Predicting Energy Consumption using Machine Learning and Deep Learning

MSc Research Project
Data Analytics

Sana Inamdar
Student ID: x22217061

School of Computing
National College of Ireland

Supervisor: Dr. Catherine Mulwa

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Sana Inamdar
Student ID:	x22217061
Programme:	Data Analytics
Year:	2018
Module:	MSc Research Project
Supervisor:	Dr. Catherine Mulwa
Submission Due Date:	12th August 2024
Project Title:	Predicting Energy Consumption using Machine Learning and Deep Learning
Word Count:	5419
Page Count:	20

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Sana Salim Inamdar
Date:	12th August 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Predicting Energy Consumption using Machine Learning and Deep Learning

Sana Inamdar
x22217061

Abstract

Forecasting the increasing demand of electricity has become important these days due to the increase in demand of Energy consumption.The increase in population is also the main reason for the increase in the demand of energy consumption. If this Exponential increasing demand for electricity is not fulfilled it can result in Blackouts and to prevent the blackout forecasting Energy Consumption should be done well in advance. In our study ,PJM electric grid dataset has been used.This dataset contain information about different companies based in the United states which has been spread over 13 different states. Based on this data we have selected Machine Learning and Deep Learning techniques that will be performed. In Machine learning the algorithms that are performed are Linear Regression, Random Forest and XG Boost whereas in deep learning the algorithms that are performed are LSTM and RNN.In terms of Machine Learning algorithms the model that performed well was Random Forest with accuracy of 99.91% and error score of 996 Mw,whereas in Deep learning algorithms the model that performed well with good accuracy of 99.46% and less error score of 5818 Mw was LSTM.

1 Introduction

1.1 Background and Motivation

Electricity is an important part of life in the modern era and important to the economy of the country. It is used in the day-to-day life for different purposes such as lighting, heating, cooling, and refrigeration and for operating appliances, computers, electronics, machinery and many others. According to United States Energy Information Administration(EIA) [1], the electricity consumption in the United States was about 4 trillion kilowatt hours (kWh) in 2022 and it was the highest record till date since 1950, this increase was observed between the years 1950 to 2022 as shown in fig 1.1.The total electricity includes retail sales and direct use, retail sales of electricity consumption is done by consumers and direct use is used by commercial or industrial sectors.3.5% of total electricity consumption was used by the industrial and commercial sectors in the year 2022 and this caused the consumption of electricity in the US to increase by 3.2% which was higher in comparison to 2021 in which the consumption of residential sector was increased by 2.6% and the commercial sector by 4.7%.

[1] https://www.eia.gov/electricity/data/browser/

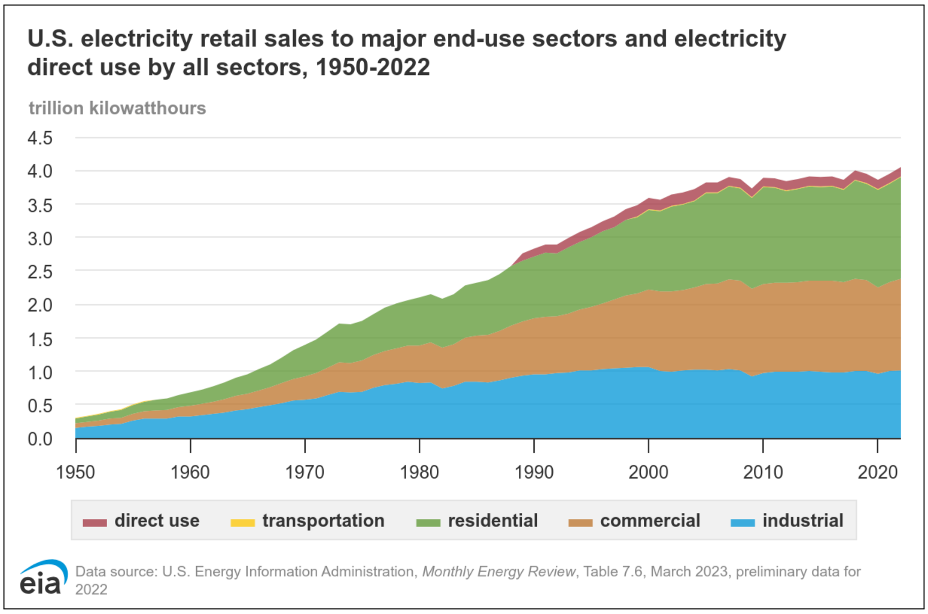

Figure 1: Statistics of Energy Consumption

The tremendous increase in population has led to an increase in demand for renewable and non-renewable natural resources. This exponential increase generates a huge amount of risk of depletion of the resources in the coming years, which creates a necessity to search for its alternative solution. The solution to this problem will be usage of renewable energy resources. Other resources suggest that electricity generation will rise to 71,000 TWh by 2050 wherein 88% of the production will be from renewable source of energy which will exhibit more of carbonization. The International Renewable Energy Agency has limited global warming to 1.5°C and has proposed global carbon neutral goal and thus accurate forecast of electricity is needed due of the increasing demand for the electricity.

