play a crucial role. Techniques like Xavier/Glorot initialization are designed to maintain signal magnitudes during forward and backward passes.

ii. Batch Normalization:

Batch normalization normalizes the inputs of each layer, reducing internal covariate shift and mitigating the vanishing gradient problem.

iii. Skip Connections and Residual Networks:

Architectures like skip connections or residual networks (ResNets) introduce shortcut connections that allow gradients to bypass certain layers during backpropagation, helping to address the vanishing gradient issue.

iv. Use of Non-Saturating Activation Functions:

Non-saturating activation functions, such as Rectified Linear Unit (ReLU) or variants, are less prone to saturation and help mitigate the vanishing gradient problem.

v. Gradient Clipping:

Gradient clipping involves setting a threshold for the gradient values. If the gradients exceed this threshold, they are scaled down, preventing them from becoming too small.

vi. Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) Networks:

For recurrent neural networks (RNNs), architectures like LSTMs and GRUs are designed to better capture long-range dependencies and mitigate the vanishing gradient problem.

The application of Deep Learning models for Time Series Forecasting along with their benefits are listed in **Table *4***.

Table 4 : Deep Learning Models - Application for Time Series Forecasting

Method	Benefits	Application Areas	Limitations
Traditional (Fan-Kung et al, 2025) (Keydana, 2020)	Simple architecture, fast training, effective for univariate time series	Stock price prediction, weather forecasting	Limited ability to capture temporal dependencies
MLP (Brownlee, 2020)	Can handle multivariate time series, flexible, good for short-term forecasting	Sales forecasting, energy consumption prediction	Requires careful data preparation, struggles with long-term dependencies
CNN (Brownlee, 2020)	Captures local patterns, effective for multivariate time series, scalable	Financial market prediction, anomaly detection	Requires large datasets, computationally intensive
RNN (Muqaddas, 2024)	Captures temporal dependencies, suitable for	Time Series prediction Natural Language processing, Speech	Prone to vanishing gradient problem, computationally expensive

Method	Benefits	Application Areas	Limitations
	sequential data, variants like LSTM/GRU	recognition, traffic forecasting	

3.12 Comparison of Statistical Machine and Deep Learning models

Having understood the architecture of Deep Learning models, the comparison of the Statistical machine learning models and Deep Learning models is summarized in **Table *5***.

Table 5 : Comparison of Statistical Machine Learning and Deep Learning Models

Characteristic	Statistical Machine Learning Model	Deep Learning Model
Definition	A statistical machine learning model is a mathematical model that is built using statistical methods to learn from data.	A deep learning model is a type of machine learning model that is inspired by the structure and function of the human brain.
Model structure	Statistical machine learning models are typically simple in structure, with a small number of parameters.	Deep learning models are typically complex in structure, with many layers of interconnected neurons.
Learning process	Statistical machine learning models learn by fitting a mathematical function to the data.	Deep learning models learn by iteratively adjusting the weights of the connections between neurons.
Prediction accuracy	Statistical machine learning models can achieve high prediction accuracy on some tasks, but they may not be able to capture complex relationships in the data.	Deep learning models can achieve state-of-the-art prediction accuracy on a wide range of tasks, including image classification, natural language processing, and machine translation.

Characteristic	Statistical Machine Learning Model	Deep Learning Model
Interpretability	Statistical machine learning models are typically more interpretable than deep learning models.	Deep learning models can be difficult to interpret, as the decisions they make are based on complex interactions between neurons.
Computational requirements	Statistical machine learning models are typically less computationally demanding than deep learning models.	Deep learning models can be very computationally demanding, especially for large datasets.
Applications	Statistical machine learning models are widely used in a variety of applications, including fraud detection, risk assessment, and customer segmentation.	Deep learning models are also widely used in a variety of applications, including image recognition, natural language processing, time series forecasting and machine translation.

3.13 LSTM Deep Learning Models

Long Short Term Memory (LSTM) (**Figure 23**), a type of RNN finds application in areas of Deep Learning. LSTM in particular can detect patterns both in short and long term in sequence of inputs and therefore well suited to classifying, processing and making predictions, like a sequence of words in a sentence, in time series data because we are unsure that patterns / lags (Gu et al., 2019; Salvi, 2019; Siami-Namini & Namin, 2018; Song et al., 2020).

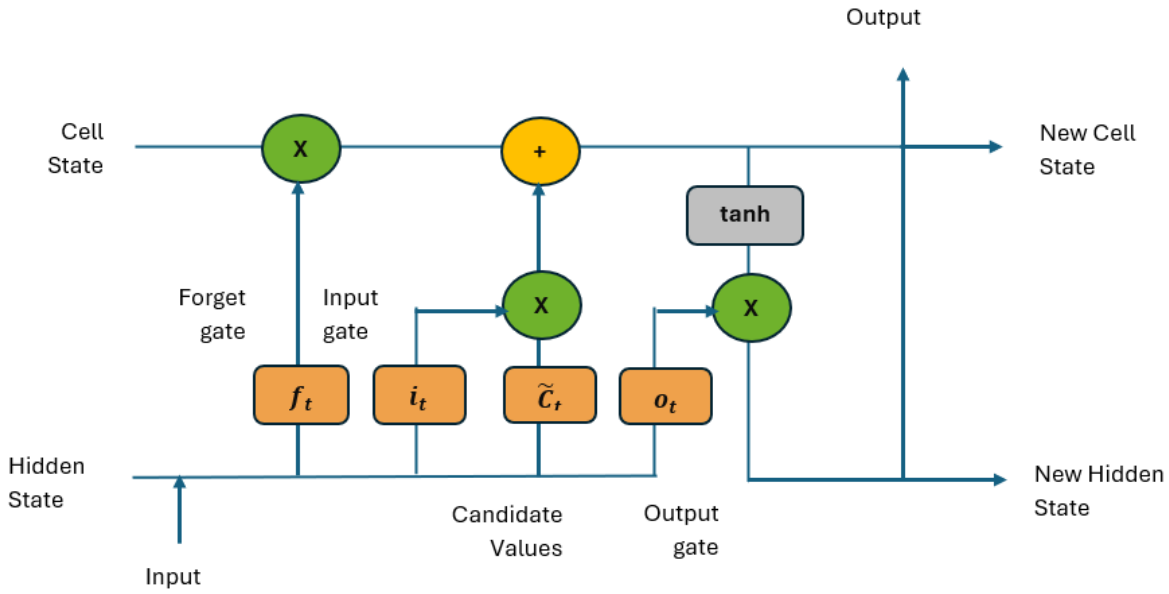


Figure 23 : Structure of LSTM Networks (Courtesy : Kulshrestha et al)

Long short term memory (LSTM) a type of RNN and finds application in areas of Deep Learning. LSTM in particular can detect patterns both in short and long term in sequence of inputs (**Figure 24**) and therefore well suited to classifying, processing and making predictions, like a sequence of words in a sentence, in time series data because we are unsure that patterns / lags.

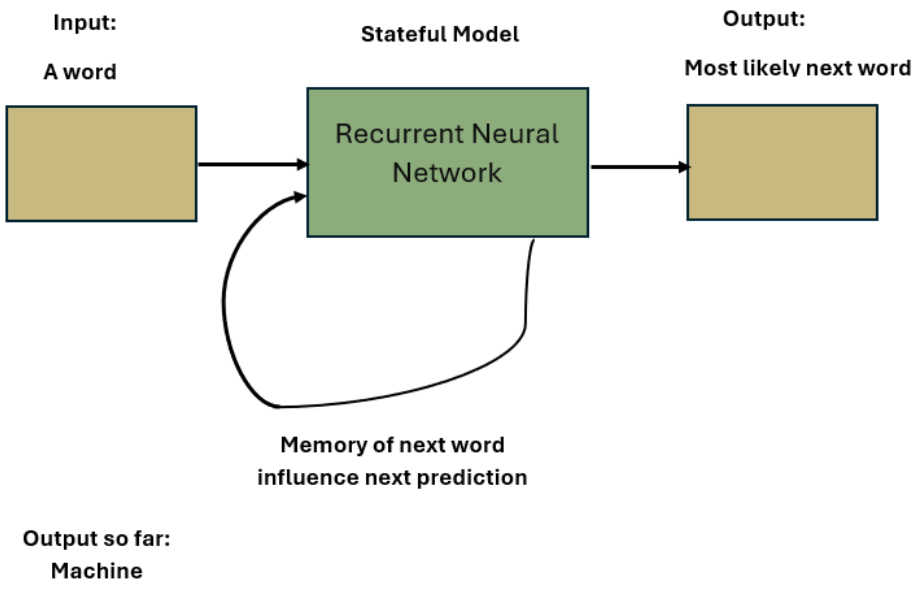


Figure 24 : LSTM for Short Sequence Sentences

LSTMS are able to do so because they are able to keep contextual information related to input sequences by providing a loop which enables information to flow from one stage to another. An LSTM unit is composed of a cell, an input gate, output gate and forget gates. In order to overcome the problem of vanishing gradient which occurs in RNN due to backpropagation, LSTMs were developed. They however still suffer from the exploding gradient problem.

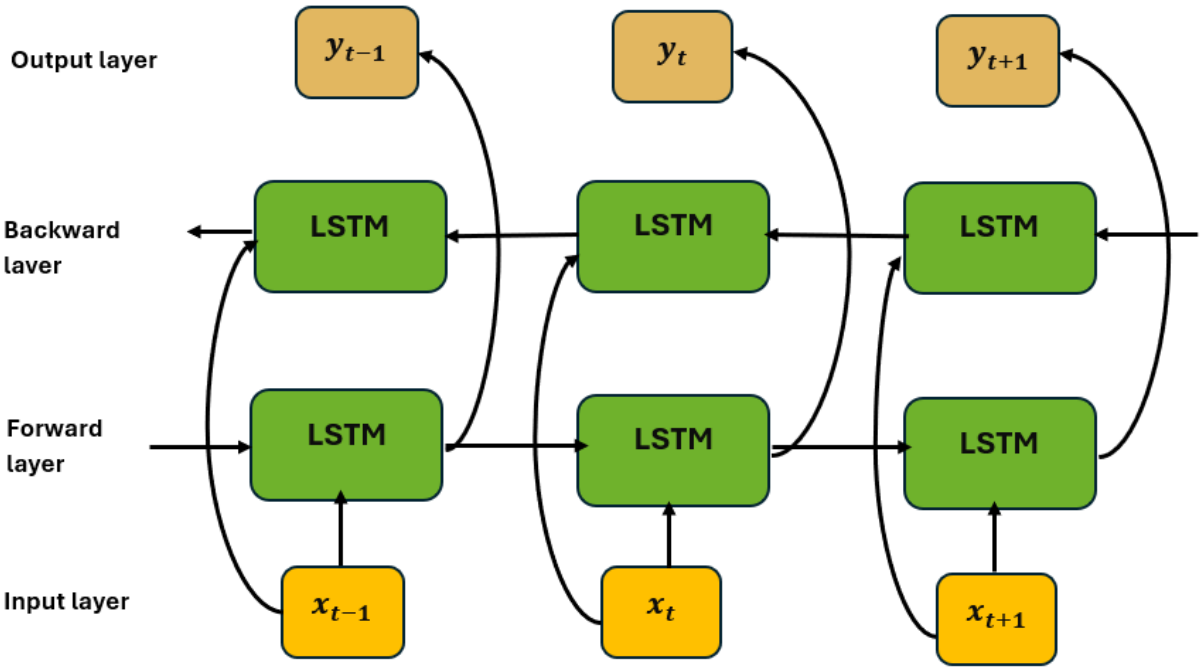


Figure 25 : Structure of Bidirectional LSTM (Adapted from Yildirim, 2018.)

ARIMA models yields **better** results in short term forecasting use cases, whereas **LSTM** yields **better** results for long term scenarios. Also, the number of times a model is trainied, known as "epoch" in deep learning, has no effect on the performance of the trained forecast model and it exhibits a truly random behavior.

LSTMs can be used to forecast **time-series**. **RNN's** (especially **LSTM's**) are known to be good at extracting patterns in input feature space, where the input data spans over long sequences. Owing to the gated architecture of **LSTM's** have the ability to manipulate its memory state, thus they are ideal for these types of problems

An extension of the traditional LSTMs is the Bi-directional LSTMs **Figure 25**. They improve performance on a sequence classification problems. They connect two hidden layers of opposite directions to the same output. In this form of degenerative learning, the information about the past and the future states are simultaneously passed on to the output layer. This helps the LSTM understand the context better (Kulshrestha et al., 2020). The various types LSTM are listed in **Table *6***.

Table 6 : Various types of LSTM

S.No.	Type of LSTM	Characteristic feature	Application use cases
1.	Vanilla LSTM	This type of LSTM has a single hidden layer of LSTM units, and an output layer used to make a prediction.	Sequence prediction
2.	Stacked LSTM	Multiple hidden LSTM layers can be stacked one on top of another in what is referred to as a Stacked LSTM model	Sequence prediction
3.	Bidirectional LSTM	LSTM model to learn the input sequence both forward and backwards and concatenate both interpretations.	Sequence prediction
4.	CNN LSTM	The CNN LSTM architecture involves using Convolutional Neural Network (CNN) layers for feature extraction on input data combined with LSTMs to support sequence prediction	Automatically extracting and learning features from one-dimensional sequence data such as univariate time series data
5.	ConvLSTM	This type of LSTM related to the CNN-LSTM is the ConvLSTM, where the convolutional reading of input is built directly into each LSTM unit.	This type of LSTM has a single hidden layer of LSTM units, and an output layer used to make a prediction.

3.14 Data Transformation (Regularization, Normalization and Optimization)

Normalization

Normalization is an approach which is applied during the preparation of data in order to change the values of numeric columns in a dataset to use a common scale when the features in the data

have different ranges. Normalization is a scaling technique in which values are shifted and rescaled so that they end up ranging between 0 and 1. It is also known as Min-Max scaling.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Where Xmax and Xmin are the maximum and the minimum values of the feature respectively.

- In case when the value of X is the minimum value in the column, the numerator will become 0, and hence X' is 0
- In the converse scenario, when the value of X is the maximum value in the column, then numerator will be equal to the denominator and thus the value of X' is 1
- In case the value of X is between the minimum and the maximum value, then the value of X' will fall between 0 and 1.

Standardization

Standardization is an alternative scaling method that centers values around the mean and adjusts them to have a unit standard deviation. This process ensures the attribute's mean is zero, resulting in a distribution with a standard deviation of one.

$$X' = \frac{X - \mu}{\sigma}$$

μ is the mean of the distribution and σ is the standard deviation of the distribution. Kindly note that in this case, the values in this case are not restricted to a particular range.

Regularization

Regularization is a method that introduces small adjustments to the learning algorithm to enhance the model's ability to generalize. This improvement ultimately leads to better performance when the model is applied to unseen data..

When to use Normalization and when to use Standardization?

- Normalization is beneficial when the data distribution deviates from a Gaussian distribution. This is particularly advantageous in algorithms such as K-Nearest Neighbors and Neural Networks, which do not make assumptions about the data distribution.
- Standardization, conversely, is advantageous when the data approximately follows a Gaussian distribution, though this assumption is not strictly required. Unlike normalization, standardization does not impose a specific range on the data. Therefore, outliers in the data remain unaffected by standardization.

Ultimately, the decision to apply normalization or standardization depends on your specific problem and the machine learning algorithm in use. There is no definitive rule dictating when to choose one over the other. It's often beneficial to experiment with both normalized and standardized data alongside raw data to determine which approach yields the best performance for your model.

A recommended practice is to fit the scaler using the training data and subsequently apply it to transform the testing data. This approach prevents any potential data leakage during the model evaluation phase. Additionally, scaling the target values is typically unnecessary.

3.15 Bayesian Optimization

In machine learning, hyperparameter optimization, or tuning, involves selecting the best set of hyperparameters for a learning algorithm. Hyperparameters are parameters that govern the learning process's behavior, unlike other parameters (typically node weights) that are learned from the data.

Model Hyperparameters therefore are those properties that govern the training process. They include variables that determine the network structure (for example, Number of Hidden Units) and the variables which determine how the network is trained (for example, Learning Rate), Number of Epochs, Activation Functions and the like. The benefits of using a good hyperparameters are:

1. Easy to manage a large set of experiments by hyper parameter tuning
2. Efficiency is search across the space of all possible hyperparameters

In machine learning / deep learning the process of finding the optimal set of parameter is called as hyper parameter optimization. Some of the most commonly used algorithms are:

1. Grid search
2. Random Search
3. Bayesian optimization

Grid Search is rather a brute force method which requires that two sets of hyperparameters are created, namely Learning Rate and Number of layers. The grid search algorithm trains for all combinations using two set of these hypermeters and measures the performance of these hyperparameters using cross-validation. Random Search uses probabilistic model for selecting a set of samples and arrives at the hyperparameter settings.

Bayesian optimization uses an automatic model tuning approach to fine tune the hyperparameters. This method uses a surrogate model, such as Gaussian Process (**Figure 26**) for approximating the objective function. The posterior distribution for this function are taken.

During this optimization, two choices are made:

1. Selection of prior functions that express assumptions about the function being optimized. Here Gaussian Process prior is chosen
2. The acquisition function must be chosen.

A Gaussian process defines the prior distribution over functions that can be converted into a posterior. Here, the following must be noted:

1. Gaussian process is then used as a prior Bayesian inference
2. Computing the posterior enables the Gaussian process to make predictions for the unseen test cases.

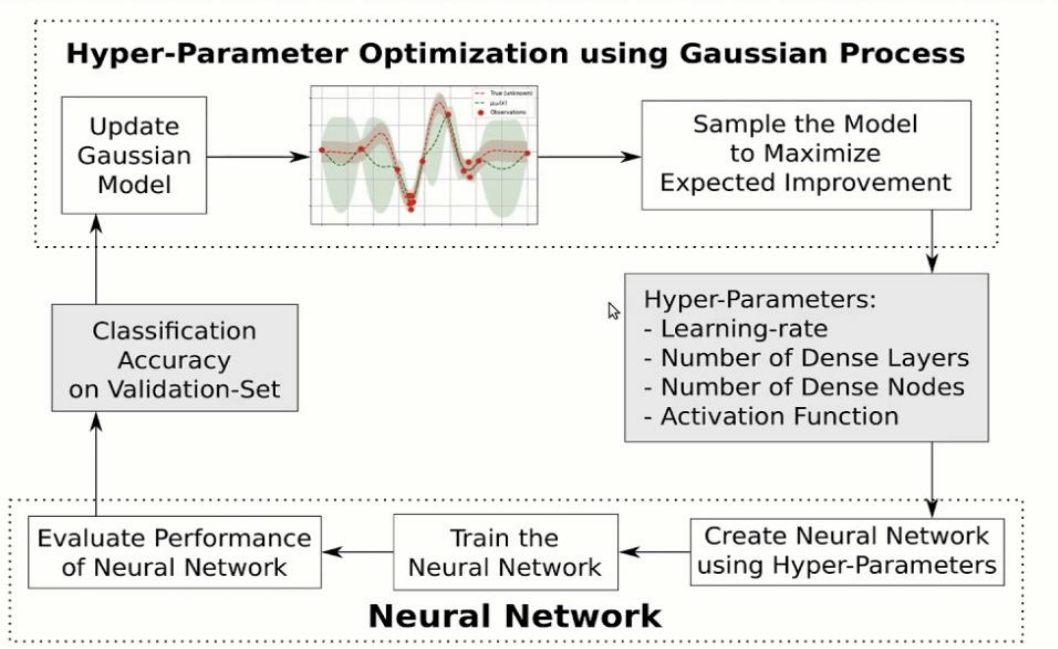


Figure 26 : Hyperparameter Optimization (Courtesy Edward Ortiz, Medium)

3.16 Explainable Artificial Intelligence (XAI)

Explainable AI (XAI) is the practice of designing and developing artificial intelligence systems in a way that allows humans to understand and interpret the decisions made by these systems. The goal of Explainable AI is to make AI algorithms and models more transparent, interpretable, and accountable. This is particularly important as AI systems become increasingly complex and are applied in critical areas such as healthcare, finance, and criminal justice.

