CAPSTONE PROJECT REPORT

(Project Term January-May 2023)

FORECASTING GLOBAL ENERGY CONSUMPTION USING DIFFERENT MACHINE LEARNING ALGORITHMS

Submitted by

Satya Krishna Prahalladh Raju Datla	**Registration Number: 11907950**
Jashwanth Reddy Yerragunta	**Registration Number: 11906192**
Vijay Deepak Karnatakapu	**Registration Number:11903057**
Karunakar Jerripothula	**Registration Number:11910298**
Krishna Vamsi Pothuri	**Registration Number:11902088**
Avinash Korada	**Registration Number:11915141**

Project Group Number: CSERGC0141

Course Code: CSE445

Under the Guidance of

Ms. Savleen Kaur

Assistant Professor

School of Computer Science and Engineering

LOVELY PROFESSIONAL UNIVERSITY

Transforming Education Transforming India

PAC FORM

TOPIC APPROVAL PERFORMA

School of Computer Science and Engineering (SCSE)

Program : P132-H::B.Tech. (Computer Science & Engineering) (Hons.)

COURSE CODE : CSE445 **REGULAR/BACKLOG :** Regular **GROUP NUMBER :** CSERGC0141

Supervisor Name : Savleen Kaur **UID :** 18306 **Designation :** Assistant Professor

Qualification : ________________________ **Research Experience :** ________________________

SR.NO.	NAME OF STUDENT	Prov. Regd. No.	BATCH	SECTION	CONTACT NUMBER
1	Korada Avinash	11915141	2019	K19SD	6303049570
2	Jerripothula Karunakar	11910298	2019	K19EX	6302238563
3	Datla Satya Krishna Prahalladh Raju	11907950	2019	K19PT	8074631215
4	Yerragunta Jashwanth Reddy	11906192	2019	K19YG	7814117173
5	Karnatakapu Vijay Deepak	11903057	2019	K19GA	6281422543
6	Pothuri Krishna Vamsi	11902088	2019	K19EX	8328215129

SPECIALIZATION AREA : Intelligent Systems-I **Supervisor Signature:** ___________________

PROPOSED TOPIC : Forecasting Global Energy Consumption using Machine Learning Algorithms

Qualitative Assessment of Proposed Topic by PAC		
Sr.No.	**Parameter**	**Rating (out of 10)**
1	Project Novelty: Potential of the project to create new knowledge	7.50
2	Project Feasibility: Project can be timely carried out in-house with low-cost and available resources in the University by the students.	7.50
3	Project Academic Inputs: Project topic is relevant and makes extensive use of academic inputs in UG program and serves as a culminating effort for core study area of the degree program.	7.50
4	Project Supervision: Project supervisor's is technically competent to guide students, resolve any issues, and impart necessary skills.	8.00
5	Social Applicability: Project work intends to solve a practical problem.	8.00
6	Future Scope: Project has potential to become basis of future research work, publication or patent.	7.50

PAC Committee Members		
PAC Member (HOD/Chairperson) Name: Harjeet Kaur	UID: 12427	Recommended (Y/N): Yes
PAC Member (Allied) Name: Sudha Shanker Prasad	UID: 21826	Recommended (Y/N): Yes
PAC Member 3 Name: Upinder Kaur	UID: 27954	Recommended (Y/N): Yes

Final Topic Approved by PAC: **Forecasting Global Energy Consumption using Machine Learning Algorithms**

Overall Remarks: Approved

PAC CHAIRPERSON Name: 13714::Dr. Prateek Agrawal **Approval Date:** 06 May 2023

5/8/2023 1:12:30 PM

DECLARATION

We here by declare that the project work entitled "Forecasting Global Energy Consumption using different Machine Learning Algorithms" is an authentic record of our own work carried out as requirements of Capstone Project for the award of B.Tech degree in Computer Science & Engineering from Lovely Professional University, Phagwara, under the guidance of Savleen Kaur, during January to May 2023. All the information furnished in this capstone project report is based on our own intensive work and is genuine.

Project Group Number: CSERGC0141

Name of Student 1: Satya Krishna Prahalladh Raju Datla
Registration Number: 11907950 (Signature of Student 1)
Date:

Name of Student 2: Jashwanth Reddy Yerragunta
Registration Number: 11906192 (Signature of Student 2)
Date:

Name of Student 3: Vijay Deepak Karnatakapu
Registration Number: 11903057 (Signature of Student 3)
Date:

Name of Student 4: Karunakar Jerripothula
Registration Number: 11910298 (Signature of Student 4)
Date:

Name of Student 5: Krishna Vamsi Pothuri
Registration Number: 11902088 (Signature of Student 5)
Date:

Name of Student 6: Avinash Korada
Registration Number: 11915141 (Signature of Student 6)
Date:

CERTIFICATE

This is to certify that the declaration statement made by this group of students is correct to the best of my knowledge and belief. They have completed this Capstone Project under my guidance and supervision. The present work is the result of their original investigation, effort and study. No part of the work has ever been submitted for any other degree at any University. The Capstone Project is fit for the submission and partial fulfillment of the conditions for the award of B.Tech degree in Computer Science & Engineering from Lovely Professional University, Phagwara.

Signature and Name of the Mentor: Savleen Kaur

Designation: Assistant Professor

School of Computer Science and Engineering,
Lovely Professional University,
Phagwara, Punjab.

Date :

ACKNOWLEDGEMENT

Executing such a project under the guidance of our mentor Ms. Savleen Kaur, Assistant Professor, Lovely Professional University for giving us a chance to explore new ideas and expand our knowledge on a topic. With such an opportunity, we are thankful to our mentor and other respected persons who are involved in completing this project.

We would like to thank everyone involved directly or indirectly in writing the research paper and now this report for our project.

Regards,

SATYA KRISHNA PRAHALLADH RAJU DATLA

JASHWANTH REDDY YERRAGUNTA

VIJAY DEEPAK KARNATAKAPU

KARUNAKAR JERRIPOTHULA

KRISHNA VAMSI POTHURI

AVINASH KORADA

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION

Forecasting global energy consumption is a crucial task for policymakers, energy companies, and investors as it helps in making informed decisions regarding energy production and consumption. Machine learning algorithms have emerged as powerful tools for analysing energy data and making accurate predictions. In this topic, we will explore the application of various machine learning algorithms for forecasting global energy consumption. We will discuss the pros and cons of different models, their suitability for various types of data, and the challenges associated with energy consumption forecasting. This topic will provide insights into the techniques used in energy consumption prediction and their potential to contribute to sustainable energy planning and management.

Energy is an essential resource for the modern civilisation, powering everything from homes, businesses, transportation and manufacturing. The increase in demand for energy raised concerns on availability of traditional energy sources such as fossil fuels and their impact on environment. Renewable resources offer a feasible solution to these challenges, but the transition to more sustainable energy system is more complex and require careful planning and investment. Machine learning plays a critical role in optimising the use of both renewable and non-renewable resources and reducing waste and emissions.

Renewable resources are energy sources such as solar, wind, hydro and geothermal energy that can replace non-renewable resources. They are considered sustainable energy sources because they are not finite and do not produce greenhouse gas emissions or other pollutants. Machine learning can be used to analyse and predict patterns in energy consumption and resource consumption.

Solar power is one of the widely used renewable energy source, with solar panels now being used in homes and businesses around the world. It is the fastest growing renewable energy sources in the modern society. It is particularly popular in sunnier regions, such as Australia, where the cost of solar power is now competitive with traditional fossil fuel-based energy sources. Wind power is another popular renewable energy source in regions with high wind speeds. In many parts of Europe and the United States, wind turbines are used to produce energy. It can also be combined with other energy sources such as solar power to provide more reliable and stable energy.

Predictive models can be developed to optimise the operation of renewable energy systems based on historical data and weather patterns. The predicted data can be used to schedule energy storage systems, switch between renewable and non-renewable energy sources and manage energy consumption. Machine learning can be used to improve efficiency and credibility of renewable systems. Machine Leaning algorithms can be used to optimise placement of wind turbines to maximise their efficiency and minimise their impact on local wildlife. Predictive models can be developed to optimise the production of oil and gas with the help of geological data and production history. It can lead to improved efficiency and reduced environmental impact.

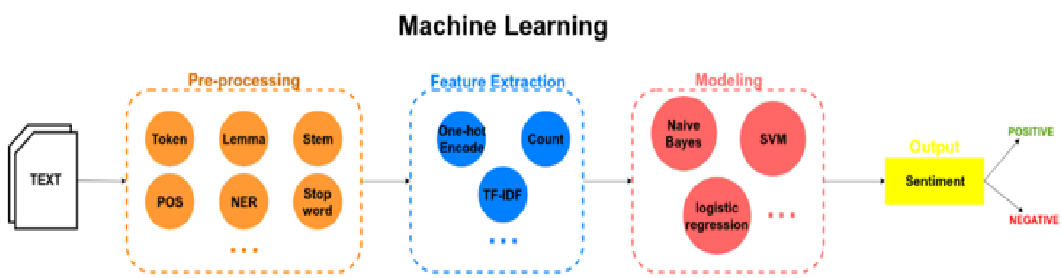

Figure 1: Process of Machine Learning

A machine learning algorithm, a mathematical model or set of guidelines, enables a computer to acquire knowledge from data and make recommendations or judgements with no need to be programmed beforehand. The three primary types of machine learning algorithms are: -

1.Supervised Learning
2.Unsupervised Learning
3.Reinforcement Learning

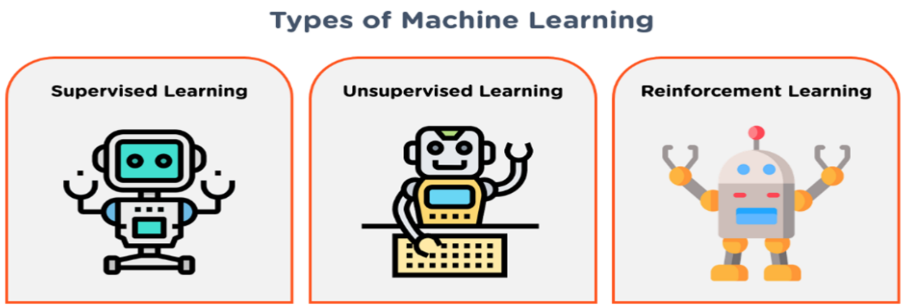

Figure 2: Types of machine learning

Supervised Learning:

The machine is trained using a labelled dataset, which comprises of input variables and associated output variables, in supervised learning, a way of machine learning. Aim of this algorithm is to find an algorithm that transforms inputs into outputs and can generate accurate predictions about fresh, uncontaminated data. Problems with supervised learning often fall into one of two categories: classification or regression.

In a regression problem, the output variable is a fixed value such as price or gold. Finding the best line that minimizes the difference between the estimate and the true value is the goal of repeated use of the regression problem, called linear regression. In classification problems, categorical values such as labels or categories are produced. The method that is frequently used to solve binary classification problems when there are only two possible values for the output variable is logistic regression. For multivariate classification problems, alternative methods are widely used including decision trees, random forests, and support vector machines (SVM).

Building a supervised learning model entails the following steps:

1. **Data preparation**: Entails gathering, cleaning, and dividing the data into training and testing sets.
2. **Feature selection**: This process entails deciding which data elements are most crucial for the algorithm to be trained on.
3. **Model selection**: Based on the properties of the data and the issue, this step entails selecting the right method and its hyperparameters.

4. **Training the model**: entails entering training data into the algorithm and modifying the parameters of the model to reduce the discrepancy among expected and actual results.
5. **Evaluating the model**: This entails assessing the model's performance using the testing data and making any required modifications to increase its accuracy.

Many different industries, including finance, healthcare, and marketing, use supervised learning extensively for tasks like forecasting consumer behaviour, spotting fraud, and identifying diseases. Supervised learning algorithms can generate precise predictions on new data by learning from prior data, allowing companies and organisations to make better decisions and perform better.

Types of Supervised Learning Algorithms:

- **Linear Regression:** For forecasting continuous output values, linear regression is a straightforward and widely used approach. It depicts the relationship as a linear function between the input characteristics and the output variable.
- **Logistic Regression:** The classification process known as logistic regression estimates the likelihood that a given occurrence will fall into a certain class. A logistic function is used to model the connection among the input characteristics and the variable that is output.
- **Decision Trees:** For both classification and regression applications, decision trees provide a flexible and understandable approach. To create predictions, they acquire a set of if-then rules based on the input attributes.
- **Random Forest:** Multiple decision trees are combined in the Random Forest algorithm to produce more precise predictions. Using arbitrarily chosen portions of the training data and characteristics, several decision trees are generated.
- **Support vector machines (SVMs):** are a potent technique for classification problems, especially for datasets with intricate boundaries for decisions. They acquire an optimal decision boundary for the margin between classes.
- **Naive Bayes:** For text classification problems, the approach known as probabilistic Naive Bayes is frequently employed. It employs the Bayes theorem to determine the probability of each class and makes the assumption that the characteristics are not dependent on one another.

- **K-Nearest Neighbours (k-NN):** is a straightforward, not parametric approach that may be used for both classification and regression applications. Finding the k closest training cases in the domain of features and making a forecast based on their output labels is how it predicts the output label.

There are many additional supervised learning algorithms to consider, but these are just a few of the most popular ones. The kind of issue, the quantity and quality of the data, and other elements like accessibility, precision, and computing complexity all influence the solution that is chosen.

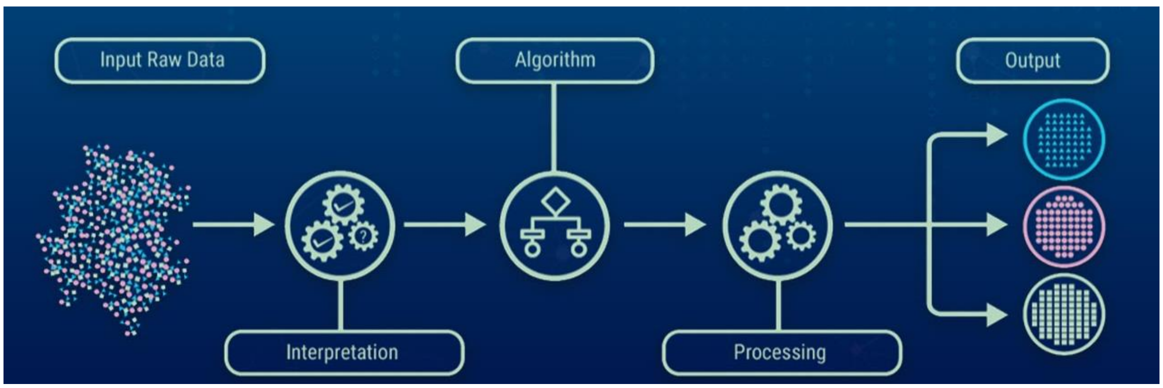

Figure 3: Process of supervised learning

Unsupervised Learning:

Unsupervised learning is a sort of machine learning in which the computer is trained using just input variables from an unlabelled dataset. Without being given a specific result to forecast, the algorithm's objective is to find patterns or structure in the data, such as clusters or relationships.

Unsupervised learning issues mostly fall into two categories dimensionality reduction and clustering. In situations involving clustering, the algorithm creates clusters out of related data points depending on how similar the input variables are to those points. K-means, hierarchical clustering, and DBSCAN are a few popular clustering techniques.

In situations involving dimensionality reduction, the approach minimises the dimensionality of high-dimensional data while preserving the majority of the original variance. For examining and visualising large, complicated data sets, this is helpful. A popular approach for reducing dimensionality is Principal Component Analysis (PCA).

There are various steps involved in creating an unsupervised learning model:

1. **Data preparation:** Entails gathering, cleaning, and preparing the data for analysis.
2. **Feature engineering**: This technique entails scaling or altering the data to make it appropriate for the algorithm.
3. **Model selection**: Based on the qualities of the data and the issue, the right method is selected.
4. **Training the model**: entails supplying the algorithm with data and modifying the model's parameters to arrive at the best outcome.
5. **Evaluating the model**: This entails assessing the model's performance using different metrics and making any required modifications.

Unsupervised learning is frequently used for tasks like grouping customer groups, spotting anomalies, and extracting useful characteristics from complicated data sets in a variety of industries, including marketing, picture and audio recognition, and natural language processing. Unsupervised learning algorithms may assist businesses and organisations in gaining useful insights and improving decision-making by identifying patterns and structure in the data.

Types of Unsupervised Learning Algorithms:

Unsupervised machine learning algorithms are those that learn from unlabelled data, or data that has not been associated with output labels that relate to it. The following are some popular unsupervised learning algorithms:

- **Clustering:** Using the clustering technique, similar instances are grouped together based on how similar they are to one another in the feature space. The two popular clustering techniques are hierarchical clustering and K-means.
- **Principal Component Analysis (PCA):** This is a dimensionality reduction method that reduces the amount of information in data with high dimensions while preserving as much of the original variance as feasible.

- **Independent Component Analysis (ICA)**: It is a method for blind source separation that divides mixed signals into their constituent parts. It is frequently employed in image analysis and processing of signals.
- **Association Rule mining:** An approach used to find patterns and connections in huge datasets is called association rule mining. To find frequently occurring goods, analysis of market baskets typically uses this technique.
- **Autoencoders:** A particular class of neural network that can learn to encode input data into a lower-dimensional representation and then decoded it to restore it to its original form is called an autoencoder. These are frequently employed for anomaly detection and dimensionality reduction.
- **Generative Adversarial Networks (GANs):** These are a form of deep learning model that learn to produce fresh data in a manner similar to the data used for training. They consist of a discriminator and a generator, two neural networks that are trained concurrently in a gaming environment.

There are numerous additional unsupervised learning algorithms as well, but these are some of the most popular ones. The kind of problem, the quantity and quality of the data, and additional elements like interpretability and computational complexity all influence the solution that is chosen.

Reinforcement Learning:

Reinforcement learning is a type of machine learning that teaches an agent to make decisions in a specific environment by gathering data in the form of rewards or punishments. The method aims to maximise the overall compensation over time.

In reinforcement learning, the agent interacts with the environment by acting, and the environment reacts by sending a signal that indicates whether the action was successful or unsuccessful in the form of a reward. The broker wants to find the policy the case-to-action map that yields more profit than forecast.

One of the fundamental challenges in reinforcement learning is the trade-off between exploration and exploitation. The computer programme must investigate its surroundings to choose the best course of action, but it must also take use of what it has learnt to increase the total payoff. The algorithms Q-learning, SARSA, and Deep Q-Networks (DQNs) are some of the most popular ones used in reinforcement learning.