In the research paper by Swan and Ugursal (2009), the authors reviewed different types of energy consumption techniques in the residential sector and they identified two approaches the top-down and bottom-up approach. As the residential sector consumes more amount of energy and the characteristic of residential sectors are complex comprehensive models are required to assess the consumption of energy. In the study by Yan et al. (2019) utilised LSTM technique to predict the consumption of energy and leveraged the correlation analysis to select the highly relevant inputs for it. They classified the nodes as strong, weak and medium to determine the best accurate forecasting method. Prediction of energy consumption is vital to forecast the future demand of electricity. As the demand for energy has increased accurate prediction is needed to avoid any haphazard. The current process is complicated as it has many challenges involved in it.

Now a days Machine Learning and Deep Learning techniques are popular as they are effective in forecasting the prediction of consumption of energy. Time series methods are commonly employed for predicting outcomes in this context. Support Vector Regression and Artificial Neural Networks are among the most utilized techniques for such predictions. To enhance energy prediction methods, this thesis will utilize 14 years of hourly energy usage data from an open-source dataset on the PJM electrical grid, a Regional Transmission Organization spanning 13 states in the United States.

This thesis explores the application of ML and DL algorithms to energy consumption forecasting. It involves simulating various ML algorithms, such as Linear Regression, Random Forest, and XG Boost, alongside DL algorithms like LSTM and RNN. To enhance energy prediction methods, this thesis will utilize 14 years of hourly energy usage data from an open-source dataset on the PJM electrical grid, a Regional Transmission Organization spanning 13 states in the United States.

1.2 Research Question, Project Objectives and Contributions

The research project focuses on comparison of machine learning and deep learning algorithms, and to check which one suits the best in predicting energy consumption prediction. Hence the research question and sub-research question is as follows:

Research Question: To what extent can machine learning(Random Forest,XG Boost,Linear Regression) and deep learning algorithms(LSTM,RNN) enhance/improve prediction of energy consumption in terms of accuracy to support the states in USA?

Sub-Research Question: Which machine learning and deep learning algorithm has the least error score,to enable the efficient model for prediction of energy consumption?

Below is the Summary for Project Objectives:

Table 1: Summary of Project Objectives

ID	Name	Description	Evaluation Method
1	Data Collection	Collect energy consumption data.	Exploratory Data Analysis
2	Data Preprocessing	Clean data, handle missing values.	Data Cleaning
3	ML Models	Implement ML models for prediction.	Accuracy, RMSE, MAE, MAPE
3.1	Random Forest	Use Random Forest model.	Accuracy, RMSE, MAE, MAPE
3.2	XGBoost	Apply XGBoost model.	Accuracy, RMSE, MAE, MAPE
3.3	Linear Regression	Use Linear Regression as baseline.	Accuracy, RMSE, MAE, MAPE
4	DL Models	Implement DL models.	Accuracy, RMSE, MAE, MAPE
4.1	LSTM	Use LSTM networks.	Accuracy, RMSE, MAE, MAPE
4.2	RNN	Apply RNN for time-series.	Accuracy, RMSE, MAE, MAPE
5	Model Evaluation	Evaluate all models.	Accuracy, RMSE, MAE, MAPE

2 Related Work

2.1 Critical review for Energy Consumption using Deep Learning Models

In the research conducted by Olanrewaju and Mbohwa (2017), the authors discuss regression analysis and Artificial Neural Network. The data used by the researchers for their research is from the year 1995 to 2009 is about the energy consumption and population. The researchers also made a note of how the economic growth and population affects the consumption of energy. It later founded a strong relation between factors such as GDP, population and many others. Later the results obtained from the performed models showed that both the regression and ANN model performed well, but ANN had more superior results than the other model. The researchers concluded their paper by stating that population plays a vital role when in prediction of energy consumption.

Whereas the other research paper by Abraham et al. (2022), the authors focus on the increasing need for precise forecasts of energy consumption and the relevance of deep learning algorithms. These forecasts are cardinals for electricity distributors in their efforts to prevent blackouts and efficiently manage energy resources. The integration of renewable energy sources, which introduce substantial variability in power supply due to their reliance on weather conditions, further exacerbates the complexity of accurate predictions. This article emphasizes the need for developing a model capable of accurately forecasting energy consumption across diverse settings, including small businesses and residential areas. It thus makes the case for deep learning algorithms, mainly due to the precision of their output. The methodology encompasses data collection, preprocessing, predictive modeling using deep learning techniques, and subsequent performance evaluation. Data collection involves the deployment of smart meters within selected buildings to capture energy usage data. Preprocessing steps are crucial to mitigating anomalies within the dataset, thereby enhancing the reliability of the predictive models. The study focuses on two models for prediction purposes: Long Short-Term Memory and Convolutional Neural Network, with a particular emphasis on the first model's proficiency in handling sequential data and retaining information over extended periods, thus rendering it more apt for this specific deployment.