Key aspects of Explainable AI include:

Interpretability: Making the internal workings of AI models understandable, ensuring that users, stakeholders, and decision-makers can comprehend how the model arrives at its predictions or decisions.

Transparency: Providing visibility into the factors, features, or data points that significantly influence the AI model's output. Understanding the model's decision-making process helps build trust in its results.

Fairness and Bias Mitigation: Addressing and mitigating biases in AI systems to ensure fair and equitable outcomes. Explainable AI helps identify potential biases and allows for adjustments to be made.

User-Friendly Explanations: Presenting explanations in a way that is accessible to users, even if they don't have a technical background. This promotes broader understanding and acceptance of AI technologies.

Contextual Explanations: Providing explanations that take into account the broader context of a decision, helping users understand not just the "what" but also the "why" behind AI predictions.

Explainable AI is crucial for various applications, especially in domains where decisions have significant consequences, such as healthcare diagnoses, loan approvals, or legal judgments. It enables better accountability, ethical use of AI, and user trust in the technology. Researchers and practitioners continue to work on developing and improving techniques for explain ability in AI models.

3.17 Measuring the Accuracy of Models and Model Evaluation

Forecasts are inherently fallible, rarely achieving perfect accuracy. Hitting the exact target is extremely challenging, and even a minor deviation is to be expected. The choice of an appropriate error metric is largely determined by the specific characteristics of the data, the objectives we aim to achieve, and the unique context of our business.

Performance Indices (Bias, Precision and Accuracy)

In order to compare forecasts, there are various measures, namely Bias, Precision and Accuracy (Walther & Moore, 2005). Bias quantifies the disparity between the expected value of an estimate and the true value of the parameter being estimated. It is represented by the parameter Sum of Errors (SE). Precision, on the other hand, assesses the reliability and consistency of results obtained independently or in multiple measurements of the same quantity, without reference to a theoretical or true value..

Accuracy is the closeness of agreement between a measured value and a true or accepted value. Measurement error is the amount of inaccuracy(Y. Wang & Wu, 2012a). There are various well

established metrics for measuring the bias, variance and accuracy (**Table *7***) of machine learning models (Walther & Moore, 2005)

Table 7 : Measures of errors: Bias, Variance and Accuracy

S.No.	Measures of Errors	Description	Measure
1	Bias	Closeness to real value	Mean Errors (ME)
2	Precision	Degree of consistency among independent measurement	Variance
3	Accuracy	Closeness of agreement between measured values	Mean Square Errors (MSE)
			Mean Absolute Errors (MAE)
			Root Mean Square Error (RMSE)
			Mean Absolute Percentage Error (MAPE)

In order to compare forecasts, there are various measures, namely Bias, Precision and Accuracy (Walther & Moore, 2005) **Figure 27**. Bias is a measure of how far the expected value of the estimate is from the true value of the parameter being estimated. The parameter Sum of Errors (SE) represents the Bias. Precision denotes the extent to which a result can be reliably determined, independent of any theoretical or true value. It reflects the consistency and agreement among independent measurements of the same quantity, indicating the reliability and reproducibility of the outcome.

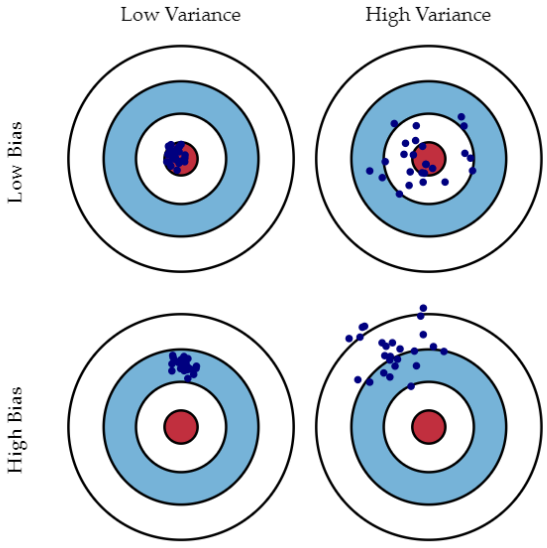


Figure 27 : Graphical representation of bias and variance (Source http://scott.fortmann-roe.com/docs/BiasVariance.html)

As is customary in the literature, the provided forecasts are statistically analysed using the Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean absolute error (MAE), Mean absolute percentage error (MAPE) (Karathanasopoulos et al., 2016b) (**Table 9**).

Table 8 : Measures of Performance

Variance	**Accuracy**
Variance $\frac{1}{N} * \sum_{i=1}^{n}(y_i - \bar{y})^2$	MAE = $\frac{1}{N} * \sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$ MSE = $\frac{1}{N} * \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ RMSE = $\sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{N}}$ $MAPE = \frac{1}{N}\sum_{i=1}^{n}\left(\frac{\lvert y_i - \hat{y}_i\rvert}{y_i} \times 100\right)$ $RMAPE = \sum_{i=1}^{n}\frac{\left(\frac{\lvert y_i - \hat{y}_i\rvert}{y_i} \times 100\right) * y_i}{y_i}$

where MAE = Mean Absolute Error, MSE=Mean Square Error and RMSE = Root Mean Square Error

y_i = Actual value

$\hat{y}_i$ = Predicted Value

$\bar{y}$ = Average of all actual values

N = No. Of Observations

For comparison of models, we have also employed and additional measure for accuracy Resistant-MAPE (R-MAPE) (**Table 9**) in our study which is a better fit for comparing accuracy of time series models(Montaño Moreno et al., 2013).

Table 9 : Comparison of MAPE and R-MAPE

Characteristic	MAPE	R-MAPE
Sensitivity to outliers	More sensitive	Less sensitive
Interpretability	Easier to interpret	More difficult to interpret
Suitability for different applications	Good for forecasting applications where the actual values are relatively similar in magnitude	Good for forecasting applications where the actual values vary widely in magnitude or where there are outliers
Examples of applications	Forecasting sales for individual products in a retail store	Forecasting demand for electricity or water

Acceptable levels of MAPE and RMSE for a time series forecasting model vary depending on the specific application and the industry. However, in general, the following guidelines can be used:

MAPE: A MAPE of 10% or less is generally considered to be highly accurate, while a MAPE of 20% or less is considered to be good. MAPE between 20% and 50% is considered reasonable. A MAPE of greater than 50% is generally considered to be poor or inaccurate forecasting (Montaño Moreno et al., 2013) .

RMSE: The acceptable level of RMSE for a particular application may be higher or lower depending on the specific needs of the business (Shcherbakov et al., 2013a). Additional factors that should be kept in mind to consider when evaluating the reasonableness of RMSE for demand forecasting:

- **Frequency of the forecasts**: More frequent forecasts (e.g., daily or weekly) will generally have higher RMSE than less frequent forecasts (e.g., monthly or quarterly). This is because more frequent forecasts are more difficult to predict accurately.
- **Level of aggregation**: Forecasts for more aggregated data (e.g., total demand for a product category) will generally have lower RMSE than forecasts for less aggregated data (e.g., demand for a specific product). This is because aggregated data is less noisy and easier to predict.
- **Horizon of the forecasts**: Forecasts for shorter horizons (e.g., next week or next month) will generally have lower RMSE than forecasts for longer horizons (e.g., next year or next

five years). This is because shorter-term forecasts are less subject to unexpected changes in the environment.

An RMSE that is less than or equal to the standard deviation of the historical data is generally considered to be good. An RMSE that is greater than the standard deviation of the historical data is generally considered to be poor. This bodes from the fact that the standard deviation of the historical data is a measure of how much the data varies. If the RMSE of a forecast is less than or equal to the standard deviation of the historical data, then it means that the forecast is at least as good as the historical average. However, if the RMSE of a forecast is greater than the standard deviation of the historical data, then it means that the forecast is not as good as the historical average. For comparative studies, the model with the least RMSE measure will be the best option.

Bias-Variance Trade off

The property of a machine learning model (supervised) that enforces a tradeoff between how "flexible" the model is and how well it performs on unseen data is known Bias – Variance trade off (**Figure 28**).

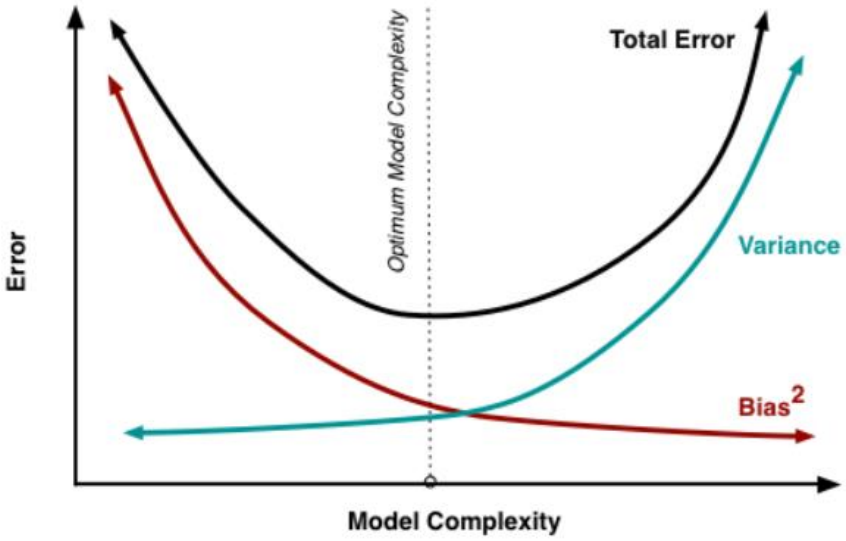


Figure 28 : Graphical representation of bias and variance tradeoff (Source: http://scott.fortmann-roe.com/docs/BiasVariance.html

The latter is known as a models generalization performance. The basic model in machine learning is given by the following:

$$Y=f(X)+\epsilon$$

This indicates that the response vector, Y, is expressed as a function, f, of the predictor vector, X, potentially nonlinear, with a series of error terms distributed normally, having a mean of zero and a standard deviation of one.

When determining the optimal statistical machine learning approach, it's imperative to characterize the relative performance among models. This necessitates comparing the known values of the underlying relationship with those predicted by a model that has been estimated.

The first term on the right-hand side denotes the variance of the estimate across numerous training sets. It quantifies the extent to which the average model estimation deviates as different training data are employed. Notably, a model with high variance suggests overfitting to the training data.

The squared bias, represented by the middle term, delineates the disparity between the averages of the estimate and the true values. A model with high bias fails to capture the underlying behavior of the true functional form effectively. This is akin to using linear regression to model a sine curve, where regardless of how well the model fits the data, it cannot adequately capture the inherent nonlinearity of a sine curve.

The last term, known as the irreducible error, serves as the minimum lower limit for the test Mean Squared Error (MSE). Since we only have access to the training data points, along with the associated randomness of the error values (ϵ), we cannot aspire to achieve a "more accurate" fit beyond what the variance of the residuals offers.

The ultimate objective in machine learning is always to minimize the expected test MSE, necessitating the selection of a statistical machine learning model that simultaneously exhibits low variance and low bias.

Comparing ARIMA models

The goodness of fit of ARIMA models can be compared using following criteria:

a) The Akaike Information Criteria (AIC)

b) The Corrected Akaike Information Criteria (AICc)

c) Schwartz Information Criteria (SIC)

d) Mean Absolute Percent Error (MAPE)

e) Root Mean Square Error (RMSE) and

f) Mean Absolute Deviation / Error (MAD/E)

g) Goodness of Fit (AIC/BIC)

3.18 Conclusion

The recent advancement of Artificial Neural Networks and especially the Deep Learning ANN in the recent years have established Recurrent Neural Networks (RNN) and specifically the LSTM Neural Networks as a credible alternative to statistical machine learning models for time series forecasting. Standalone and hybrid models of LSTM Neural Networks can also be used for forecasting Crack spread price and LPG Demand 1-3 months ahead. Coupled with relevant metrices for evaluation of forecasting models, this research aims to develop models.

Chapter 4

Crack spread forecasting

"Deep learning is not just a tool; it's a new way of thinking about problem-solving, especially in forecasting complex and dynamic systems."

- Ian Goodfellow, Computer Scientist

4.1 Introduction

The price difference (gap) between crude oil and refined goods distillates (such gasoline, diesel, jet fuel, etc.) significantly impacts the profit margins of refiners in the petroleum industry(Steven et al., 2014b). Since the refining process "Crack" crude oil into its primary refined products, this spread is commonly referred to as the "Crack Spread".

Indian Oil refiners import about 85% of the required crude oil to meet India's growing energy demand. Refined products such as LPG, petrol, diesel, and kerosene are sold through various distribution channels across the country. The profitability of Indian Oil Marketing Companies (OMCs) depends on the cost of imported crude oil and the selling price of products, both linked to international benchmarks. For profitability calculations, Indian Oil companies use crude prices from Dubai and product prices from the Singapore exchange. Modeling and predicting crack spreads are crucial for optimizing supply chain operations and hedging against crude oil price volatility.

4.2 Conceptual model

The review of literature in the domain of very synonymous Crude Price Forecasting and Cracks Forecasting leads us to the understanding that the crude prices are very volatile and therefore there is no straight forward method for forecasting the crude price. It is observed that short term crude price forecasts are more accurate than long term forecasts. Owing to the presence of both linearity and non-linearity in the crude prices, a hybrid model is expected to yield better forecasting results. A combination of the statistical / econometric model such as ARIMA or ETS and a non-linear

ANN based model is expected to yield better results. This understanding, therefore has been the guiding framework (Figure **29**) for development of a model for cracks forecasting.

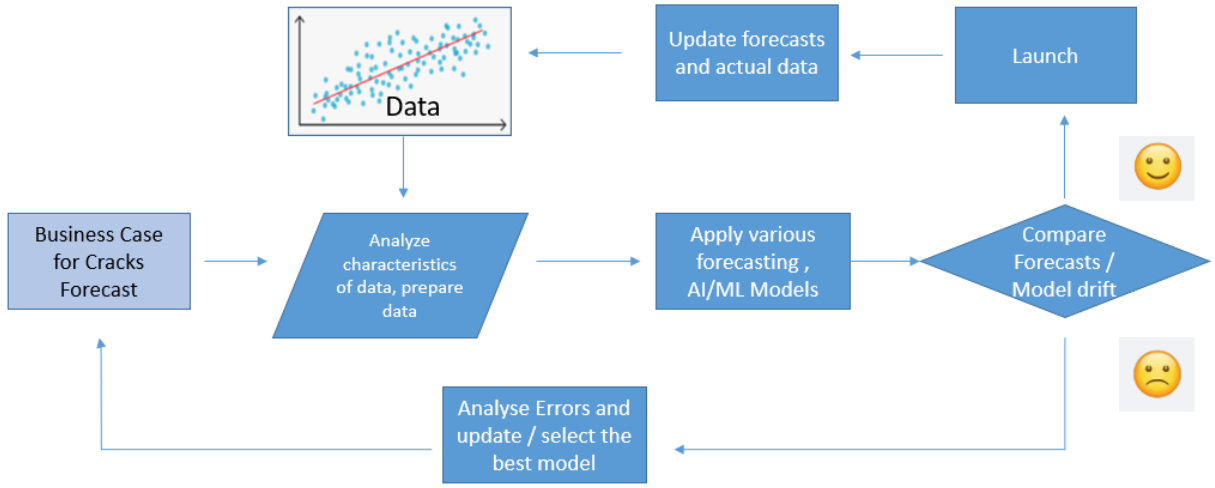


Figure 29: Framework for Crack Spread Forecasting

The performance of the model so developed will be assessed and compared employing the prevalent error metrics and using the metrices Mean, Variance, Standard Deviation, Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Resistant-MAPE (R-MAPE).

4.3 Methodology

The hybrid forecasting method developed in this research is a 3 stage process (**Figure 30**). Stage 1 is used to generate forecast based on the linearity present in the time series data using econometric models. For this purpose, ARIMA/ ETS techniques are employed. The residuals of the ARIMA / ETS techniques typically resemble 'white noise' which means that the model has captured all the relevant patterns in the data.

However, the non-linearity that is present in residuals may contain additional information that can be used to improve the accuracy of the model. In this context, Stage 2 attempts to exploit the non-linearity present in the residuals by employing ANN models and derive additional information that can increase the accuracy of the forecasts.

Therefore in the Stage-3 of the process, the forecasts generated from Stage 1 is combined with the forecasted residual generated in Stage 2 to compute the final forecast

The schematic flowcharts of the entire process are presented below:

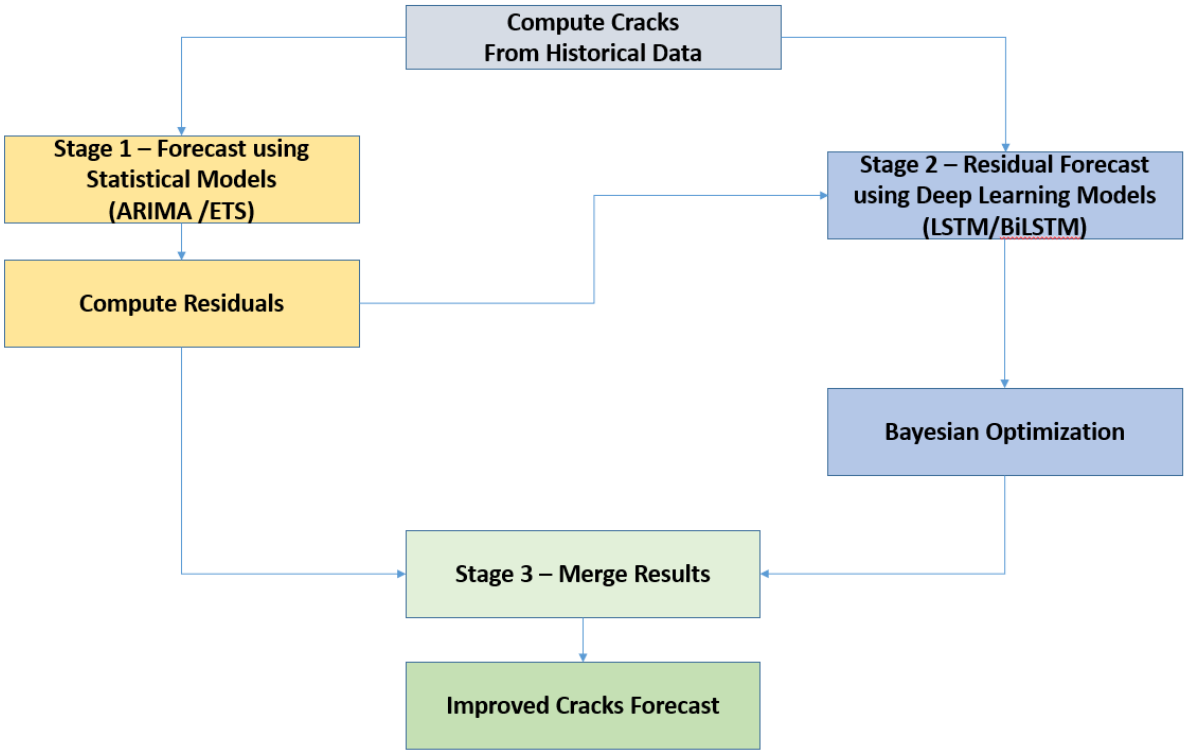


Figure 30 : Overall view of the Hybrid ARIMA + BiLSTM Model

Stage 1 - Forecasting cracks price using ARIMA

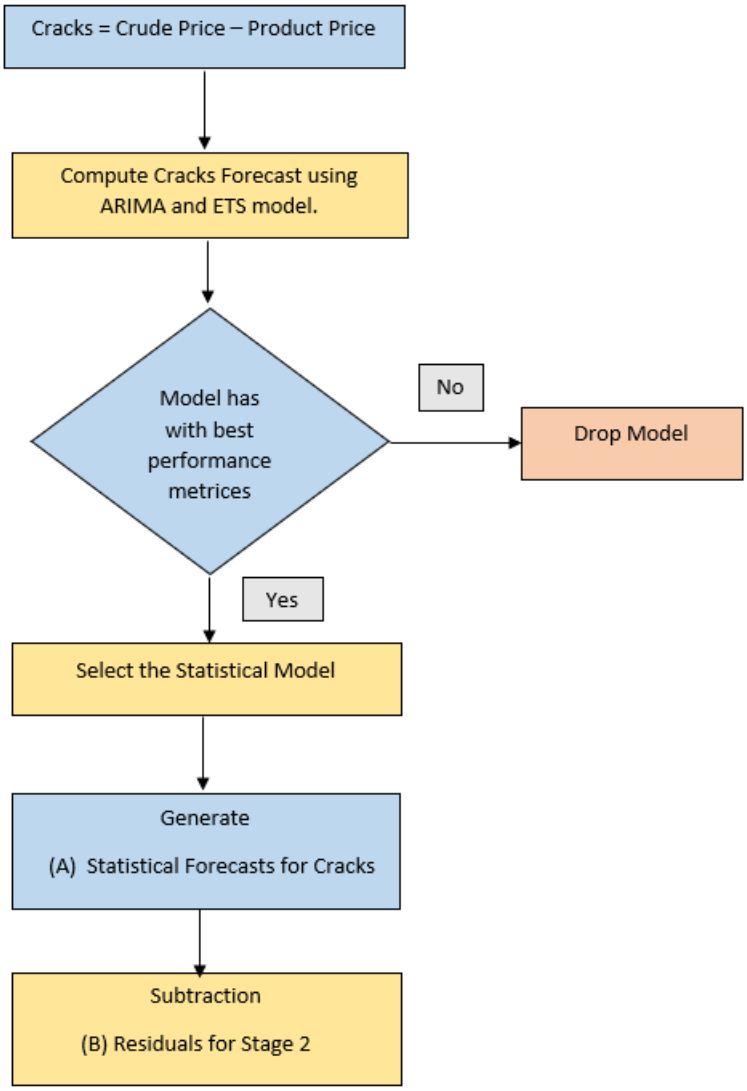


Figure 31 : Details of Stage 1

1. Data preprocessing

A time series may contain noisy and missing values. The noisy values are smoothed and the missing values are replaced using an appropriate (averaging, imputation etc) technique. A time-series data exhibits heteroscedasticity if the dependent variable changes significantly from the beginning to the end of the series. The method to remove heteroscadacity is to take the logarithm of the time series.