Building a reinforcement learning model entails the following steps:

1. **Defining the environment**: This entails specifying the environment's guidelines and limitations, as well as the activities that the agent is permitted to conduct.
2. **Creating the incentive function**: This entails creating the reward function, which informs the agent whether or not its activities were successful.
3. **Educating the agent**: This entails providing it with knowledge about surrounding factors and rewards, then using an algorithm to modify its behaviour to maximise the anticipated aggregate benefit.
4. **Assessing the agent**: This entails evaluating the drug's efficacy using fresh, unrecognised data and making any required modifications.

For activities like managing a robot arm, playing video games, and operating a self-driving

automobile, reinforcement learning is widely employed in a variety of industries. Reinforcement learning algorithms can make judgements in real-time, adapt to shifting circumstances, and maximise incentives by learning from past.

Figure 4: Process of Reinforcement Learning

2. LITERATURE REVIEW

Energy which is an indispensable component for smooth operation of many things around the world. Progress of humans wouldn't have been such successful if it isn't for energy fueling it. Prediction of energy demand is important for any nation for proper energy planning. For prior forecasting of energy consumption many prediction models were derived using different factors ranging from socio-economical to geographical and environmental.

Kadir et al., developed an MLP Artificial Neural Network model for forecasting electricity usage by considering various socio-economic parameters like population and imports as independent factors. Authors concluded that using random data set than deterministic data set has shown better results [5]. An ANN model has been developed by Szoplik for forecasting the natural gas usage. After trial-and-error MLP ANN with 22-36-1 architecture gave better results with MAPE being as low as 5.6% **[6].** In Iran to predict the consumption of renewable energy ANN has been proposed and comparison with conventional and fuzzy regression models has been done. The results showed a high accuracy of about 99.9% and low MAPE for ANN when compared to other traditional regression methods [7].

An ANN with MLP has been proposed to predict the consumption of energy on long-term basis in Greek and the results showed similarity with that of SVM but a variation has been observed with the values of linear regression [8]. Most recently comparison of three different algorithms Random Forest, Long Short-Term Memory and XGBoost has been done to predict energy consumption in Korea. Researchers applied the algorithms in two periods. One before COVID-19 pandemic and the other after the slowdown in spread of COVID-19 and concluded that LSTM showed better accuracy in period 1 and Random Forest performed best in period 2 [9]. In the year 2004 ABDUSSELAM et al. proposed fuzzy logic algorithm to predict the consumption of water in the Turkish city of Istanbul on monthly basis by considering the previous three months consumption values as independent variables with an error rate being less than 10% [4]**.**

For the estimation of consumption profile of natural gas for the following day an LSTM which is a recurrent neural network has been proposed [1]**.**In their research, Lipan Fan et al. provided a data mining and support vector machine-based model for forecasting energy usage in residential structures [10]. Mahjoub et al. innovative LSTM network-based solution for predicting periodic energy usage was tested using actual measurement data from a French city. The outcomes demonstrated that the suggested strategy performed better than conventional forecasting techniques like ARMA [11].Yinlong et al. tested three algorithms for forecasting annual energy and electricity usage in UK. The three algorithms are Multiple Linear regression, LS-SVM and Neural Network with backpropagation training. Among these LS-SVM performed well compared to others [12].

As such LS-SVM algorithm, gave higher accurate values than traditional regression analysis and neural network in forecasting electricity consumption in Turkey [13]. Volkan et al. proposed auto-regressive models for forecasting the need of energy sourced from fuel in Turkey [14]. Electricity Consumption has been forecasted using linear regression model in Italy by considering historical electricity consumption and GDP as input parameters [15]. Authors proposed an SVR model to predict electricity consumption of Turkey by using socio-economic factors like population, GDP, imports, and exports. An individual forecasting model of each independent variable has been created to predict the future electricity consumption and predicted that for the year 2026 consumption would be 284.9 TWh [16].

A MLP neural network has been modeled to get the estimated coal usage values till 2030 [17]. An ANN has been proposed for predicting monthly wind power generation in India [18]. A hybrid CNN-LSTM was modeled for the prediction of consumption of residential energy in Korea [19]. Neural network and conditional demand analysis has been compared to predict the sudden changes in terms of residential energy usage in Canada and it was concluded that neural networks work better with socio-economic factors [20] . LSTM and SVM had minor difference in estimated values of energy [21]. Continuing the trend of using LSTM Wang et al. used it for predicting energy consumption with periodicity [22].

3. SCOPE OF THE STUDY

3.1 DESCRIPTION OF STUDY

The Main prospective of this study is to develop a predictive model for energy consumption which can be modified to forecast future energy demand accurately. The study aims to enhance understanding of energy consumption patterns, identify influential factors, and provide insights for energy planning, resource allocation, and policy-making.

- **Data Collection:** Relevant data on energy consumption and related factors are collected from reliable sources such as energy agencies, government reports, research publications, and industry databases. The data may include historical energy consumption records, socio-economic indicators, climate data, population demographics, energy prices, and technological variables.
- **Data Pre-processing:** The collected data is cleaned, organized, and prepared for analysis. This involves handling missing values, outlier detection and treatment, data normalization, and feature engineering. Time series data may also be aggregated or transformed to capture relevant temporal patterns.
- **Exploratory Data Analysis:** Descriptive analysis and visualizations are performed to gain insights into the characteristics of the energy consumption data. This step involves identifying trends, seasonality, cyclical patterns, and correlations between energy consumption and other variables. The analysis helps in understanding the data distribution and identifying potential predictors.
- **Feature Selection:** Statistical methods such as correlation analysis or importance of methods are used to select the most important features to estimate energy. This step aims to reduce dimensionality and focus on variables that have a significant impact on energy demand.
- **Model Development:** Various predictive modelling techniques are explored to develop a robust energy consumption prediction model. This may include regression-based approaches neural network models.
- **Model Training and Validation:** The selected prediction model is trained using historical energy consumption data and validated using appropriate evaluation metrics such as mean squared error (MSE), mean absolute error (MAE), or root mean squared

error (RMSE). The model is fine-tuned by adjusting hyperparameters and validating against different time periods to ensure its generalizability.

- **Prediction and Forecasting:** The trained model is then used to make energy consumption predictions for future time periods. This allows for forecasting energy demand based on different scenarios, policy interventions, or changes in influential factors. Uncertainty analysis or confidence intervals may be incorporated to assess the reliability of the predictions.
- **Model Evaluation:** Evaluate the performance of the prediction model by comparing the estimated benefit values of with actual observations over the lifetime of the model. The metrics measured the accuracy, reliability, and robustness of prediction models.
- **Interpretation and Insights:** The study concludes by interpreting the results, providing insights into the factors influencing energy consumption, and discussing the implications for energy planning, policymaking, and sustainability. The findings may highlight opportunities for energy efficiency measures, renewable energy integration, demand-side management, or infrastructure development.

3.2 LIMITATIONS OF THE STUDY

Forecasting global energy consumption using machine learning algorithms has become an active research area in recent years. While these techniques have shown promising results, there are still several limitations that need to be considered. Here are some of the limitations:

1. **Data availability and quality:** The accuracy of the prediction model is very high due to the availability and quality of data used for training. Energy consumption data can be challenging to obtain and may be subject to errors or missing values, which can affect the accuracy of the forecasts.
2. **Lack of standardized methodology:** There is no standardized methodology for forecasting energy consumption, which can lead to variations in the results obtained using different machine learning algorithms.
3. **Complexity of energy systems:** Energy systems are complex, dynamic, and influenced by various factors such as economic conditions, policy changes, and technological

advancements. These factors can make it difficult to develop accurate forecasting models.

4. **Limited model interpretability**: Some machine learning algorithms, such as neural networks, are known for their black-box nature, making it difficult to interpret the reasoning behind the predictions they generate. This can limit their usefulness in providing insights into the drivers of energy consumption.
5. **Limitations of historical data:** Energy consumption patterns may change over time due to various factors, such as shifts in energy policies or technological advancements. Historical data may not always be an accurate representation of future trends, making it challenging to develop accurate forecasts.
6. **Computational resources:** Some machine learning algorithms require significant computational resources, which may not be readily available in some contexts, particularly in developing countries.
7. **Overfitting and underfitting:** Overfitting is when the model is close to the training data and gets worse for new data. Controversy arises when the model is too simple and does not show differences in the data. It can be difficult to strike the correct balance between generalization performance and model complexity.

3.3 FUTURE RESEARCH DIRECTIONS

To overcome the limitations of the study future research can explore Hybrid models and Data augmentation.

- **Hybrid models:** Combining multiple machines learning algorithms or integrating machine learning with traditional forecasting methods can improve the accuracy and robustness of energy consumption forecasts. Hybrid models can leverage the strengths of different methods to overcome the limitations of individual models.
- **Data augmentation:** Generating synthetic data can be used to supplement limited or incomplete historical data, improving the accuracy of forecasting models. Data augmentation techniques can also be used to simulate different scenarios and evaluate the impact of external factors on energy consumption.

3.4 PROJECT PLAN

- **Define the problem**: Clearly define the problem and the goals of the project. Identify the target variable(s) and the dataset(s) that will be used for the analysis.
- **Collect and preprocess data**: Collect the relevant datasets and preprocess them for analysis. This might involve cleaning, merging, and transforming data, as well as dealing with missing values and outliers.
- **Explore and visualize data**: Explore the data to gain insights into the patterns, trends, and relationships between the variables. Use visualizations to help understand the data and identify any outliers or anomalies.
- **Feature engineering**: Create new features or extract relevant features from the data that might help improve the accuracy of the machine learning algorithms. This might involve feature scaling, dimensionality reduction, or feature selection techniques.
- **Train and test models**: Select appropriate machine learning algorithms, such as regression models, time series models, or neural networks, and train them on the preprocessed data. Use cross-validation techniques to evaluate the performance of the models and fine-tune their parameters.
- **Refine and improve**: Continuously refine and improve the model over time by collecting new data, fine-tuning the parameters, or experimenting with new algorithms.

4. OBJECTIVES

1. To study and analyze existing approaches for Energy consumption.
2. Pre-processing, data cleaning and feature engineering in R.
3. Train and Evaluate the Machine learning algorithm to analyze the total global energy consumption and energy intensity improvement.
4. Show the trend of global nuclear energy generation and consumption.
5. Prediction of dependency on renewable resources than non-renewable resources.
6. Analysis of energy supply from renewable resources.

- **Understanding energy consumption patterns:** Analyzing global energy consumption helps identify patterns and trends in energy usage across different countries, regions, sectors, and energy sources. This understanding can provide insights into the factors influencing energy consumption, such as economic development, population growth, industrial activities, and technological advancements.
- **Assessing energy sources:** Analyzing global energy consumption involves evaluating the mix of energy sources used worldwide, including fossil fuel such as coal, oil, and natural gas), renewable energy (such as solar, wind and hydro), and nuclear energy. Understanding the distribution and proportion of different energy sources helps assess the sustainability, environmental impact, and resilience of global energy systems.
- **Assessing renewable energy capacity and generation:** Analyzing the energy supply from renewable resources involves evaluating the installed capacity and actual generation from renewable energy sources, such as solar, wind, hydroelectric, geothermal, and biomass. This assessment helps understand the current contribution of renewables to the overall energy mix and their potential to meet future energy demands.
- **Evaluating renewable energy growth and trends:** Tracking the growth and trends of renewable energy supply provides insights into the rate of adoption and the pace of renewable energy deployment globally. Analyzing factors such as investment trends,

policy support mechanisms, technological advancements, and market dynamics helps identify the drivers and barriers for renewable energy expansion.

- **Comparing renewable energy share and transition targets:** Analyzing the share of renewable energy in the overall energy mix allows for comparisons between different countries or regions. Evaluating the progress towards renewable energy targets, such as national renewable energy goals or international commitments (e.g., Paris Agreement), helps assess the effectiveness of policies and measures aimed at promoting renewable energy adoption and facilitating energy transition.
- **Identifying regional and country-specific dynamics:** Analyzing renewable energy supply requires considering regional and country-specific dynamics, including resource availability, policy frameworks, regulatory environments, and socio-economic factors. Understanding these dynamics helps identify the unique opportunities and challenges associated with renewable energy development in different regions and countries.

5. SOFTWARE REQUIREMENT ANALYSIS

R is a software environment and programming language for statistical computation and graphics. Ross Ihaka and Robert Gentleman of the University of Auckland in New Zealand developed it in the early 1990s.

R is often used for applications including data processing, statistical modelling, visualisation, and machine learning. It is especially well-liked in the data science and statistical analysis fields. For these objectives, R offers a large variety of built-in functions and libraries in addition to a strong and flexible language for writing new functions and analyses.

A library in R is a group of pre-written programmers that may be used to increase the language's capability. Functions, data sets, and other objects that may be imported into an R session and utilized for a particular job or analysis can be found in libraries.

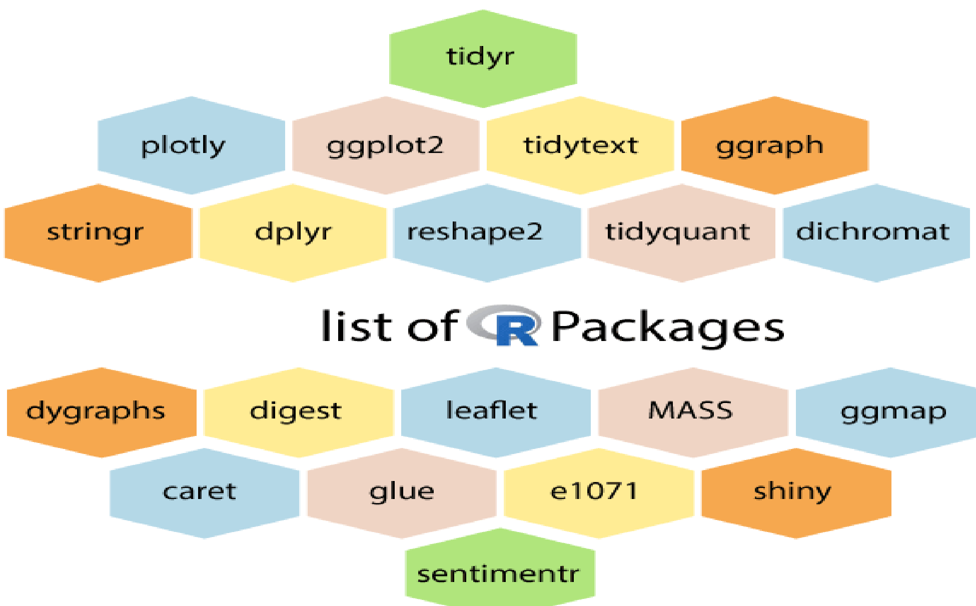

Figure 5: List of R Packages

5.1. LIBRARIES USED IN THE PROJECT

5.1.1. dplyr

'dplyr' offers functions for working with "window functions" like 'row_number()', 'cumsum()', and 'lag()' in addition to these fundamental ones. These functions include functions for joining data frames ('left_join()', 'inner_join()', 'full_join()', etc.) and joining data frames.

One of 'dplyr's' advantages is that it offers a standard syntax for various data manipulation activities, which makes it simple to create understandable and succinct code. Using 'dplyr', for instance, the following code may filter rows from a data frame, choose columns, group by one variable, and compute summary statistics for another variable:

library(dplyr)

5.1.2. ggplot2

A robust and adaptable framework for producing visuals and visualizations is offered by the R programme ggplot2. It is founded on the ideas of the "Grammar of Graphics" (thus the "gg" in the name), which offers a systematic approach to thinking about data visualization and permits highly configurable and expressive images.

For making various visuals, such as scatter plots, line charts, bar charts, histograms, and more, the ggplot2 programme offers a large range of functions. You may alter the colour, fonts, labels, and layout of your graphics as well as any other element of their outward look.

The "layered" method of ggplot2's visual creation is one of its advantages. As a result, you may add different layers of graphic components to a plot, such as points, lines, and text, and alter each layer independently. By merging several layers, you may also quickly construct complicated visualizations like faceted plots and heat map displays.

5.1.3. neuralnet

An R package called neuralnet offers tools for building and evaluating neural networks. Neural networks are a type of machine learning model that are based on the structure and function of the human brain. They may be applied to a range of tasks, including classification, regression, and prediction.

A straightforward and adaptable interface for building and training neural networks in R is provided by the neuralnet package. You may define the number of layers, the number of nodes in each layer, and the activation function used for each node in the neural network's structure. Additionally, it offers choices for defining the convergence criterion, the number of iterations, and the training method.

5.2. TABLEAU PREP

Data cleansing and preparation software called Tableau Prep is provided by Tableau, a business intelligence and data visualisation software provider. Tableau Prep is a tool created to assist customers in getting their data ready and clean for analysis in Tableau Desktop, another one of the company's products.

Users using Tableau Prep may execute data preparation activities including cleaning, converting, and merging data by connecting to a wide range of data sources, including databases, spreadsheets, and cloud services. With the use of a drag-and-drop method, it offers a straightforward and intuitive interface for visually browsing and editing data.

Tableau Prep has a number of important features, including:

- **Visual Data Preparation:** Tableau Prep offers a drag-and-drop user interface that makes exploring and modifying data graphically simple. Using a number of built-in procedures, users may quickly combine, filter, aggregate, pivot, and transform data.
- **Data Connectors:** Tableau Prep offers connectors for a variety of data sources, including databases, spreadsheets, cloud services, and big data platforms, enabling users to quickly connect to and access their data.

- **Data Cleaning:** To assist consumers in purification and standardizing their data, Tableau Prep contains a variety of built-in cleaning procedures, including deduplication, data type conversion, and data validation.

- **Data Integration:** Tableau Prep makes it simple to mix and integrate data from several sources, enabling users to do cross-functional analysis and discover information that may not be accessible through examination of a single source of data.

- **Automation:** Tableau Prep contains automation tools that let users plan data refreshes and build recurring data flows.

Tableau Desktop, a different service offered by Tableau, is intended to be used in conjunction with Tableau Prep. Users may quickly connect to the data in Tableau Desktop to build visualizations, visualizations, and statistics after the data has been created in Tableau Prep.

Figure 6: Tableau Prep

6. METHODOLOGY

6.1. ANN

A particular kind of machine learning algorithm called an ANN is modelled after the composition and operation of the human brain. ANNs are made up of several linked nodes, or neurons, arranged in layers. Every neuron takes in inputs, processes them, and then produces an output that is sent to the layer below.

A set of inputs and associated outputs are given to an ANN during training, and it modifies its internal parameters (weights) to reduce the discrepancy between the expected and actual outputs. Calculating the gradient of the error with regard to the weights and making the necessary adjustments are steps in the backpropagation process.

ANNs are useful for a variety of tasks, such as pattern recognition, prediction, classification, and optimisation. They are capable of learning intricate nonlinear correlations between inputs and outputs. They have been used in fields including speech recognition, picture recognition, natural language processing, and financial forecasting.