According to the article, the LSTM model is a variant of recurrent neural networks and is renowned for its capacity to manage long-term dependencies, making it particularly suitable for time-series data tasks such as energy consumption forecasting. The study makes use of a three-layer LSTM network to continuously train and extract features from hourly energy consumption data, addressing the vanishing gradient problem commonly associated with standard recurrent neural networks. According to this article, CNN models are traditionally used for tasks involving spatial hierarchies in data, such as image recognition. However, this study considers CNN models applied to time-series data, albeit with less success compared to LSTM. Its performance was limited by its comparatively inferior capability in capturing temporal dependencies, which the LSTM networks handled more effectively.

The study concludes that LSTM models significantly outperform CNN models in predicting energy consumption, primarily due to their superior handling of sequential data. The LSTM model demonstrated a remarkably low error rate of 0.45%, in contrast to the 4.27% error rate observed in the CNN model. These findings suggest that LSTM models are particularly well-suited for applications where accurate prediction for forecasting time-dependent phenomena is essential. Hence, the article advocates for the use of LSTM models in the context of energy forecasting.

2.2 Investigating Machine Learning models for Energy Consumption

In the research conducted by Wu and Chu (2021), the authors focus on the Sampling strategy analysis for energy consumption and the relevance of machine learning algorithms. This article focuses the need for developing a model capable of accurately predicting energy consumption across diverse settings.

The methodology consists of data collection, preprocessing, correlation analysis using machine learning techniques, and subsequent performance evaluation. Data collection involves the deployment of smart meters within selected buildings to capture energy usage data which has several parameters such as temperature, humidity, weather conditions and many others. Preprocessing steps are crucial to handle anomalies within the dataset, to enhance the reliability of the predictive models. The steps included in the preprocessing are cleaning the data, synchronizing timestamps and finding incomplete records. Later is the correlation analysis where the researchers have analysed the correlation coefficient between the variables. The study focuses on machine learning models for prediction purposes: Linear regression, Support Vector Regression and Random Forest with a particular emphasis on the Mean Absolute Error and Mean squared error. To verify the results the researchers used the sampling strategy for all the models.

According to the article, machine learning models were applied on the data where the data was split into the training and test data sets.Amongst which 95% of the data was used o train the models and the remaining 5% was used to validate our machine learning models.The study concludes that Random Forest models significantly outperform other machine learning models in predicting energy consumption. The Random Forest model demonstrated a remarkably lesser mean absolute error rate of 0.15%, in contrast to the other models. These findings suggest that Random Forest model is particularly well-suited for applications where accurate prediction for forecasting is essential. Hence, the article suggests the use of Random Forest models model in the context of energy forecasting.

On the other hand, in the study conducted by Smpokos et al. (2018), the authors focus on the weather conditions regarding the energy consumption of data centres. This article focuses on how important the role of data centers is as it is needed to cool down the huge IT infrastructure.

The methodology consists of data collection, data preprocessing and feature selection and its extraction. Real time data has been collected from the test be which is a part of the IOT platform which has several parameters such as temperature, rain fall , atmospheric pressure conditions and many others. Preprocessing steps are crucial to handle anomalies within the dataset, to enhance the reliability of the predictive models. The steps included in the preprocessing are cleaning the data, synchronizing timestamps and finding the null values. Later is the feature extraction and selection where the researchers have analyzed the main weather parameters and the features with good correlation coefficients are picked, to remove the coefficients with least coefficient backward elimination process has been applied. The study focuses on machine learning models for prediction purposes: Multi variable Linear regression model with a particular emphasis on the Mean Absolute Error and Mean squared error. To verify the results the researchers used the sampling strategy for all the models.

According to the article, machine learning models were applied on the data where the data was split into the training and test data sets. The study concludes that several weather parameters had correlation with the consumption of energy in the datacenters, where wind chill and dew point temperature had the highest correlation. The R squared values for wind chill was highest as in comparison to the dew point temperature. The study concludes that to achieve good accuracy several parameters were combined with the highest correlation.