2. **Finding the right set of parameters a,b,g for ARIMA / ETS:**

   i. Stationarity test Augmented Dickey Fuller (ADF) Test

   ADF is a statistical hypothesis test, its null hypothesis essentially asserts non-stationarity. Therefore when we run the test, low p-values indicate stationarity. If a time series has a unit root process, then the lagged level series (Yt-1) will provide no relevant information in predicting the change Yt. In contrast, when the process has no unit root, it is stationary and hence exhibits reversion to the mean - so the lagged level will provide relevant information in predicting the change of the series and the null of a unit root will be rejected.

   ii. Auto Correlation Factor (ACF) and Partial Auto Correlation Factor (PACF)

   ACF is an auto-correlation function which gives us values of auto-correlation of any series with its lagged values. It describes how well the present value of the series is related with its past values. A time series can have components like trend, seasonality, cyclic and residual. ACF considers all these components while finding correlations hence it's a 'complete auto-correlation plot'.

   PACF is a partial auto-correlation function. It finds correlation of the residuals (which remains after removing the effects which are already explained by the earlier lag(s)) with the next lag value hence the term 'partial' and not 'complete' as we remove already found variations before we find the next correlation. So if there is any hidden information in the residual which can be modeled by the next lag, we might get a good correlation and keep that next lag as a feature while modeling.

4.4 Forecasting residuals using LSTM Deep Learning

During the prediction in stage 1, the linearity in the time series data is harnessed to make the forecasts. The objective of the second stage of the model is to derive additional information from the non-linearity in the time series. The residuals remaining after Stage-1 are essentially white noise. This is where we employ non-linear techniques such as ANN to derive any possible information and improve the forecast. The steps performed in the second stage (**Figure *32***) are:

i. Splitting the Residual time series in Training and Test series
ii. Generating forecasts for the desired period using the Bi-Directional LSTM RNN model.

iii. Optimizing the hyperparameters using Bayesian optimization

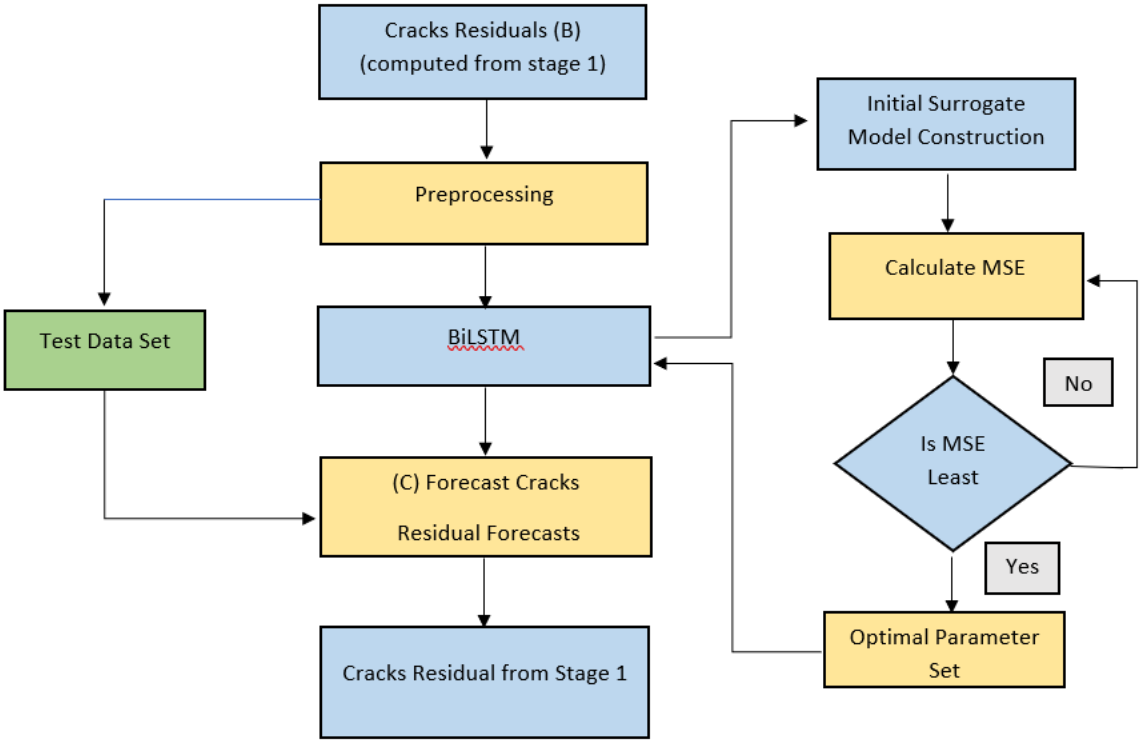


Figure 32 : Details of Stage 2

4.5 Data Processing (Source, Selection and Pre-processing)

The price of Crude oil has undergone tectonic shifts during the past few decades. The **Figure *33*** depicts the price of Crude Oil from 1980s till date.

Figure 33 : Crude Prices for the last 7 decades (Courtesy: Macrotrends)

The volatility (rise and fall) and major shifts in price levels are clearly discernible and attributable to major events such as Gulf War (1990-91), The Y2K Bubble (1998-200), Sub-prime lending crisis in USA (2008), Excess shale gas production by US (2014-15) and the COVID-19 pandemic in 2020 to name some of the few major events. If for the purpose of model development and training the complete set of data ranging from 1990 till date is used, then it will result in models which wrongly would factor extraneous events and their implications which have no systemic bearing in forecasting of the prices of Crude and products for operational purposes.

To exclude the extreme effects of such drastic shifts in prices owing to extraneous events as outlined above, data for this research has been chosen from the window corresponding to June 2015 till Dec 2019. During this period, there is an absence of any political, economic or any such event and the prices are purely reflective of the market dynamics namely the demand, supply, and prices.

1. Data Selection

This research uses data sourced from daily feeds provided by Platts comprising of all types of traded Crude and Petroleum products. For the purpose of model development, Brent Crude oil prices and Gasoline 0.05 % Singapore prices have been used. The prices for the Crude and Products are plotted in USD $/Barrel in **Figure *34***. The price of crack spreads on a given day is determined by subtracting the difference between the daily closing product prices in US dollars from the price of crude in US dollars. Whereas for the purpose of planning and GRM calculation, various crack spreads namely the 5:3:2 and 3:2:1 crack spread are used. The 5:3:2 Crack spread denotes that 5 units of Crude oil is used to produce 3 units of Gasoline and 2 units of Gas oil. The GRM is thus calculated basis on the yield (i.e., percentage of Gasoil and Gasoline produce per gallon of Crude oil and computed proportionately. However, for the purpose of simplicity of model development only one Product (Gasoline) has been used to calculate the gasoline crack based on Brent Crude oil prices. This scenario closely reflects the reference crude and product prices being used by Indian Oil companies. The model so developed can be easily extended to other crack spread or any combinations if so desired.

Data from the training set make up 80% of the total, whereas data from the test set make up 20% of the total.

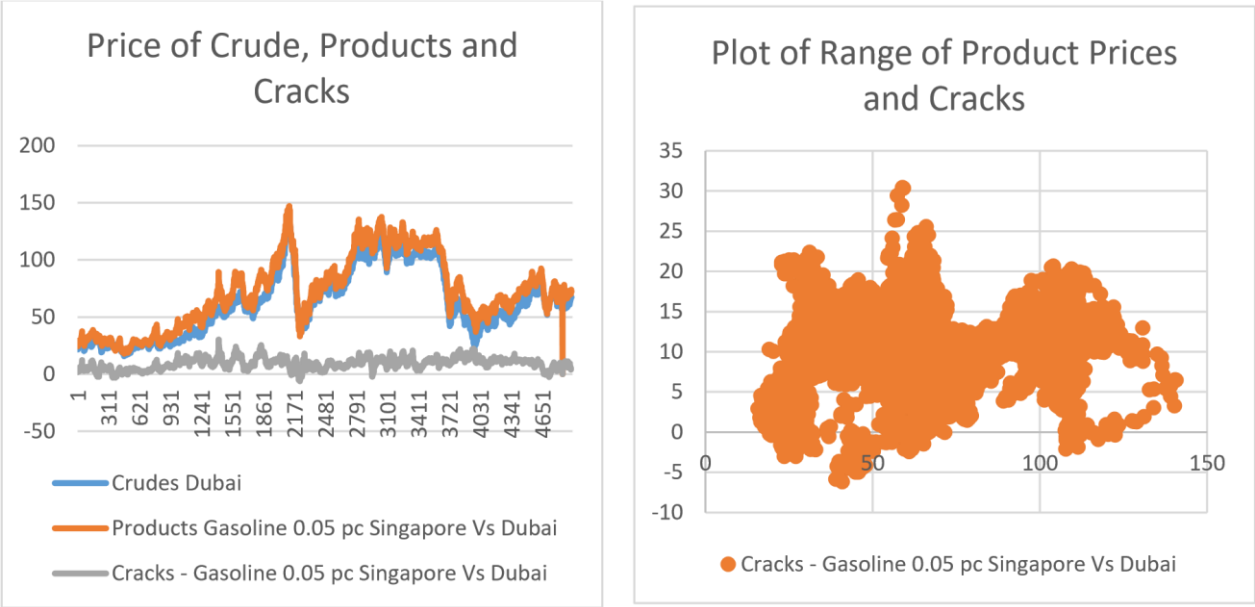


Figure 34 : Plot of range of products prices and cracks

2. Data Preprocessing

A time series could have missing and irregular numbers. Using appropriate method, the noisy data are smoothed, and the missing values are replaced. Heteroscedastic data is when a time series' dependent variable significantly varies from the beginning to the end of the series. By taking the time series' logarithm, heteroscedastic behavior can be eliminated.

In order to assess the predictability of time series crack spread data is used, the ADF statistics for the crack time series data is computed and is listed in **Table *10*** below:

Table 10 : ADF Statistics for Crack Spead Time Series Data

S.No.	Parameter	Value
1.	ADF Statistics	-6.160728
2.	p-Value	0.0000
3.	Critical value of 1%	-3.432
4.	Critical Value of 5 %	-2.982
5.	Critical Value of 10%	-2.567

The Augmented Dickey-Fuller (ADF) test is a statistical test which is used to check whether a time series is stationary (i.e., its statistical properties like mean and variance do not change over time).

For ADF Test, the Null and Alternate hypothesis are :

H_0 (Null Hypothesis): The time series has a unit root (i.e., it is non-stationary).

H_1 (Alternative Hypothesis): The time series does not have a unit root (i.e., it is stationary).

The ADF Test Statistic has a negative value of -6.160728; the lower (more negative), the stronger the rejection of H_0. The Critical Values for thresholds at different confidence levels (1%, 5%, 10%) is used to compare with the ADF test statistic which in our case is negative at all these levels. The p-value depicts the probability that the series is non-stationary. If the p-value is > 0.05, then we fail to reject the null hypothesis H_0 → Data is non-stationary (needs further transformation). If p-value is ≤ 0.05 → Reject H_0 → Data is stationary (can be used for ARIMA modeling).

The test results therefore clearly indicate that the data is stationary and hence we can proceed with model fitting using ARIMA.

During the deep learning training process, data is standardized. When the actual forecasts are generated at the end of the modelling process, this process is reversed. The Google TensorFlow framework and various Python libraries like Keras, Pandas, etc. are utilized to implement the deep learning techniques employed in this study. The kerastuner software serves as a foundation for the Bayesian optimization-based hyperparameter training for Deep Leaning and Optimization Model.

4.6 Modelling

The available dataset is divided into two halves for the purpose of training the LSTM models, with the training set comprising 80% of the data and the testing set the remaining 20% the testing set When building and training the LSTM network, various hyperparameters are used in addition to the lag size parameter, including the number of hidden layers, epoch size, batch size, number of neurons, learning rate, and dropout rate. In the following stage, an LSTM network is built and trained using the training set that corresponds to each set of hyperparameters that was valued using

Bayesian optimization in the previous stage. The grid search method is used to turn the time series into a set of instances with an input-output format based on the lag size that is chosen. An example of the produced instances with lag size = 10 is shown in Table 5. After that, a train set and a test set are created from the instances that were produced. The test set comprises 20% of all the examples in each experiment. An LSTM network is constructed and trained for each combination of hyperparameters. The Adam algorithm serves as the optimizer for all networks (Kingma & Ba, 2014). Furthermore, the mean squared error (MSE) is utilized as the loss function for every constructed LSTM network.

The final crack price forecast is arrived at by adding the forecast obtained in stage (1) with the residual forecast obtained in stage (2). The performance of the model so developed is computed based on the prevalent error metrics and using the metrices Mean, Variance, Standard Deviation, Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Resistant-MAPE (R-MAPE).

1. **Generating instances based on lag size**

   Based on the lag size selected in the previous step, the time series is converted into a set of input and output format. This data reshaping task is required to employ an LSTM.

2. **Splitting instances into train and test sets**

   For the purpose of training the LSTM models, the available dataset is split into 2 parts with 80% data comprising the training set and the remaining 20% comprising the testing set.

3. **Modelling**

   Except for the lag size parameter, the remaining hyperparameters including number of hidden layers, epoch size, batch size, the number of neurons, learning rate and dropout rate are used in configuring and training the LSTM network. Having determined the values of these parameters using the Bayesian optimization in the previous step, in this step, corresponding to each hyperparameter combination, a LSTM network is created and trained using the training set.

4. **Selecting the best model**

   Based on the lag size selected using the grid search method, the time series is transformed into a set of instances with the input-output format.

For instance, for lag size = 10, a subset of the created instances is shown in Table 5. Then, the created instances are split into a train and test set. For all experiments, the test set contains 20% of all instances.

5. **Model configuration and training**

For each combination of hyperparameters, an LSTM network is designed and trained. For all networks, the Adam algorithm (Kingma & Ba, 2014) is used as the optimizer. In addition, the mean of squared error (MSE) as loss function is for all designed LSTM networks.

4.7 Hyperparameters for tuning

Hyperparameter tuning involves generating a list containing combinations of hyperparameters for LSTM which yields the best accuracy. Some of the hyperparameters that can be tuned are:

i. Lag size: The lag size parameter has significant impact on the performance of time series forecasting (Ribeiro et al., 2011). Therefore, it is crucial to test the performance of a model using different lag sizes.

ii. The number of hidden layers: It is shown that deep neural network architectures have better generalization than single layer architectures (Hermans & Schrauwen, 2013; Utgoff & Stracuzzi, 2002). As a result, it is important to investigate the performance of a model with more than one layer.

iii. The number of neurons: Selecting the optimal number of neurons for each layer in the LSTM network is not a straightforward task. If the number of neurons is very small, the LSTM will not able to memorize all necessary information to perform prediction optimally. Also, if the number of neurons is very high, the LSTM will overfit on the training instances and will not demonstrate suitable generalization to accurately forecast the test set (Reimers & Gurevych, 2017).

iv. Dropout rate: Applying dropout rate on the LSTM influences the performance of the resulting model by enhancing the model generalization (Reimers & Gurevych, 2017; Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014).

v. Learning rate: Learning rate is an important hyperparameter which adjusts the extent of change to the model weights. Choosing the best transforming data into supervised learning rate is essential to obtain a model with high performance (Reimers & Gurevych, 2017).

vi. Batch size: Adjusting batch size is another factor in determining the performance of the LSTM model. Hence it is significant to find an optimal value for batch size (Shi et al., 2019).

vii. Epoch size: One training epoch is referred to a single iteration over all training instances (Reimers & Gurevych, 2017). If the number of training epochs is too small, the model will not capture the patterns of training instances. Also, if the epoch number is too large, the model will suffer from overfitting. Therefore, finding a suitable epoch number is vital in achieving a model with high performance.The hyperparameters used for arriving at the results are listed in **Table *11*** below:

Table 11 : Hyperparameters Bounds for Bayesian Optimization

S.No.	Hyperparameter	Value
1	Lag	-0.31
2.	No. of Hidden Layers	59.02
3.	No. of neurons for each layer	0.68
4.	Dropout Rate	0.67
5.	Learning Rate	0.83
6	Batch Size	0.78
7.	Epoch size	10
8.	Activation Function	Relu

4.8 Forecasts Aggregation

The final cracks price forecast is arrived at by adding the forecast obtained in stage (1) with the residual forecast obtained in stage (2) **Figure *35***.

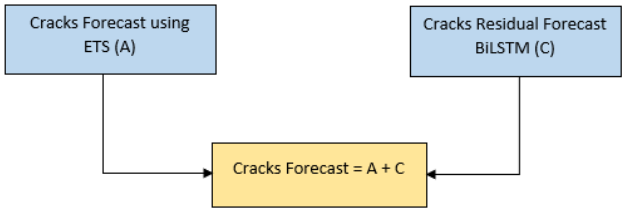


Figure 35 : Details of stage 3

4.9 Comparing ARIMA and ETS with MLP, LSTM and BiLSTM

As a baseline, forecasts were separately computed using statistical models namely ARIMA, ETS and ANN Deep Learning models namely MLP, RNN models LSTM and BiLSTM (**Figure *36***).