6.2. Boruta

Boruta is a feature selection algorithm used in the machine learning to choose the most relevant Features which are in a dataset. The algorithm is designed to handle datasets with a large number of features (also known as predictors or independent variables) and can be used with various machine learning models.

The Boruta algorithm is an extension of the Random Forest algorithm, a popular ensemble learning technique. The algorithm works by comparing the importance of each feature in the original dataset to the importance of randomized versions of the same dataset. The randomized versions of the dataset are created by randomly permuting the values of each feature. The randomization process ensures that any irrelevant features are assigned low importance scores by the Random Forest model.

Figure 7: Boruta

The Boruta algorithm works in the following steps:

1. Create a duplicate set of the original dataset, and randomize the values of the features in the duplicated set.
2. Train a Random Forest model on the original dataset and the randomized duplicate set.
3. Calculate the feature importance scores for each feature in the original dataset using the Random Forest model.
4. Compare the feature importance scores of each feature in the original dataset to the maximum feature importance score of the corresponding features in the randomized duplicate set.
5. If a feature has a higher importance score than the maximum importance score of its corresponding features in the randomized duplicate set, it is considered significant and is marked as "Confirmed."
6. If a feature has an importance score that is not significantly higher than the maximum importance score of its corresponding features in the randomized duplicate set, it is marked as "Rejected."
7. Repeat steps 1-6 until all features have been either confirmed or rejected.

After running the Boruta algorithm, the most relevant features can be selected for use in a machine learning model. The confirmed features are considered important for the model, while the rejected features are considered unimportant and can be removed from the dataset. The Boruta algorithm can be used with various machine learning models which includes linear regression, logistic regression, and support vector machines.

6.3. MLP

Multi-Layer Perceptron (MLP) is a kind of Artificial Neural Network (ANN) which is used in machine learning for classification and regression tasks. It consists of one input layer, one or more hidden layers, and one output layer.

The input layer gets the input data, which is then processed into the hidden layers. The hidden layers consist of nodes or neurons that perform computations on the input data.

Every neuron present in the hidden layer receives inputs from the previous layer and a weighted sum of the inputs, followed by the application of an activation function. The activation function introduces non-linearity into the network and allows the network to learn complex patterns in the input data.

The output layer produces the final output of the network, which is typically a classification or regression result. The output layer consists of one or more neurons, depending on the type of task. For classification tasks, the output layer typically has one neuron per class, and the output of each neuron represents the probability of the input data belonging to that class. For regression tasks, the output layer has one neuron that produces a continuous output.

The backpropagation approach is used to train the MLP algorithm, which modifies the weights of the neurons in the network to reduce the discrepancy between the expected output and the actual output. In order to minimize the error, the backpropagation method calculates the gradient of the error with respect to the network weights and modifies the weights in the opposite direction of the gradient.

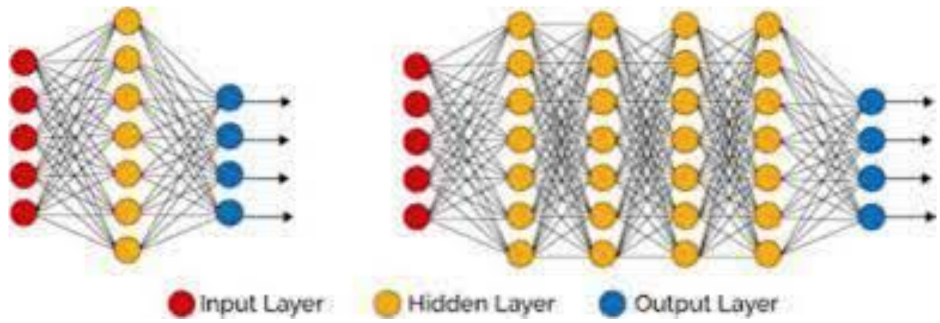

Figure 8: Multi-Layer Perceptron

MLP with ANN is used in various machine learning applications, including image and speech recognition, NLP, and predictive modeling. MLP with ANN is popular because of its ability to learn high complex patterns in the input data, making it suitable for tasks that require high accuracy and robustness.

7. DATA PREPARATION

Data has been sourced from the website owid energy, using the tableau prep the data of different countries has been added to form the specific dataset which is used for this project. Null Values and all the unwanted columns were removed.

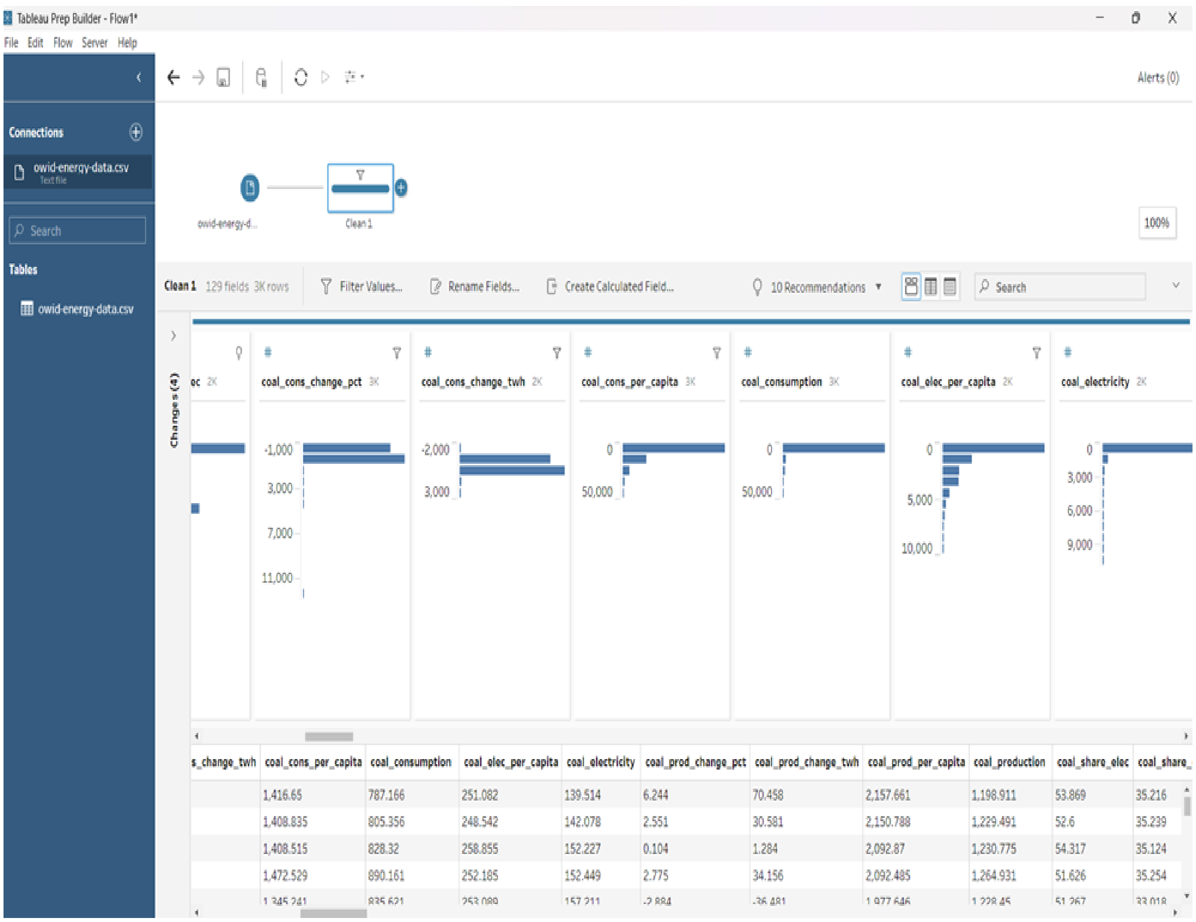

Figure 9: Preparation of data using Tableau prep

S.no	Column	Description
1	country	Geographic location
2	year	Year of observation
3	population	Population
4	gdp	Total real gross domestic product, inflation-adjusted
5	coal_cons_change_twh	Annual change in coal consumption, measured in terawatt-hours
6	coal_consumption	Primary energy consumption from coal, measured in terawatt-hours
7	coal_electricity	Electricity generation from coal, measured in terawatt-hours
8	coal_prod_change_twh	Annual change in coal production, measured in terawatt-hours
9	coal_production	Coal production, measured in terawatt-hours
10	coal_share_elec	Share of electricity generation that comes from coal.
11	coal_share_energy	Share of primary energy consumption that comes from coal
12	gas_cons_change_twh	Annual change in gas consumption, measured in terawatt-hours
13	gas_consumption	Primary energy consumption from gas, measured in terawatt-hours
14	gas_electricity	Electricity generation from gas, measured in terawatt-hours
15	gas_prod_change_twh	Annual change in gas production, measured in terawatt-hours
16	gas_production	Gas production, measured in terawatt-hours
17	gas_share_elec	Share of electricity generation that comes from gas
18	gas_share_energy	Share of primary energy consumption that comes from gas
19	nuclear_cons_change_twh	Annual change in nuclear consumption, measured in terawatt-hours
20	nuclear_consumption	Primary energy consumption from nuclear power, measured in terawatt-hours
21	nuclear_electricity	Electricity generation from nuclear power, measured in terawatt-hours
22	nuclear_share_elec	Share of electricity generation that comes from nuclear power
23	nuclear_share_energy	Share of primary energy consumption that comes from nuclear power
24	oil_cons_change_twh	Annual change in oil consumption, measured in terawatt-hours

25	oil_consumption	Primary energy consumption from oil, measured in terawatt-hours
26	oil_electricity	Electricity generation from oil, measured in terawatt-hours
27	oil_prod_change_twh	Annual change in oil production, measured in terawatt-hours
28	oil_production	Oil production, measured in terawatt-hours
29	oil_share_elec	Share of electricity generation that comes from oil
30	oil_share_energy	Share of primary energy consumption that comes from oil
31	renewables_cons_change_twh	Annual change in renewable energy consumption, measured in terawatt-hours
32	renewables_consumption	Primary energy consumption from renewables, measured in terawatt-hours
33	renewables_electricity	Electricity generation from renewables, measured in terawatt-hours
34	renewables_share_elec	Share of electricity generation that comes from renewables
35	renewables_share_energy	Share of primary energy consumption that comes from renewables
36	solar_cons_change_twh	Annual change in solar consumption, measured in terawatt-hours
37	solar_consumption	Primary energy consumption from solar, measured in terawatt-hours
38	solar_electricity	Electricity generation from solar, measured in terawatt-hours
39	solar_share_elec	Share of electricity generation that comes from solar
40	solar_share_energy	Share of primary energy consumption that comes from solar
41	wind_cons_change_twh	Annual change in wind consumption, measured in terawatt-hours
42	Wind_consumption	Primary energy consumption from wind, measured in terawatt-hours.

43	wind_electricity	Electricity generation from wind, measured in terawatt-hours
44	wind_share_elec	Share of electricity generation that comes from wind.
45	wind_share_energy	Share of primary energy consumption that comes from wind

8. IMPLEMENTATION

ANN model has been derived to implement and predict the various consumption values. Derived ANN model has been used to predict the consumption of energy of varied sources both renewable and non-renewable.

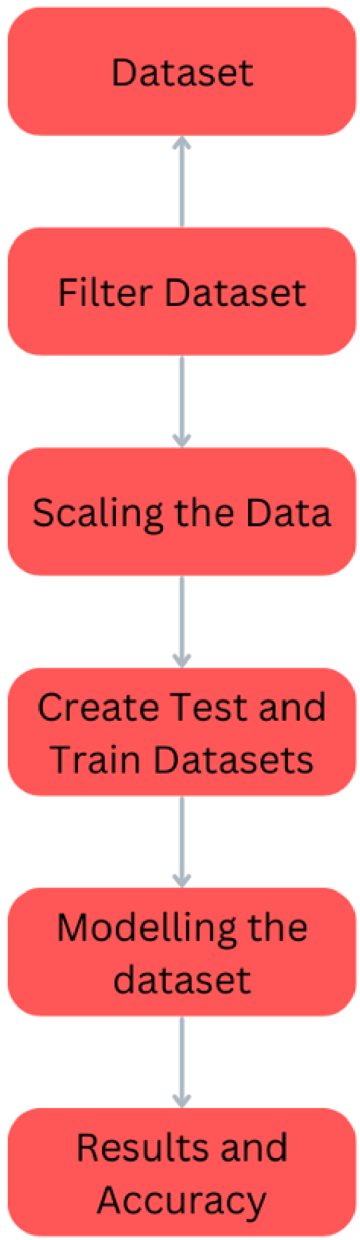

Figure 10: DFD

Important Independent variables are selected by using Boruta algorithm which is based on Random Forest to get better accuracy**.**

Depending upon the dependent variable different independent variables are used to predict the outcome.

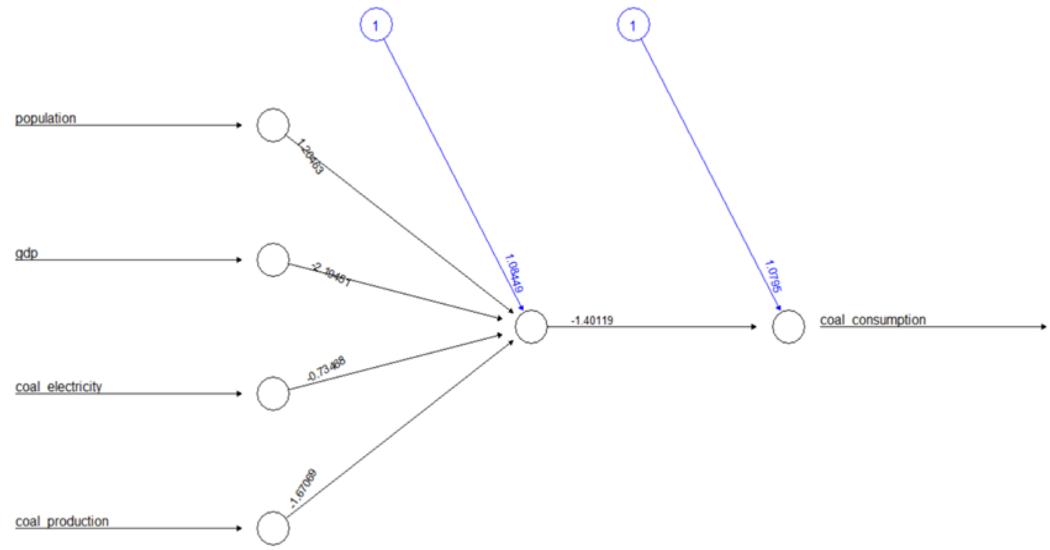

Figure 11: Model Plot

The above Image shows for Coal Consumption 4 dependent variables are used namely Population, GDP, Coal Electricity and Coal Production.

9. RESULTS AND ANALYSIS

Accuracy Table: -

S. No	Dependent Variable	Accuracy
1	Coal Consumption	85% - 90%
2	Gas Consumption	75% - 80%
3	Oil Consumption	80% - 90%
4	Nuclear Consumption	70% - 80%
5	Solar Consumption	60% - 70%
6	Wind Consumption	65% - 73%
7	Hydro Consumption	68% - 75%

Analysis of Non – renewable energy sources like Oil, Gas, Coal and Nuclear consumption has been predicted using the model.

Global Coal Consumption:

Using the model global coal consumption has been predicted.

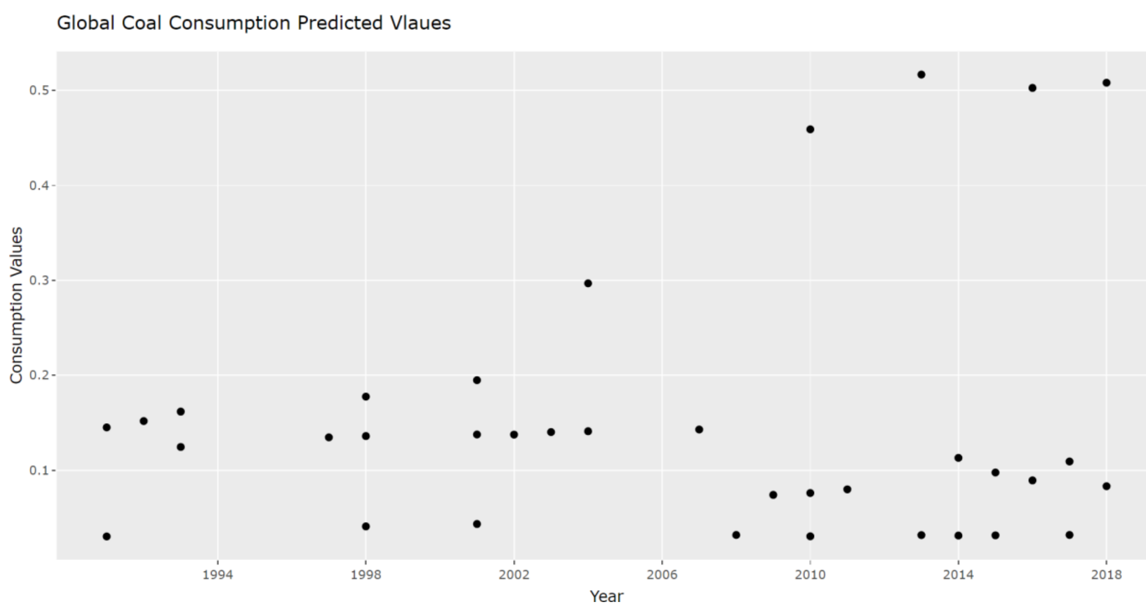

Figure 12: Global Coal Consumption

The above graph depicts the results of the predicted values of Global Coal Consumption. Using the model Individual Country wise prediction has been performed.