3 Energy Consumption Methodology Approach used and Data Pre-processing

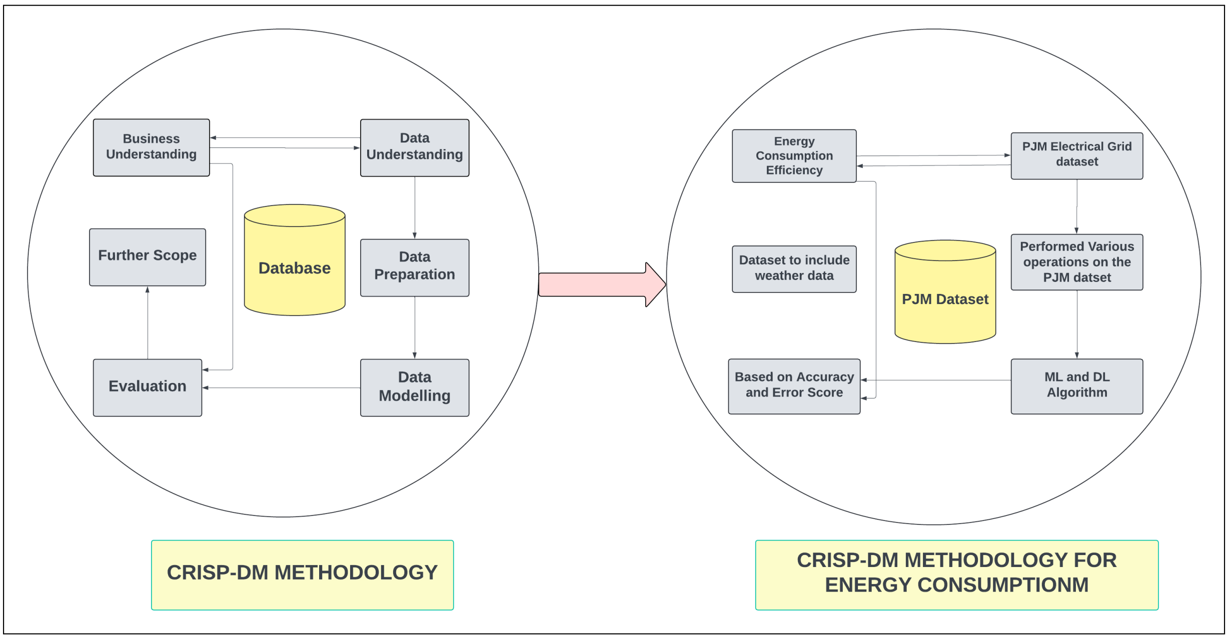

Figure 2: CRISP DM Methodology for Energy Consumption

This Section talks about the algorithms used and overall end-to-end process as shown in the fig 2. To understand this process, the project uses the CRISP-DM Methodology. The first step which is the crucial one is understanding the business, followed by it is data understanding, data preparation, data modeling or assessing the models, evaluation, and future scope or deployment.

3.1 Business understanding

The process of Business understanding is when organizations decide the values, missions and their goals and the important factors that can affect them. Here when predicting energy consumption, the main factors that can be looked into are to identify areas where the electric consumption can be reduced which can help save the operational cost and increase the efficiency.

3.2 Data understanding

Next comes data understanding, the dataset contains information about the energy consumption of several states in the United States, the files contain information about usage, date and time of different electric companies that come under PJM. Once the data understating is completed and overview about the features that our dataset contains is know the data needs to be prepared.

This research project uses dataset from Kaggle[2] and it is also accessible from the website of PJM .The dataset contains information about the PJM electric grid which is Regional Transmission Organization in the United States and is spread over 13 states. There are several electric companies under the PJM grid such as AEP(American Electric Power),ComEd(Commonwealth Edison), The Dayton Power and Light Company(DAYTON), Duke Energy Ohio/Kentucky (DEOK), Dominion Virginia Power (DOM), Duquesne Light Co. (DUQ), East Kentucky Power Cooperative (EKPC), FirstEnergy (FE), Northern Illinois Hub (NI), PJM East Region: (PJME), PJM West Region: (PJMW), PJM Load Combined: (PJM_Load). The dataset contains information about the above companies and each dataset has two attributes Datetime and the consumption of energy (MW).The datatype of Datetime is object and that of consumption of energy(MW) is float as show in table 2.

Attributes	**Type**
Datetime	Object
Consumption of energy (MW)	Float

Table 2: Attributes and their Types

3.3 Data preparation

In order to run several models for the prediction of consumption of energy data preparation is needed, where in data is cleaned, NA values are filled with the average value and many other steps are performed.In order to use all these files to get the desired results we need to prepare the data. There are certain steps involved in these and they are described as following.

1. Data Collection : The PJM dataset consists of 12 csv files representing different power grids in the United States. In order to predict the energy consumption, it is important to concatenate all the csv files. As all the individual files have two columns it is difficult to combine them based on this information, and so to ease this process, we have added an extra column named electric_company to all the files, which makes combining all the csv files easier as shown in figure 3. The data set now has the total number of rows, Datetime, mw_energy_consumption, electric company.

2. Data Description: The combined dataset now has 10,90,167 rows and 3 columns. The Table 3 prvoides detail description of attributes of the combined dataset.

3. Data Tranformation: Table 4 shows the type of data for all attributes in the dataset. The data type of attribute "datetime" is object. The data type of the "datetime"

[2]https://www.kaggle.com/datasets/robikscube/hourly-energy-consumption

	Datetime	mw_energy_consumption	electric_company
45725	2009-01-13 17:00:00	19389.0	AEP
21482	2015-02-18 04:00:00	2348.0	EKPC

Figure 3: Final result post concatenating

Attributes	**Type**
Datetime	Object
Consumption of energy (MW)	Float
Electric Company	Object

Table 3: Attributes and their Types

attribute was changed to datetime so that more information can be derived from the attribute value.

Attributes	**Type**
Datetime	datetime64[ns]
Consumption of energy (MW)	Float
Electric Company	Object

Table 4: Attributes and their Types

4. Data Pre-processing: In this step, the dataset was analysed to determine whether it contains any null or missing values, as it might impact the performance of model. As shown in figure 4, the dataset does not contain any values

```
df1.isna().sum()

datetime                 0
mw_energy_consumption    0
electric_company         0
dtype: int64
```

Figure 4: Checking Null values

3.3.1 Feature Engineering

To better understand the data and to get a good outcome, let us delve more into our dataset. Our dataset consists of three columns, datetime, energy consumption (mw), and electric company. We will now extract the following from the datetime column:

1. Date: The purpose of extracting the date from the datetime column is to have a view of energy consumption without considering the time.

