International Journal of Energy Economics and Policy

ISSN: 2146-4553

available at http: www.econjournals.com

International Journal of Energy Economics and Policy, 2024, 14(1), 496-506.

Forecasting High Speed Diesel Demand in India with Econometric and Machine Learning Methods

Aarti Mehta Sharma[1]*, Saina Baby[2], Varsha Raghu[3]

[1]Symbiosis Centre for Management Studies, Bengaluru, Symbiosis International (Deemed University), Pune, Maharashtra, India. [2]Symbiosis Institute of Business Management Bengaluru, Symbiosis International (Deemed University), Pune, Maharashtra, India, [3]Freelance Researcher, India. *Email: director@scmsbengaluru.siu.edu.in

Received: 13 September 2023 **Accepted:** 15 December 2023 **DOI:** https://doi.org/10.32479/ijeep.15572

ABSTRACT

According to International Energy Agency (IEA), India is expected to surpass China by 2024 to become the second largest consumer of oil in the world followed by the United States. High-Speed Diesel (HSD) has the biggest share in the total petroleum products consumed in India accounting for around 38% of the total consumption. Considering the volatile global oil market and an oil import dependency ratio of more than 80% during the last 4 years, the probability of supply disruptions is high in the Indian context. As any uncertainty about the supply of diesel can affect the smooth functioning of the economy and may create inflationary pressures. Accurate forecasting of HSD demand will be essential for appropriate supply management arrangements. Artificial Neural Networks (ANN) with Multi-Layer Perceptron (MLP) and extreme learning machines is used for forecasting diesel demand in this study. Demand forecasting has been carried out using monthly HSD demand data drawn from the "Indiastat" database for the period 1991-2022. Comparison of ANN with traditional forecasting methods of Autoregressive Integrated Moving Average (ARIMA) and Exponential Smoothing has also been undertaken in this study. This study identifies the deep learning technique of ANN with MLP as the best diesel demand forecasting technique.

Keywords: High Speed Diesel, Forecasting, Time Series, ANN, Accuracy, ARIMA, Exponential Smoothing
JEL Classifications: Q41, Q47, C53

1. INTRODUCTION

India is the third largest consumer of petroleum products in the world (Bloomberg, 2022). India is expected to surpass China by 2024 to become the second largest consumer of oil in the world followed by the United States(International Energy Agency, 2019). The consumption of petroleum products in India increased by 142% between 1997-1998 and 2021-2022 (Petroleum Planning and Analysis Cell, 2022). A detailed break up of total petroleum consumption into product wise categories is essential to understand the proportion of various products in total consumption. High-Speed Diesel (HSD) accounts for the highest proportion in consumption basket of petroleum products with a share of 37.5% in 2021-2022. HSD consumption increased from 36.07 million metric tonnes to 76.69 million metric tonnes between 1997-1998 and 2021-2022 registering a growth rate of 112.6%. Motor Spirit (MS) popularly known as Petrol accounts for the second highest proportion in total consumption.

Trends in consumption of HSD, MS and total petroleum products based on the data from Indiastat database are given in the Figure 1 below;

Total consumption of petroleum products and also the consumption of both HSD and MS show continuously increasing trend except during the Covid period (2019-2020 and 2020-2021). Comparison of the growth rates in demand and share in total consumption for both the products are given in the Table 1 below;

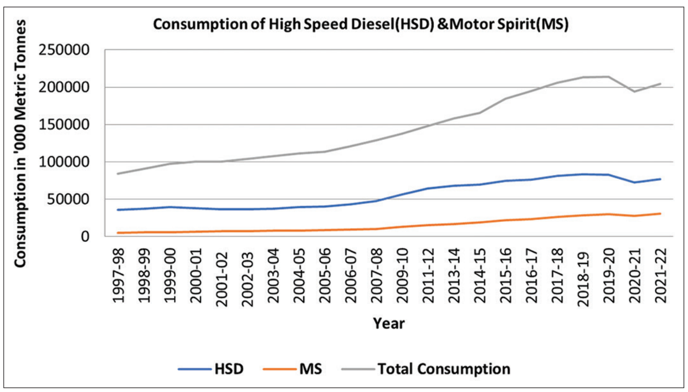
While HSD registered an average growth rate of 5.06% during the above-mentioned period, the demand for MS grew at a higher rate of 10.08%. Though HSD still accounts for the highest share

Table 1: HSD and MS demand trends

Year	HSD demand growth rate (%)	MS demand growth rate (%)	HSD share in total consumption (%)	MS share in total consumption (%)
2005-2006	1.36	4.80	35.50	7.6
2006-2007	6.73	7.40	35.53	7.7
2007-2008	11.12	11.26	36.97	8.0
2009-2010	17.99	24.06	40.81	9.3
2011-2012	15.13	16.96	43.71	10.1
2013-2014	5.58	14.25	43.16	10.8
2014-2015	1.54	11.37	41.94	11.5
2015-2016	7.54	14.53	40.42	11.8
2016-2017	1.85	8.78	39.07	12.2
2017-2018	6.64	10.14	39.32	12.7
2018-2019	3.03	8.06	39.18	13.3
2019-2020	−1.11	5.98	38.58	14.0
2020-2021	−11.97	−6.69	37.42	14.4
2021-2022	5.47	10.30	37.55	15.1

Source: Authors' calculation based on data available on PPAC website. HSD: High-speed diesel

Figure 1: High-speed diesel and MS consumption trends



Source: Petroleum planning and analysis cell

in total consumption, there has been a decline in its share over the years. This can be attributed to factors like the push towards renewables in the farm sector, government's policy drive towards electric vehicles, stricter emission norms, decline in the share of diesel-powered passenger vehicles due to the narrowing price differential between HSD and MS etc. MS on the other hand records an increase in its share in total consumption. Increase in share of petrol engine passenger vehicles from 42% in 2013 to 83% in 2021 could be a major reason for increase in demand for MS in the country (Petroleum Planning and Analysis Cell, 2022).

Even though there has been a drop in HSD consumption growth rate, the clamour for HSD is going to increase in the coming years due to multiple factors. According to Sanjiv Singh, Indian Oil Corporation, Chairman "We are a diesel driven economy." According to a working group constituted by "Ministry of Petroleum and Natural Gas," the demand for diesel is expected to double by 2030 (Economic Times, 2019). Positive demographics, low vehicle penetration, proposed huge investments in infrastructure etc. are expected to push the demand for diesel in the future (Money Control, 2019). Diesel plays a much bigger role in supporting economic growth compared to petrol. In India, 71% of the freight transportation is done through road using trucks- Heavy and Light Commercial Vehicles (HCVs and LCVs). The HCV and LCV segments account for 64.2% of the total diesel demand in the country (Petroleum Planning and Analysis Cell, 2022). Any shortage in the availability of diesel will lead to paralysis in the transportation resulting in supply chain issues and inflation. While the world was grappling with Covid-19, economies slowed down and with that there was a deep dip in the demand for both petroleum products. The demand and consumption of both increased in the year 2022 and will continue to do so. Keeping in mind the projections made, it is imperative to ensure that sufficient supply of HSD is available to meet the growing demand from the economy.