United States:

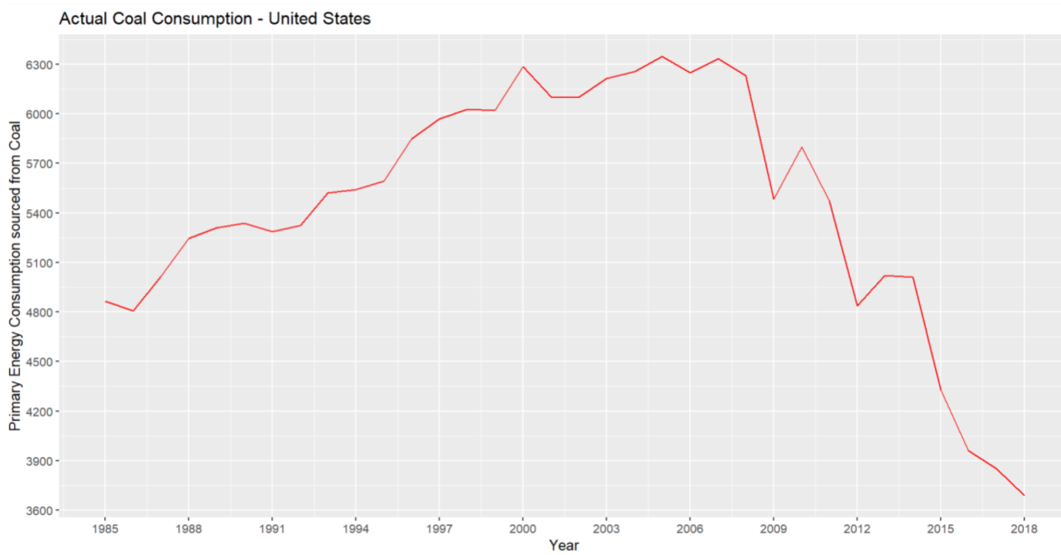

Figure 13: US – Actual Coal Consumption

The above graph depicts the actual coal consumption of United States from 1985 – 2018. As we can see from the plot the decrease in the coal consumption has begun in the year 2008 and decreased year on year which can be attributed to the climate goal the country has pledged.

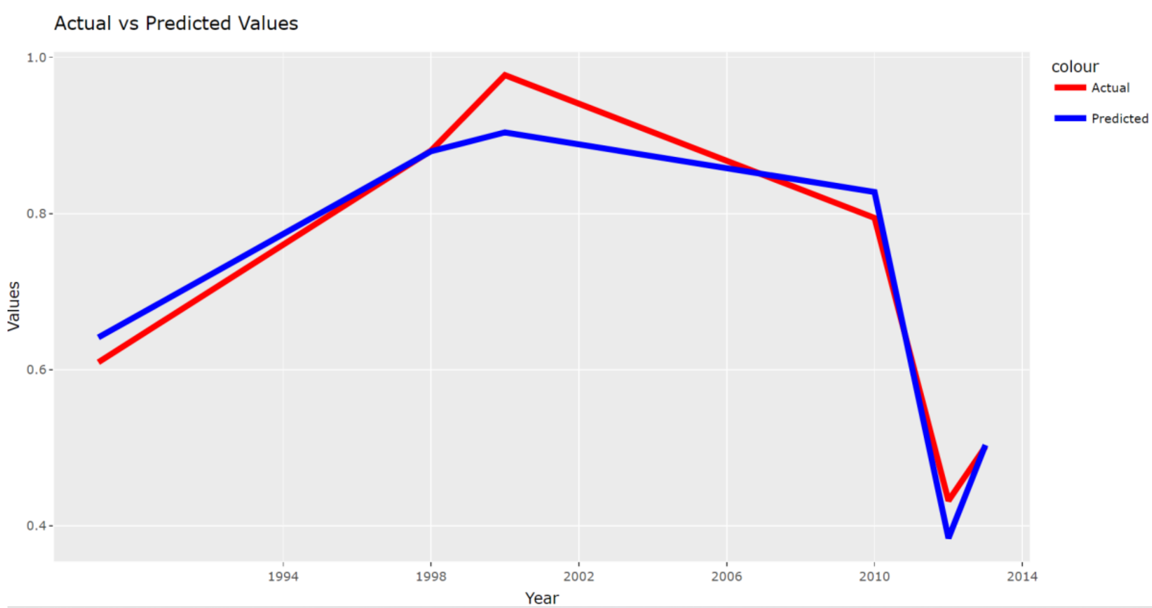

Figure 14: US - Actual vs Predicted

Actual vs Predicted values were depicted in the above plot.

United Kingdom:

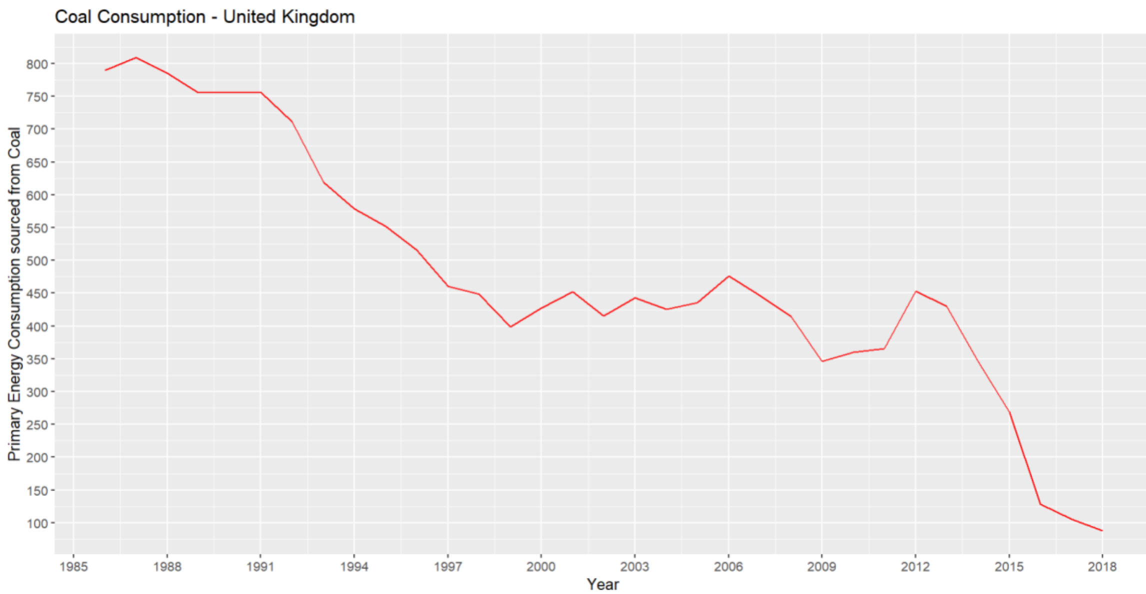

Figure 15: UK – Coal Consumption

The above graph depicts the actual coal consumption of United Kingdom from 1985 to 2018.

And it can be observed from the plot that the gradual decrease in the consumption of coal has started in 1990s itself.

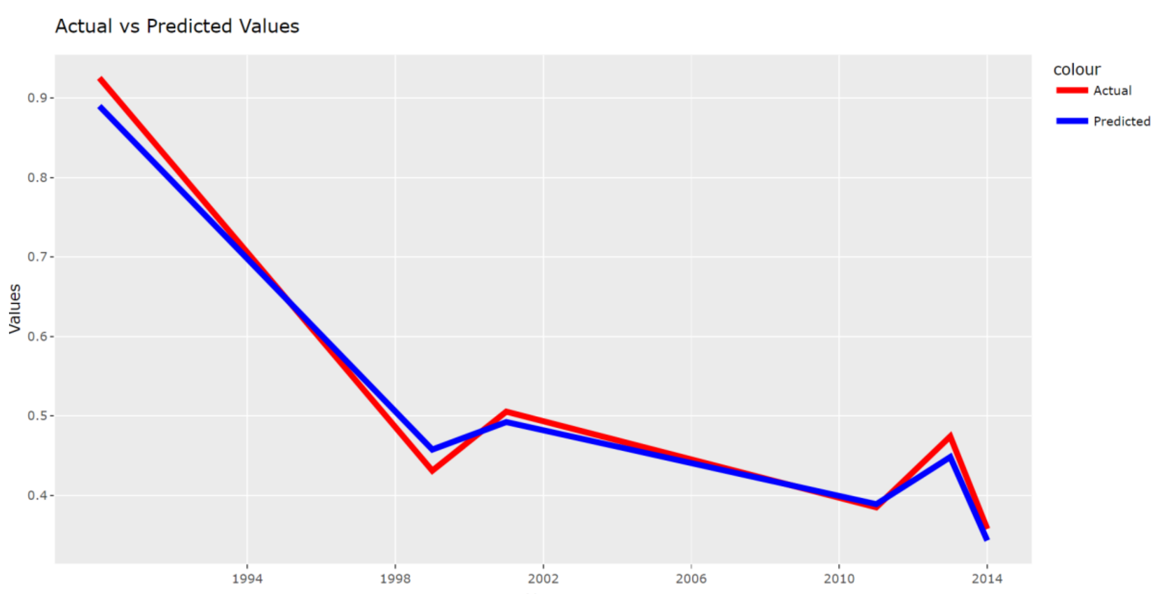

Figure 16: UK - Actual vs Predicted

Actual vs Predicted values were depicted in the above chart.

Russia:

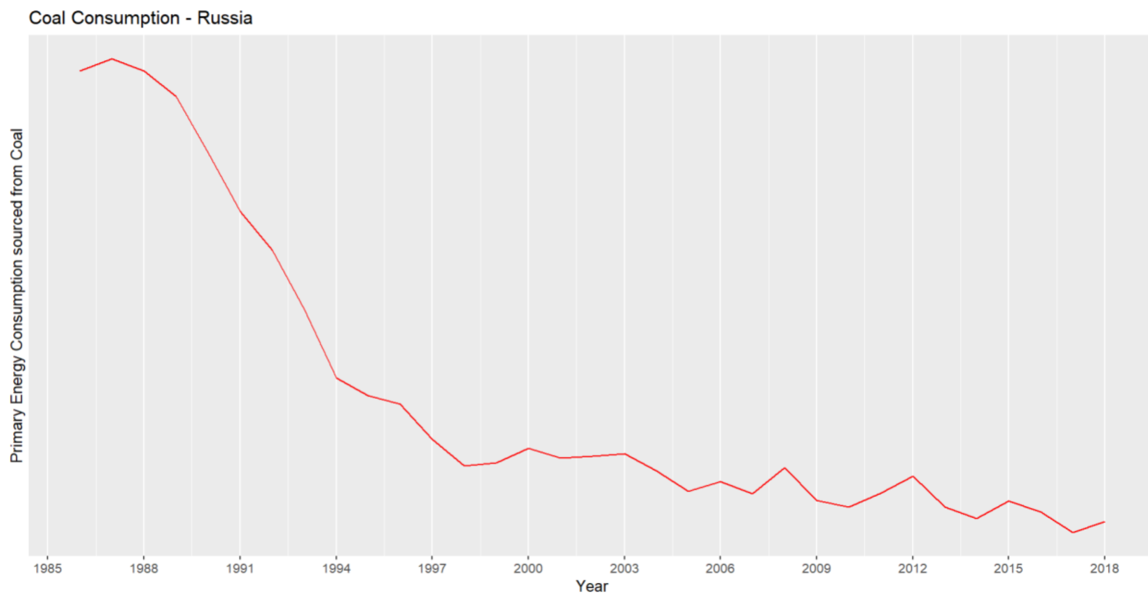

Figure 17: Russia – Coal Consumption

The above graph depicts the actual coal consumption for the years 1985 to 2018.As we can see from the plot, between 1985 to 1998 the consumption of coal has been gradually reduced year on year. And from 2000 to 2018 steady decrease and increase of the consumption has been deducted.

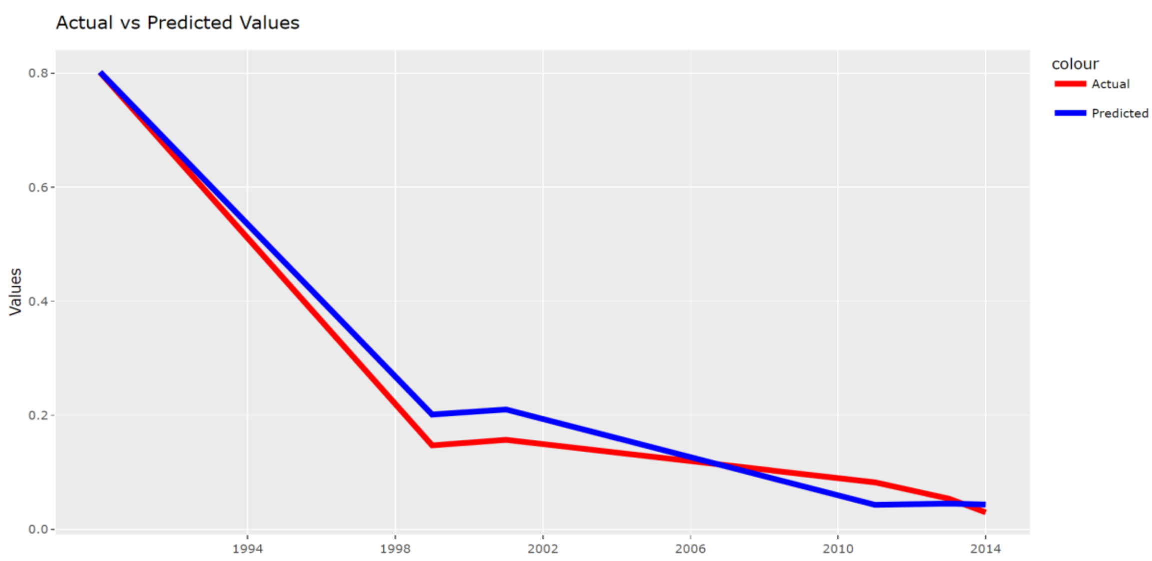

Figure 18: Russia - Actual vs Prediction

The above graph depicts the actual vs predicted values of Russia's Coal Consumption by the model.

China:

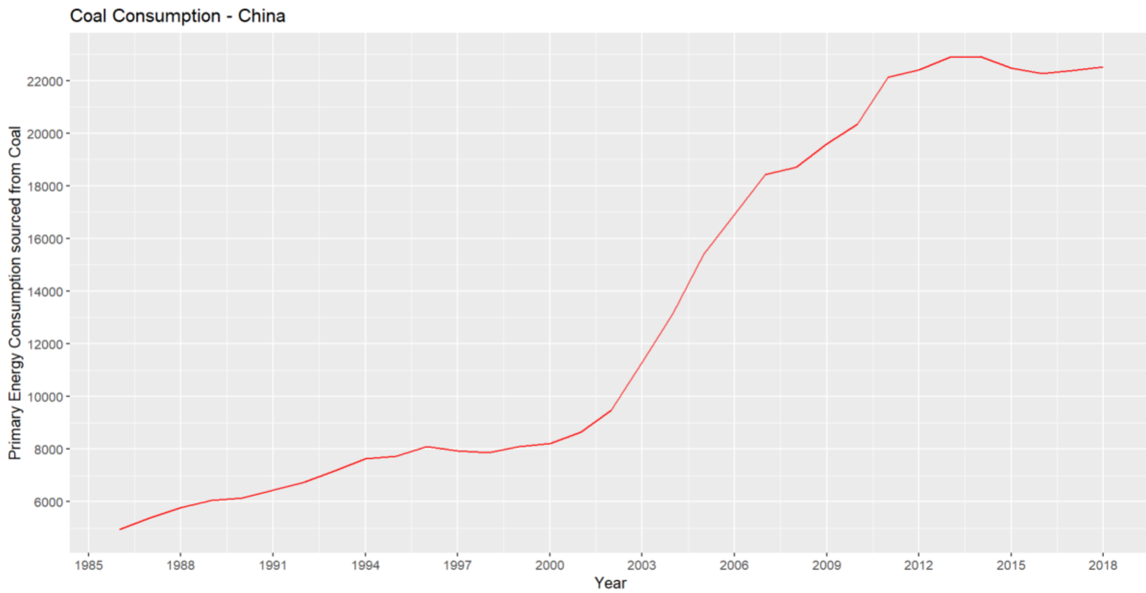

Figure 19: China – Coal Consumption

The above graph depicts the actual coal consumption of China . As we can see from the plot there has been a tremendous increase in the consumption of coal for energy from 1990 and went on increasing which can be attributed to the China's economical development. This graph shows how China's energy has been dependent on coal.

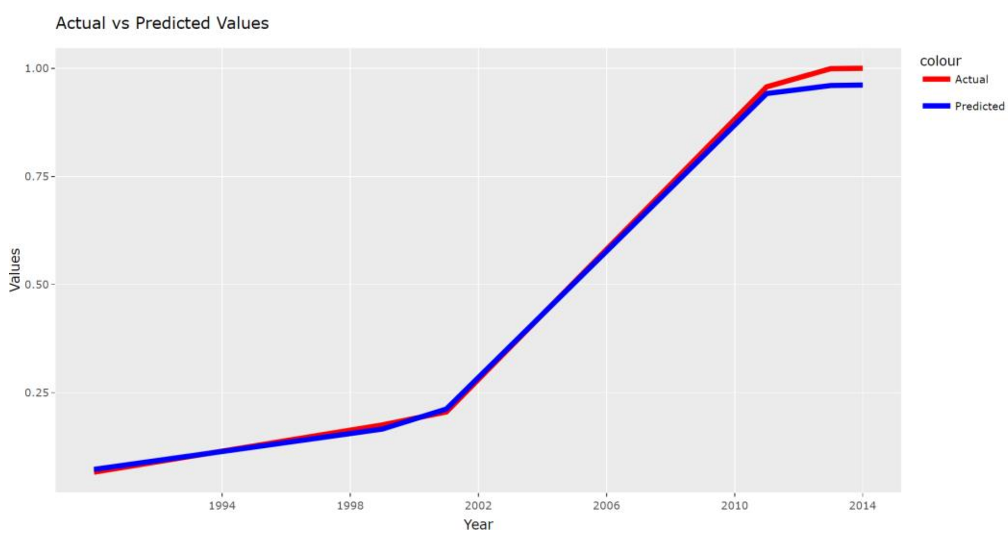

Figure 20: China – Actual vs Predicted

The above graph shows the comparison of actual values and the values predicted by the model.