2. Year: The purpose of extracting year from the datetime column is to have a view of energy consumption over the years.

3. Month: The purpose of extracting Month from the datetime column is to have a view of energy consumption over the months which can help in knowing the seasonal patterns and trends.

4. Hour of the day: The purpose of extracting Hour from the datetime column is to have a view of energy consumption in an entire day.

We have also defined seasons, holidays and day of the week. Defining holidays will determine whether it was a holiday or not. By breaking down the datetime column into the above columns the dataset now has columns such as datetime mw_energy_consumption, electric_company,date,year, month ,hour_of_day,season, holidays, day_of_week. Table 5 provides detailed description of all the columns and their datatypes.

Column Name	**Data Type**
`datetime`	`datetime64[ns]`
`mw_energy_consumption`	`float64`
`electric_company`	`object`
`date`	`datetime64[ns]`
`year`	`int64`
`month`	`int64`
`hour_of_day`	`int64`
`season`	`object`
`holidays`	`object`
`day_of_week`	`int64`

Table 5: Datatype of all columns

3.4 Exploratory Data Analysis

1. Univariate Analysis

   The main variable of our dataset mw_energy_consumption and the plot is as shown below 5

2. General Analysis

   Here we have plotted the Bar Plot and KDE plot 6 below for the energy Energy consumption of all the companies,followed by plot of all the seasons and plot of holidays vs normal days.

3. Bivariate Analysis

   Here we have plotted energy consumption and hour of the day plot 7 to see the view of energy consumption throughout the day,wherein we see the energy conusmption increases in the morning and decreases again with again increase in the afternoon.