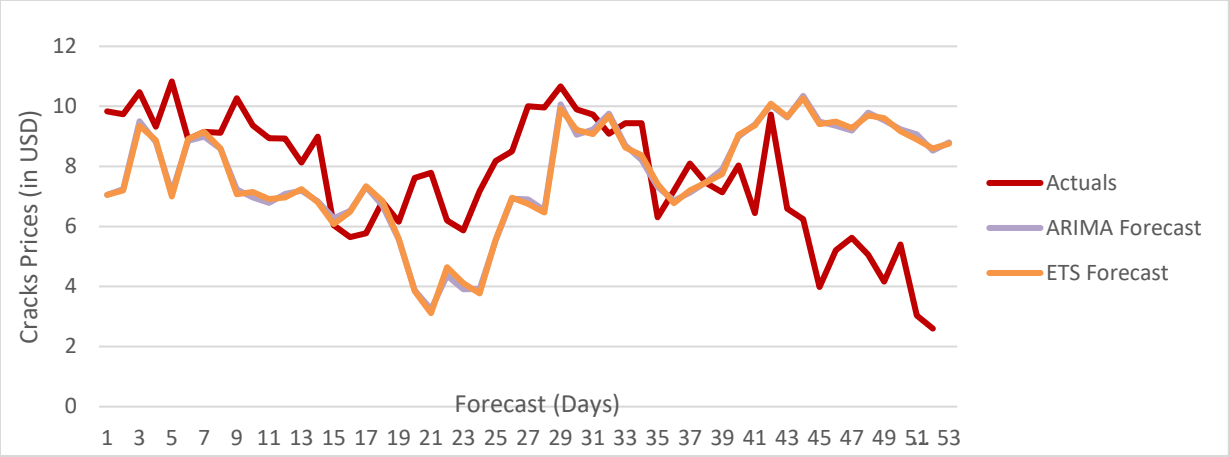


Figure 36 : Plot of Actual Crack Spreads and Forecast using Statistical (ARIMA and ETS) Methods)

As is evident from figure above, the forecasts computed using statistical techniques ARIMA and ETS denote similar behaviour (**Figure *37***).

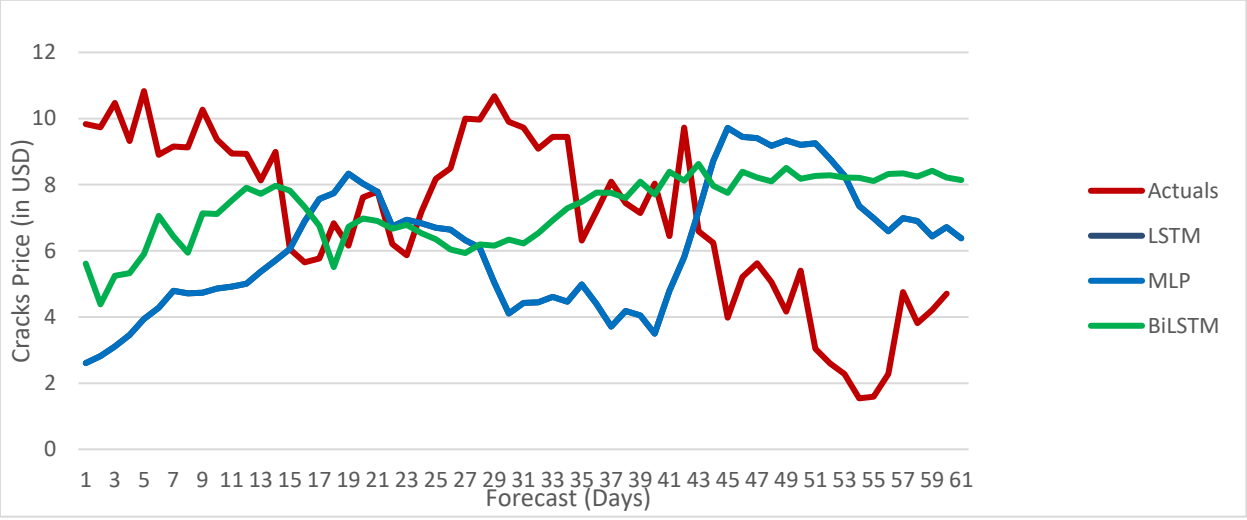


Figure 37 : Plot of Actual Crack Spread and Forecast using Deep Learning (MLP, LSTM, BiLSTM) Methods

However, forecasts generated using deep learning methods MLP, LSTM and BiLSTM (**Figure *38***) are exhibit different behavior.

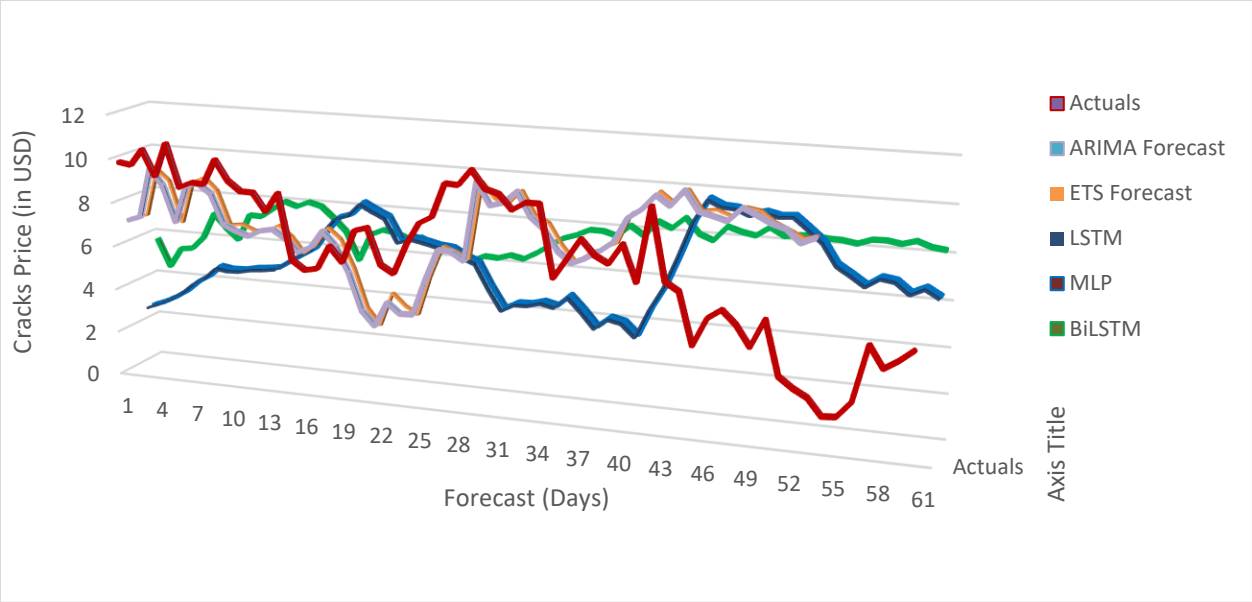


Figure 38 : Plot of Actual Crack Spreads and Forecast using Statistical and Deep Learning Techniques together

A combined plot of Actual crack spread and forecasts using all the techniques under study is plotted in Figure above. The tabulation of performance metrices related to these forecasts is provided in **Table *12*** below.

Table 12 : Comparison of Errors between Statistical ARIMA, ETS Deep Learning (MLP/LSTM/BiLSTM) Models

	ARIMA	ETS	MLP	LSTM	BiLSTM
VAR	6.78	6.72	17.01	11.06	5.87
SD	2.60	2.59	4.12	3.30	2.42
MSE	6.84	6.77	30.61	14.72	6.15
RMSE	2.63	2.66	5.52	3.88	2.60
MAE	2.24	2.23	3.87	3.00	2.03
MAPE	35.29	35.21	49.27	112.97	34.32
RMAPE	29.21	29.31	48.39	119.64	26.55

It can be inferred from the table above that ARIMA and ETS methods provide a very similar performance amongst the statistical techniques but the performance of deep learning methods varies considerably. BiLSTM exhibits the best performance amongst the deep learning techniques and exhibits better MSE and RMSE. However, its MAPE , RMAPE performance is inferior to

ARIMA and ETS methods. On an individual basis ETS, LSTM and BiLSTM have RMSE scores less than their Standard Deviation, meaning that they are generating better forecasts. Viewed from the perspective of MAPE / RMAPE none of the methods when considered individually produce good quality forecasts (i.e., their individual MAPE / RMAPE are greater than 20)

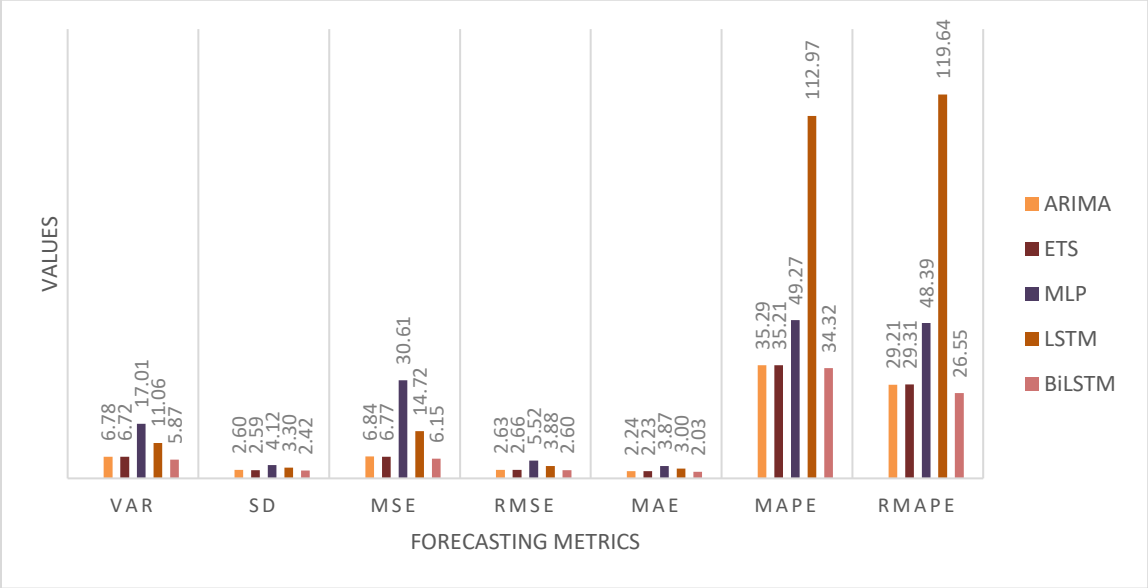


Figure 39 : Bar Plot of Accuracy Measures for all methods studied

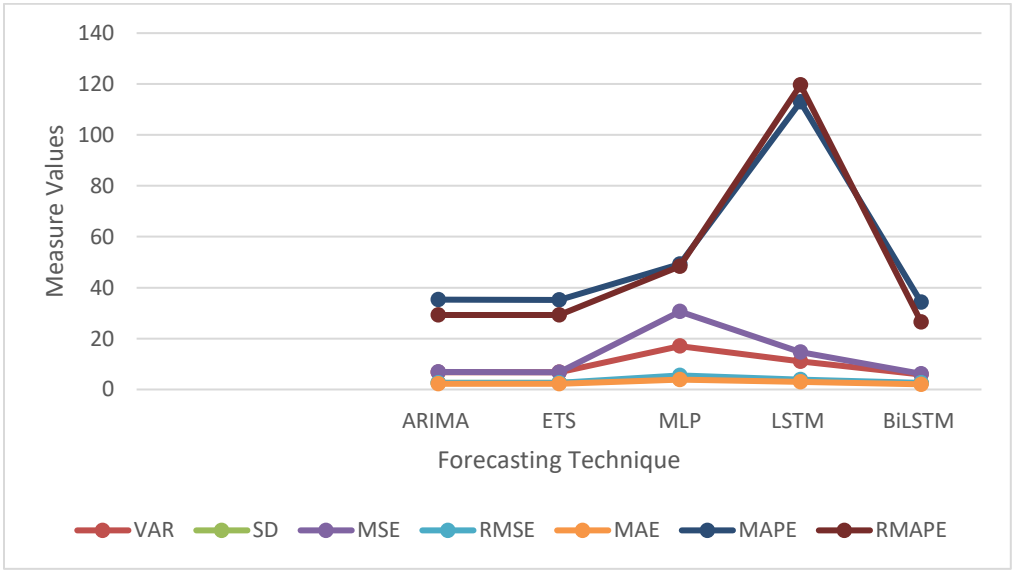


Figure 40 : Line Plot of Accuracy Measures for all methods studied

Bar plots and Line plots of the various metrices are provided in **Figure *39*** and **Figure *40*** above.

To improve the forecasting method, as mentioned in the earlier section of this paper, we have taken ARIMA and ETS both as statistical method and BiLSTM method as the deep learning technique for further studies **Figure *41***.

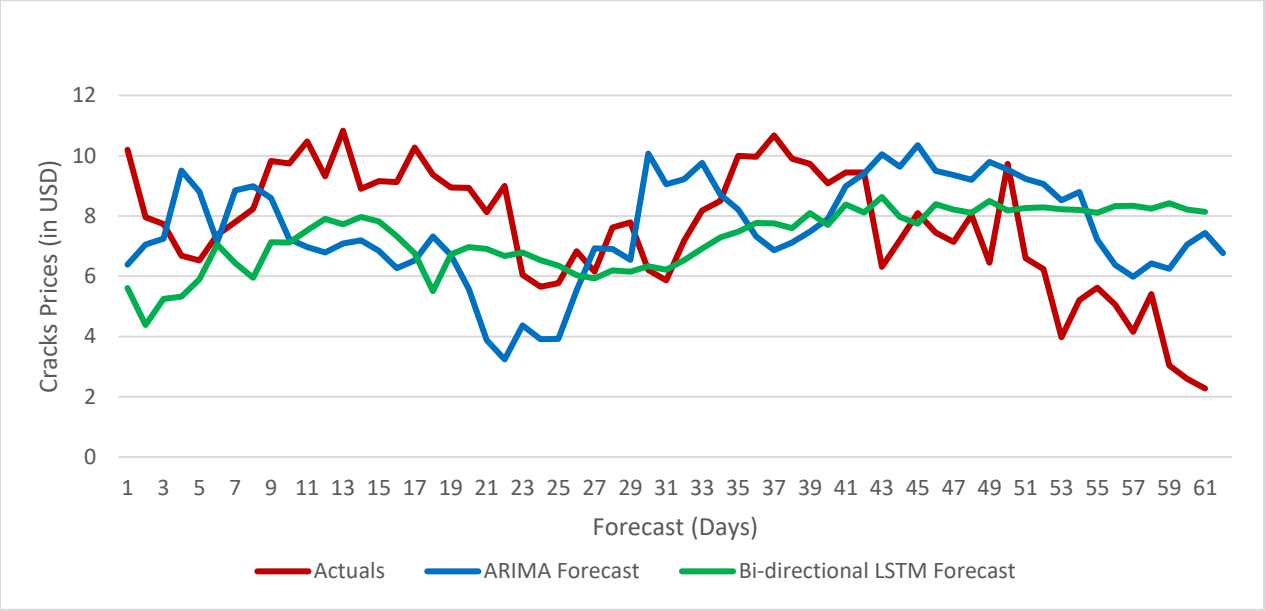


Figure 41 : Plot of Actual Crack Spread and Forecast using Statistical (ARIMA) and Deep Learning (BiLSTM) Methods

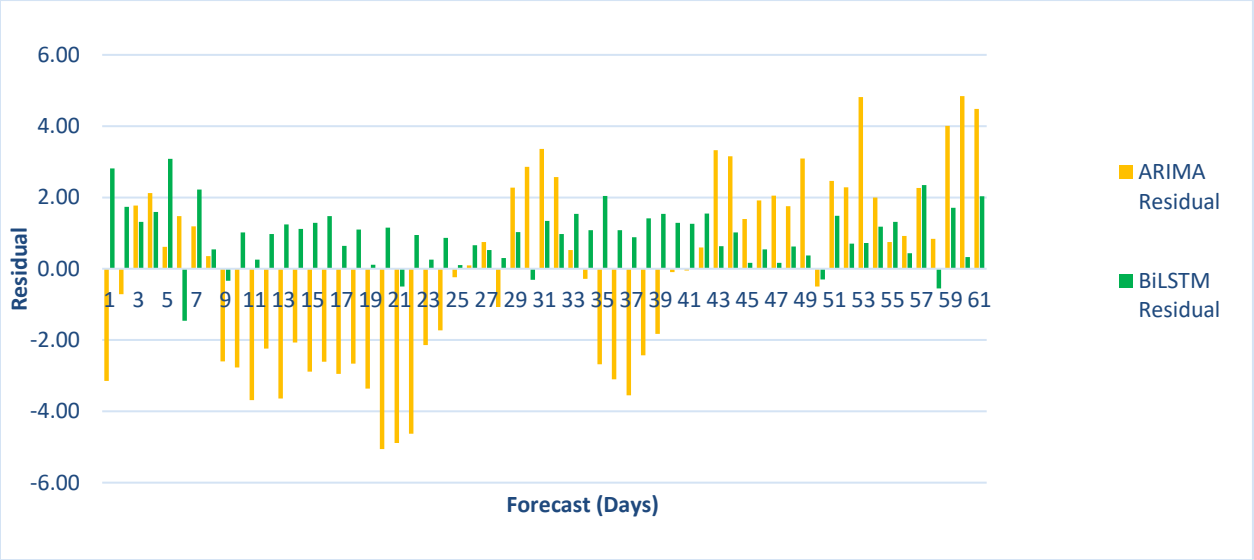


Figure 42 : Residual plot of Actual Crack Spreads and Forecast using Statistical (ARIMA) and Deep Learning (BiLSTM) Methods

The **Figure *42*** provides comparison of ARIMA and BiLSTM forecasts plotted side by side.

The results of 2 stage forecasts using ARIMA and ETS for stage 1 and BiLSTM for forecasting residuals in Stage 2 and finally calculating the forecasts is listed in **Table *13*** below.

Table 13: Comparision of Errors between ETS+LSTM and Hybrid LSTM Models

	ARIMA	BiLSTM	ARIMA + BiLSTM Residual	ETS	ETS + BiLSTM Residual
VAR	6.78	5.87	2.26	6.72	2.78
SD	2.60	2.42	1.50	2.59	1.67
MSE	6.84	6.15	2.25	6.77	2.80
RMSE	2.63	2.60	1.50	2.66	1.66
MAE	2.24	3.00	1.26	2.23	1.39
MAPE	35.29	34.32	19.89	35.21	23.54
RMAPE	29.21	26.55	16.4	29.31	18.61

We observe that there is a considerable improvement in forecasts using ARIMA + BiLSTM when compared to using ARIMA, ETS and BiLSTM individually. The RMSE score of the ARIMA + BiLSTM is less than the SD and MAPE, RMAPE scores are also less than 20, which is the threshold for the quality of the forecasts to be considered good.