India:

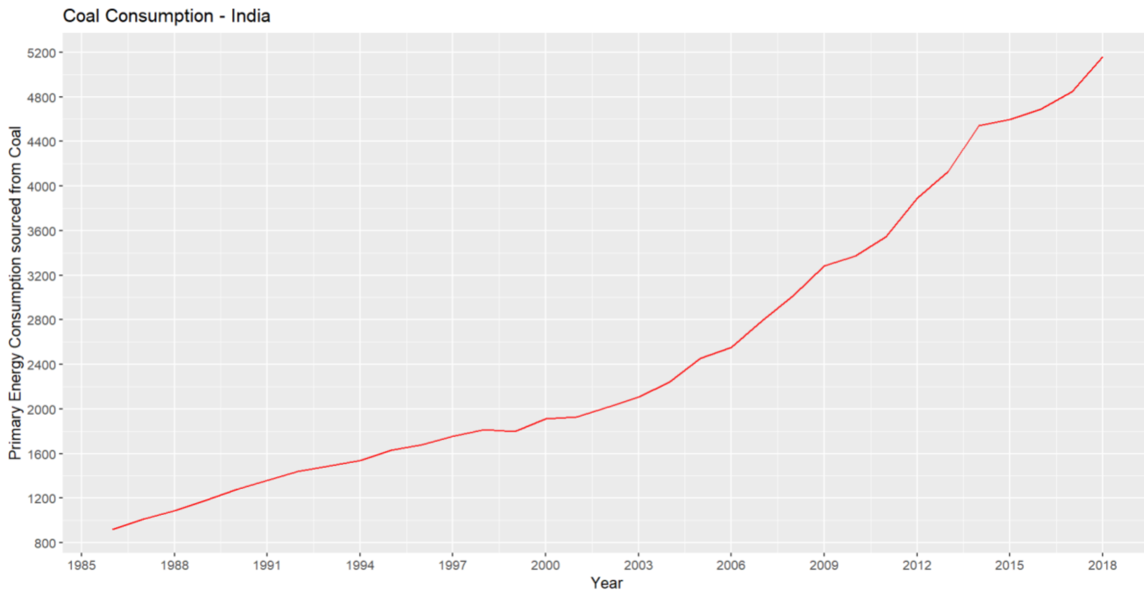

Figure 21: India – Coal Consumption

The above graph depicts the actual coal consumption of India for the years 1985 to 2018. We can clearly see from the plot that there has been a gradual increase in the consumption of coal from 1990 and further increase in consumption from 2003 which can be well attributed to the India's growth in economic terms and more industrial presence.

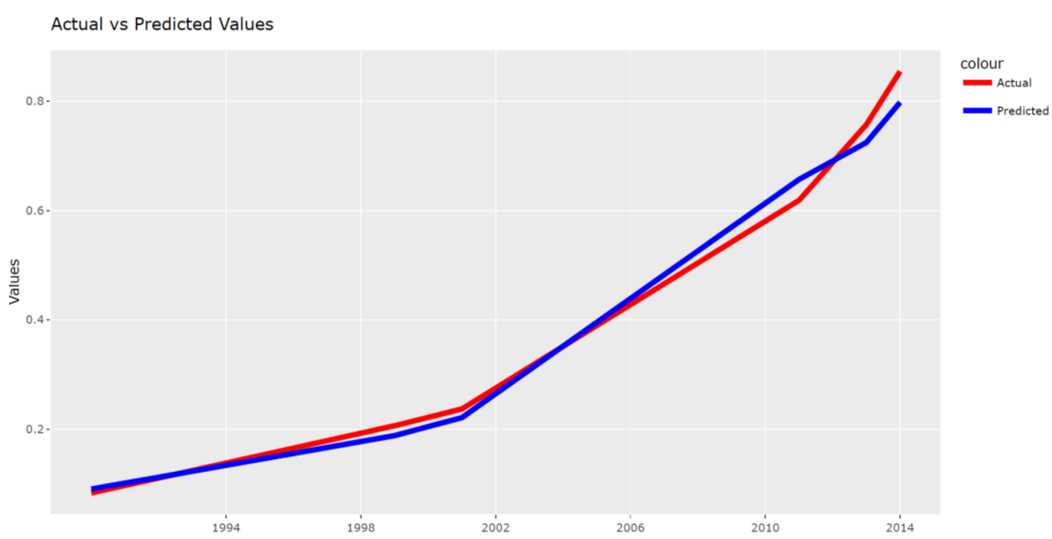

Figure 22: India – Actual vs Predicted

The above graph depicts the actual vs predicted values of coal usage given by the model. We can see from the China's and India's plots that the consumption of coal has been increasing in developing nations when compared to developed nations.

Global Oil Consumption: -

Using the derived model prediction of global oil consumption has been performed.

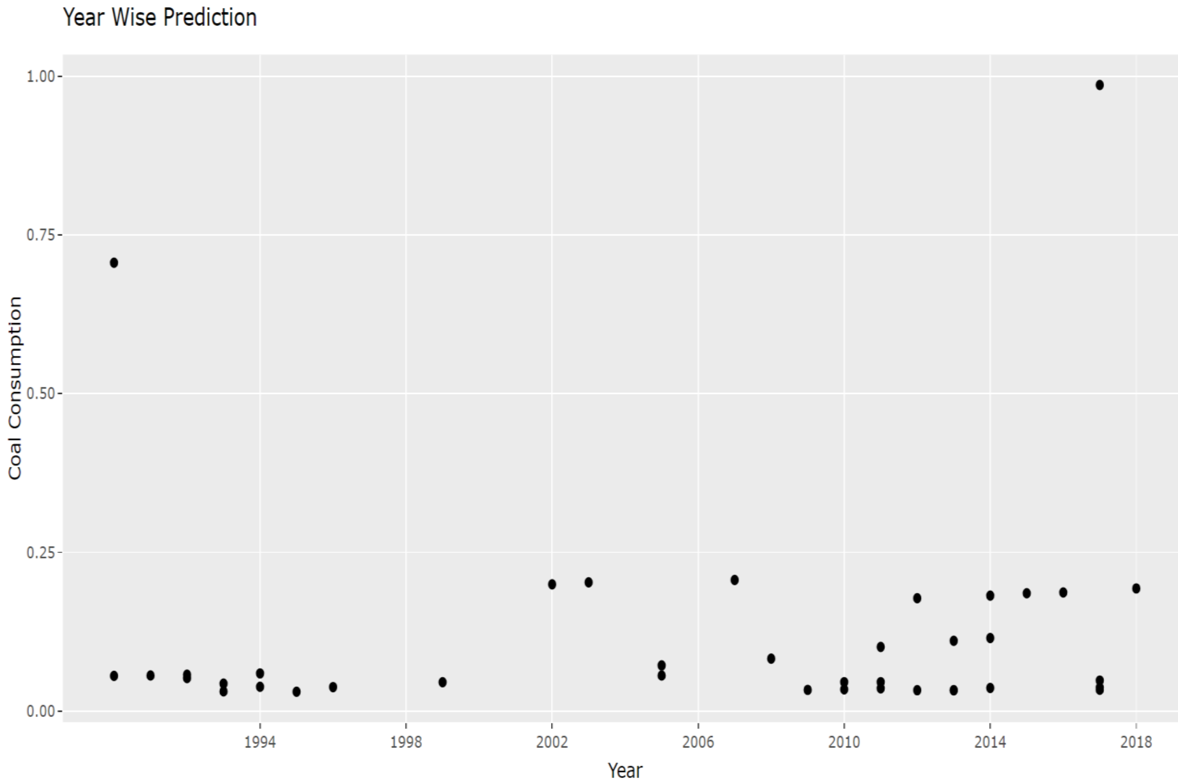

Figure 23: Global Oil Consumption

The above graph shows the values of global oil consumption predicted by the ANN model. Using the model individual country wise prediction has been performed.

United States:

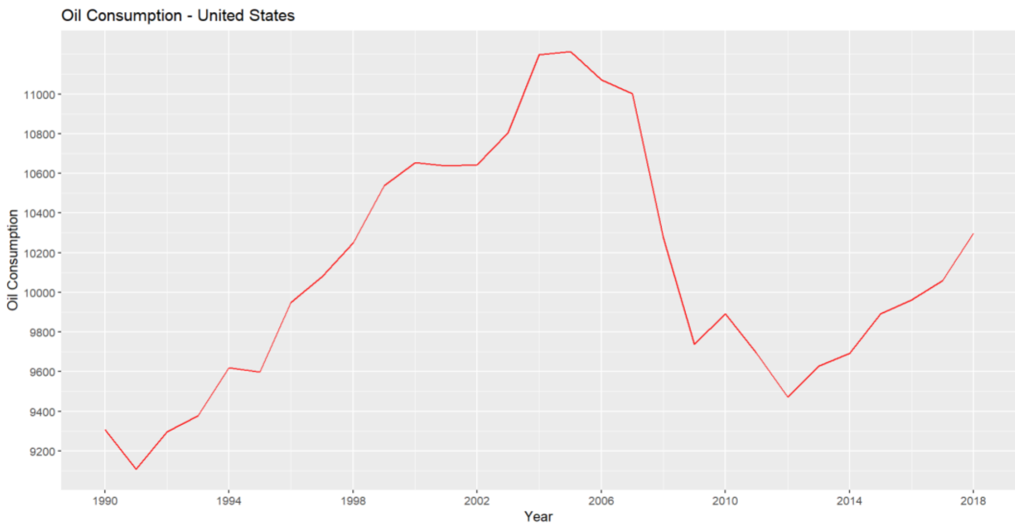

Figure 24: US – Oil Consumption

The above plot depicts the actual Oil Consumption of United States from 1990 to 2018. As we can see from the plot that there is a gradual increase in oil consumption from 1990 to 2000 and gradual decrease from 2005 to 2010 which can be attributed great economic recession.

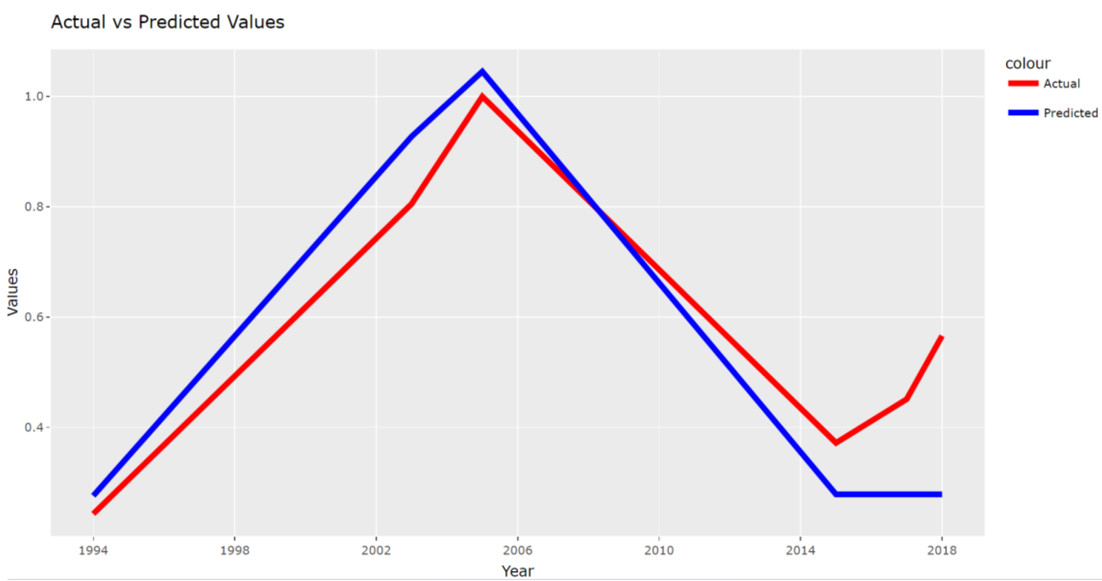

Figure 25: US – Actual vs Predicted

The above graph shows the comparison of actual and predicted values of model for oil consumption.

United Kingdom:

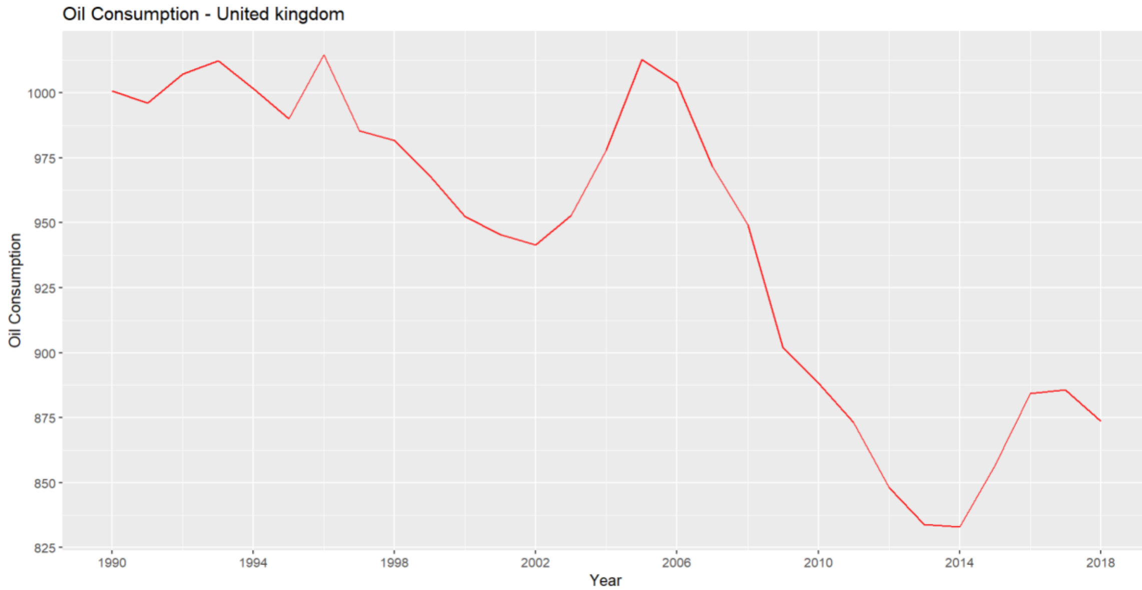

Figure 26: UK – Oil Consumption

The above graph shows the actual oil consumption of UK for the years 1990 to 2018. As we can see from the plot that there is a huge decrease in consumption of oil from the year 2006 to 2014.

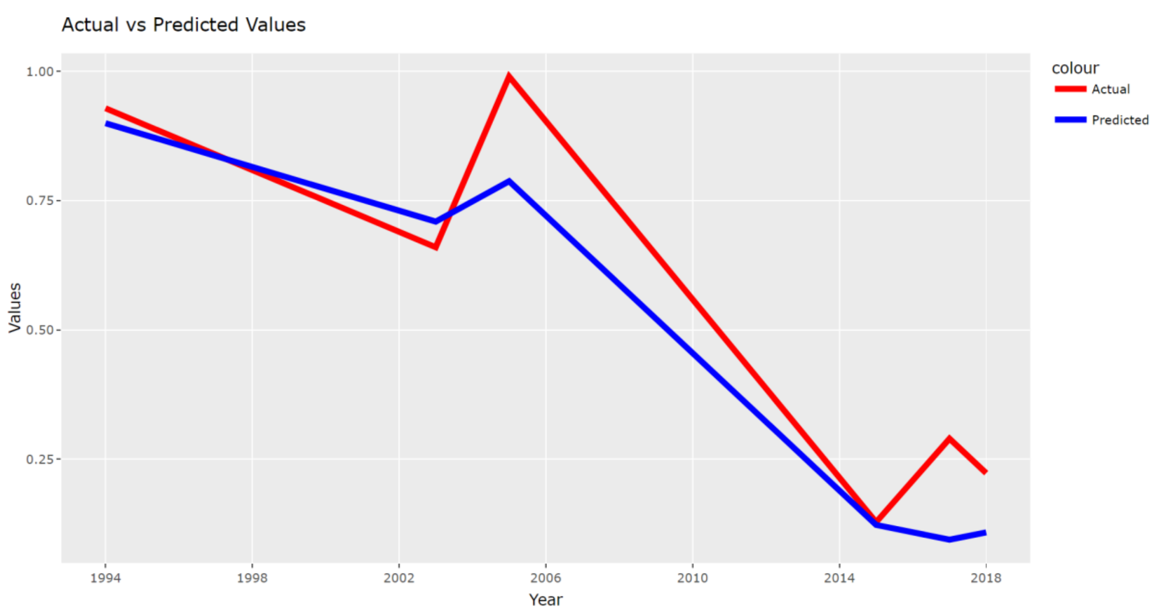

Figure 27: UK – Actual vs Prediction

Russia:

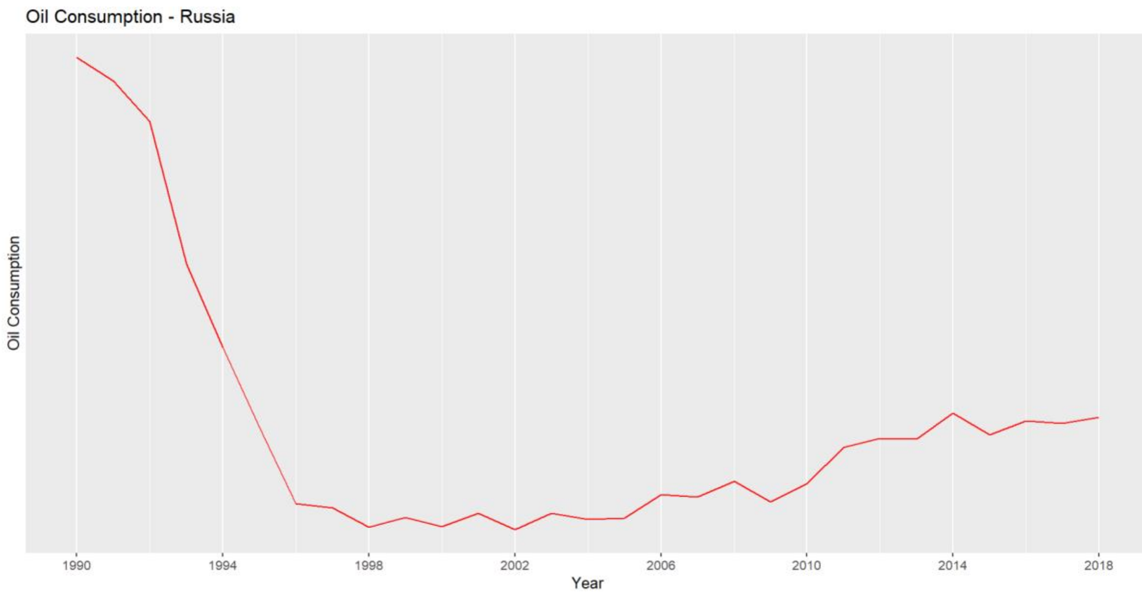

Figure 28: Russia – Oil Consumption

The above graph depicts the actual oil consumption of Russia for the years 1990 to 2018. As we can observe from the plot consumption decreased tremendously in the span of 4 years from 1990 to 1994 and remained constant for the period 1995 to 2007.