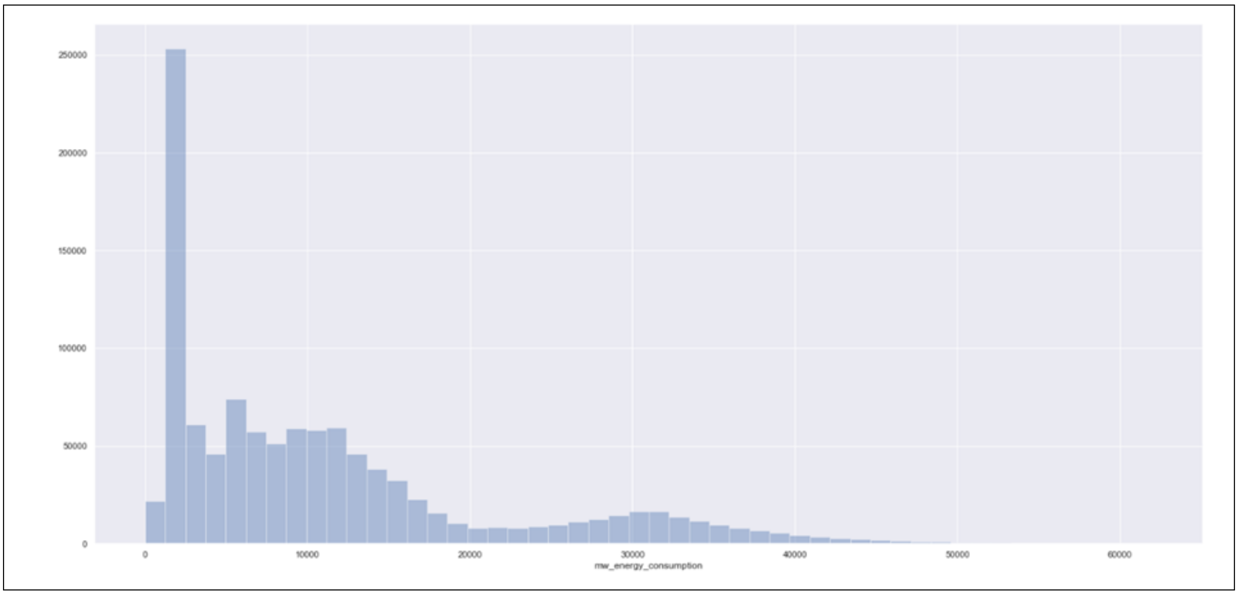

Figure 5: Plot for Energy consumption

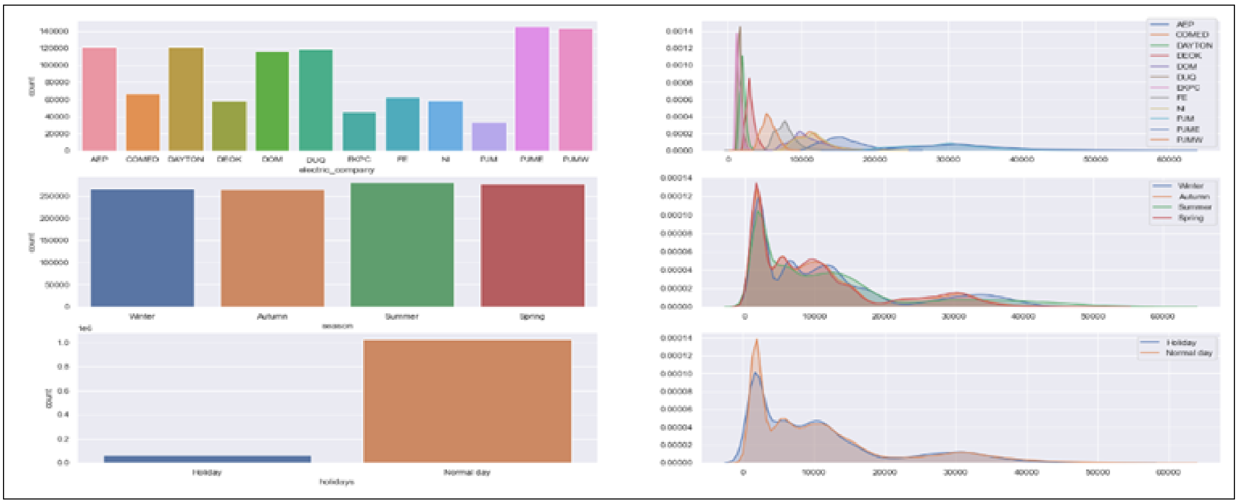

Figure 6: Bar and KDE plot for Energy Consumption

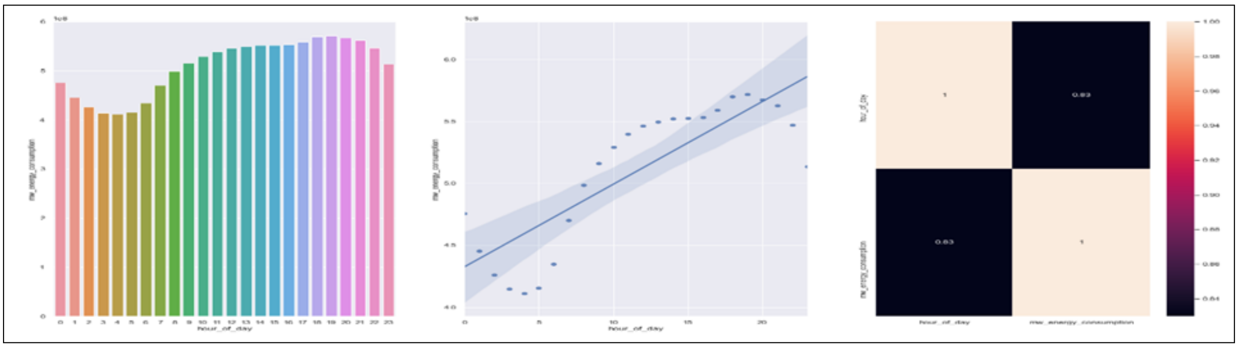

Figure 7: Hour of the day plot

4. Seasonal Plot

   Here we have plotted energy consumption and season to see the view of energy consumption as shown below 8throughout the yeat,wherein we see the energy conusmption in smore in the summers and winter and the second plot which is spread over number of years also give the same observation.

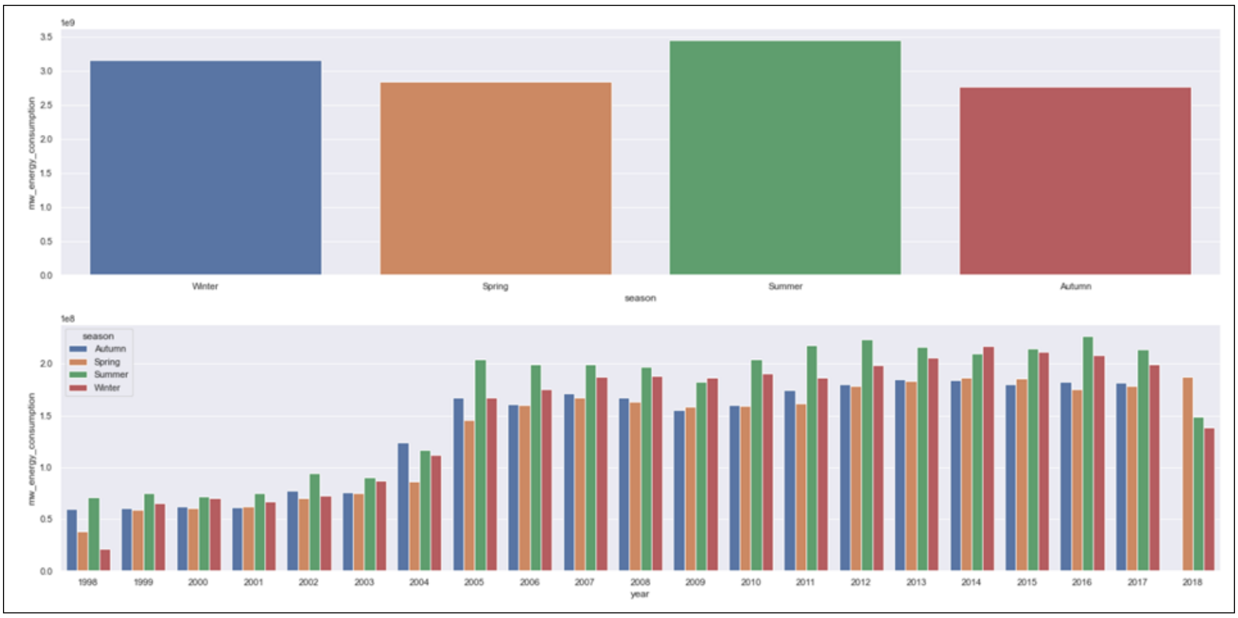

Figure 8: Seasonal Plot

5. Holiday vs Normal day plot

   Here we have plotted energy consumption vs holidays as shown below 9 to see the view of energy consumption on a normal day and a holiday,wherein we see the energy consumption in more in on a normal day and very less on a holiday.The second plot is graph of holidays spread over the years which also gave us the same observation.