Though India is the third largest oil consumer, it never imported petrol or diesel because of its surplus refining capacity. India has been traditionally importing crude oil and then refining it into various petroleum products. State refiners like Bharat Petroleum, Indian Oil Corporation, Hindustan Petroleum Corporation Ltd and Mangalore Refinery and Petrochemicals Ltd account for 60% of the crude oil refining capacity. Private players and Joint venture firms do the remaining 40% of the refining activity (The Mint, 2018). India has been heavily depended on crude oil import to meet its ever-increasing demand for various petroleum products. Import dependency ratio has been continuously increasing since 2011-2012 and for the year 2021-2022, 85.7% of the requirement is met by import (Petroleum Planning and Analysis Cell, 2022). Such a heavy dependence on import for meeting the demand for essential products like petrol and diesel puts the country in a precarious situation. The volatility in global crude oil prices and exchange rate fluctuations increase the risk of being a heavy import dependent country.

Given the globalized nature and interconnectedness of economies around the world, it is not possible for a country like India to influence neither global crude oil prices nor the exchange rates. Given the current geographical factor endowments, it is also not possible for India to initiate large-scale oil exploration. The only viable option is to accurately predict the demand for petroleum products and ensure their on-time supply in the market. Any disruption in the supply of these essential fuels will reduce the momentum of India's economic growth. Therefore, it is imperative to develop accurate forecasting techniques to assess the demand for HSD in India. This paper aims at identifying an accurate method for predicting HSD demand in India, by employing machine learning techniques and traditional forecasting methods.

This paper is divided into four sections. Part two provides a comprehensive review of empirical studies for energy demand forecasting conducted in the global and Indian contexts and identification of the research gap and definition of the objective of this study. Part three comprises of a detailed explanation of the forecasting methods used in this study - Artificial Neural Network, exponential smoothing and ARIMA. Data analysis and inference is included in part four of the paper. Part five includes summary of findings, recommendations and conclusion.

2. LITERATURE REVIEW

This segment consists of a review of various empirical studies conducted in Indian and global contexts to predict the demand for various energy products. These studies have used various machine learning methods, time series and other econometric forecasting models to predict the demand for energy products. A comprehensive review of these papers provides a fair understanding of the relative merits and demerits of various forecasting methods. The later papers have used machine learning techniques like neural networks, fuzzy logic etc. and the older ones have used ARIMA, regression etc.

While there have been a large number of studies on models that can be used to predict energy, there has been no clear-cut answer to which is the best model (Tamba et al., 2018). Learning from the repetitive patterns in time series is important, especially when such recurrences are not periodic. Recent research in time series data has incorporated the use of deep learning methods like artificial neural networks (Dubnov, 2022; Ferretti and Saletta, 2022; Góez et al., 2022) for data which is not repetitive. The accuracy of the models is around 50%. These methods are although new, still interesting.

Various econometric modelling methods were performed on univariate time series for forecasting natural gas consumption in Pakistan in a study (Hussain et al., 2022).The authors forecasted the demand till the year 2030 and ascertained that ARIMA was the best way to forecast the demand for natural gas across various sectors like agriculture, industry, household etc.

A study (Li et al., 2022) undertaken in the Chinese context analyzed various ways in which oil consumption could be predicted. The authors used several tools like ARIMA, grey models, neural networks and Grey multivariable Verhulst model to forecast the consumption of oil. The neural network model was considered to be the most accurate followed by the Grey model in prediction performance. The authors also suggest that China should move away from fossil fuels to clean energy and that the same model can be used to predict the demand for clean energy.

A comparison was made between statistical, econometric and machine learning models for predicting the demand for natural gas in Korea (Shin and Woo, 2022). According to the authors, moving towards carbon neutrality and sustainability requires accurate prediction of natural gas demand in Korea. The authors compared the techniques of Random forests and artificial neural networks and identified that the latter gives more accurate results. Structural Vector Auto Regression (SVAR) model was used by a study to deduce the relationship among oil price, economic growth and import demand for developing countries, using India as a model (Dash et al., 2018). The advantage of SVAR is that, the relationships among the various variables can be observed in detail and not just of the independent variables on the output variables. Coefficients of elasticity were hence calculated. The study found that the import demand goes down with increase in oil prices, but overall, the demand tend to be relatively inelastic.

A study was undertaken for comparing five models used by five different organizations to forecast demand for and emissions from transport fuel in India (Paladugula et al., 2018). The five models have considered independent assumptions and used different data sources. As a result, the demand forecasted by all the five models are different - with some being high and some being low. The study was done to establish the fact that due to fundamental differences in the assumptions by each of the model, the demand forecasted becomes substantially different, leading to uncertainty. Population and GDP were mainly used to arrive at the figures, although the individual institutions differed in the way in which they considered their data - for example, some considered it fuel wise and so on. It was found that, the models predicted a higher figure in general, especially with service demand for freight as compared to passenger transport. The models displayed differences in the results of total demand, various sub-sects of energy consumption breakdown and emissions.

A study on predicting the energy needs of China and India used various linear and non-linear methods based on the grey model and found it to be very accurate. The paper also discussed the projections showing that India's energy needs will surpass those of China by the year 2026 and hence there is need to have good modelling techniques to be prepared for the surge in demand (Wang et al., 2018).

A comprehensive review of literature was undertaken to compare nine different time series models of demand forecasting (Debnath and Mourshed, 2018). Initially, the authors have provided an extensive review of the applications of these techniques used in various empirical studies. In the end, the authors of this paper have outlined the advantages and disadvantages of the various techniques. The paper also included an evaluation of hybrid techniques-techniques that combine aspects of various techniques and that have been proved to be better than single techniques in forecasting demand. Some advantages of ANN include the fact that it can optimally deduce associations between variables without clear coding of the relationships. However, it has the problem of over-fitting and it often becomes difficult to distinguish between an over-fit and a good fit. ARIMA is more flexible in that as it strikes a middle ground and can provide a good fit for any time series data. It is, however, time consuming.

An exploratory study was conducted to analyze the trends in demand of three resources-gasoline, diesel and LPG in India (Sen and Sen, 2016). The increase in demand for oil resources in India has been looked at through the lens of fall in oil prices, increase in motorization and per capita income and the "Make in India" policy. Overall, the increase in demand for oil resources has seen significant in the past few years. Diesel has been used mainly in the agricultural sector to fuel tractors and freight transportation. Although the growing demand for diesel was affected during the floods in South India, the demand has generally been strong. Gasoline is mainly used in private vehicles and the increase in motorisation has led to increase in gasoline demand. Policies aimed at achieving a greener environment, such as imposing green cess taxes, introduction of greener fuels like CNG also will have an effect on fuel demand.