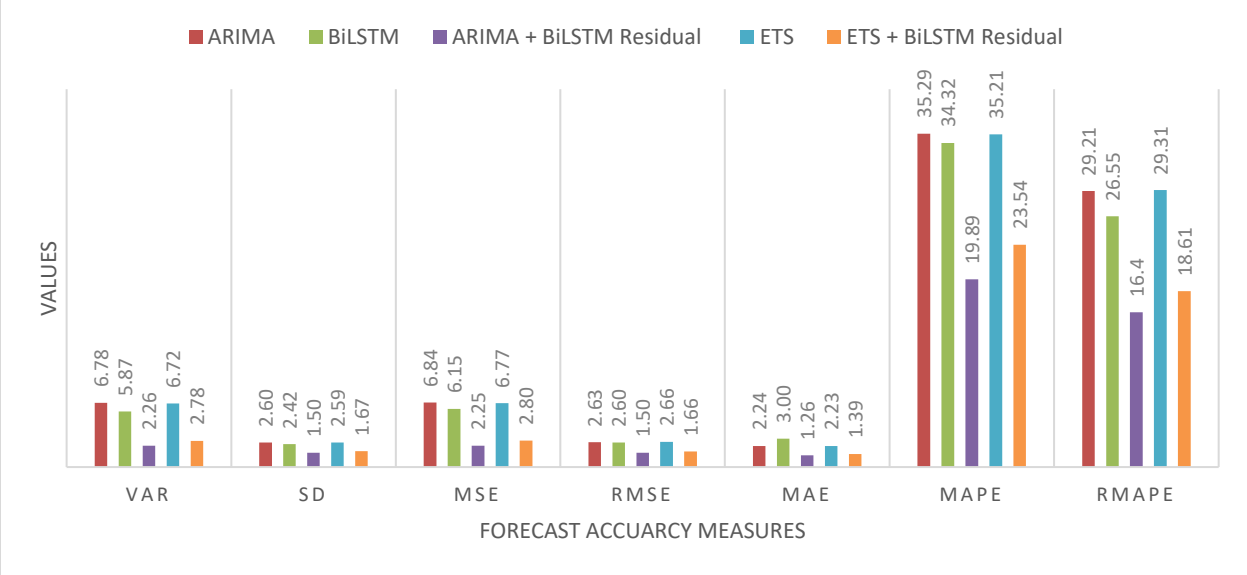


Figure 43 : Plot of Actual Crack Spread and Statistical forecast (ARIMA and ETS) Methods)

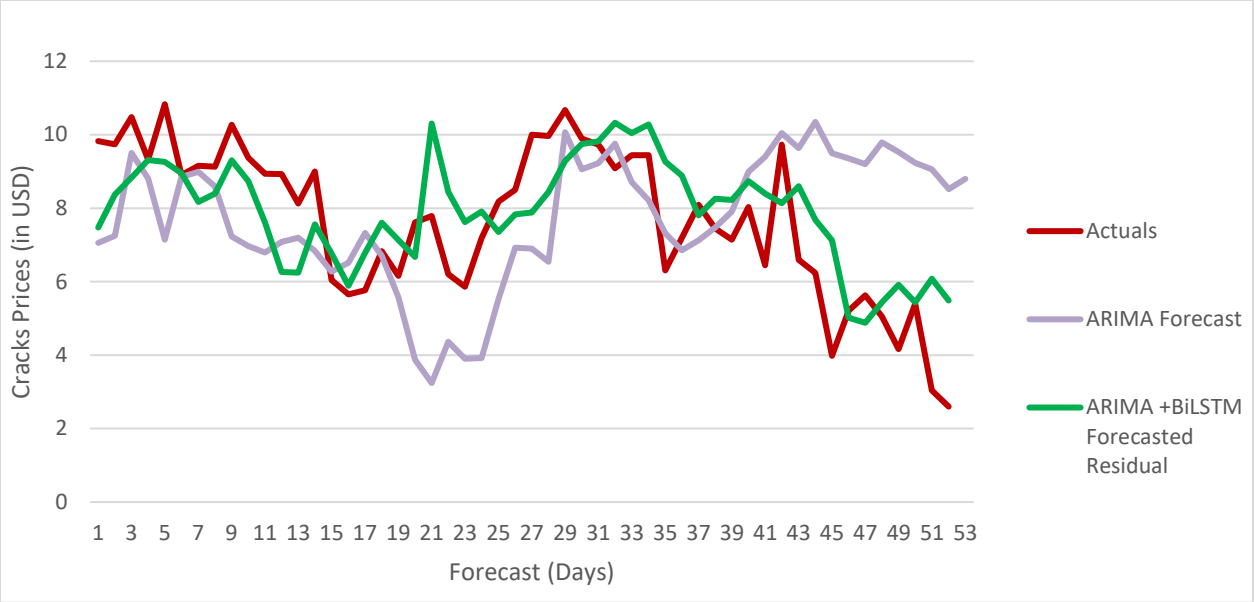


Figure 44 : Plot of Actual Crack Spreads and Forecast using Statistical (ARIMA and ETS) Methods)

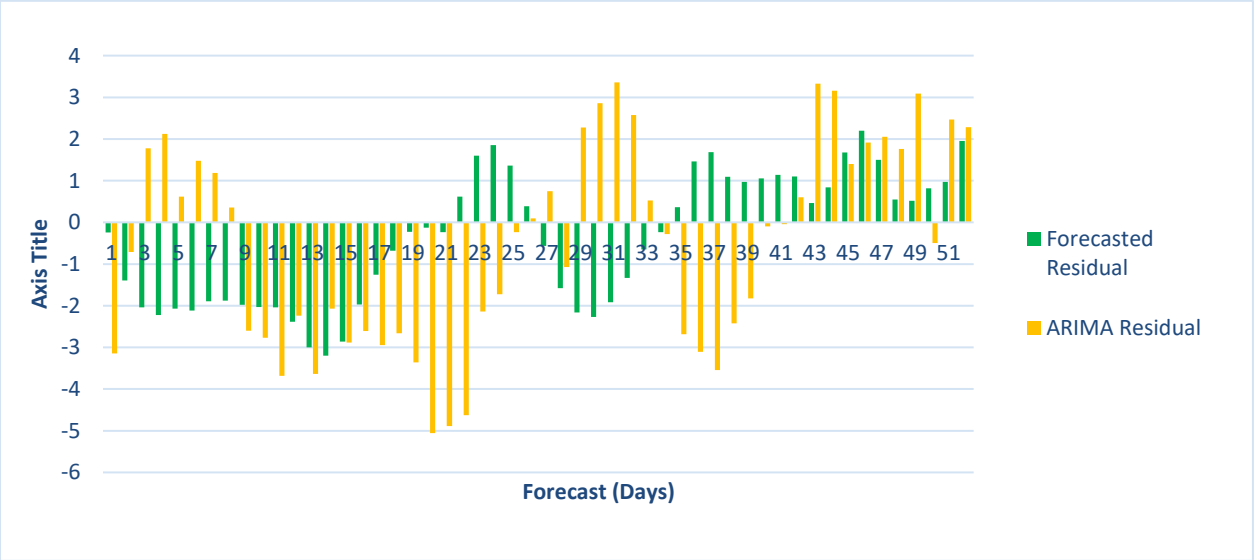


Figure 45 : Comparison plot of ARIMA and ARIMA + BiLSTM forecasts.

Ranking of the various models (Standalone and Hybrid)

The **Table 14** provides the ranking of all the models standalone (ARIMA, ETS, MLP, LSTM, BiLSTM) and hybrid models ARIMA+ BiLSTM and ETS + BiLSTM models. The performance metrices in the order of significance are MAE, RMSE, MAPE and RMAPE.

Table 14 : Ranking of Crack Spread Forecasting models

Model	Standard Deviation	MAE Score	RMSE Score	MAPE Score	RMAPE Score	Rank
ARIMA+BiLSTM	1.5	1.26	1.5	19.89	16.4	1
ETS+BiLSTM	1.67	1.39	1.66	23.54	18.61	2
BiLSTM	2.41	2.03	2.6	34.32	26.55	3
ETS	2.59	2.23	2.66	35.21	29.31	4
ARIMA	2.6	2.24	2.63	35.29	29.21	5
LSTM	3.3	3	3.88	112.97	119.64	6
MLP	17.01	3.87	5.52	49.27	48.39	7

A combination of statistical forecasting technique ARIMA and BiLSTM for forecasting the price of crack offers the best performance. This approach brings together the forecasting strengths of two techniques namely ARIMA, a statistical technique which addresses well the linearity present the data and BiLSTM, a deep learning method, which captures the non-linearity present in the data. Combining these techniques yields improvement in the accuracy of the forecast beyond what each of the techniques individually yield.

Hyper Parameter Tuning

Hyperparameter tuning using Bayesian optimization offers significant improvement in training times in comparison to other methods of optimization like Random search and Grid search. The best results are obtained with 4 Dense Layers, Epoch =10 and activation function = ‘relu’.

Table 15 : The Hyperparameters for the best result in the LSTM Model

S. No.	Hyperparameters	Value
1	Lag	-0.31
2.	No. of Hidden Layers	5
3.	No. of neurons for each layer	9
4.	Dropout Rate	0.67
5.	Learning Rate	0.001
6.	Batch Size	0.78
7.	Epoch size	10
8.	Activation Function	Relu

4.10 Conclusion

Crack spread forecasting is less researched a topic as compared to Crude price forecasting. Most of the existing research on Crack spread forecasting is oriented towards short term forecast with a horizon of few days and is primarily intended for Traders and hedge fund managers. Various techniques such as Vector Auto Regression, Multiple regression, neural networks, General Brownian Motion and ARFIMA has been employed for the purpose of forecasting crack spreads. Time series analysis has been the dominant technique adopted for most of the research on crude oil and crack spread forecasting. On visual examination, it can be deduced that the Crack spread time series has both linear and non-linear component. Any technique employed which is primarily aimed for modelling either the linear or non-linear component for forecasting will invariably be unable to capture the nuances of the other component. This is evident in the results discussed earlier in this chapter.

This research attempts to develop a hybrid model for improving forecast accuracy which employs statistical forecasting technique ARIMA for forecasting the linear component and ANN Deep Learning BiLSTM model for forecasting the non-linear component. This study is based on and extends a similar approach proposed by (Panigrahi & Behera, 2017) by employing Deep Learning LSTM in place of the Multi-Layer Perceptron (MLP) to the domain of crack spread forecasting. For comparison, various Deep Learning Architectures have been studied for hybridization and the results were obtained with BiLSTM. The results indicate that this hybridization results in generation of forecasts with better accuracy compared to individual methods which model linear or non-linear components, respectively.

Deep Learning techniques have a history of success when employed in pattern recognition applications like image categorization, self-driving automobiles, and similar ones. The Multi-Layer Perceptron (MLP), Convolution Neural Network (CNN), and Recurrent Neural Networks (RNN) Deep Learning architectures all perform differently for a given set of tasks, though Long Short-Term Memory (LSTM) neural networks, a type of RNN, are known to perform better on short and long sequences of input streams in contrast to other Deep Learning Neural Network Architectures. In this use case a comparative study of performance of all these Deep Learning architectures on Time Series data, Crack spread prices was undertaken and we observed a that the behavior was congruent on the stated lines i.e., BiLSTM models perform better than MLP and

classical LSTM models. Bayesian optimization has been used to increase the rate of convergence of models and has been found to be effective in reducing run times for large combinations of model and their configurations.

This study can also be extended to include crude / crack futures which is also represented as time series and combination of these series can be expected to increase the efficacy of forecast still better.

The crude oil constitutes 85% of the input cost of a refinery operation. Indian Oil companies, unlike their western counterparts, are integrated oil refining and marketing companies. With limited options to engage in pricing of petroleum products and limited mandate to engage in hedging as a safeguard, managers in Indian oil companies are left with limited options to improve their profitability.

Oil refineries procure and process a 'basket of crude' depending upon the design of the refinery to process a particular type of crude (sweet / sour crude), the configuration of the refinery for producing various distillates (LPG, Gasoil, etc.) and suitable alternatives of crude available at a lesser cost among other factors. The crude(s) thus selected amongst the alternatives available by comparing the forecasted Crack spreads for each of the available options and ensuring the highest forecasted crack spreads can immediately lead to increase in cost optimization and profitability as measured using Gross Refining Margin (GRM). The calculation for computing the impact of unit percentage improvement in forecast accuracy to GRM is convoluted, however rough estimates indicate that it can lead to increase in around 0.1- 0.2 % annual profit for Indian Oil refining and marketing companies amounting to approximately 20-40 crores annually. Crack spread forecasts with higher accuracy, therefore becomes central to this theme.

This improvised and easy to implement method therefore provides a robust and inexpensive way for generating crack spread forecast with higher accuracy to Managers of Oil companies to aid their planning and procurement decisions.

Chapter 5

LPG Demand Forecasting

"Deep learning transforms forecasting from a task of prediction into a realm of understanding and learning from data patterns."

- Juergen Schmidhuber, Computer Scientist

5.1 Introduction

Consumption of LPG in India has been on a steady rise, with 27.6 million metric tonnes consumed in the fiscal year 2020-21, marking a 5% increase from the prior year. This upward trend is projected to continue, thanks to supportive government initiatives. As per the study done by (PPAC, 2022), the number of active LPG domestic customers in India has more than doubled from 14.9 crores in 2015 to 30.5 crores as of April 2022. The number of LPG distributors has also significantly increased, with 25269 in April 2022, compared to 9686 in 2010 and 15930 in 2015. Moreover, there are now 202 LPG bottling plants in India, leading to a 99.8% LPG coverage in 2021, compared to just 56.2% in 2015. The demand for clean cooking fuels is driving India's LPG market, especially among rural and urban households. The Pradhan Mantri Ujjwala Yojana (PMUY) beneficiaries scheme implemented by the government is expected to remain a key driver of LPG market growth in India. In conclusion, the increase in LPG demand, coupled with government initiatives, is forecasted to fuel market growth and lead to significant developments in the foreseeable future.

In urban and semi-urban areas of India, LPG is the most commonly used fuel for cooking. However, the demand for LPG exceeds what domestic refiners produce. To meet this demand, Oil Marketing Companies (OMCs) import bulk LPG and distribute it across the country through their network of LPG bottling plants and distributors. The amount of LPG imported is determined based on projected demands, which can be cyclical. Accurate LPG demand forecasts are crucial for effective supply chain planning, especially given the long lead times involved (Agarwal et al., 2021).

OMCs in India source bulk LPG from their refineries or other OMCs for bottling at their plants. If neither of these options is available, OMCs obtain bottled LPG directly from other OMCs. LPG is packaged into 14.2 Kg cylinders for domestic use and 19 Kg cylinders for commercial use. OMCs use a secondary distribution model by supplying domestic-use cylinders to distributors responsible for selling and distributing LPG to households. Commercial-use LPG accounts for less than 8% of total LPG sales by OMCs (Petroleum Planning and Analysis Cell (PPAC), MoPNG, 2021).

Due to insufficient domestic production, India relies heavily on LPG imports to meet demand. To make effective import decisions, it is essential to accurately estimate the demand shortfall. This forecasted demand can then be fed to an optimisation model which can optimally plan the import requirement and schedule them over a period. This model can take the predicted demands and plan out the supply of LPG to the vast distributor network.

Studies in the past have attempted to forecast the Demand for LPG using Time Series modelling (Lopes et al., 2017). A recent study in Indonesia employed Auto-Regressive Forecasting Technique for estimating the monthly house(Nurochman & Moeis, 2021)esia. (Nurochman & Moeis, 2021). In case of a recent study in Cameroon to determine the household LPG demand, it was found that price, income and urbanization are the important factors determining the demand (Sapnken et al., 2023) .

The objective of this research is to develop a robust LPG demand forecast model for 30 days period ahead for use in Indian context utilizing Time Series analysis and a range of Statistical and AI/ Machine Learning methods. These models have the potential to enhance the supply chain of Oil Marketing companies by delivering accurate forecasts. This research extends earlier research in LPG demand forecasting by employing deep learning models to generate more accurate models as compared to statistical models.

5.2 Conceptual model

LPG demand forecasting for domestic consumption requires predicting the quantity of filled LPG cylinders that OMCs have to deliver in the upcoming months. This specific type of LPG utilized by households is sold in 14.2 kg cylinders and are supplied by oil marketing companies through their Distributors located throughout India. The demand for LPG cylinders at each distributor level is combined into a total aggregate demand for LPG at the national level. The demand for domestic

cylinders on any given day can be determined based on the total number of LPG refill bookings made by customers that day. These customers are typically loyal to their chosen distributor and are unlikely to switch to another distributor of the same or different oil marketing company without good reason. The customer's choice of LPG distributor is usually based on their proximity to the distributor, and the distributors' layout is coordinated among all the oil marketing companies to ensure that customers have limited options. Additionally, each household can only have one LPG connection at a time, which means that customer turnover is low, and demand data for households is easily available.

5.3 Methodology

Oil companies rely solely on analysing past sales data to predict future LPG sales. This process is executed at the consumer level and involves several steps. The base demand for the current year is determined by the actual demand of the previous year for existing customers. For new connections added in the first six months of the previous year, a yearly demand is estimated and added to the base demand. For new connections added in the last six months of the previous year, the average yearly demand of existing customers is taken and added to the base demand. Furthermore, the projected demand for the All-India Level planning area is calculated by projecting a percentage-based growth of demand due to the addition of new customers. This projected percentage is reviewed and approved at the provincial and Head office levels to arrive at the estimate for the upcoming months' All India LPG demand.

5.4 Data Collection and Data Modelling

The sales data of the delivered refill booking is collected and aggregated at various levels of the supply chain. The data is initially collated at the office level, followed by the state and region level, and ultimately at the country level. The sales data belongs to the domestic LPG sales consisting of 14.2Kg cylinders which account for above 90% of the total sale in the country. To ascertain the actual LPG requirement, the forecasting data may be multiplied by 14.2 Kg to arrive at the value. The sales data is acquired on a daily basis, spanning from April 2019 to March 2023, for a total of 1458 data points. The data consists of 2 columns Date and cumulative Sales Quantity of that date. The data downloaded is structured and indexed without any missing values as data of all the dates

as stated above is available. ADF test (Dickey & Fuller, 1979) was performed to test the stationarity of LPG sales data, the results of which are tabulated in **Figure *46***.

```
Results for Augmented Dicky-Fuller Test:
Test Statistic                  -4.849489
p-value                          0.000044
#Lags Used                      20.000000
Number of Observations Used    1437.000000
Critical Value (1%)              -3.434909
Critical Value (5%)              -2.863553
Critical Value (10%)             -2.567842
dtype: float64
Stationary
```

Figure 46: ADF Test Result on data for stationary check

Determining if a time series dataset is stationary is essential for forecasting and modelling of time series data. Stationary time series maintain consistent statistical properties such as mean, variance, and autocorrelation over time, while non-stationary time series exhibit statistical characteristics that change over time. The Augmented Dickey-Fuller (ADF) test is a statistical tool used to determine if a particular time-series dataset is stationary or non-stationary. The ADF statistic is calculated by comparing the dataset's characteristics to the critical value. If the calculated ADF statistic is lower than the critical value, we can conclude that the time series is stationary. Conversely, if the ADF statistic is greater than the critical value, we can conclude that the time series is non-stationary. In this study, the LPG Sales dataset is found to be stationary using the ADF test. Rolling Mean and Rolling Standard Deviation plot on the data was generated on the data to check the variance in the data collection and found that both mean and standard deviation are constant without any variation **Figure 47**.

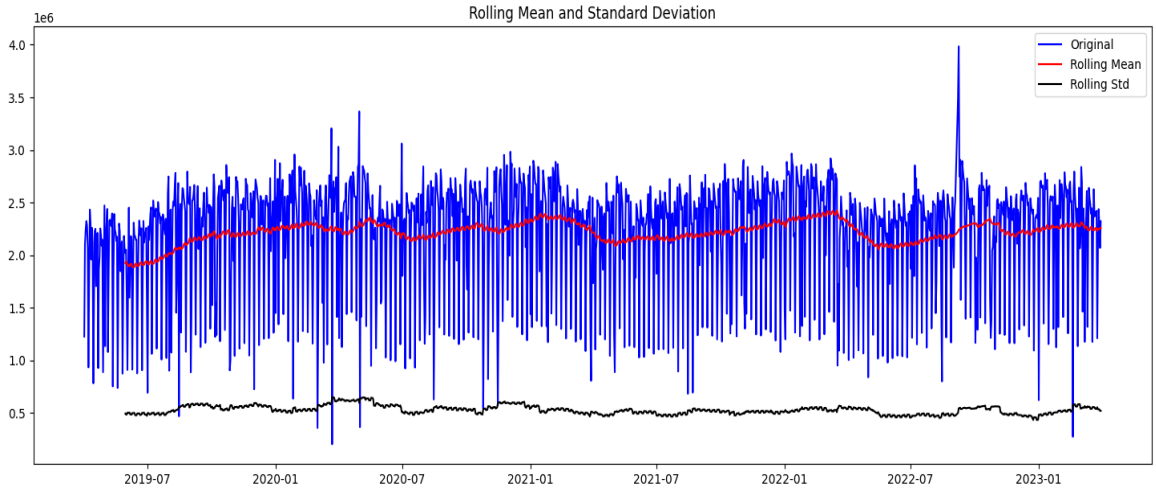


Figure 47: Mean and Standard Deviation trend of data

The data is also checked for the Trend, Seasonality and Residual components. Analysing the output, it was found that the sales data does not have any seasonality or cyclicality and has a linear trend **Figure 48**.