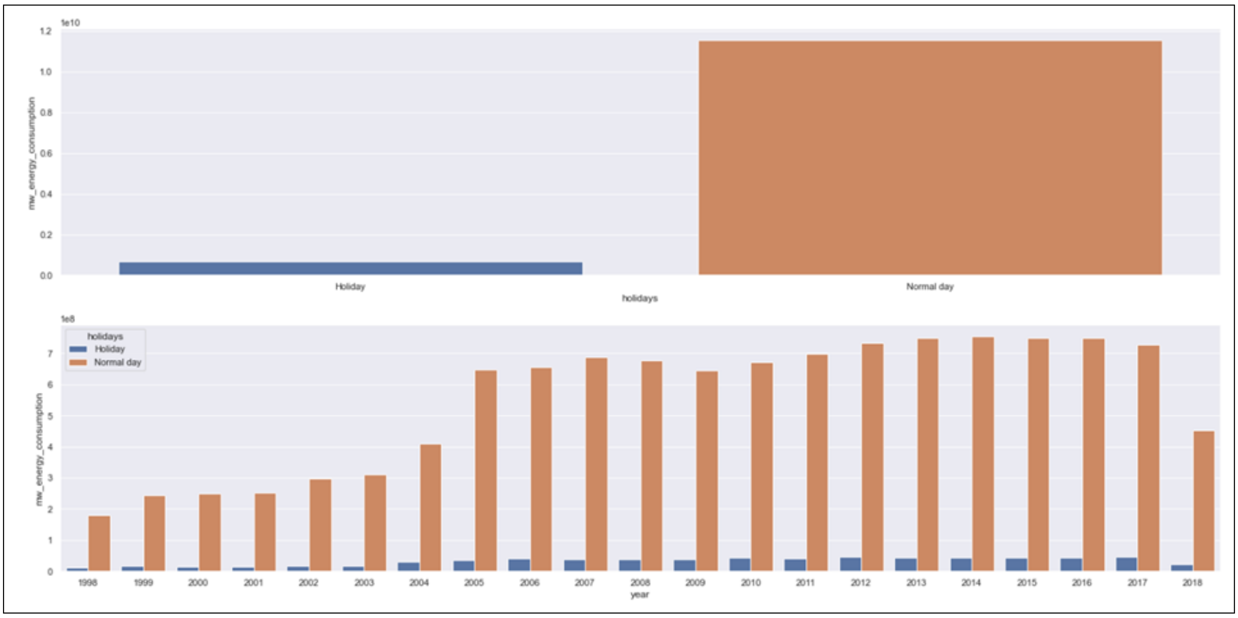

Figure 9: Holiday vs Normal Day Plot

6. Day of the week plot

   Here we have plotted energy consumption and day of the week 10as shown below to see the view of energy consumption throughout the week,wherein we see the energy consumption in more on the weekends than that of the weekdays. The second plot is which is spread over number of months also gives the same observation,where 0=Monday,5=Saturday and 6=Sunday.

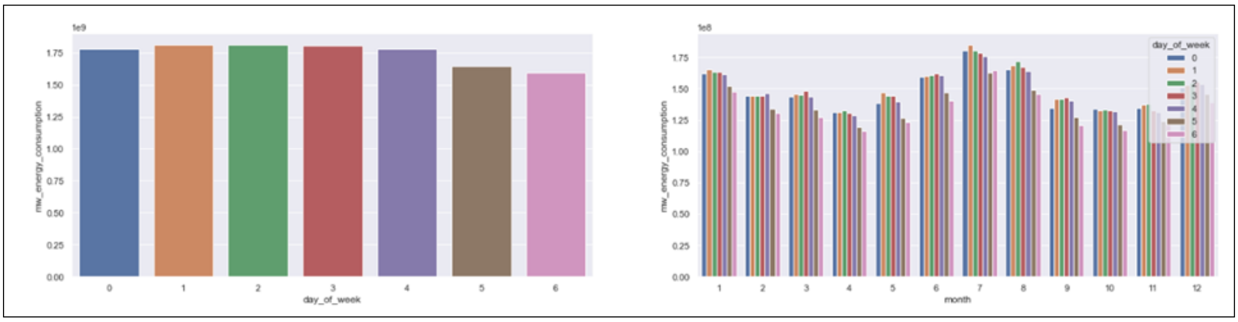

Figure 10: Day of the week plot

7. Multivariate Analysis

   Let us look at the Numerical and Categorical Variables in a heatmap plot.