Another study carried out in the Indian context detailed the determination of price and income elasticity for both gasoline and diesel (Hasanov, 2015). In addition to testing the elasticity (in short and long run), the project also deduced if the response in demand for both the fuels with change in price are symmetric- that is, with increase and decrease in price, whether the extent of change in demand for gasoline and diesel is the same. The author has tested different models such as Partial Adjustment Model (PAM), distributed lag (DL) and Autoregressive Distributed Lag (ARDL), but mainly focused on the aspect of cointegration. While in the case of diesel, they found cointegration among the variables, they could not ascertain cointegration among the variables for gasoline. This led to the conclusion that while there exists a long-term relationship of demand for diesel with price and income, no such relationships exist for gasoline. Gasoline was price elastic and income inelastic in the short run, while the opposite was true for diesel. In the long run too, diesel is relatively inelastic to price and elastic to income changes. On checking the symmetry of response for demand, it was revealed that the extent of change in demand for diesel was symmetric in either direction with increase or decrease in price. The same was true for gasoline demand as well. Transport fuels (diesel and gasoline) turned out to be price inelastic and with an income elasticity of unity.

A study which was done in Ghana used ARDL approach with an "Unrestricted Error Correction Model (UECM) to find the coefficients of elasticity for crude oil demand with respect to price, income, exchange rates, population growth rate and domestic oil production (Marbuah, 2014). The logarithmic forms of the variables were used in the analysis. The model is dynamic in the sense that it incorporates lags in the time series. The demand was found to increase with increase in price (postulated by the author to be due to the lack of alternative fuel resources) in the short term, while it the relationship becomes inelastic in the long term. The strongest driver of demand was found to be economic activity. Real Effective Exchange Rate and population growth drive demand for crude oil, while the production of crude oil domestically does not have too large an effect on the demand.

A study which was done in the Indian context calculated coefficients of elasticity of demand for crude oil, petroleum and diesel with respect to income level (Gross Domestic Product) and price of the energy resource (Agrawal, 2012). The model used is ARDL with inclusion of error correction. Coefficients of elasticity of demand for three commodities: Crude oil, Petroleum and Diesel were calculates using a log-linear approach as log linear models are better suited for non-stationary data. Once the coefficients were calculated, demand growth equations were determined to use these coefficients to forecast demand for the three commodities till the year 2025. To forecast demand, three scenarios of GDP (positive, neutral and negative) and three scenarios each of crude oil, petroleum and diesel prices were taken into consideration. In the case of petroleum, a few dummy variables had to be included to account for anomalous years with respect to price, out of which, only one was found statistically significant. Overall, it was concluded that by 2025, demand for crude oil would increase by 100%, for diesel by 110% and for petroleum by 165%.

A study was conducted in the Chinese economy to forecast for energy demand by the years 2020 and 2030 (Shan et al., 2012). The model used is LEAP-Long-range Energy Alternatives Planning System). The model is based on technological and environmental database, and hence is a bottom-up approach. The authors have considered six scenarios based on three different possible economic scenarios and two different scenarios of development in non-fossil, alternative forms of energy. The studies revealed that in all the scenarios, economic growth rate will increase and then eventually slow down. The study also included a look into the input-output relationships in energy conversion and consumption processes. All in all, the study found that the energy consumption would become cleaner in the future no matter the economic scenario, with the slated decrease in CO_2 emissions being achieved in all three scenarios.

A comparison study was undertaken in the South Korean context to evaluate the merits of Artificial Neural Networks (ANN) against Multiple Linear Regression (MLR) models (Geem, 2011). Five independent variables were considered as affecting the energy demand, which were used in different permutations and combinations to build the models. The study has also explored if oil price and number of vehicle registrations can be better substitutes for GDP and population, respectively. The study evaluated three models for multiple linear regression and three for Artificial Neural network-each model differing in the combination of variable considered. The study evaluated three models for multiple linear regression and three for Artificial Neural Network-each model differing in the combination of variables considered. The models included the following combinations: (a) GDP, population and passenger transport (b) GDP, number of vehicle registrations and passenger transport (c) oil price, number of vehicle registrations and transport (d) oil price, population and passenger transport. As for the ANN, a feed forward network with back-propagation algorithm for error was used. As explained by the author, in this scenario, the Root Mean Square Error (RMSE) values are more relevant as compared the R squared values as a high R squared value just establishes the linearity of relationship between the given variables whereas in this situation, the relationship is more likely to be complex and dynamic than a simple linear one. It is, however, much more important that the error in prediction be less, which is what the ANN achieves comparatively more efficiently than the MLR model.

A study was undertaken to assess the effects of renewable energy usage on economic growth in India. A Structural Vector Autoregression (SVAR) model has been employed

(Tiwari, 2011). The model was used to study the relationship between RES (Renewable Energy Source consumption, in this paper, hydroelectric power consumption was considered as a representative), GDP per capita and CO_2 emissions. The results revealed that a positive shock on RES had a positive effect on the GDP and a negative effect on CO_2 emissions (except in the 1st year) for the 20 years under consideration. Also, one positive impact on GDP showed positive effects on CO emissions for the entire duration of 20 years under consideration. Further, to test the cause of lags/errors in data, a variance decomposition was generated.

A comprehensive review of eight different methods used for forecasting demand was undertaken to assess the relative merits and demerits of various models (Li et al., 2010). The authors segregated the models both on the basis of complexity and also on whether they were used to deduce a relationship between the dependent and the independent variables over a time period or to establish a relationship between a variable's value in the past, using it to predict its future value. Surprisingly, it was found that the simpler models gave a lesser error for this particular data. In fact, the quadratic model was found to be one of the best suited for both seasonal and, non-seasonal data. The authors concluded that the best model need not always be the most complex one.

Adaptive Network Based Fuzzy Interference System (ANFIS) was used in this cross-country study covering the United States of America (USA), India, Russia and Brazil for more accurate prediction of demand for oil (Azadeh et al., 2010). It combines fuzzy logic with traditional elements like pre and post processing of data. Usually, the input variables to be used in a model are selected by Trial and Error (TEM), but this is cumbersome since it generates many possible permutations and combinations of the available variables. To avoid this, ACF (Autocorrelation Function) was used to select the variables. Once the input variables have been arrived at, the next step is to see how much of a lag there is, between what is calculated and what is observed for these time series. To bring about stationarity in the data so that there isn't too vast a lag, the data is pre-processed or normalized. The model that was relatively error free was chosen and trained using training data and validated using test data. The model was applied to real life scenarios-wherein, it was used to calculate the demand for oil in USA, Russia, India and Brazil.