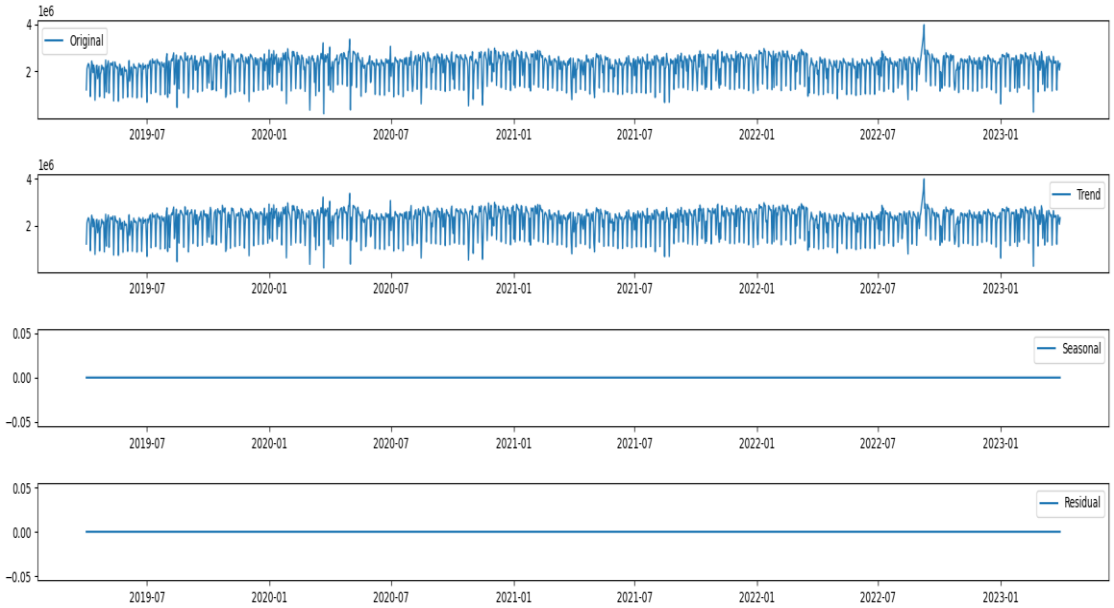


Figure 48: Trend, Seasonality and Residual check on data

For undertaking LPG demand forecasting, first 1428 data points were used as training dataset, while the remaining 30 data points served as testing dataset. The models were implemented in Python programming language. The outcomes were then carefully compared and analysed for both forecasting purposes and to determine the accuracy of each model on the dataset.

5.5 Selection of Forecasting Modelling Techniques

Time series forecasting has been a very extensively researched field in the domain of Energy forecasting, stock markets and inventory demand forecasting. With the advent of Deep Learning models, a study was undertaken to compare the relative performance of Statistical, Machine Learning and Deep Learning model (Makridakis et al., 2023). The multi-step forecasting techniques ARIMA and ETS were considered as reference models whereas MLP technique was considered as the candidate for Machine Learning Model. The study also considered DeepAR and LSTM as the Deep Learning Models for study. Accordingly, these techniques were also employed in the studies. Additionally, Simple Auto-Regression (AR) model was used as a baseline method for our studies based on our research in other areas. Whereas in the study hybrid Statistical and Deep Learning models in the study, the hybrid models to keep the model building exercise simple. One of the learnings from the study was hyperparameter tuning which is an essential part of modelling Deep Learning techniques. Likewise, Bayesian hyper-parameter tuning to arrive at the best possible hyperparameter combinations for this study.

1. Forecasting Methodology using statistical methods (AR/ARIMA/ETS)

For the purpose of the study, day wise aggregate of LPG refill bookings on all India basis of 14.2 kg cylinder is taken as the base demand for LPG. Time series forecasting technique was used to forecast the demand for 30 days in future.

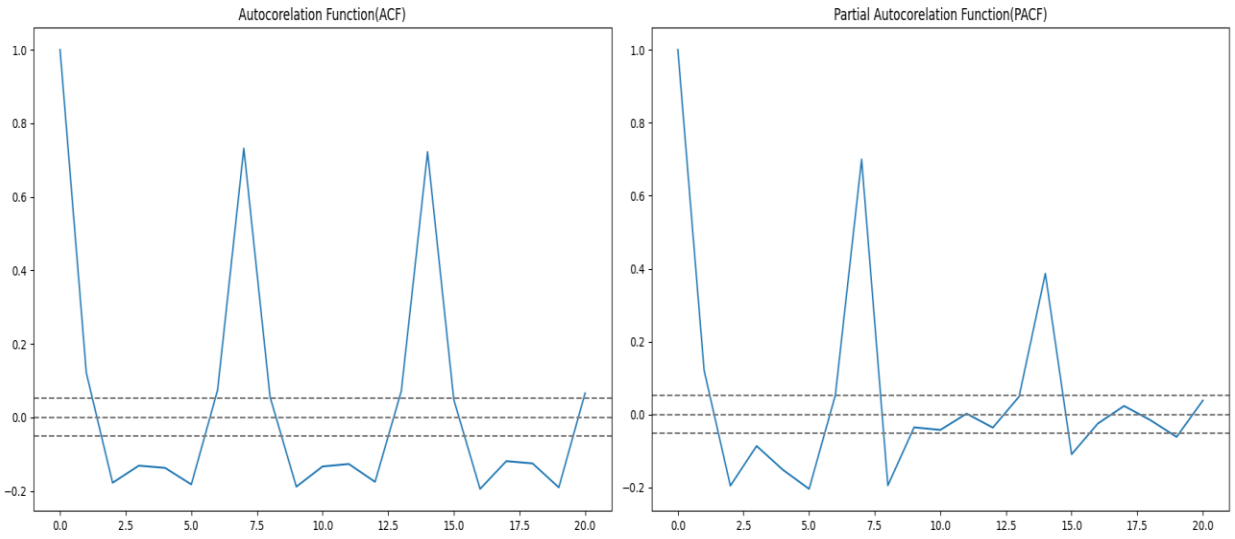


Figure 49: ACF / PACF Plot of Data

In the above ACF plot, , we observe the following:

i. Significant Autocorrelation at Specific Lags: There are several lags where the autocorrelation coefficients extend beyond the blue shaded area (confidence interval). Specifically, a significant positive autocorrelations at lags around 1, 6, and 13 is observed. This suggests that there's a pattern in the data where values are positively correlated with their values a certain number of time periods in the past.

ii. Potential Seasonality: There is a clear oscillating pattern, alternating between positive and negative correlations suggesting a seasonal pattern.

In the above PACF plot, we can observe the following:

i. Significant Peak at around Lag 6 and 13: The most striking feature is the large, significant peak at lag 6 and 13. This peak extends well beyond the confidence interval (the dashed lines).

ii. No Significant Autocorrelation Beyond Lag 13: After lag 13, the partial autocorrelation coefficients fall within the confidence interval, suggesting no significant direct relationship between the current data point and its values at those higher lags.

Note: The second ACF and PACF plot is shorter because in the acf and pacf methods parameter values for lags = 20 is specifically assigned.

. Various scenarios need to be tested out along with the respective Akaike Information Criterion (AIC) values before the best model can be chosen.

2. Forecasting Methodology using Deep Learning methods (FFNN/LSTM/BiLSTM)

Deep Learning models namely FFNN, LSTM and BiLSTM were developed in Python programming language. The forecasts are generated for 30 days ahead. The sales data post the data pre-processing stage was fed into the Deep Learning models. The hyperparameters were tuned using Bayesian optimization. The performance of the models was compared using metrices namely Mean Error (ME), Standard Deviation (SD), Mean Square Errors (MSE), Root Mean Square Errors (RMSE) and Mean Absolute Percentage Errors (MAPE). These metrices were then tabulated and plotted for comparison.

The optimized models generated using Bayesian Optimization is depicted in **Figure 50**.

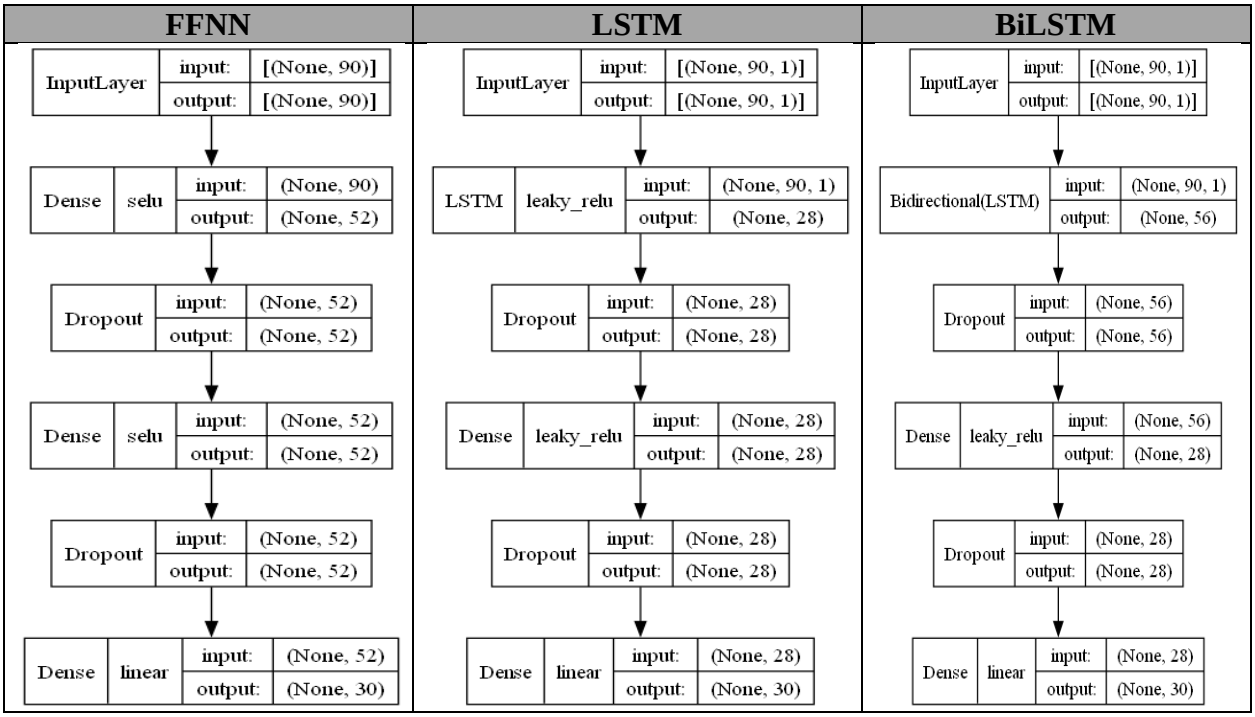


Figure 50: Plot of Neural Network Model Architecture of Deep Learning Models

5.6 Results and Findings

When analyzing data distribution, Mean Error and Standard Deviation (SD) are commonly used. However, when it comes to time series forecasting, different other metrics are utilized to compare machine learning models, such as MSE, RMSE, and MAPE. The choice of metric depends on the specific application and industry. Nonetheless, the following guidelines are typically observed:

MAPE: A MAPE of 10% or less is considered highly accurate, while a MAPE of 20% or less is considered good. A MAPE ranging from 20% to 50% is deemed reasonable, and a MAPE greater than 50% is considered poor or inaccurate forecasting (Montaño Moreno et al., 2013).

RMSE: The acceptable level of RMSE varies depending on the business's needs(Shcherbakov et al., 2013b). When evaluating the reasonableness of RMSE for demand forecasting, the following factors should be considered:

i. **Frequency of the forecasts:** More frequent forecasts (e.g., daily or weekly) tend to have higher RMSE than less frequent forecasts (e.g., monthly or quarterly) because they are more challenging to predict accurately.

ii. **Level of aggregation:** Forecasts for more aggregated data (e.g., total demand for a product category) tends to have lower RMSE than forecasts for less aggregated data (e.g., demand for a specific product) because aggregated data is less noisy and easier to predict.

iii. **The horizon of the forecasts:** Forecasts for shorter horizons (e.g., next week or next month) tend to have lower RMSE than forecasts for longer horizons (e.g., next year or next five years) because shorter-term forecasts are less subject to unexpected changes in the environment. An RMSE that is less than or equal to the standard deviation of historical data is generally considered good, while an RMSE greater than the standard deviation of the historical data is considered poor. This is because the standard deviation of historical data is a measure of how much the data varies. If the RMSE of a forecast is less than or equal to the standard deviation of the historical data, it means that the forecast is at least as good as the historical average. However, if the RMSE of a forecast is greater than the standard deviation of the historical data, it means that the forecast is not as good as the historical average. When conducting comparative studies, the model with the least RMSE measure is typically considered the best option.

1. Performance of Statistical ML Models (AR/ARIMA/ETS)

The plot of original and prediction of all the statistical models used in the study is depicted below.

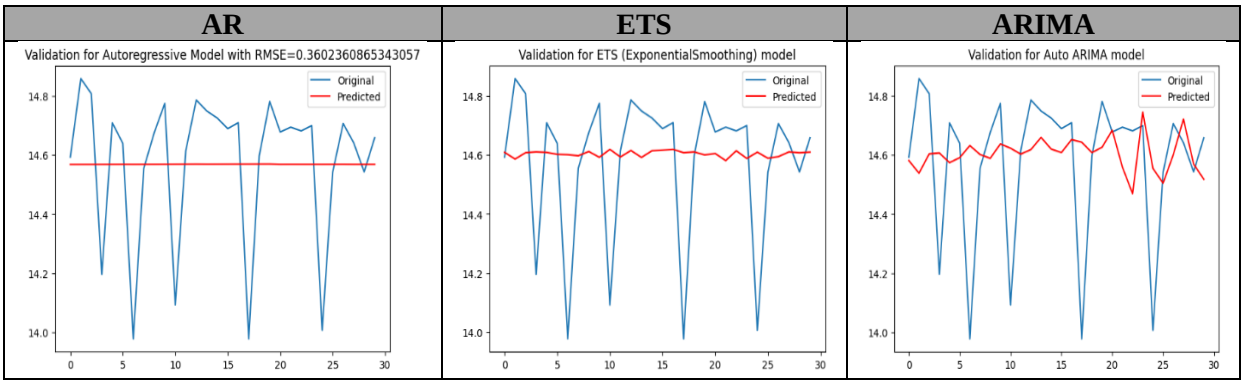


Figure 51: Prediction by statistical methods

Based on the results in Table 16 and it is observed that ETS model generates forecast with minimum error rate of MAPE at 1.206 while RMSE,MSE, SE error values are minimum in AR model. However, the MAPE metric value for the AR Residual technique is the highest. Although MAPE is easier to interpret and better suited for Linear forecasting, it is more sensitive to outliers. Hence, when dealing with conflicting scenarios, more emphasis is given to MSE / RMSE while interpreting and comparing the performance of the models.

Table 16: Performance metrices for statistical models

Header	AR	ARIMA	ETS
Mean	-0.010	0.260	0.026
Standard Deviation	0.254	0.261	0.257
Mean Square Error (MSE)	0.062	0.067	0.064
Root Mean Square Error (RMSE)	0.250	0.258	0.254
Mean Absolute Percentage Errors (MAPE)	1.298	1.222	1.206

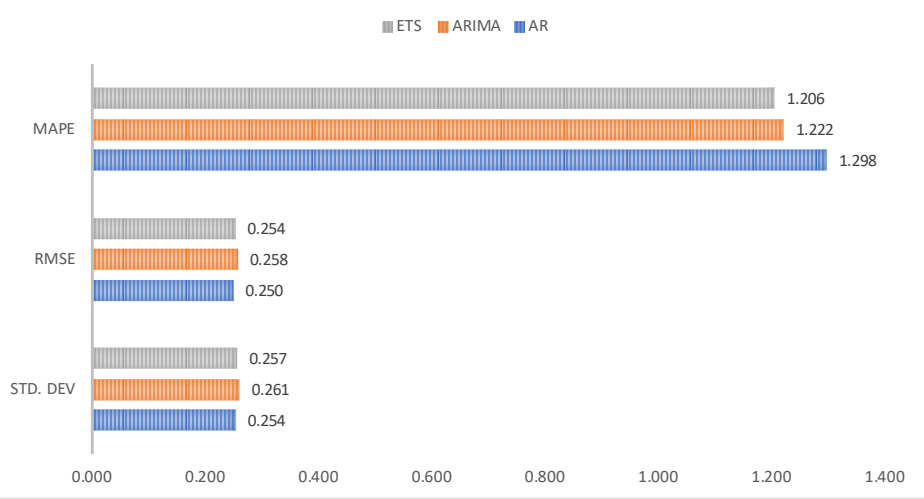


Figure 52: Comparison of Performance Metrices for Statistical Models

2. Performance of ANN - Deep Learning Models (FFNN/LSTM/BiLSTM)

The **Figure *53*** plots the validation for original and prediction of all the deep learning models used in the study.

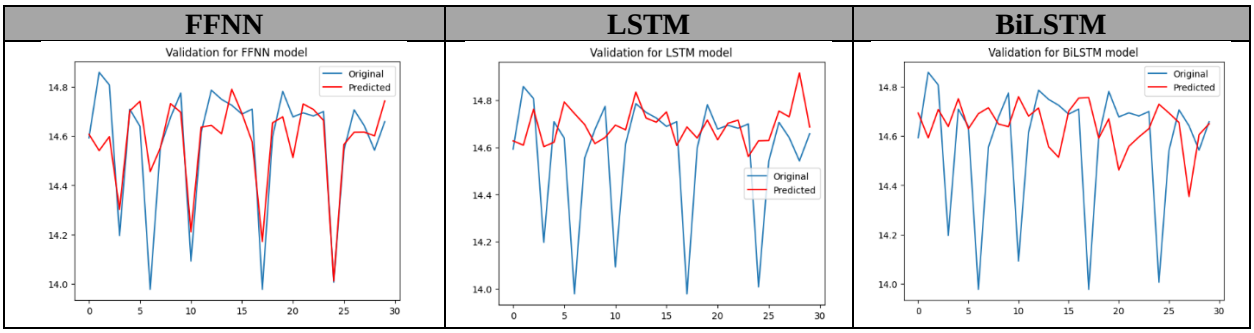


Figure 53: Validation of forecasts by Deep Learning Models

Training loss and validation loss were plotted against the number of training epochs to visualize how the model is performing as it is trained. In general, the goal was to minimize both training

loss and validation loss. However, a model with low training loss may not necessarily have low validation loss. This is because the model may be overfitting the training data. The graph plots graph of training and validation loss for the FFNN, LSTM, BiLSTM models used the study is depicted in **Figure *54***.

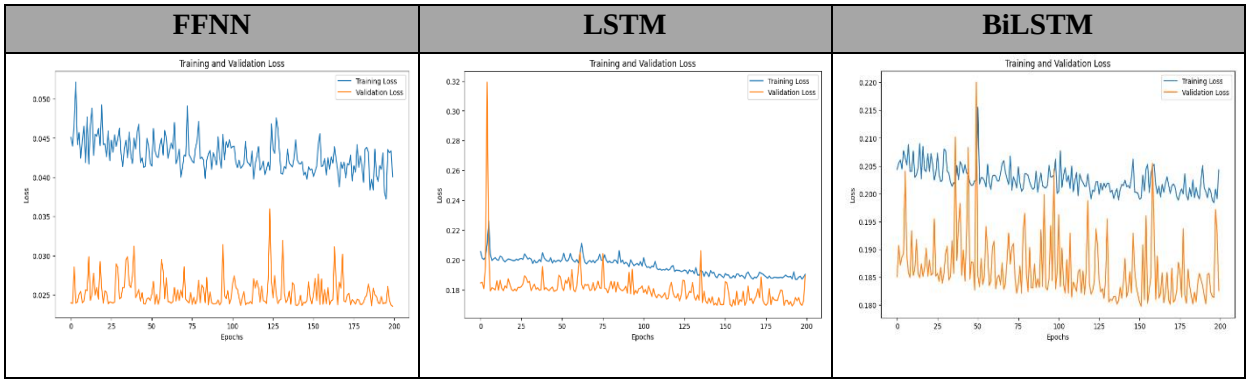


Figure 54: Training and Validation loss in Deep Learning Models with epoch=200

In the tests conducted with epoch = 200, it was observed that training loss and validation loss for LSTM models were very similar implying that the model is generalizing well to unseen data. FFNN and BiLSTM on the other hand had differences in training and validation loss implying that the model was overfitting the data. To overcome this cross validation and smaller batch size was used. Analysing the output of the deep learning models used for the forecasting (**Table 17 / Figure *55***), it can be observed that **Feed Forward Neural Networks (FFNN)** model exhibited the minimum values of MSE, RMSE and MAPE metrices. Thus, in this case it can be considered that the FFNN model gives the best performance with respect to the forecast accuracy.