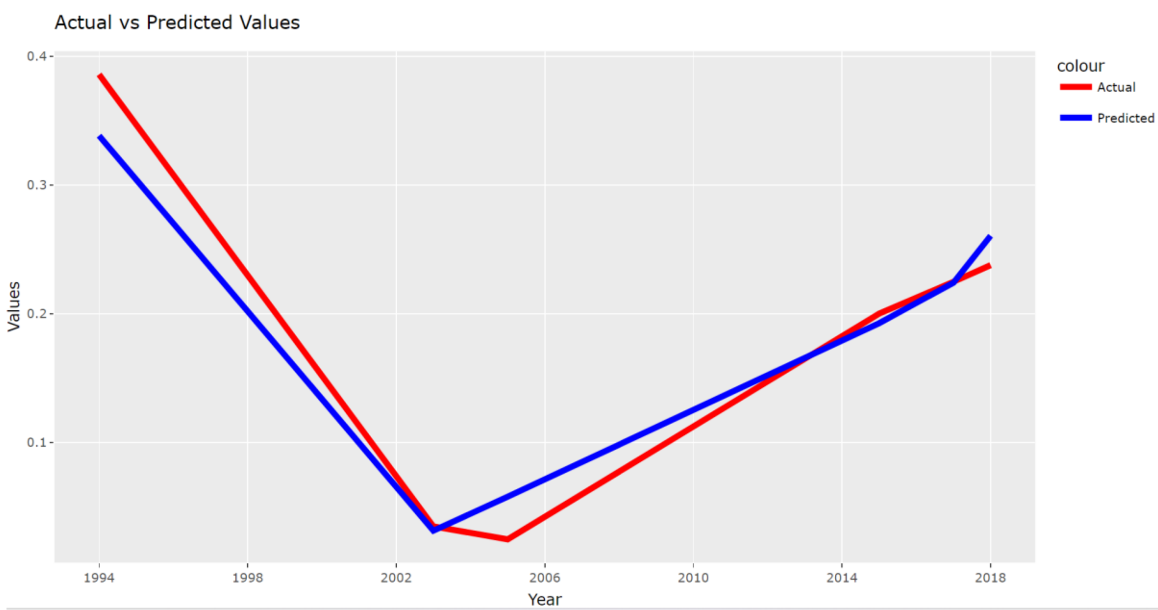

Figure 29: Russia – Actual vs Prediction

The above graph depicts the actual and predicted values comparison.

China: -

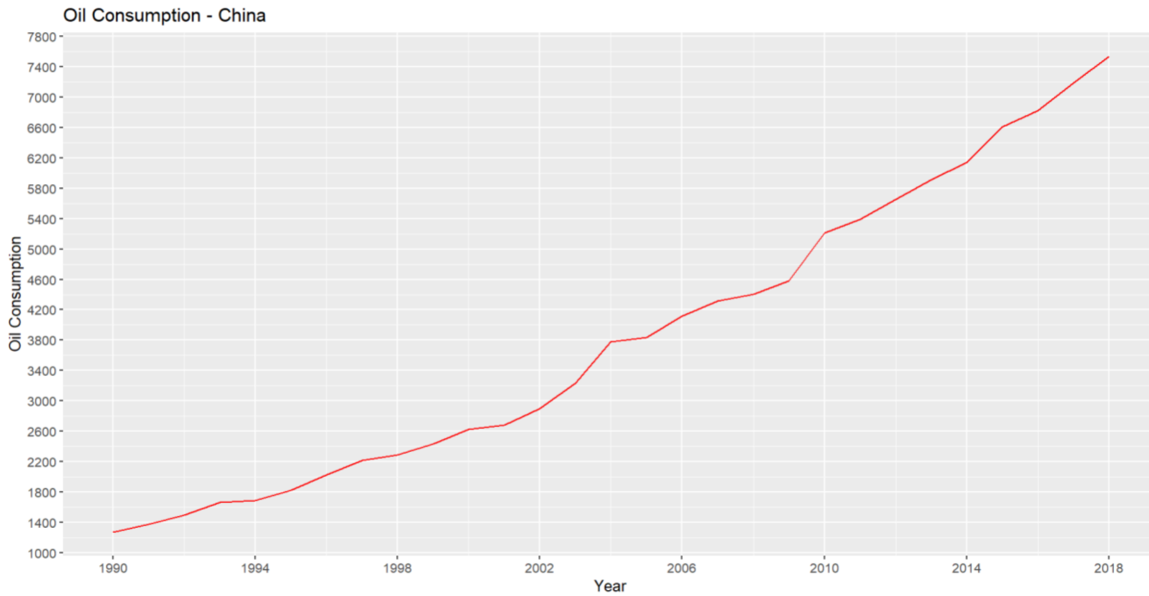

Figure 30: China – Oil Consumption

The above graph depicts the actual oil consumption of China for the years 1990 to 2018. As we can see from the graph there is a steady raise in usage of oil energy.

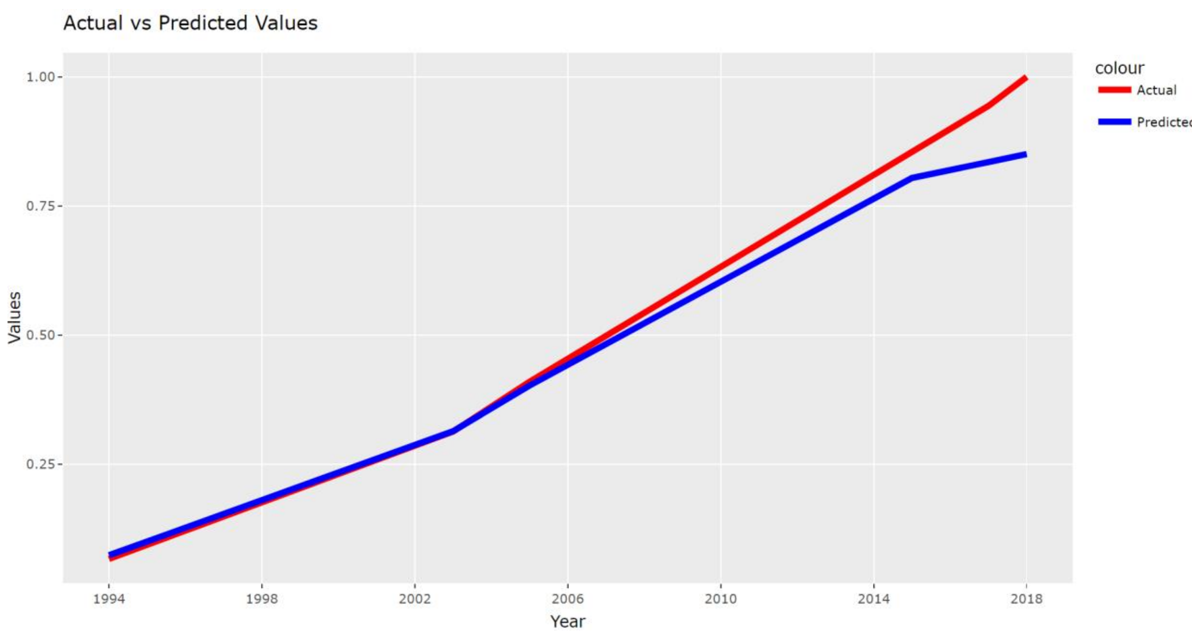

Figure 31: China – Actual vs Prediction

The above graph depicts the comparison of actual and predicted values of oil consumption for China.

India: -

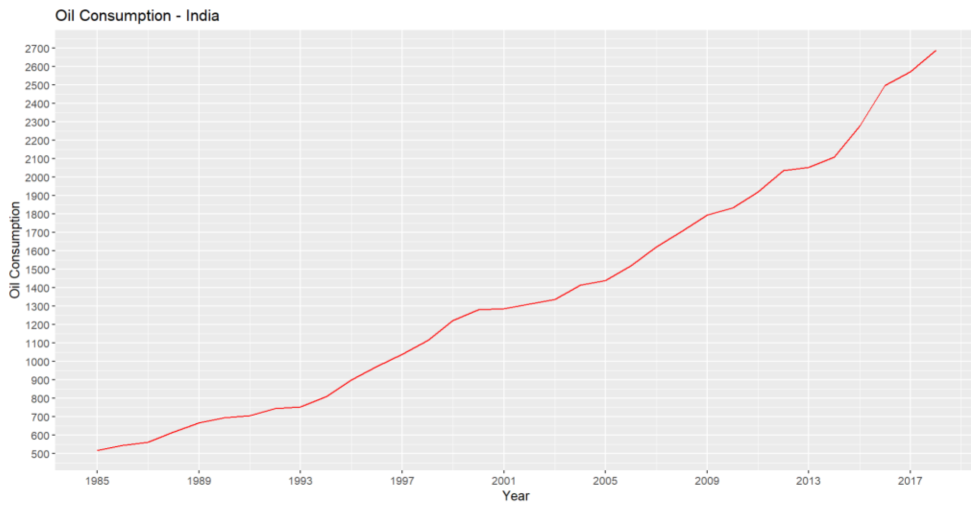

Figure 32: India – Oil Consumption

The above graph depicts the actual oil consumption of India for the years 1985 to 2018. As we can see from the graph there is a tremendous increase in the oil consumption from 1990 to 2000’s.

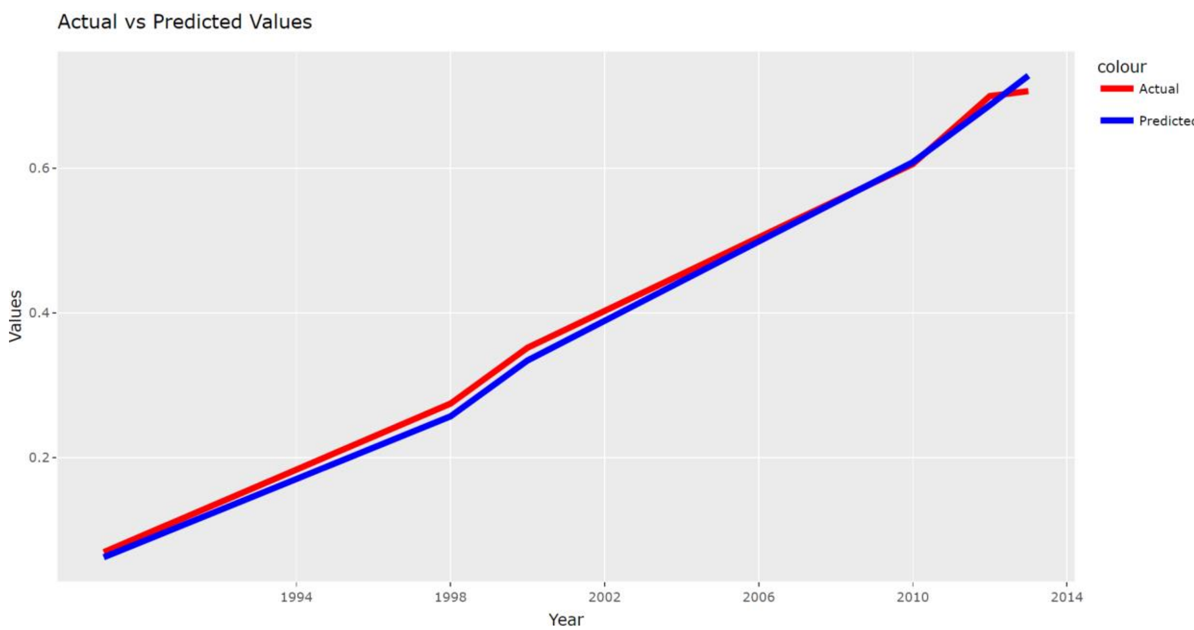

Figure 33: India – Actual vs Prediction

The above graph depicts the actual vs predicted graph.

Global Gas Consumption:

With the help of model global gas consumption has been predicted.

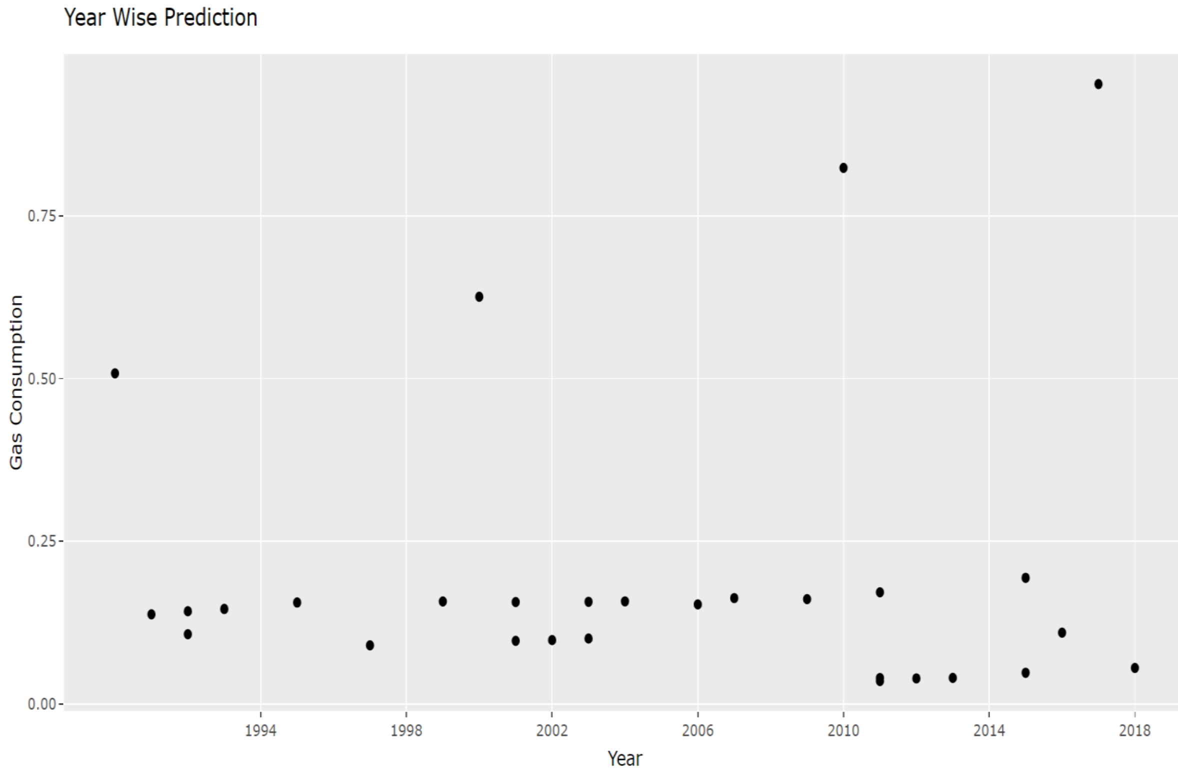

Figure 34: Global Gas Consumption

The above graph shows the year wise predicted values of global gas consumption. Using the model individual country wise consumption has been predicted.

United States:

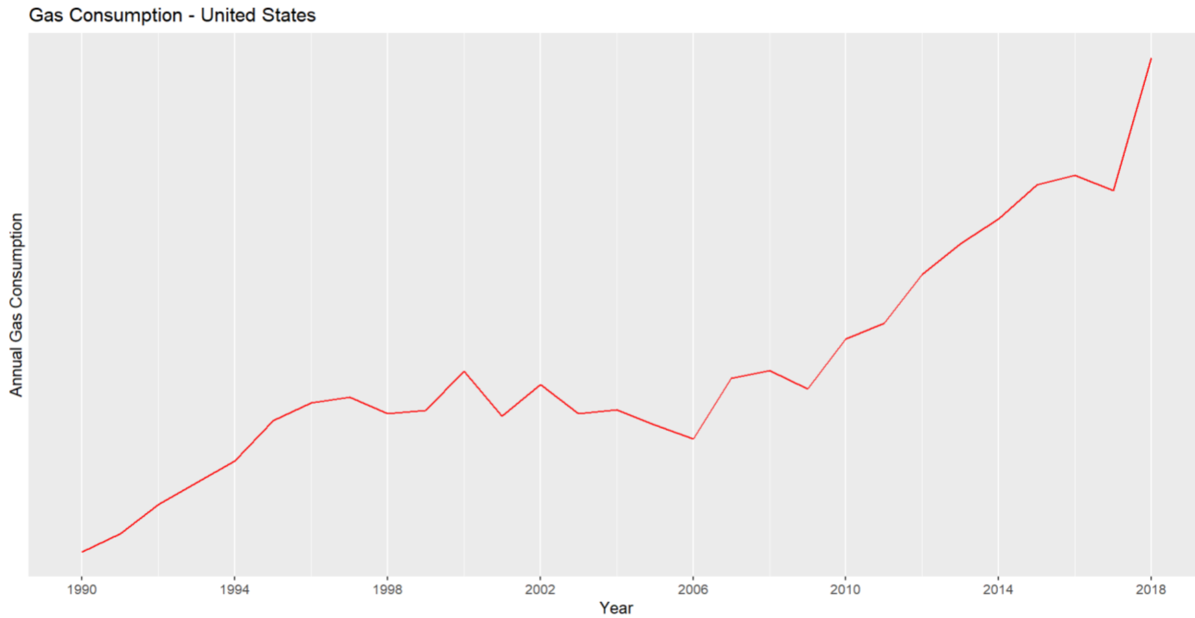

Figure 35: US – Gas Consumption

From 2010 to 2017 steady increase in consumption has been noticed from the plot.

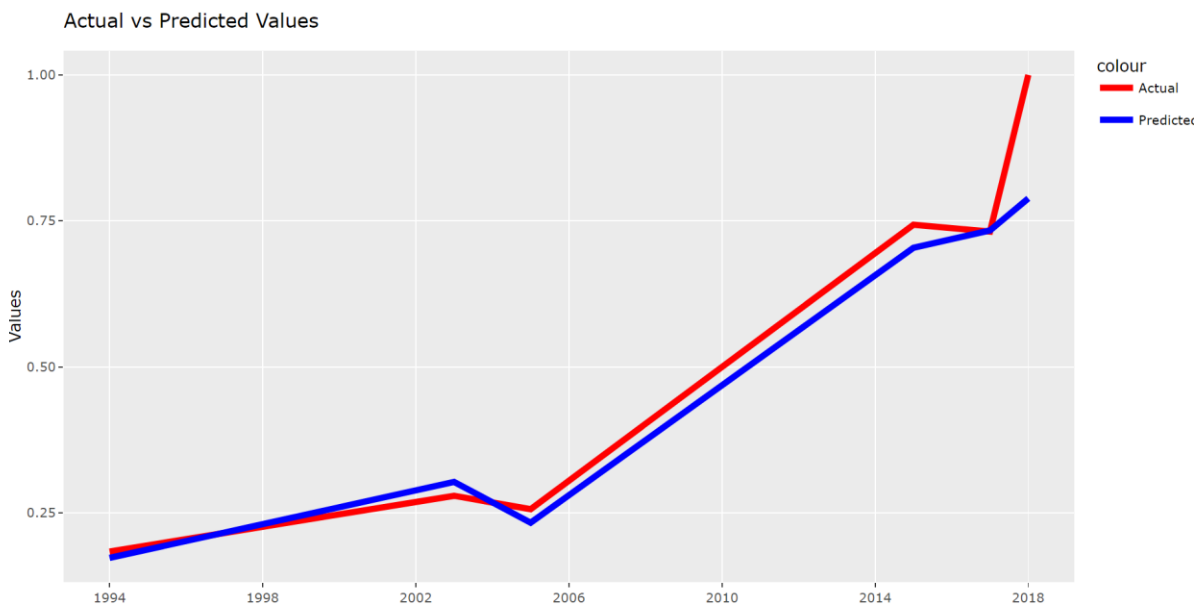

Figure 36: US – Actual vs Prediction

The above graph shows the comparison between the actual and predicted values by the model.