In a study conducted in Indonesia cointegration approach was used to look at the price and income elasticities of petroleum consumption and gasoline consumption (Sa'ad, 2009). When 2 time series are cointegrated, it rules out the possibility that any correlation between them may be spurious. Here, an ARDL/Unrestricted Error Correction model was used-as in, all the indictors – such as price, efficiency of usage, per capita GDP etc. could be specified. Upon analysis, the null hypothesis of no cointegration was rejected using the F statistic. Next, the elasticities were determined. To gain insight into the extent to which the predicted scenario may be close to the actual, several parameters were tested, such as AIC, R squared method, SBC and HQIC. SBC was accepted as it passed the test to be used as a viable parameter. In conclusion, it was established that, the consumption of petroleum products was more sensitive to the income level than it was to the change in price of the commodity. Since the time series of petroleum products consumption and per capita gasoline consumption were cointegrated, by extension, this means that the consumption of gasoline in Indonesia was more sensitive to change in income levels than change in prices.

The authors have used an Artificial Neural Network to form mathematical equations for forecasting Net Energy Consumption (NEC) with population, gross generation, installed capacity and years as inputs (Sözen et al., 2006). ANN works just like a human neural network. It has input layers, hidden layers (wherein, back-propagation of errors and weighing happens) and output layers. Just like the neurons in a human brain are activated by passing on impulses, here, there is an "activation function"-a sigmoidal wave equation. An evolutionary algorithm was used to zero in on the number of neurons and the neural architecture. In the end, a successful equation was developed to model the future demand for energy resources in Turkey. Since ANN was used in building this equation, flexibility could be incorporated in the model to provide leeway for future uncertainties.

Feed forward, Artificial Neural Network (ANN) model with a back -propagation algorithm was used in this study to forecast transport energy demand (Murat and Ceylan, 2006). The study has taken into consideration, socio-economic factors, namely, GNP, population and vehicles as independent variables. These three factors are the inputs to this neural. Using ANN, with testing data, the total average minimum relative error was found to be 13%, significantly lower than other methods. The model, predicted that the demand for transport energy would increase steadily post 2010 till when, the effects of the economic crises of Turkey were predicted to last.

The review of literature clearly reveals the use of multiple energy demand forecasting techniques including traditional (time series and causal) and soft computing methods. Traditional time series methods like AR, ARIMA etc. have been extensively used by researchers for energy demand forecasting (Al-Saba and El-Amin, 1999), (Murat and Ceylan, 2006). Regression models like ARDL, MLR and SVAR also have been applied in energy demand forecasting (Dash et al., 2018), (Agrawal, 2012; Hasanov, 2015; Marbuah, 2014). Traditional forecasting methods are limited in their ability to incorporate the uncertainty and dynamism in the environment. Soft computing methods, which are based on artificial intelligence, are more flexible and adaptive in incorporating uncertainties in model building. AI based forecasting methods mimic the ability of human mind to learn in an environment of uncertainty and complexity. Deep learning techniques which focus on creating networks that are capable of learning unsupervised from data is an emerging field in energy demand forecasting. Artificial Neural Network (ANN), which is a deep learning technique, started gaining popularity in the field of energy demand forecasting since 1980s. Many studies in the literature provide support to the fact that ANN method minimizes forecasting errors compared to traditional demand forecasting methods (Agrawal, 2012; Al-Saba and El-Amin, 1999). Energy demand forecasting in India has been done mostly using traditional demand forecasting methods which are not very effective in including uncertainties in the economic environment into model building. Therefore, it would be interesting to use ANN for estimating the demand for diesel in India. We expect the forecasting model using ANN to be more accurate compared to traditional models. The objective of this study is to develop a suitable forecasting model for estimating the demand for diesel in India.

3. MATERIALS AND METHODS

In this paper we have forecasted the values of diesel consumption through Artificial Neural Networks (multilayer perceptions and extreme learning machines), Exponential smoothing and ARIMA using the R software. Artificial neural networks are one of the main tools used in machine learning. ANN is a machine learning tool which replicates the human brain learning process. Neural networks contain input and output layers along with a hidden layer. Hidden layer contains units which convert the input into something that can be used by the output layer. Neural networks have the capacity to identify complex patterns and make the machines to spot them.

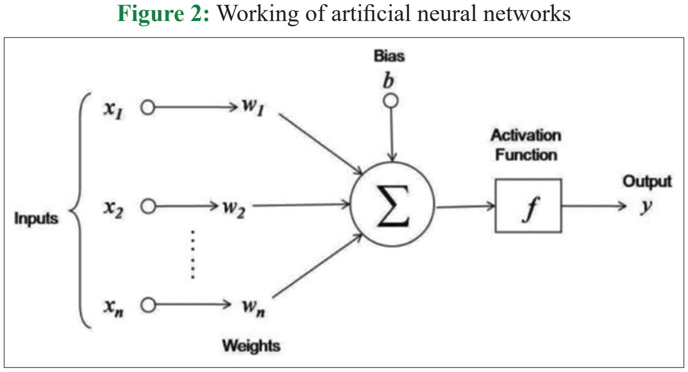
Mathematical models are generated by computational systems in ANN method. “The artificial neuron has a number of input channels, a processing stage, and one output that can fan out to multiple other artificial neurons” (Puri et al., 2015). The working is shown in Figure 2.

$Y = \sum$ (Weight * Input) + error

Input variables in the models are $x_1,x_2.....x_n$. Respective input weights are represented by $w_1,w_2.....w_n$. The bias in the model is indicated by ‘b’ which is summed with the weighted inputs to form the net inputs. Bias and weights are the adjustable parameters of a neuron. Time series modelling is purely dependent on the idea that past behavior can be used to predict future behavior. Neural networks have become an important method for time series forecasting. Artificial Neural Networks (ANN) usually work on multiple inputs, and just recently are being used for univariate time series analysis. Here we have used the programming language R to model the series. R works on packages and we have used two packages to forecast the values.

ANNs are classified into 2 types - feedforward and feedback artificial neural networks. Feedforward neural network are non-

Figure 2: Working of artificial neural networks



Source: Medium.com

cyclical in nature. Neurons in each layer is connected only to the next layer. In this case the signal will travel sequentially in one direction towards the output layer.

Feedback neural networks are cyclical in nature. Signals travel in both directions by introducing loops in the network. Network’s behaviour change can be caused by feedback cycles over time depending on the input. In this paper we have analysed the series by using two feed-forward algorithms:

1. Multi Layer Perceptrons (MLP)
2. Extreme Learning Machines (ELM).

Multi Layer Perceptrons (MLP): MLP is a type of feedforward artificial neural network. A MLP contains input, output and hidden layers. Output and hidden Layers use a nonlinear activation function. Backpropagation learning technique has been used in MLP for training. The package used for mlp in R are “NNFOR” and “Forecast.”

Extreme Learning Machines (ELM): ELMs are a type of feedforward neural networks. These have been used for clustering, classification, regression, and feature learning. Single or multiple layers of hidden nodes makes Feature learning possible. Adjustment of hidden node parameters is not required in this case. The creators of this neural network names them as ELM as the learning power of this models is thousand times faster than backpropagation-based learning techniques. The package used for elm in R is “Forecast.”