Table 17: Performance metrices for Deep Learning Models

Header	FFNN	LSTM	BiLSTM	Deep AR
Mean	0.000	0.112	0.066	0.015
Standard Deviation	0.142	0.261	0.298	0.442
Mean Square Error (MSE)	0.019	0.078	0.090	0.189
Root Mean Square Error (RMSE)	0.139	0.280	0.300	0.435
Mean Absolute Percentage Errors (MAPE)	0.666	1.232	1.384	2.207

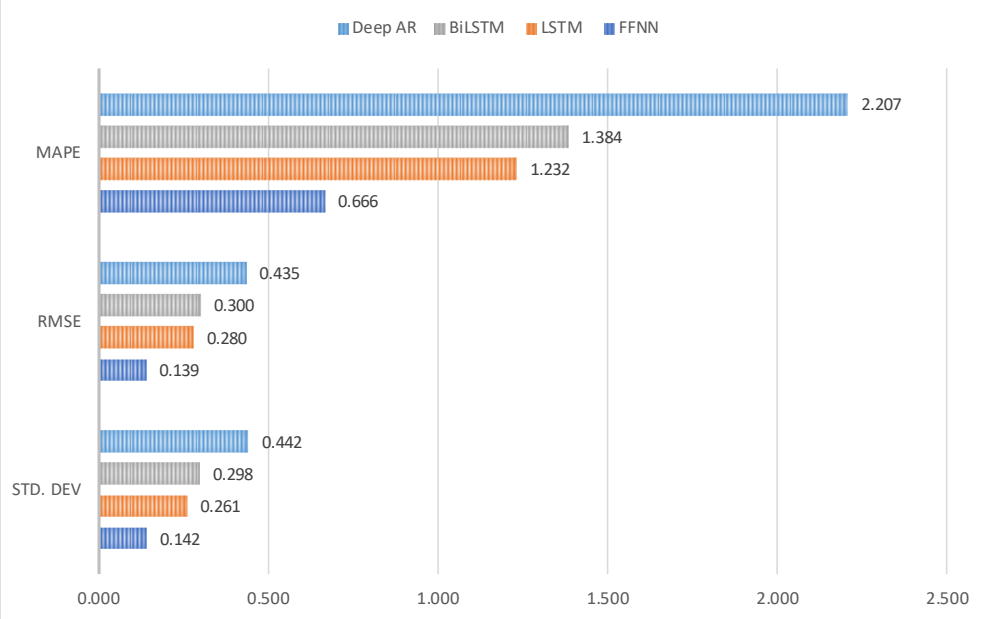


Figure 55: Performance Metrices for Deep Learning Models

The combination of hyperparameters delivering the best results is listed in **Table 18** below:

Table 18: The Hyperparameters for the best result in the FFNN Model

S. No.	Hyperparameter	Value
1	Lag	90
2	No. of Hidden Layers	2
3	No. of neurons for each layer	90,90,52,52
4	Dropout Rate	0.1
5	Learning Rate	0.001
6	Batch Size	32
7	Epoch size	200
8	Activation Function	selu

5.7 Comparison of Statistical ML and ANN Deep Learning Models

RMSE and MAPE has been used in conjunction to interpret the performance of machine learning models for univariate LPG sales time series data for forecasting demand using statistical and deep learning models to derive complementary insights into the model's accuracy. RMSE measures the average squared difference between the predicted and actual values. It is a good metric for measuring the overall accuracy of a model, but it is sensitive to outliers. MAPE measures the average absolute percentage difference between the predicted and actual values. It is a good metric

for measuring the accuracy of a model on a relative basis, but it can be affected by outliers, and it is not defined when the actual values are zero. By using both RMSE and MAPE in situations like predicting LPG demand, we get a more comprehensive picture of the model's performance. A model which has both the least RMSE and MAPE (**Figure *56***) therefore would be the most accurate model amongst the options chosen for the study.

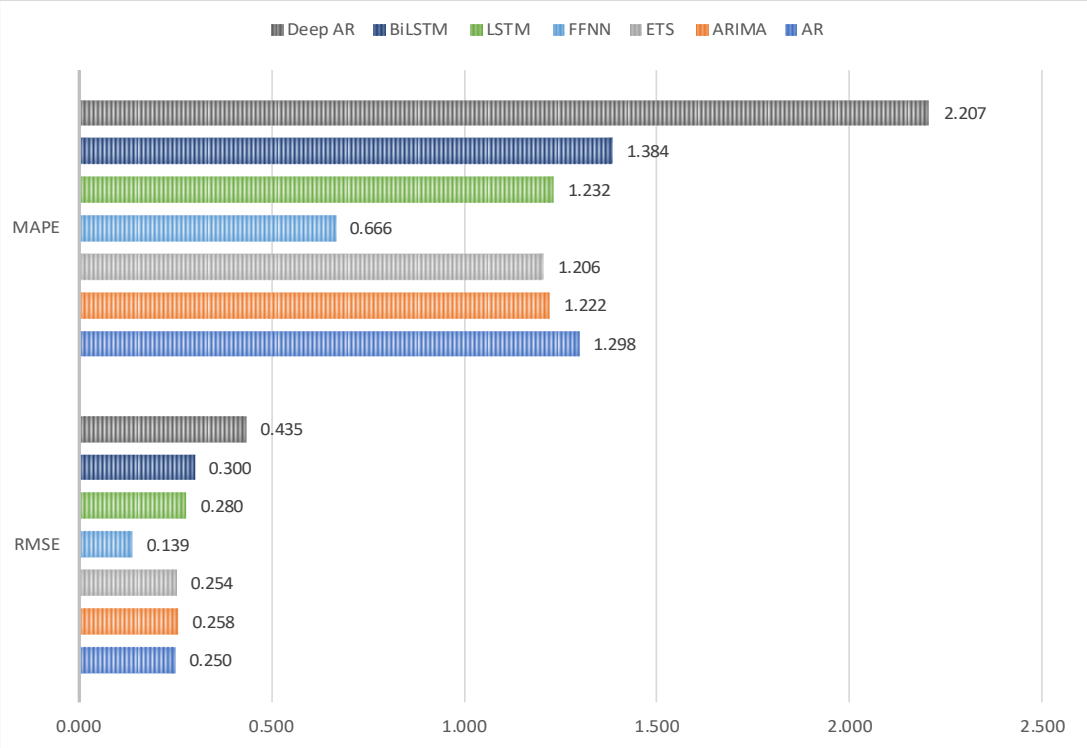


Figure 56: Comparision of Statistical ML Models Vs Deep Learning Models

Both the RMSE and MAPE scores of forecasts done using all the models studied is listed **Table 19** below.

Table 19: Comparison between forecasting techniques on RMSE/MAPE score

Name of Model	RMSE	MAPE	Rank
FFNN	0.13920	0.66567	1
AR	0.24993	1.29791	2
ETS	0.25359	1.20582	3
ARIMA	0.25796	1.22189	4
LSTM	0.27964	1.23178	5
BiLSTM	0.30029	1.38371	6
DeepAR	0.43455	2.20687	7

It is evident from the values above that the **Forward Neural Networks (FFNN)** model exhibited the least RMSE and MAPE scores and accordingly can be adjudged the best performing model.

5.8 Conclusion

This study aimed to evaluate the effectiveness of statistical machine learning techniques such as AR, ARIMA, ETS versus recently developed Deep Learning models such as FFNN, LSTMs etc. The study found the deep learning **Feed Forward Neural Networks (**FFNN) model outperforms all the other models in our study for LPG demand forecasting.

In the realm of LPG demand forecasting through time series analysis, deep learning model FFNN stood out as superior to both statistical & other deep learning and hybrid models for a variety of reasons. One critical factor is the relatively low volatility of LPG demand data, which tends to remain relatively consistent daily. Furthermore, given the longer forecast time horizons (30 days), and the fact that the data being studied is limited to a single variable (namely, daily sales of LPG refill cylinders), this model proved highly effective.

While utilizing a DeepAR model that is solely trained on one time series, it may not exhibit the same level of efficacy as the more traditional forecasting algorithms, such as ARIMA, ETS or other Deep Learning models namely FFNN, LSTM, BiLSTM. Nonetheless, if the dataset contains hundreds of related time series, DeepAR is known to be more effective than the conventional methods. This behaviour has been exhibited is our study as well. It is important to consider that the overall number of observations within all training time series should amount to at least 300.

Statistical Time Series modelling techniques are a very matured field which have been in use and practice over decades. Decision makers find this technique simple to use and the accuracy of the

forecasts generated are good for decision making. The skills required for modelling and usage of this technique is easily acquired and can be rapidly put to use. Deep Learning techniques on the other hand uses neural networks with multiple layers in analysing intricate patterns and relationships for forecasting workloads. In this use case of time series data for LPG demand forecasting, as expected the above exercise demonstrates that forecasts generated by deep learning technique outperformed the statistical models such as AR/ ARIMA / ETS. For this specific use case where there is only one time series, volatility of the data is less and forecasts are desired for 30/60 periods (forecast steps) ahead, the deep learning forecasting technique (FFNN) yielded better results.

Based on the findings above it can be concluded that in situations where the data is characterized by low frequency or volatility and forecast horizon being long (to the order of more than 1 month) ANNs technique FFNN outperforms all other models. The deep learning technique, which uses sophisticated and autonomous algorithms, can better handle data with more structured patterns and longer periods. Further initiatives can be undertaken to study the nuances of applying ANN Deep Learning models to forecast time series data for a still longer period which will yield much more accurate results. One such option to enhance the efficacy of forecasting models is hybridization of techniques within and across genres which can be explored. By combining the strengths of ANNs and statistical machine learning, researchers can leverage the power of both approaches to achieve even better results. This hybridization can lead to improved accuracy, adaptability, and robustness in analyzing distinct types of data. The hybridization of statistical and deep learning techniques has the potential to deliver superior results in various other domains. By combining the strengths of both approaches, researchers can leverage the interpretability and robustness of statistical techniques with the predictive power and flexibility of deep learning.

Furthermore, the emergence of newer deep learning techniques, such as transformers (Generative Pre-trained Transformers) , presents exciting opportunities for enhancing forecast accuracy. These techniques excel in handling multiple time series data, making them particularly well-suited for applications involving complex and interconnected datasets. The integration of transformers enables the modeling of long-term dependencies and captures intricate patterns in time series data, resulting in more accurate predictions. The DeepAR technique utilizes recurrent neural networks to capture temporal dependencies using multiple time series and incorporates probabilistic

forecasting, which provides valuable uncertainty estimates. By embracing these advanced deep learning techniques and combining them with statistical methods, researchers can unlock the potential for improved forecasting accuracy. This hybrid approach allows for more accurate predictions, better capturing the underlying dynamics and complexities of the data. In the context of LPG Demand forecasting, multiple time series can be explored and be integrated along with the demand time series to explore the impact and increase forecast accuracy.

In conclusion, this research highlights the importance of understanding the specific characteristics and requirements of the data when selecting the appropriate modeling technique. Tailoring the choice of approach based on the nature of the data will lead to more accurate predictions and better decision-making in various fields such as finance, economics, and risk management.

Chapter 6

Conclusion

"The development of full artificial intelligence could spell the end of the human race. It would take off on its own, and re-design itself at an ever-increasing rate. Humans, who are limited by slow biological evolution, couldn't compete and would be superseded."

- Stephen Hawking,

6.1 Summary

Practicing managers engaged in the planning and optimization of petroleum product supply chains have consistently recognized the critical importance of obtaining highly accurate forecasts for petroleum product prices, including crude and distillates/crack spread, as it directly influences Gross Refining Margin (GRM) and overall profitability.

The field of crack spread forecasting has seen limited interest compared to crude price forecasting, particularly with an emphasis on short-term forecasts from traders and hedge fund managers. Existing methods include Vector Auto Regression, Multiple regression, neural networks, General Brownian Motion, and ARFIMA, primarily employing time series analysis. However, visual examination reveals both linear and non-linear components in crack spread time series, challenging methods that focus solely on one aspect.

This thesis has proposed a hybrid model to enhance forecast accuracy, combining ARIMA for the linear component and ANN Deep Learning BiLSTM for the non-linear component. Building on a similar approach, this study compares various deep learning architectures, concluding that BiLSTM models outperform MLP and classical LSTM models in crack spread forecasting. Bayesian optimization is employed for faster convergence, proving effective in reducing run times for model configurations.

The research can be extended to include crude/crack futures, anticipating increased forecast efficacy. In the context of Indian oil companies, where crude oil constitutes a significant input cost, improving profitability is challenging due to limited pricing and hedging options. The study emphasizes the importance of accurate crack spread forecasts for cost optimization and

profitability, with potential annual profit increases of 0.1-0.2%. The proposed hybrid model is presented as an easy to implement, practical and cost-effective solution for oil company managers to enhance crack spread forecast accuracy, aiding in planning and procurement decisions.

For forecasting LPG demand use-case, in this thesis, the efficacy of various statistical machine learning techniques are studied, including AR/ARIMA/ETS, in comparison to deep learning models such as FFNN and LSTMs for LPG demand forecasting. It is found that the FFNN Deep Learning model outperforms other models in terms of accuracy. This is attributed to low volatility of LPG demand data, which remains relatively consistent on a daily basis. While traditional forecasting algorithms like ARIMA/ETS and even other deep learning models might be less effective for longer forecast horizons and single-variable datasets, the FFNN model demonstrates better effectiveness.

This study emphasizes that statistical time series modeling techniques are mature and widely used due to their simplicity and accuracy. However, for LPG demand forecasting with low volatility and longer forecast horizons, deep learning techniques, particularly FFNN, are more effective. In scenarios with low-frequency data and extended forecast horizons, deep learning outperforms other models. Further exploration is recommended for applying deep learning models for even longer forecast periods, suggesting potential improvements through hybridization of techniques across genres. Combining the strengths of statistical and deep learning methods can enhance accuracy, adaptability, and robustness.

The emergence of newer deep learning techniques, such as transformers, presents opportunities to improve forecast accuracy, especially in handling multiple time series data. Integrating transformers enables modeling long-term dependencies and capturing intricate patterns, providing more accurate predictions. Advanced deep learning techniques can be combined with statistical methods to unlock improved forecasting accuracy. This hybrid approach allows for more accurate predictions, particularly in understanding complex and interconnected datasets.

This study underscores the importance of selecting modeling techniques based on the specific characteristics and requirements of the data. Tailoring the approach to the nature of the data enhances prediction accuracy, facilitating better decision-making in various fields, including finance, economics, and risk management.

6.2 Limitations of study

The studies undertaken as part of this research thesis have been drawn from the context of Indian Refiners. The crude / product selection chosen therefore is specific to how crack spread / profitability is computed in the Indian context. Likewise, the demand forecasting of LPG has also been modelled in the Indian context. However, the techniques / frameworks that have been developed are universal. A small amount of adjustment in approach / methodology with respect to nuances of the use case will need to be undertaken when techniques are being applied in other geographical contexts. The studies undertaken have attempted to generate forecast of medium duration i.e 1-3 months. However, forecasts over a larger time horizon (3monh – 6 month) can also be extremely useful and would therefore require improvisation of the techniques developed herein.

6.3 Recommendations and Future Research

This study establishes that hybrid models, amalgamating statistical and AI techniques, yield superior results. Future extensions could explore the effectiveness of these steps for multi-step/period forecasting. Additionally, newly discovered ANN architectures like Autoencoders, Transformers, and Deep AR can be studied for potential improvements in forecast accuracy. Combining these techniques with optimization methods, such as portfolio optimization, holds promise for sustaining profitability. The recent advances in Artificial Neural Network ecosystems, GPUs, TPUs, programming languages, and machine learning libraries have accelerated research in petroleum product demand, price forecasting, and supply chain planning. While this study focuses on univariate time series and short-term forecasts, future research could extend the hybrid model for medium- and long-term forecasts, analyze its performance for multistep forecasting, explore the effectiveness of other deep learning models, consider a portfolio approach for diverse forecasting horizons, investigate the impact of bagging on ARIMA models and hybrid models, and delve into hyperparameter tuning methods using Bayesian optimization.

The hybrid model, combining ARIMA and BiDirectional LSTM, can also be tested for medium and long-term forecasts. The performance of these models for multistep forecasting can be studied, along with the effectiveness of combining other deep learning models. A portfolio approach for different forecasting horizons and studying the impact of bagging on ARIMA models and hybrid

models can provide further insights. The recent advancements in Artificial Neural Network ecosystems and related technologies are expected to result in promising products and processes in the immediate to near future.

Bibliography

Agarwal, R., Mittal, N., Patterson, E., & Giorcelli, M. (2021). Evolution of the Indian LPG industry: Exploring conditions for public sector business model innovation. *Research Policy*, *50*(4). https://doi.org/10.1016/j.respol.2020.104196

Agnolucci, P. (2009). Volatility in crude oil futures: A comparison of the predictive ability of GARCH and implied volatility models. *Energy Economics*, *31*(2), 316–321. https://doi.org/10.1016/j.eneco.2008.11.001

Alquist, R., & Kilian, L. (2010). What do we learn from the price of crude oil futures? *Journal of Applied Econometrics*, *25*(4), 539–573. https://doi.org/10.1002/jae.1159

Apergis, N., & Vouzavalis, G. (2018). Asymmetric pass through of oil prices to gasoline prices: Evidence from a new country sample. *Energy Policy*, *114*, 519–528. https://doi.org/10.1016/j.enpol.2017.12.046

Barbosa, G., Sousa, D., Besarria, C. da N., Lima, R., & de Jesus, D. P. (2024). Asymmetry in the prices of crude oil and diesel and gasoline prices in Brazil. *Journal of Economic Studies*, *51*(1), 71–95. https://doi.org/10.1108/JES-08-2022-0437

Baumeister, C., Kilian, L., & Zhou, X. (2013). *Are Product Spreads Useful for Forecasting? An Empirical Evaluation of the Verleger Hypothesis*.

Baumeister, C., Kilian, L., & Zhou, X. (2018). ARE PRODUCT SPREADS USEFUL for FORECASTING OIL PRICES? AN EMPIRICAL EVALUATION of the VERLEGER HYPOTHESIS. *Macroeconomic Dynamics*, *22*(3), 562–580. https://doi.org/10.1017/S1365100516000237

Bhutto, F. A., Ghumro, I. A., & Lakhan, A. B. (2023). Causal Relationship Between Spot Price & Future Derivative Prices – Case of Crude Oil. *Research Journal for Societal Issues*, *5*(2), 396–408. https://doi.org/10.56976/rjsi.v5i2.126

Box, G. E. P., & Jenkins, G. M. (1976). *Time Series Analysis - Forecasting and Control*. Holden-Day, Inc.