United Kingdom:

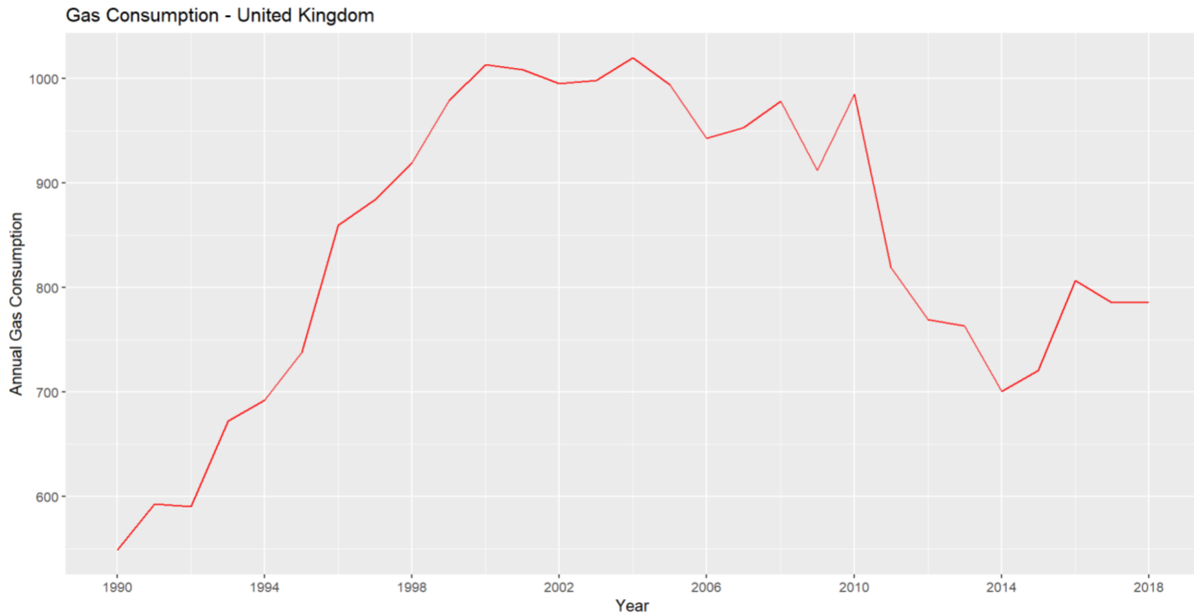

Figure 37: UK – Gas Consumption

The above graph shows the actual gas consumption of UK for the years 1990 to 2018. As we can see from the graph the consumption has risen from 1990 to 2002.

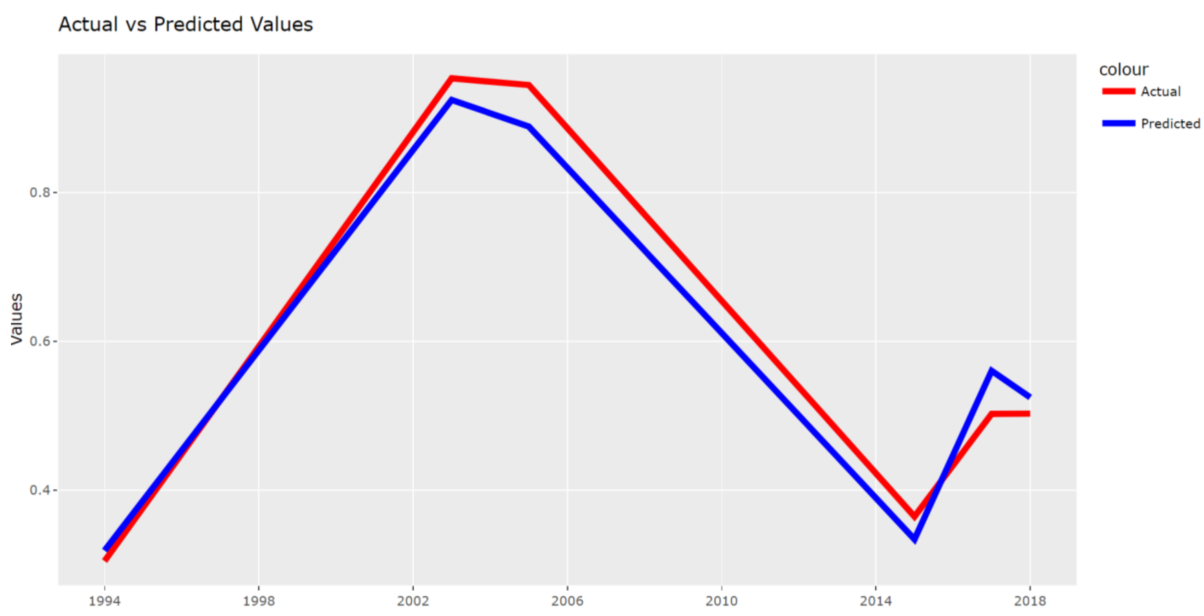

Figure 38: UK – Actual vs Prediction

The above plot illustrates the comparison between actual and predicted values.

Russia:

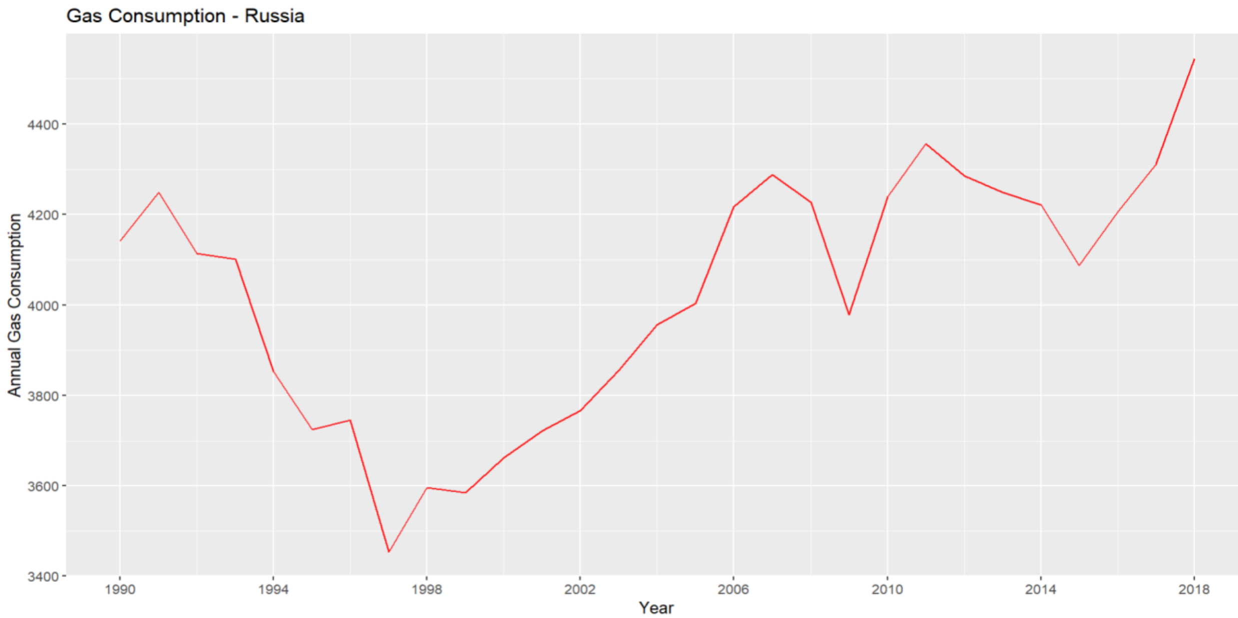

Figure 39: Russia – Gas Consumption

The above graph shows the actual consumption of gas for the years 1990 to 2018.

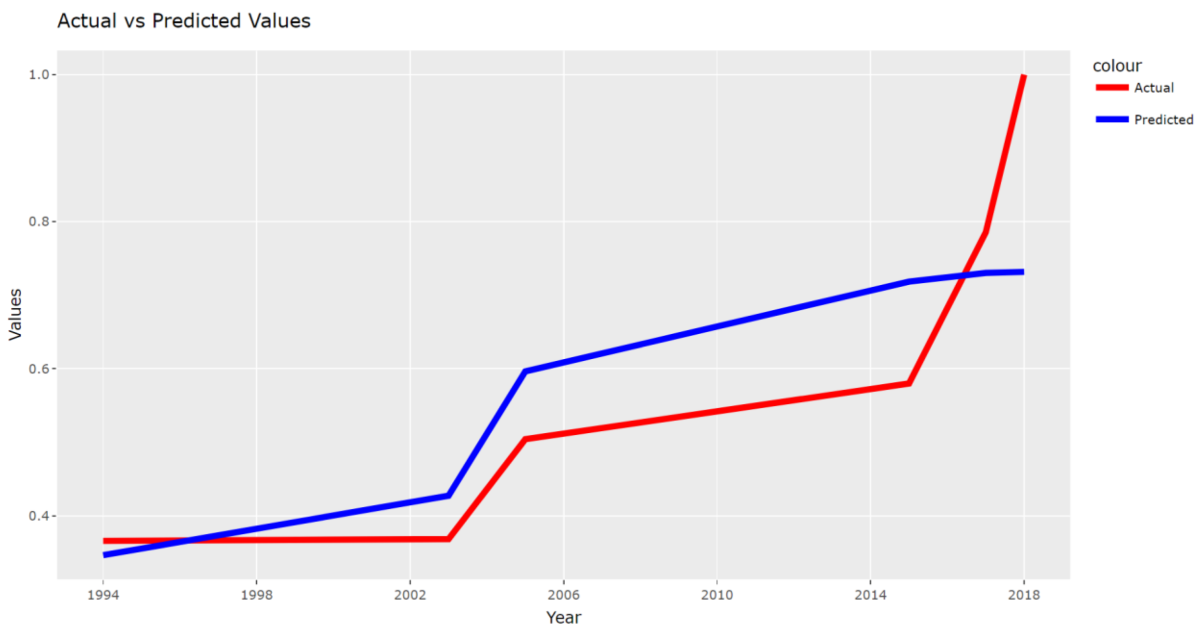

Figure 40: Russia – Actual vs Prediction

The above plot illustrates the comparison between actual and predicted values.

China:

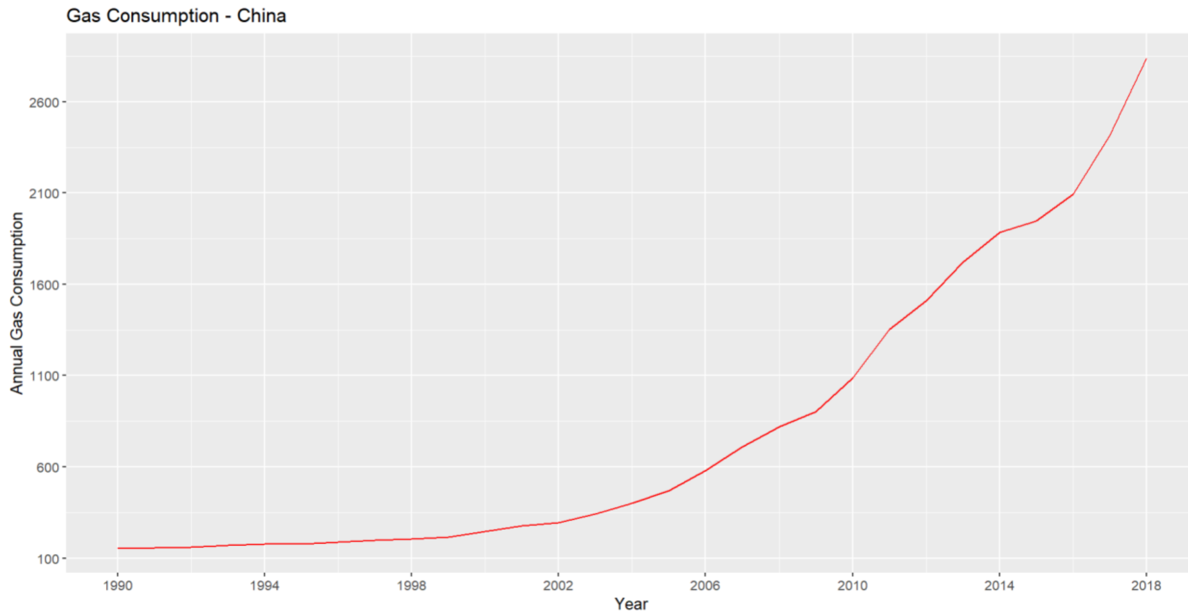

Figure 41: China – Gas Consumption

The above graph shows the actual consumption of gas in China for the years 1990 to 2018.

As we can see from the plot there is a constant increment in the consumption.

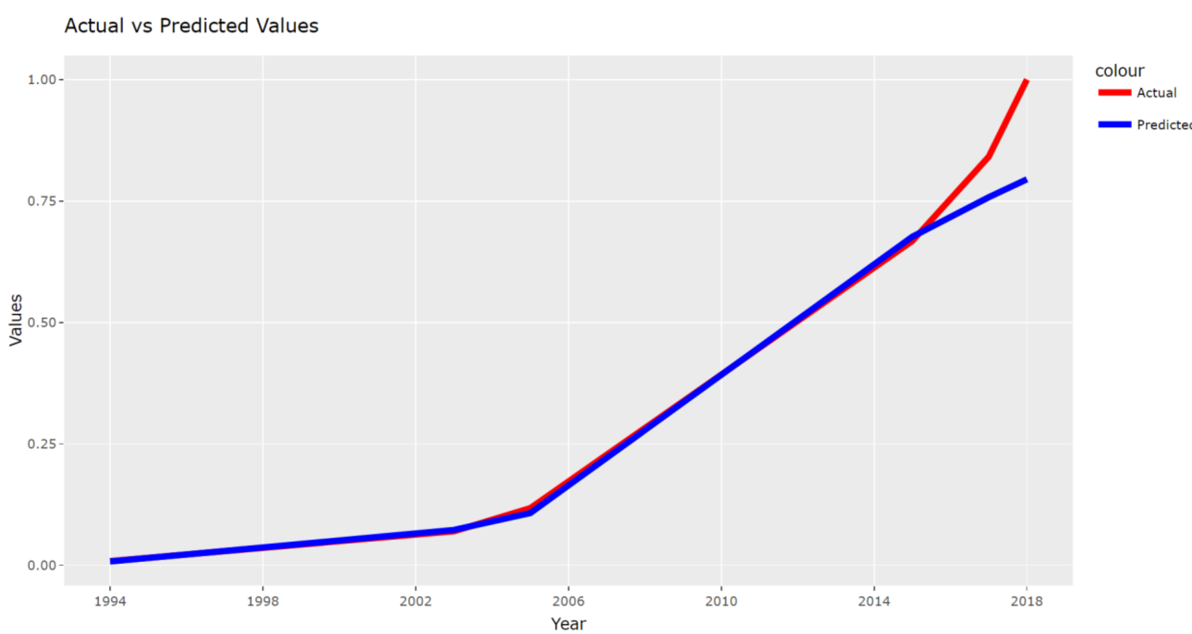

Figure 42: China – Actual vs Prediction

The actual and predicted values are compared in the graph above.

India:

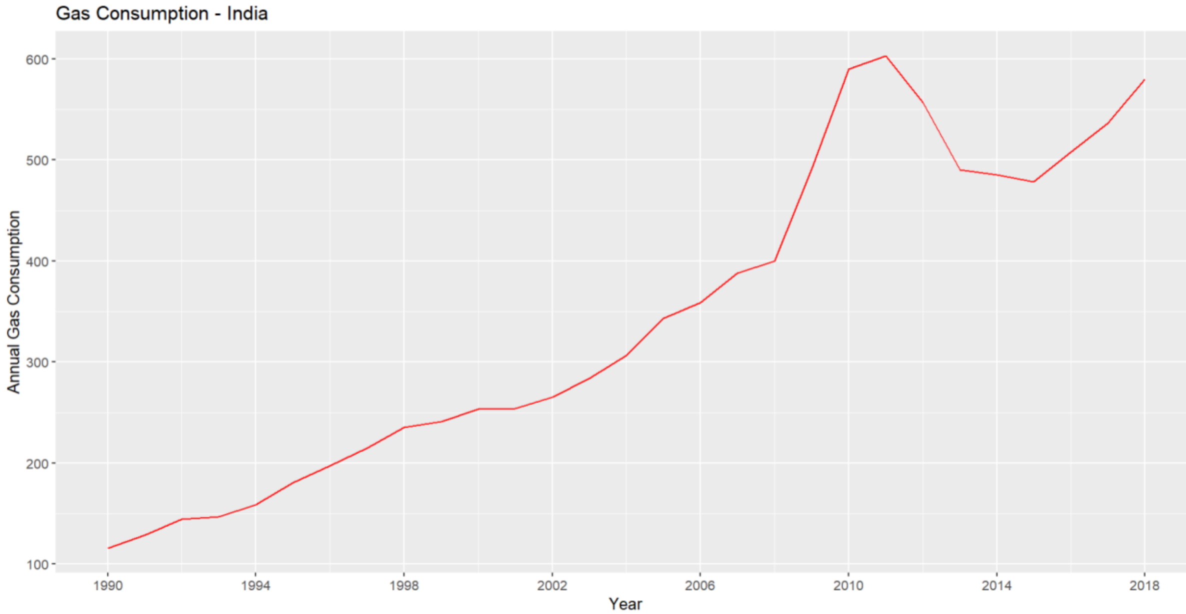

Figure 43: India – Gas Consumption

The above graph shows the actual gas consumption of India for the years 1990 to 2018.