ARIMA, like exponential smoothing is a more traditional form of time series forecasting. ARIMA models attempt to identify patterns in historical data. ARIMA consists of three components - The “AR” or autoregressive component which accounts for the patterns between any 1 time period and previous periods. The “MA” or moving average component measures the adaptation of new forecasts to prior forecast errors. The “I” or integrated component connotes a trend or other “integrative” process in the data. In the face of trends, the differences from 1 month to the next must be modelled rather than the monthly data themselves. In some cases, an exponential smoothing method is considered to have less errors than an ARIMA model (Litterman, 1986).

The exponential smoothing method is where newer values are given relatively greater weight compared to previous observations. If there is data from the t-observation, the forecast value at time t + 1 is:

$$St + 1 = \propto X_t + (1-\propto) St$$

Where,
St+1= forecasting value to t+1,
Xt= actual data at time t
α= parameter values between 0 and 1,
St= forecasted value to t.

The package “timeseries“was used to model by using ARIMA. In this paper, we have compared the traditional forms of forecasting of time series with the newer machine learning methods. Multiple algorithms were run and compared through the Root mean square error and Mean square error, which were computed for all four methods and forecasts were calculated for 12 time periods that is, for 12 months.

4. RESULTS

Monthly data of HSD Consumption (in "1000 Metric Tonne) was collected from the database "Indiastat" for the time period 1998-1999 to 2022-2023. For the year 2022-2023 data for the months till September 2022 were considered. There are a total of 294 observations. The yearly descriptive statistics for the entire dataset are presented in Table 2.

While examining the annual mean values we can see that there is a gradual increase in the consumption of diesel over the years, except for the Covid years (2019-2020, 2020-2021 and 2021-2022). The raw data is shown in the graph in Figure 3 and we can observe that during the beginning of our data there was a steep increase in the consumption of diesel but the rate of increase has slowed down during the latter half of the data and it has decreased drastically during the Covid years.

The modelling and forecasting for this data was done by using five methods:

- MLP using "NNFOR" package,
- MLP using "forecast package"
- ELM using "NNFOR" package
- Exponential smoothing using "forecast" package
- ARIMA "using" time series package.

4.1. Method 1: MLP using "NNFOR" Package

The first method used MLP fit with 5 hidden nodes and 20 repetitions. There were univariate lags at (1, 2, 4). The package used was "NNFOR" and the command was mlp. A graphical representation of the model is given in Figure 4.

The forecasted values for the next 12 months are given in the Table 4. A graphical representation of the forecasted values is given in Figure 5 and we can see that the forecasted values follow the trend of the earlier observations.

4.2. Method 2: MLP using "Forecast Package"

The next modelling was by using a Feed-forward neural network with a single hidden layer and lagged inputs for forecasting univariate time series. This was by using the command NNetar in the package "Forecast." The point forecasts for the next 12 months are given in Table 4. Comparison of error terms shown in Table 3 shows that the RMSE for this algorithm is the least.

A graphical representation of the data and forecasted values using NNAR is shown in Figure 6 and it shows that the forecasted values mirror the actual values.

4.3. Method 3: ELM using "NNFOR" Package

The next algorithm applied is of extreme learning machines. The algorithm was fitted with 100 hidden nodes and 20 repetitions. The input was Univariate lags of 1, 3 and 4. The Output weight estimation was arrived at by using lasso regression. Forecasts based on ELM method have been shown in Table 4 and forecast accuracy has been displayed in Table 3. Figure 7 shows the model

Figure 3: High-speed diesel consumption

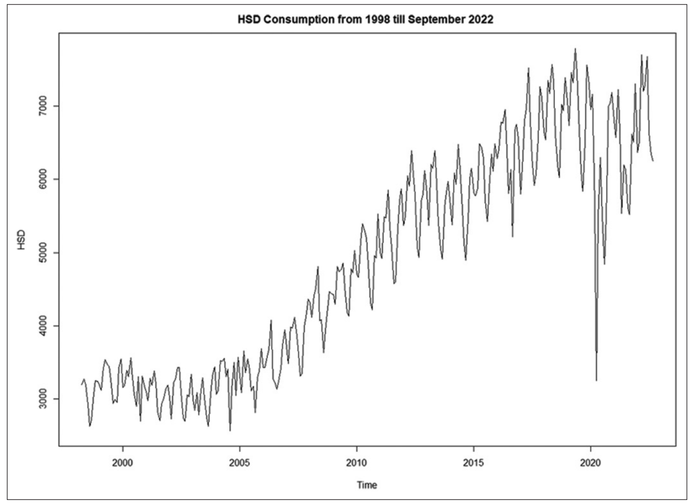


Figure 4: Graphical representation of a neural network with multi-layer perceptron

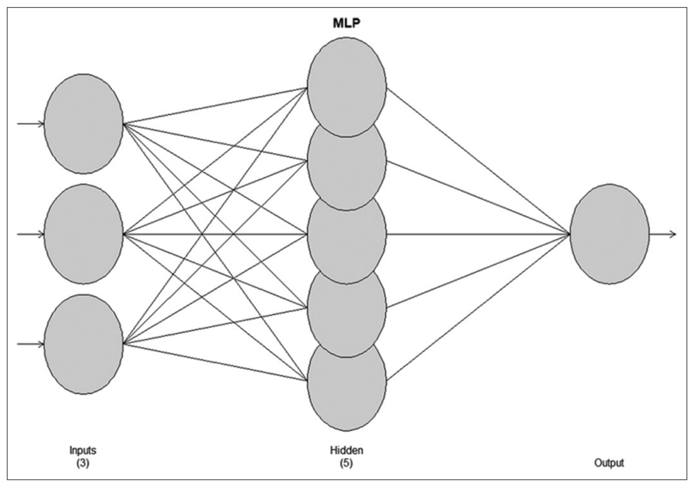


Figure 5: Graphical representation of the forecast using neural network with multi-layer perceptron

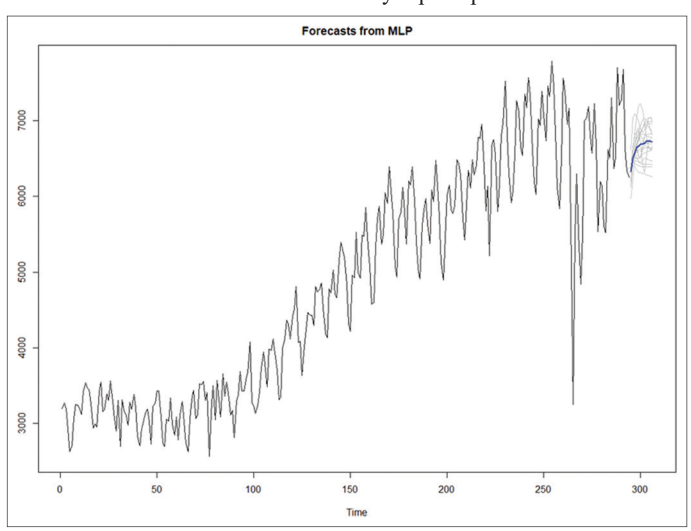


Figure 6: A graphical representation of the data and forecasted values using "Forecast" package