Burba, D. (2019). An overview of time series forecasting models. *Towards Data Science*, 1–20. https://towardsdatascience.com/an-overview-of-time-series-forecasting-models-a2fa7a358fcb

Caniato, F., Kalchschmidt, M., & Ronchi, S. (2011). Integrating quantitative and qualitative forecasting approaches: Organizational learning in an action research case. *Journal of the Operational Research Society*, *62*(3), 413–424. https://doi.org/10.1057/jors.2010.142

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. https://doi.org/10.1016/j.energy.2018.12.016

Chen, Y., He, K., & Tso, G. K. F. (2017). Forecasting Crude Oil Prices: A Deep Learning based Model. *Procedia Computer Science*, *122*, 300–307. https://doi.org/10.1016/j.procs.2017.11.373

Chesnes, M. (2016). Asymmetric Pass-Through in U.S. Gasoline Prices. *The Energy Journal*, *37*(1), 153–180. https://doi.org/10.5547/01956574.37.1.mche

Chiroma, H., Abdul-Kareem, S., Shukri Mohd Noor, A., Abubakar, A. I., Sohrabi Safa, N., Shuib, L., Fatihu Hamza, M., Ya'u Gital, A., & Herawan, T. (2016). A Review on Artificial Intelligence Methodologies for the Forecasting of Crude Oil Price. *Intelligent Automation and Soft Computing*, *22*(3), 449–462. https://doi.org/10.1080/10798587.2015.1092338

Clyde Russell. (2023, August 24). *Asia oil refining margins climb, but a peak looms*. Reuters. https://www.reuters.com/markets/commodities/asia-oil-refining-margins-climb-peak-looms-russell-2023-08-24/

Clyde Russell. (2024, May 24). *Weak Asia refining margins see swing to crude from the Americas*. Hydrocarbon Processing . https://www.hydrocarbonprocessing.com/news/2024/05/weak-asia-refining-margins-see-swing-to-crude-from-the-americas/

Correia, A., Costa E Silva, E., Lopes, C., Henriques, C., Henriques, F., Pinto, M., Monteiro, M., Lopes, R. B., & Sapata, A. (2017). *LPG Demand Forecast using Time Series*. http://www.apetro.pt

Dickey, D. A., & Fuller, W. A. (1979). *Distribution of the Estimators for Autoregressive Time Series With a Unit Root, Journal of the American Statistical Association, Vol. No. 366, pp. . 74*(March 2013), 427–431.

Diebold, F. X. (2007). *Elements of Forecasting* (Fourth Edition). Thomson - South Western.

Dunis, C. L., Laws, J., & Evans, B. (2006a). Modelling and trading the gasoline crack spread: A non-linear story. *Derivatives and Hedge Funds*, *12*(1), 126–145. https://doi.org/10.1057/9781137554178

Dunis, C. L., Laws, J., & Evans, B. (2006b). Modelling and Trading the Gasoline Crack Spread: A Non-Linear Story Practical applications. In *Journal of Business and Economic Statistics* (Vol. 19).

Ebrahimi Meimand, S., Ganjalikhan Hakemi, H., & Nizam Ismail, H. (2012). Demand Forecasting: A Review on Qualitative Methods and Surveyed Done By Delphi Technique. *GEBF*.

Ederington, L. H. (1979). The Hedging Performance of the New Futures Markets. In *THE JOURNAL OF FINANCE* (Issue 1). http://about.jstor.org/terms

Ederington, L. H., Fernando, C. S., Hoelscher, S. A., Lee, T. K., & Linn, S. C. (2019a). A review of the evidence on the relation between crude oil prices and petroleum product prices. *Journal of Commodity Markets*, *13*(September), 1–15. https://doi.org/10.1016/j.jcomm.2018.09.002

Ederington, L. H., Fernando, C. S., Hoelscher, S. A., Lee, T. K., & Linn, S. C. (2019b). Characteristics of petroleum product prices: A survey. *Journal of Commodity Markets*, *14*(September 2018), 1–15. https://doi.org/10.1016/j.jcomm.2018.09.001

Engle, R. F., & Grangeri, C. W. J. (1987). CO-INTEGRATION AND ERROR CORRECTION: REPRESENTATION, ESTIMATION, AND TESTING. In *Source: Econometrica* (Vol. 55, Issue 2).

Ewing, B. T., & Thompson, M. A. (2018). Modeling the Response of Gasoline-Crude Oil Price Crack Spread Macroeconomic Shocks. *Atlantic Economic Journal*, *46*(2), 203–213. https://doi.org/10.1007/s11293-018-9582-x

Figuerola-Ferretti, I., McCrorie, J. R., & Paraskevopoulos, I. (2020). Mild explosivity in recent crude oil prices. *Energy Economics*, *87*. https://doi.org/10.1016/j.eneco.2019.05.002

Fousekis, P., & Grigoriadis, V. (2017). Price co-movement and the crack spread in the US futures markets. *Journal of Commodity Markets*, *7*, 57–71. https://doi.org/10.1016/j.jcomm.2017.08.003

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

Gao, S., & Lei, Y. (2017). A new approach for crude oil price prediction based on stream learning. *Geoscience Frontiers*, *8*(1), 183–187. https://doi.org/10.1016/j.gsf.2016.08.002

Geman, H. (2005). *Commodities and Commodity Derivatives: Modeling and Pricing for Agriculturals, Metals, and Energy*. Wiley Finance.

Geman, H., & Smith, W. O. (2013). Theory of storage, inventory and volatility in the LME base metals. *Resources Policy*, *38*(1), 18–28. https://doi.org/10.1016/j.resourpol.2012.06.014

Gerolimetto, M. (2010). *ARIMA and SARIMA models and ARIMA*. *1*, 1–5.

Gu, Q., Lu, N., & Liu, L. (2019). A novel recurrent neural network algorithm with long short-term memory model for futures trading. *Journal of Intelligent and Fuzzy Systems*, *37*(4), 4477–4484. https://doi.org/10.3233/JIFS-179280

Gupta, V., & Pandey, A. (2018). Crude Oil Price Prediction Using LSTM Networks. *International Journal of Computer and Information Engineering*, *12*(3), 226–230.

Haidar, I., Kulkarni, S., & Pan, H. (2008). Forecasting model for crude oil prices based on artificial neural networks. *ISSNIP 2008 - Proceedings of the 2008 International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, 103–108. https://doi.org/10.1109/ISSNIP.2008.4761970

Haigh, M. S., & Holt, M. T. (2002). Crack spread hedging: Accounting for time-varying volatility spillovers in the energy futures markets. *Journal of Applied Econometrics*, *17*(3), 269–289. https://doi.org/10.1002/jae.628

Ho, S.L. and Xiw, M. (1998). The use of ARIMA Models for Reliability Forecasting and Analysis, Elsevier Science Ltd. *Science*, *35*(98), 213–216.

Hu, J. W. S., Hu, Y. C., & Lin, R. R. W. (2012). Applying neural networks to prices prediction of crude oil futures. *Mathematical Problems in Engineering*, *2012*(August 2006). https://doi.org/10.1155/2012/959040

Jammazi, R., & Aloui, C. (2012). Crude oil price forecasting: Experimental evidence from wavelet decomposition and neural network modeling. *Energy Economics*, *34*(3), 828–841. https://doi.org/10.1016/j.eneco.2011.07.018

Kamyabi, N., & Chidmi, B. (2023). Asymmetric Price Transmission between Crude Oil and the US Gasoline Market. *Journal of Risk and Financial Management*, *16*(7), 326. https://doi.org/10.3390/jrfm16070326

Karathanasopoulos, A., Dunis, C., & Khalil, S. (2016a). Modelling, forecasting and trading with a new sliding window approach: the crack spread example. *Quantitative Finance*, *16*(12), 1875–1886. https://doi.org/10.1080/14697688.2016.1211796

Karathanasopoulos, A., Dunis, C., & Khalil, S. (2016b). Modelling, forecasting and trading with a new sliding window approach: the crack spread example. *Quantitative Finance*, *16*(12), 1875–1886. https://doi.org/10.1080/14697688.2016.1211796

Kuha, J. (2004). AIC and BIC: Comparisons of assumptions and performance. *Sociological Methods and Research*, *33*(2), 188–229. https://doi.org/10.1177/0049124103262065

Kulkarni, S., & Haidar, I. (2009). *Forecasting Model for Crude Oil Price Using Artificial Neural Networks and Commodity Futures Prices*. *2*(1). http://arxiv.org/abs/0906.4838

Kulshrestha, A., Krishnaswamy, V., & Sharma, M. (2020). Bayesian BILSTM approach for tourism demand forecasting. *Annals of Tourism Research*, *83*(April), 102925. https://doi.org/10.1016/j.annals.2020.102925

Lackes, R., Börgermann, C., & Dirkmorfeld, M. (2009). Forecasting the Price Development of Crude Oil with Artificial Neural Networks. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *5518 LNCS*(PART 2), 248–255. https://doi.org/10.1007/978-3-642-02481-8_36

Lisitsa, S., Levina, A., & Lepekhin, A. (2019). Supply-chain management in the oil industry. *E3S Web of Conferences*, *110*, 1–10. https://doi.org/10.1051/e3sconf/201911002061

Lopes, C., Correia, A., Silva, E. C. E., & Monteiro, M. (2017). *LPG Demand Forecast using Time Series LPG Demand Forecast using Time Series*. *January 2019*.

Mahringer, S., & Prokopczuk, M. (2015). An empirical model comparison for valuing crack spread options. *Energy Economics*, *51*, 177–187. https://doi.org/10.1016/j.eneco.2015.06.015

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLoS ONE*, *13*(3), 1–26. https://doi.org/10.1371/journal.pone.0194889

Makridakis, S., Spiliotis, E., Assimakopoulos, V., Semenoglou, A.-A., Mulder, G., & Nikolopoulos, K. (2023). Statistical, machine learning and deep learning forecasting methods: Comparisons and ways forward. *Journal of the Operational Research Society*, *74*(3), 840–859. https://doi.org/10.1080/01605682.2022.2118629

Matsuoka, K., & Hamori, S. (2021). Forecasting WTI Futures Prices Using Recurrent Neural Networks. *Review of Integrative Business and Economics Research*, *10*(1), 34–50.

Montaño Moreno, J. J., Palmer Pol, A., Sesé Abad, A., & Cajal Blasco, B. (2013). El índice R-MAPE como medida resistente del ajuste en la previsiońn. *Psicothema*, *25*(4), 500–506. https://doi.org/10.7334/psicothema2013.23

Moroff, N. U., Kurt, E., & Kamphues, J. (2021). Machine Learning and Statistics: A Study for assessing innovative Demand Forecasting Models. *Procedia Computer Science*, *180*, 40–49. https://doi.org/10.1016/j.procs.2021.01.127

Murat, A., & Tokat, E. (2009a). Forecasting oil price movements with crack spread futures. *Energy Economics*, *31*(1), 85–90. https://doi.org/10.1016/j.eneco.2008.07.008

Murat, A., & Tokat, E. (2009b). Forecasting oil price movements with crack spread futures. *Energy Economics*, *31*(1), 85–90. https://doi.org/10.1016/j.eneco.2008.07.008

Mushtaq, R. (2011). Testing Time Series Data for Stationarity. *Ssrn*, 1–19. http://ssrn.com/abstract=1911068

Narayan, P. K., & Narayan, S. (2007). Modelling oil price volatility. *Energy Policy*, *35*(12), 6549–6553. https://doi.org/10.1016/j.enpol.2007.07.020

Nguyen, H. S. (n.d.). *LUT School of Business and Management Master's Programme in Strategic Finance and Analytics Financial Time Series Forecasting with Long Short-Term Memory (LSTM): A Comparative Experiment between Deep Learning and Econometrics*.

Niknam, A., Zare, H. K., Hosseininasab, H., Mostafaeipour, A., & Herrera, M. (2022). A Critical Review of Short-Term Water Demand Forecasting Tools—What Method Should I Use? In *Sustainability (Switzerland)* (Vol. 14, Issue 9). MDPI. https://doi.org/10.3390/su14095412

Nurochman, A., & Moeis, A. O. (2021). *EasyChair Preprint Auto-Regressive Integrated Moving Average (ARIMA) Forecasting for Monthly Household LPG Demand in Indonesia Auto-Regressive Integrated Moving Average (ARIMA) Forecasting for Monthly Household LPG Demand in Indonesia*.

Panigrahi, S., & Behera, H. S. (2017). A hybrid ETS–ANN model for time series forecasting. *Engineering Applications of Artificial Intelligence*, *66*(July), 49–59. https://doi.org/10.1016/j.engappai.2017.07.007

Petroleum Planning and Analysis Cell (PPAC), MoPNG, G. (2021). Summary of LPG Profile 2. Number of Bottling Plants and Bottling Capacity Number of LPG Distributors Packed domestic LPG Sales. *PPAC*.

Pindyck, R. S. (1999). The long-run evolution of energy prices. In *The Energy Journal* (Vol. 20).

Pirrong, Craig. (2014). *Commodity Price Dynamics A Structural Approach*. Cambridge University Press.

Poitras, G., & Teoh, A. (2003). Trading & Regulation Volume Nine Number Two 2003 Derivatives Use. In *Trading & Regulation* (Vol. 9, Issue 2). Henry Stewart Publications.

PPAC. (2022). *PPAC's Snapshot of India's Oil & Gas data*. PPAC. www.ppac.gov.in

Reichsfeld, D., & Roache, S. (2011). Do Commodity Futures Help Forecast Spot Prices? *IMF Working Papers*, *11*(254), 1. https://doi.org/10.5089/9781463923891.001

Sagheer, A., & Kotb, M. (2019). Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing*, *323*, 203–213. https://doi.org/10.1016/j.neucom.2018.09.082

Salvi, H. (2019). Long Short-Term Model for Brent Oil Price Forecasting. *International Journal for Research in Applied Science and Engineering Technology*, *7*(11), 315–319. https://doi.org/10.22214/ijraset.2019.11050

Sapnken, F. E., Kibong, M. T., & Tamba, J. G. (2023). Analysis of household LPG demand elasticity in Cameroon and policy implications. *Heliyon*, *9*(6). https://doi.org/10.1016/j.heliyon.2023.e16471

Shcherbakov, M. V., Brebels, A., Shcherbakova, N. L., Tyukov, A. P., Janovsky, T. A., & Kamaev, V. A. evich. (2013a). A survey of forecast error measures. *World Applied Sciences Journal*, *24*(24), 171–176. https://doi.org/10.5829/idosi.wasj.2013.24.itmies.80032

Shcherbakov, M. V., Brebels, A., Shcherbakova, N. L., Tyukov, A. P., Janovsky, T. A., & Kamaev, V. A. evich. (2013b). A survey of forecast error measures. *World Applied Sciences Journal*, *24*(24), 171–176. https://doi.org/10.5829/idosi.wasj.2013.24.itmies.80032

Siami-Namini, S., & Namin, A. S. (2018). *Forecasting Economics and Financial Time Series: ARIMA vs. LSTM*. 1–19. http://arxiv.org/abs/1803.06386

Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2019). *A Comparative Analysis of Forecasting Financial Time Series Using ARIMA, LSTM, and BiLSTM*. http://arxiv.org/abs/1911.09512

Solo, V. (1984). The order of differencing in ARIMA models. *Journal of the American Statistical Association*, *79*(388), 916–921. https://doi.org/10.1080/01621459.1984.10477111

Song, X., Liu, Y., Xue, L., Wang, J., Zhang, J., Wang, J., Jiang, L., & Cheng, Z. (2020). Time-series well performance prediction based on Long Short-Term Memory (LSTM)

neural network model. *Journal of Petroleum Science and Engineering*, *186*, 106682. https://doi.org/10.1016/j.petrol.2019.106682

Steven, L., Taylor, G., Daniel, A., & Tolleth, M. (2014a). Understanding Crude oil and Product Markets. *American Petroleum Institute*, *4*, 1–39.

Steven, L., Taylor, G., Daniel, A., & Tolleth, M. (2014b). Understanding Crude oil and Product Markets. *American Petroleum Institute*, *4*, 1–39.

Suenaga, H., & Smith, A. (2011). *Volatility Dynamics and Seasonality in Energy Prices: Implications for Crack-Spread Price Risk*. *32*(3–2). https://doi.org/10.5547/ISSNO

Sung, C., Kwon, H., Lee, J., Yoon, H., & Moon, I. (2012). Forecasting Naphtha Price Crack Using Multiple Regression Analysis. *Computer Aided Chemical Engineering*, *31*(July), 145–149. https://doi.org/10.1016/B978-0-444-59507-2.50021-4

SZUCS, D., & HASSEN, K. (2012). *Supply Chain Optimization in the Oil Industry :A Case Study of MOL Hungarian Oil and Gas PLC*. *May*, 1–64. http://www.diva-portal.org/smash/get/diva2:534394/FULLTEXT01.pdf

Thalor, Dr. M., Choudhary, R., Jangid, A., Gandhecha, D., & Bhat, R. (2021). A Review on Forecasting Models of Natural Gas. *International Journal of Scientific Research in Science and Technology*, 251–258. https://doi.org/10.32628/ijsrst218375

Thomasson, S. A. (2017). *Improving forecast performance of LPG demand*.

Vedenov, D., & Power, G. J. (2022). We don't need no fancy hedges! Or do we? *International Review of Financial Analysis*, *81*. https://doi.org/10.1016/j.irfa.2022.102060

Walther, B. A., & Moore, J. L. (2005). The concepts a literature with of species richness the performance estimators , of estimator review performance precision. *Ecography*, *28*(July), 815–829. https://doi.org/10.1111/j.2005.0906-7590.04112.x

Wang, J., & Wang, J. (2016). Forecasting energy market indices with recurrent neural networks: Case study of crude oil price fluctuations. *Energy*, *102*, 365–374. https://doi.org/10.1016/j.energy.2016.02.098

Wang, M., Zhao, L., Du, R., Wang, C., Chen, L., Tian, L., & Eugene Stanley, H. (2018). A novel hybrid method of forecasting crude oil prices using complex network science and artificial intelligence algorithms. *Applied Energy*, *220*(March), 480–495. https://doi.org/10.1016/j.apenergy.2018.03.148

Wang, Y., & Wu, C. (2012a). Forecasting energy market volatility using GARCH models: Can multivariate models beat univariate models? *Energy Economics*, *34*(6), 2167–2181. https://doi.org/10.1016/j.eneco.2012.03.010

Wang, Y., & Wu, C. (2012b). What can we learn from the history of gasoline crack spreads?: Long memory, structural breaks and modeling implications. *Economic Modelling*, *29*(2), 349–360. https://doi.org/10.1016/j.econmod.2011.11.001

Xu, S., Chen, X., & Han, A. (2008). Interval forecasting of crude oil price. *Advances in Soft Computing*, *46*, 353–363. https://doi.org/10.1007/978-3-540-77664-2_27

Yamak, P. T., Yujian, L., & Gadosey, P. K. (2019). A comparison between ARIMA, LSTM, and GRU for time series forecasting. *ACM International Conference Proceeding Series*, 49–55. https://doi.org/10.1145/3377713.3377722

Yu, L., Dai, W., & Tang, L. (2016). A novel decomposition ensemble model with extended extreme learning machine for crude oil price forecasting. *Engineering Applications of Artificial Intelligence*, *47*, 110–121. https://doi.org/10.1016/j.engappai.2015.04.016

Yu, L., Wang, S., & Lai, K. K. (2008). Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy Economics*, *30*(5), 2623–2635. https://doi.org/10.1016/j.eneco.2008.05.003

Zhan, Z., & Kim, S. K. (2024). Versatile time-window sliding machine learning techniques for stock market forecasting. *Artificial Intelligence Review*, *57*(8). https://doi.org/10.1007/s10462-024-10851-x

Zhang, X., Lai, K. K., & Wang, S. Y. (2008). A new approach for crude oil price analysis based on Empirical Mode Decomposition. *Energy Economics*, *30*(3), 905–918. https://doi.org/10.1016/j.eneco.2007.02.012

Zhu, L., & Laptev, N. (2017). Deep and Confident Prediction for Time Series at Uber. *IEEE International Conference on Data Mining Workshops, ICDMW*, *2017-Novem*, 103–110. https://doi.org/10.1109/ICDMW.2017.19