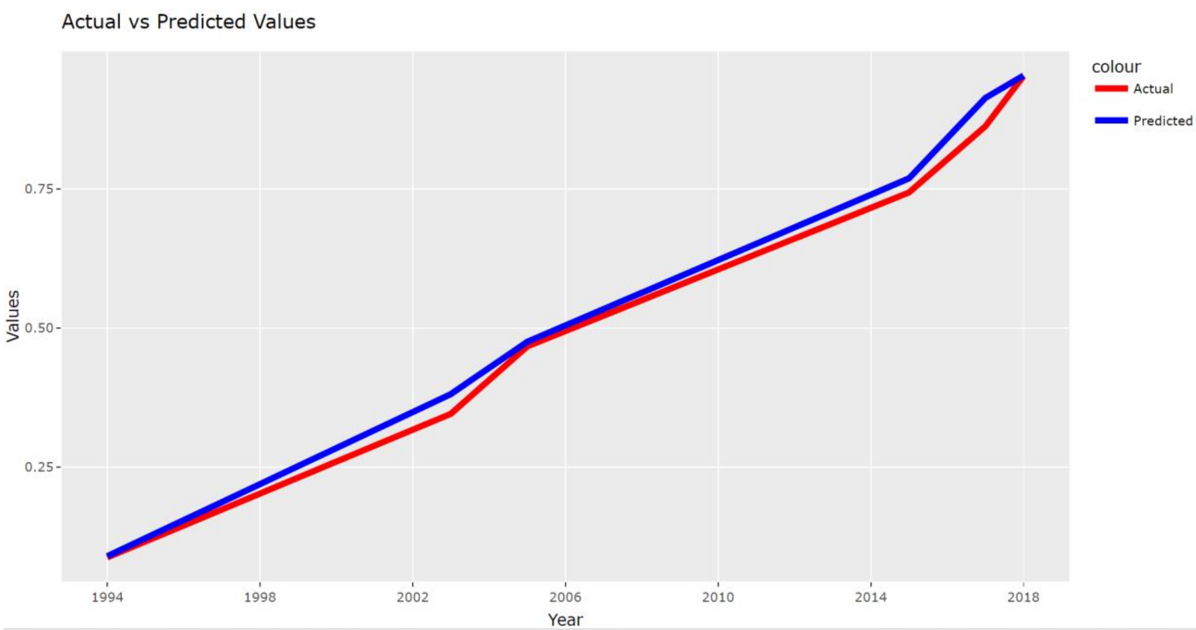

Figure 44: India – Actual vs Prediction

The above graph depicts the comparison of actual and predicted values.

Global Nuclear Consumption:

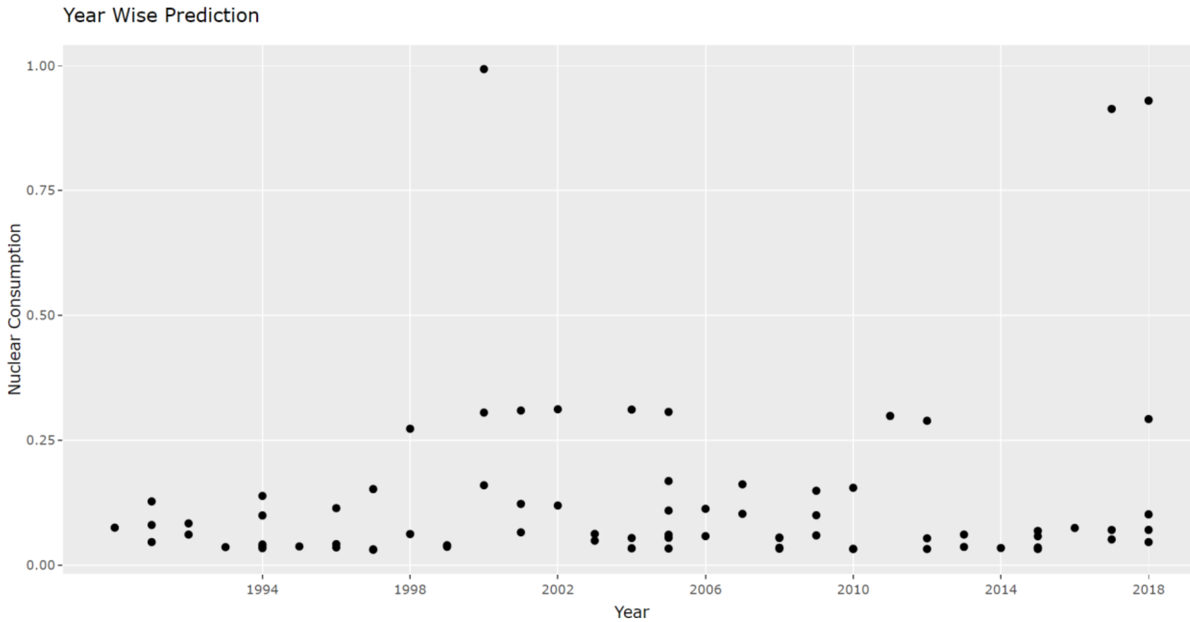

Figure 45: Global Nuclear Consumption

Using the ANN model Global Nuclear Consumption has been predicted and has been plotted as seen in the above graph.

Renewable Energy Sources:

Using the model, consumption of energy sourced from renewable's has been predicted. Global Solar Energy Consumption: -

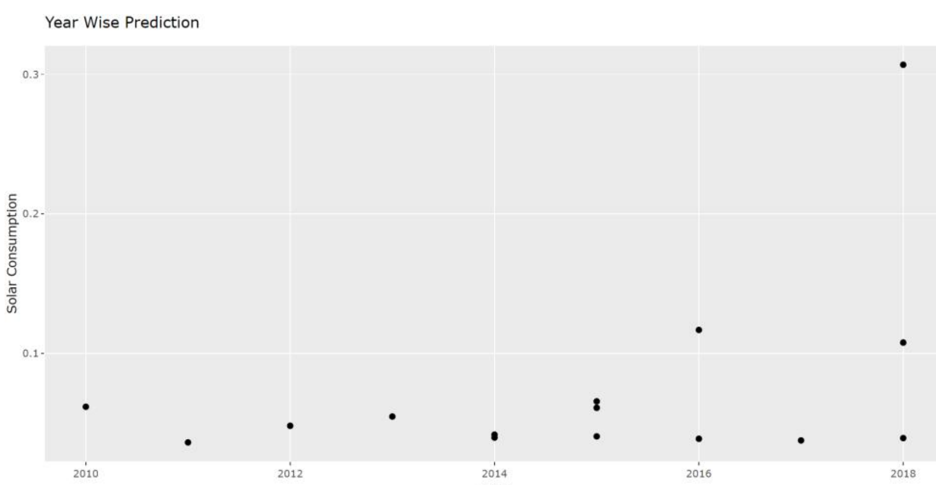

Figure 46: Global Solar Consumption

The above graph depicts the year-wise predicted values by the model

Global Wind Energy Consumption:

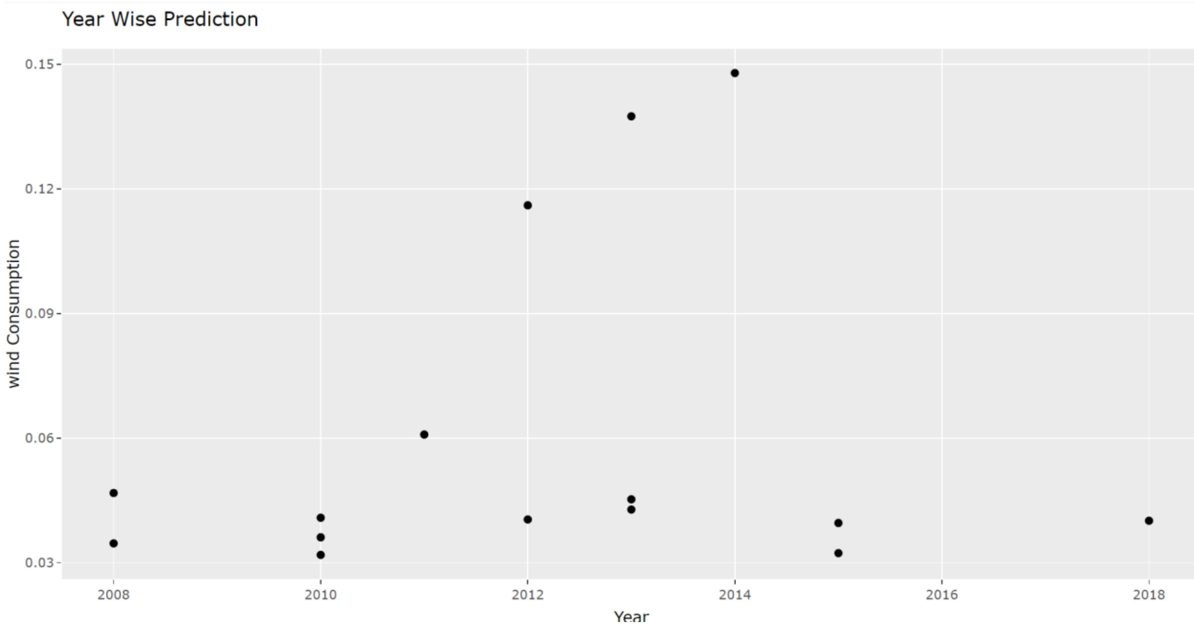

Figure 47: Global Wind Consumption

The ANN model's anticipated values for wind energy consumption are displayed in the above plot

Global Hydro Energy Consumption:

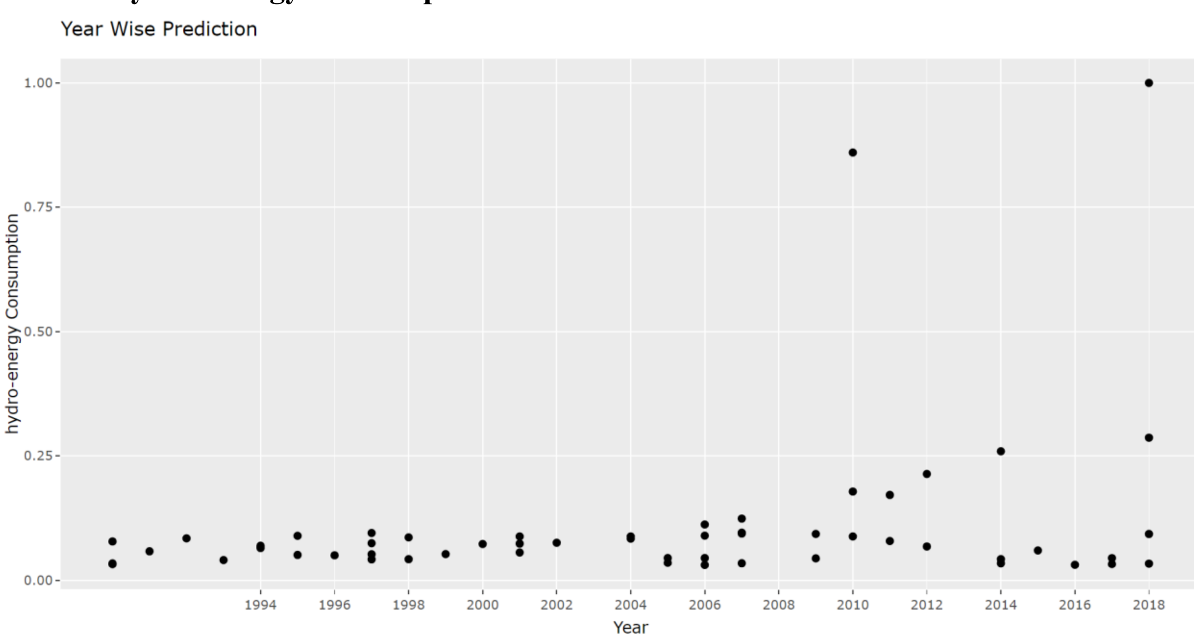

Figure 48: Global Hydro Consumption

The above shows the predicted values by the model for global hydro energy consumption.

10. CODE SNIPPETS:

1.) Reading the Data

```
Data <- read.csv(file.choose(),header = TRUE)
View(Data)
```

2.) Filtering out the data based on the dependent variable

```
 library(dplyr)
dde1= Data %>% select(country,year,population,gdp,gas_electricity,
                      gas_production,gas_share_elec,gas_share_energy,
                  gas_consumption) %>% filter(country == "United States" &
                                            year >= 1990 & year <=2018)
```

3.) Preparing Train and Test Data

```
set.seed(1234)
ind <- sample(2, nrow(dde_norm), replace = T, prob = c(0.7, 0.3))
dde_train <- dde_norm[ind == 1,]
dde_test <- dde_norm[ind == 2,]
```

4.) Model

```
library(neuralnet)
dde_model <- neuralnet(gas_consumption~., data = dde_train[,-1])
dde_model
```

5.) Model results and Accuracy

```
model_results <- compute(dde_model,dde_test[,-8])
model_results

(predicted_cons <- model_results$net.result)
cor(dde_test$gas_consumption, predicted_cons)


VALIDATION=table(dde_test[,8],predicted_cons)
(ACCURACY=sum(diag(VALIDATION))/sum(VALIDATION)*100)
```

11. CONCLUSION

Estimating energy consumption is an important challenge for all countries, particularly with the growing dependency on non-renewable energy sources. Based on the analysis of the different machine learning algorithms used for forecasting global energy consumption, it can be concluded that all the models performed reasonably well in predicting future energy consumption trends.

Energy usage has been predicted using machine learning model ANN with accuracy varying from 70% to 90%. Moreover, the analysis also suggests that the accuracy of the models can be improved by incorporating additional features related to energy consumption, such as GDP, population, and production. It is also important to note that the choice of model and input data should depend on the specific needs and context of the forecasting task.

To forecast energy consumption, the models take into account a variety of input variables, including geographic, socioeconomic, and environmental aspects. Governments and businesses may find it easier to predict future demands and make plans accordingly for a sustainable future with the help of the many models. It is absolutely important to do further research to create more precise models that can forecast long-term energy usage.

Overall, the defined ANN model offer a promising approach for forecasting global energy consumption, and further research in this area can lead to more accurate and reliable predictions, which can aid in policymaking, resource planning, and energy management.

12. REFERENCES

[1] Laib, O., Khadir, M. T., & Mihaylova, L. (2019). Toward efficient energy systems based on natural gas consumption prediction with LSTM Recurrent Neural Networks. *Energy*, *177*, 530-542.

[2] Abd El-Aziz, R. M. (2022). Renewable power source energy consumption by hybrid machine learning model. *Alexandria Engineering Journal*, *61*(12), 9447-9455.

[3] Khan, I., Han, L., Khan, H., & Kim Oanh, L. T. (2021). Analyzing renewable and nonrenewable energy sources for environmental quality: dynamic investigation in developing countries. *Mathematical Problems in Engineering*, *2021*, 1-12

[4] Altunkaynak, A., Özger, M., & Çakmakci, M. (2005). Water consumption prediction of Istanbul city by using fuzzy logic approach. *Water resources management*, *19*, 641-654.

[5] Kavaklioglu, K. (2011). Modeling and prediction of Turkey's electricity consumption using Support Vector Regression. *Applied Energy*, *88*(1), 368-375.

[6] Szoplik, J. (2015). Forecasting of natural gas consumption with artificial neural networks. *Energy*, *85*, 208-220.

[7] Azadeh, A., Babazadeh, R., & Asadzadeh, S. M. (2013). Optimum estimation and forecasting of renewable energy consumption by artificial neural networks. *Renewable and Sustainable Energy Reviews*, *27*, 605-612.

[8] Ekonomou, L. (2010). Greek long-term energy consumption prediction using artificial neural networks. *Energy*, *35*(2), 512-517.

[9] Shin, S. Y., & Woo, H. G. (2022). Energy Consumption Forecasting in Korea Using Machine Learning Algorithms. *Energies*, *15*(13), 4880.

[10] Fan, L., Fu, C., Guo, Y., Wang, Y., Wu, M., & Sun, L. (2022, July). Prediction and Analysis of Power User Energy Consumption Based on Demand Side Management. In *2022 4th International Conference on Power and Energy Technology (ICPET)* (pp. 932-937). IEEE.

[11] Mahjoub, S., Chrifi-Alaoui, L., Marhic, B., Delahoche, L., Masson, J. B., & Derbel, N. (2022, May). Prediction of energy consumption based on LSTM Artificial Neural

Network. In *2022 19th International Multi-Conference on Systems, Signals & Devices (SSD)* (pp. 521-526). IEEE.

[12] Liu, Y., & Li, J. (2022). Annual Electricity and Energy Consumption Forecasting for the UK Based on Back Propagation Neural Network, Multiple Linear Regression, and Least Square Support Vector Machine. *Processes*, *11*(1), 44.

[13] Kaytez, F., Taplamacioglu, M. C., Cam, E., & Hardalac, F. (2015). Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines. *International Journal of Electrical Power & Energy Systems*, *67*, 431-438.

[14] Ediger, V. Ş., & Akar, S. (2007). ARIMA forecasting of primary energy demand by fuel in Turkey. *Energy policy*, *35*(3), 1701-1708.

[15] Bianco, V., Manca, O., & Nardini, S. (2009). Electricity consumption forecasting in Italy using linear regression models. *Energy*, *34*(9), 1413-1421.

[16] Kavaklioglu, K., Ceylan, H., Ozturk, H. K., & Canyurt, O. E. (2009). Modeling and prediction of Turkey's electricity consumption using artificial neural networks. *Energy Conversion and Management*, *50*(11), 2719-2727.

[17] Benalcazar, P., Krawczyk, M., & Kamiński, J. (2017). Forecasting global coal consumption: An artificial neural network approach. *Gospodarka Surowcami Mineralnymi*, *33*..

[18]Mabel, M. C., & Fernandez, E. (2008). Analysis of wind power generation and prediction using ANN: A case study. *Renewable energy*, *33*(5), 986-992.

[19] Kim, T. Y., & Cho, S. B. (2019). Predicting residential energy consumption using CNN-LSTM neural networks. *Energy*, *182*, 72-81.

[20] Aydinalp-Koksal, M., & Ugursal, V. I. (2008). Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector. *Applied energy*, *85*(4), 271-296.

[21] Satan, A., Khamzina, A., Toktarbayev, D., Sotsial, Z., Bapiyev, I., & Zhakiyev, N. (2022, April). Comparative LSTM and SVM Machine Learning Approaches for Energy Consumption Prediction: Case Study in Akmola. In *2022 International Conference on Smart Information Systems and Technologies (SIST)* (pp. 1-7). IEEE.

[22] Wang, J. Q., Du, Y., & Wang, J. (2020). LSTM based long-term energy consumption prediction with periodicity. *Energy*, *197*, 117197

[23] Prieto, A., Prieto, B., Ortigosa, E. M., Ros, E., Pelayo, F., Ortega, J., & Rojas, I. (2016). Neural networks: An overview of early research, current frameworks and new challenges. *Neurocomputing*, *214*, 242-268.

[24] Raza, M. Q., & Khosravi, A. (2015). A review on artificial intelligence-based load demand forecasting techniques for smart grid and buildings. *Renewable and Sustainable Energy Reviews*, *50*, 1352-1372.

[25] Ruiz Moreno, Tobias. (2019). Crop yield prediction with ensemble algorithms and Artificial Neural Networks (ANN)