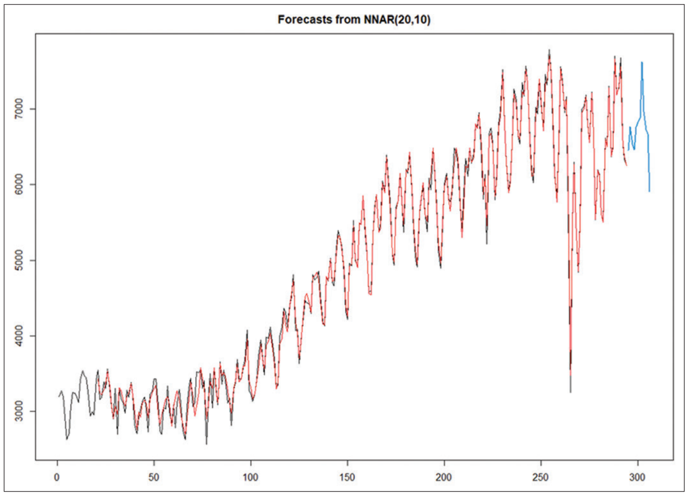


Figure 7: A graphical representation of the model for extreme learning machines using "NNFOR" package

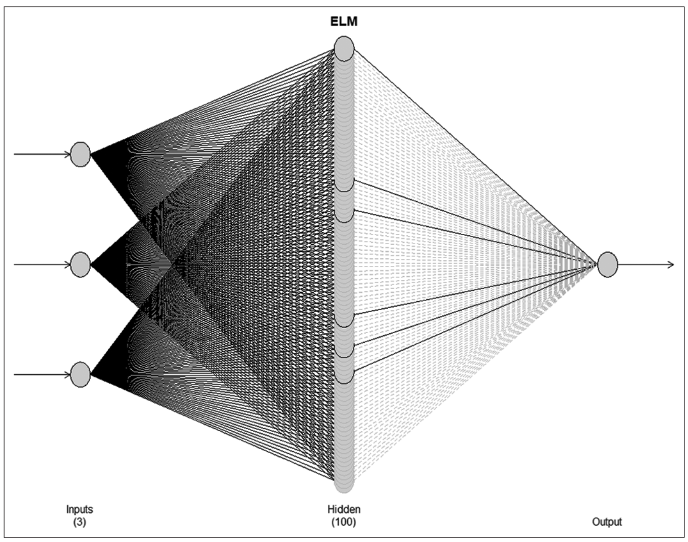


Figure 8: A graphical representation of the data and forecasted values using extreme learning machines

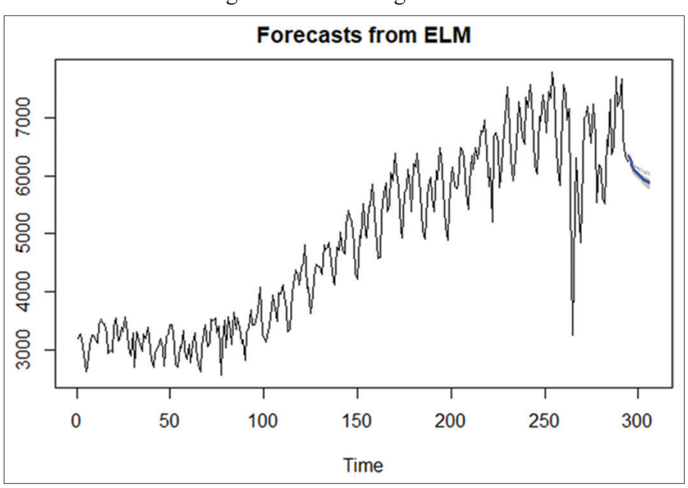


Figure 9: A graphical representation of the data and forecasted values using exponential smoothing

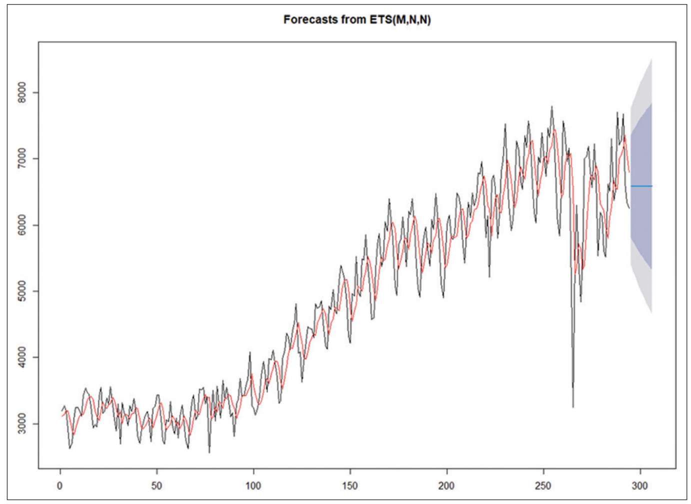
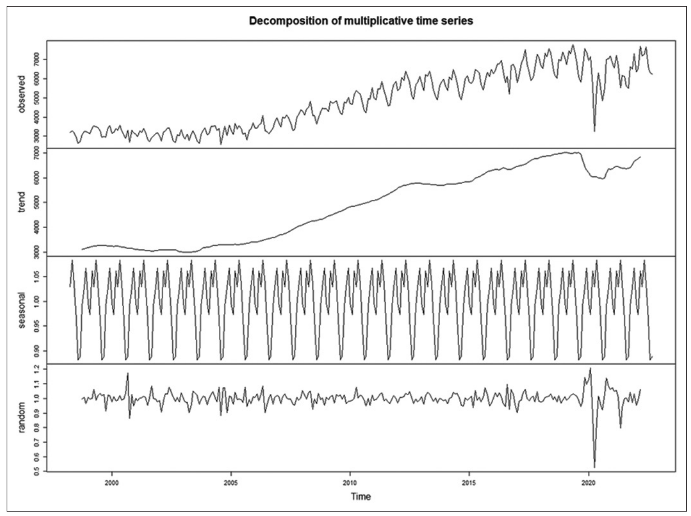


Figure 10: A graphical representation of the decomposed time series data



and Figure 8 shows the modelling of the data and the forecasts. This forecast while examining visually seems different from the forecast in Figure 5 and similar to the forecast in Figure 6.

4.4. Method 4: Exponential Smoothing using "Forecast" Package

The next modelling is done by using the econometric method of exponential smoothing. The smoothing parameters are as given below:

alpha = 0.3871 initial states: l = 3115.345 sigma: 0.0918

Figure 9 gives the modelling of the data and the forecast. While examining visually the forecast seems to be higher than all the other methods.

4.5. Method 5: ARIMA "using" Time Series Package

The data is given as monthly consumption and was hence taken as time series data. The data was decomposed and the result is in Figure 10.

As is clear from the diagram, the data has a trend and is not seasonal by nature. The random fluctuations are visible majorly in the year 2021 and to a smaller extent in 2022. The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) show that the data is non-stationary. The data was made stationary by differencing by 1. The stationarity was checked again using Augmented Dickey-Fuller test (ADF) test which showed that data is stationary (P = 0.04012). The data was modelled by using multiple algorithms and the best was ARIMA (2,1,1) with drift. This means that it autoregresses by 2 months (coefficients are 0.6151 and −0.2318), was differenced by 1, and had a moving average of 1(coefficient is −0.8602).

Table 2: Descriptive statistics

Year	Min	Max	Mean	Median
1998-1999	2625	3423	3101	3192
1999-2000	2936	3547	3275	3327
2000-2001	2696	3561	3163	3227
2001-2002	2704	3384	3045.5	3073.5
2002-2003	2693	3432	3053.833	3065
2003-2004	2623	3525	3089.667	3090
2004-2005	2565	3655	3304.333	3359
2005-2006	2815	3685	3349.25	3384.5
2006-2007	3131	4080	3574.75	3574
2007-2008	3310	4409	3972.5	4050
2008-2009	3628	4810	4309.333	4361.5
2009-2010	4129	5153	4687	4731
2010-2011	4213	5524	5005.917	4980
2011-2012	4577	6051	5396	5424
2012-2013	4938	6394	5763.583	5795.5
2013-2014	4908	6394	5697.5	5732.5
2014-2015	4899	6477	5784	5852
2015-2016	5426	6783	6220.75	6319
2016-2017	5213	6958	6334	6472
2017-2018	5921	7527	6756.167	6725
2018-2019	6027	7567	6960	7048
2019-2020	5660	7781	7547.5	7547.5
2020-2021	3252	7225	6059.41	6434
2021-2022	5516	7707	6390.92	6441
2022-2023	6255	7767	6900	6920.5

Source: Authors' calculation based on data available on PPAC website

5. RESULTS AND DISCUSSION

The comparison for the five methods is given in Table 3 and forecasted values are given in Table 4. Forecast accuracy has been compared in Table 3 using Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Mean Absolute Scaled Error (MASE).

From Table 3 it is clear that the method of MLP by using the NNFOR or the Forecast package gives us the best result as all the error terms are low for it.

Table 4 gives us the forecasts for the next 12 months starting from October 2022 onwards. Here, 295 corresponds to the predicted consumption for October 2022, 296 corresponds to November 2022 and so on. Traditionally, researchers have used exponential smoothing, moving averages and ARIMA to predict time series variables. However, the first two can be applied for very short-term predictions and ARIMA has to be further modelled to incorporate volatility. This paper digs into deep learning techniques which mimics the human learning process to learn fast and absorb volatile movements, if needed. There is now growing technical sophistication of forecasting tools but researchers are still not certain about which models surpass the other. This study is an attempt to further the movement and provide insights into the relative merit of three different algorithms in ANN.

Table 3: Comparison of different methods

Method	Packages	ME	RMSE	MAE	MAPE	MASE
Neural network (MLP)	NNFOR	**−0.10582**	**384.4121**	**295.4018**	**6.216435**	**0.854814**
Neural network (MLP)	FORECAST	0.77061	**90.45781**	**62.85262**	**1.566193**	**0.181108**
Extreme learning machines	NNFOR	**0.098**	454.1025	339.6353	7.106792	0.982814
Exponential smoothing	FORECAST	30.4825	479.7927	358.9002	7.56955	1.034162
ARIMA	FORECAST, TSERIES	53.43753	420.694	310.5999	6.574937	0.894986

Table 4: Predictions using the five different algorithms

Forecasted HSD Demandv (in 000, Metric Tonnes) - Point Forecast					
Forecast period	MLP (nnfor)	MLP (forecast)	ELM (nnfor)	ARIMA (forecast, ARIMA)	Exponential smoothing (forecast)
Oct 2022	6323.556	6800.653	6316.49	6503.179	6584.85
Nov 2022	6496.399	6734.877	6244.921	6675.77	6584.85
Dec 2022	6599.68	6518.382	6144.661	6724.398	6584.85
Jan 2023	6651.63	6387.001	6068.248	6714.3	6584.85
Feb 2023	6698.243	6562.724	6036.762	6696.816	6584.85
Mar 2023	6754.994	6652.307	6010.376	6688.403	6584.85
Apr 2023	6786.799	6446.466	5980.523	6687.281	6584.85
May 2023	6797.396	7220.897	5956.322	6688.541	6584.85
Jun 2023	6783.248	6919.775	5934.281	6689.576	6584.85
Jul 2023	6736.131	6764.917	5915.668	6689.921	6584.85
Aug 2023	6727.003	6791.156	5898.597	6689.893	6584.85
Sep 2023	6742.928	6422.686	5882.024	6689.796	6584.85

6. CONCLUSION

India is expected to surpass China by 2024 to become the second largest consumer of consumer oil in the world. A recent report by International Energy Agency says that India is one of the largest consumers of Oil and is not in a position to change its consumption pattern towards alternate energy sources- in the short to medium term time horizon (International Energy Agency, 2022).

The war in Ukraine has destabilized the supply of oil all across the world (Huang et al., 2023) and thus there is an even stronger need to have an accurate method of forecasting the increasing demand for HSD in India. Considering the diesel driven nature of our economic growth it is imperative to develop suitable supply management policies to ensure continuous availability of fuel. As the "Make in India" initiative gains momentum, the availability of Diesel is even more vital. This can happen when there is a credible forecasting tool to estimate the consumption of Diesel in the coming years.

The objective of this paper was to arrive at an accurate forecasting tool by comparing traditional forecasting tools to machine learning algorithms to predict the demand for HSD in India. We used ARIMA, Exponential Smoothing, ANN (multilayer perceptrons and extreme learning machines) to predict the demand of HSD in India. We compared the machine learning algorithms namely MLP and ELM and worked with multiple packages from the programming language R. From this study it is clear that the deep learning technique of MLP, when using the "Forecast" package or the "NNFOR" package gives us the best result. This paper also shows that while most machine learning tools were constructed to analyse big data, these same tools can be successfully applied to smaller datasets as well.

ANN is a new deep learning technique and has not been extensively used for time series forecasting. The algorithm does not explain the tool used to generate the output. It gives us the best result by using the available tools (it could be ARIMA, ARCH, GARCH, ARIMAX, Exponential Smoothing etc.) internally. The only drawback of using ANN is that while it produces a probing solution, it does not satisfy our curiosity by explaining "why" and "how". However, this paper showcases the successful application of ANN to fit and forecast time series data and this method can be applied to other time series data of commodities as well, for e.g. petrol, Liquefied Petroleum Gas etc. At this juncture, India's economy will need correct estimates of these to prepare for the future and we believe that this method offers a viable solution.

REFERENCES

Agrawal, P. (2012), India's Petroleum Demand: Empirical Estimations and Projections for the Future. Delhi: Institute of Economic Growth, University of Delhi.

Al-Saba, T., El-Amin, I. (1999), Artificial neural networks as applied to long-term demand forecasting. Artificial Intelligence in Engineering, 13(2), 189-197.

Azadeh, A., Saberi, M., Ghorbani, S. (2010), An ANFIS algorithm for improved forecasting of oil consumption: A case study of USA, Russia, India and Brazil. Journal of Scientific and Industrial Researh, 69, 194-203.

Bloomberg. (2022), India Is Bright Spot for World Oil Demand. Available from: https://www.bloomberg.com/opinion/articles/2022-11-14/india-is-bright-spot-for-world-oil-demand-lagrxccb

Dash, D.P., Sethi, N., Bal, D.P. (2018), Is the demand for crude oil inelastic for India? Evidence from structural VAR analysis. Energy Policy, 118, 552-558.

Debnath, K.B., Mourshed, M. (2018), Forecasting methods in energy planning models. Renewable and Sustainable Energy Reviews, 88, 297-325.

Dubnov, S. (2022), Predictive quantization and symbolic dynamics. Algorithms, 15(12), 484.

Economic Times. (2019), India's Diesel Demand to Double by 2029-30: Dharmendra Pradhan. Available from: https://energy.economictimes.indiatimes.com/news/oil-and-gas/indias-diesel-demand-to-double-by-2029-30-dharmendra-pradhan/67332081

Ferretti, C., Saletta, M. (2022), Do neural transformers Learn human-defined concepts? An extensive study in source code processing domain. Algorithms, 15(12), 449.

Geem, Z.W. (2011), Transport energy demand modeling of South Korea using artificial neural network. Energy Policy, 39(8), 4644-4650.

Góez, D., Soto, P., Latré, S., Gaviria, N., Camelo, M. (2022), A methodology to design quantized deep neural networks for automatic modulation recognition. Algorithms, 15(12), 441.

Hasanov, M. (2015), The demand for transport fuels in Turkey. Energy Economics, 51, 125-134.

Huang, M., Shao, W., Wang, J. (2023), Correlations between the crude oil market and capital markets under the Russia--Ukraine conflict: A perspective of crude oil importing and exporting countries. Resources Policy, 80, 103233.

Hussain, A., Memon, J.A., Murshed, M., Alam, M.S., Mehmood, U., Alam, M.N., Rahman, M., Hayat, U. (2022), A time series forecasting analysis of overall and sector-based natural gas demand: A developing South Asian economy case. Environmental Science and Pollution Research, 29(47), 71676-71687.

International Energy Agency. (2019), Oil 2019:Analysis and Forecast to 2024. Available from: https://eneken.ieej.or.jp/data/8429.pdf

International Energy Agency. (2022), World Energy Outlook 2022. Available from: https://iea.blob.core.windows.net/assets/830fe099-5530-48f2-a7c1-11f35d510983/worldenergyoutlook2022.pdf

Li, H., Liu, Y., Luo, X., Duan, H. (2022), A novel nonlinear multivariable Verhulst grey prediction model: A case study of oil consumption forecasting in China. Energy Reports, 8, 3424-3436.

Li, Z., Rose, J.M., Hensher, D.A. (2010), Forecasting automobile petrol demand in Australia: An evaluation of empirical models. Transportation Research Part A: Policy and Practice, 44(1), 16-38.

Litterman, R.B., Makridakis, S., Andersen, A., Carbone, R.F., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., & Winkler, R.L. (1986). The Forecasting accuracy of major time series methods. Journal of the American Statistical Association, 81, 262.

Marbuah, George. (2014). Understanding crude oil import demand behaviour in Ghana. Available from: https://mpra.ub.uni-muenchen.de/60436/1/MPRA_paper_60436.pdf

Money Control. (2019), India's Diesel Demand to hit Record Highs in 2019 as Country Goes to Polls. Available from: https://www.moneycontrol.com/news/business/economy/indias-diesel-demand-to-hit-record-highs-in-2019-as-country-goes-to-polls-3597191.html

Murat, Y.S., Ceylan, H. (2006), Use of artificial neural networks for transport energy demand modeling. Energy Policy, 34(17), 3165-3172.

Paladugula, A.L., Kholod, N., Chaturvedi, V., Ghosh, P.P., Pal, S., Clarke, L., Evans, M., Kyle, P., Koti, P.N., Parikh, K., Qamar, S., Wilson, SA. (2018), A multi-model assessment of energy and emissions for India's transportation sector through 2050. Energy Policy, 116, 10-18.

Petroleum Planning and Analysis Cell. (2022), India's Oil & Gas: A Ready Reckoner. Available from: https://ppac.gov.in/uploads/rep_studies/1672496831_1671948944_ready/reckoner/magazine%281%29_compressed.pdf

Puri, M., Pathak, Y., Sutariya, V.K., Tipparaju, S., Moreno, W. (2015), Artificial Neural Network for Drug Design, Delivery and Disposition. Cambridge: Academic Press.

Sa'ad, S. (2009), An empirical analysis of petroleum demand for Indonesia: An application of the cointegration approach. Energy Policy, 37(11), 4391-4396.

Sen, A., Sen, A. (2016), India's Oil Demand: On the Verge of "take-off"? OIES Paper: WPM.

Shan, B., Xu, M., Zhu, F., Zhang, C. (2012), China's energy demand scenario analysis in 2030. Energy Procedia, 14, 1292-1298.

Shin, S.Y., Woo, H.G. (2022), Energy consumption forecasting in Korea using machine learning algorithms. Energies, 15(13), 4880.

Sözen, A., Akçayol, M.A., Arcaklioğlu, E. (2006), Forecasting net energy consumption using artificial neural network. Energy Sources, Part B, 1(2), 147-155.

Tamba, J.G., Essiane, S.N., Sapnken, E.F., Koffi, F.D., Nsouandélé, J.L., Soldo, B., Njomo, D. (2018), Forecasting natural gas: A literature survey. International Journal of Energy Economics and Policy, 8(3), 216-249.

The Mint. (2018), BS 6 Fuel Rollout to Shut Indian Refineries, Push up India's Oil Imports in 2019. Available from: https://www.livemint.com/industry/dgl0icz9m5tf871ttvmhlm/wave-of-refinery-shutdowns-may-push-india-into-importing-fue.html

Tiwari, A.K. (2011), A structural VAR analysis of renewable energy consumption, real GDP and CO2 emissions: Evidence from India. Economics Bulletin, 31(2), 1793-1806.

Wang, Q., Li, S., Li, R. (2018), Forecasting energy demand in China and India: Using single-linear, hybrid-linear, and non-linear time series forecast techniques. Energy, 161, 821-831.