   Numerical Attributes in the figure 11

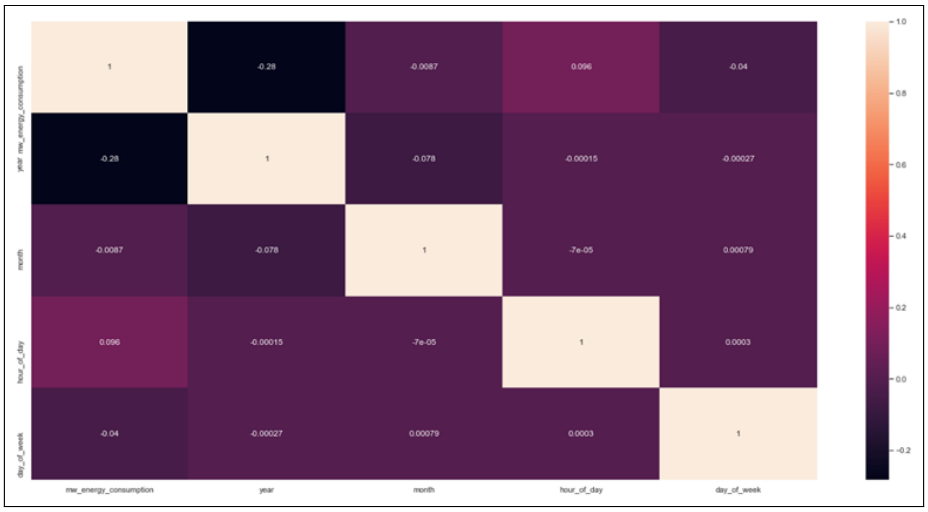

Figure 11: Heat Map for Numerical Variables

   Categorical Attributes in the figure 12

3.5 Feature Selection

We derived certain columns from the datetime column to derive more insights from our data.Now let us drop these extra columns and next step is to split the dataset into train and test models.As the range of dates are vary a lot we have determined the cut off dates for each electrical company on basis of which we have splitted the dataset into the train and test sets and store these cut off dates into a new dataframe which will help in getting better visuals.

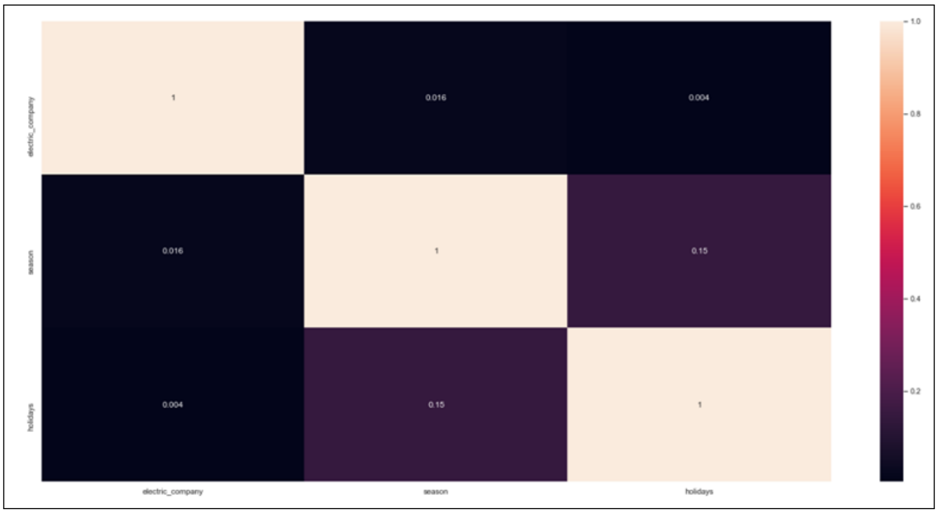

Figure 12: Heat Map for Categorical Variables

Now using the Boruta function will help in removing the unnecessary features that are less relevant and can help to identify the most prominent features in the dataset.Based on the features selected during EDA and the features selected by the boruta function the finalised features are electric_company,hour_of_day_sin, hour_of_day_cos, season_sin, season_cos,day_of_week_sin, day_of_week_cos,datetime, mw_energy_consumption.

4 Design Specification and Implementation of Machine Learning & Deep Learning Models

4.1 Design Specification

1. **Identifying Problem :** In this section 13, the core of problem of energy consumption prediction in United States was clearly identified. The primary aim was to understand the importance of accurate forecasting in order to manage energy resources efficiently across different states in United states.

2. **Survey :** In this section, existing literature related to energy consumption prediction was comprehensively reviewed. The findings of the review helped in identifying the gaps in current research and selecting specific models for this research.

3. **Data Collection and Processing :** In this section, the outline of the process for data gathering and processing is provided. The section cover sources of data and details of steps followed during data processing.

4. **Feature Selection and Model Application (ML& DL Models) :** In this section, various machine learning and deep learning models were selected based on the nature of data and findings of the literature review. The Machine Learning models include Random Forest, XGBoost, Linear Regression, and the Deep Learning mdoels include LSTM, and RNN.

5. **Result Analysis** In this section, results obtained from implementation of Machine Learning and Deep Learning models were analyzed. The The performance

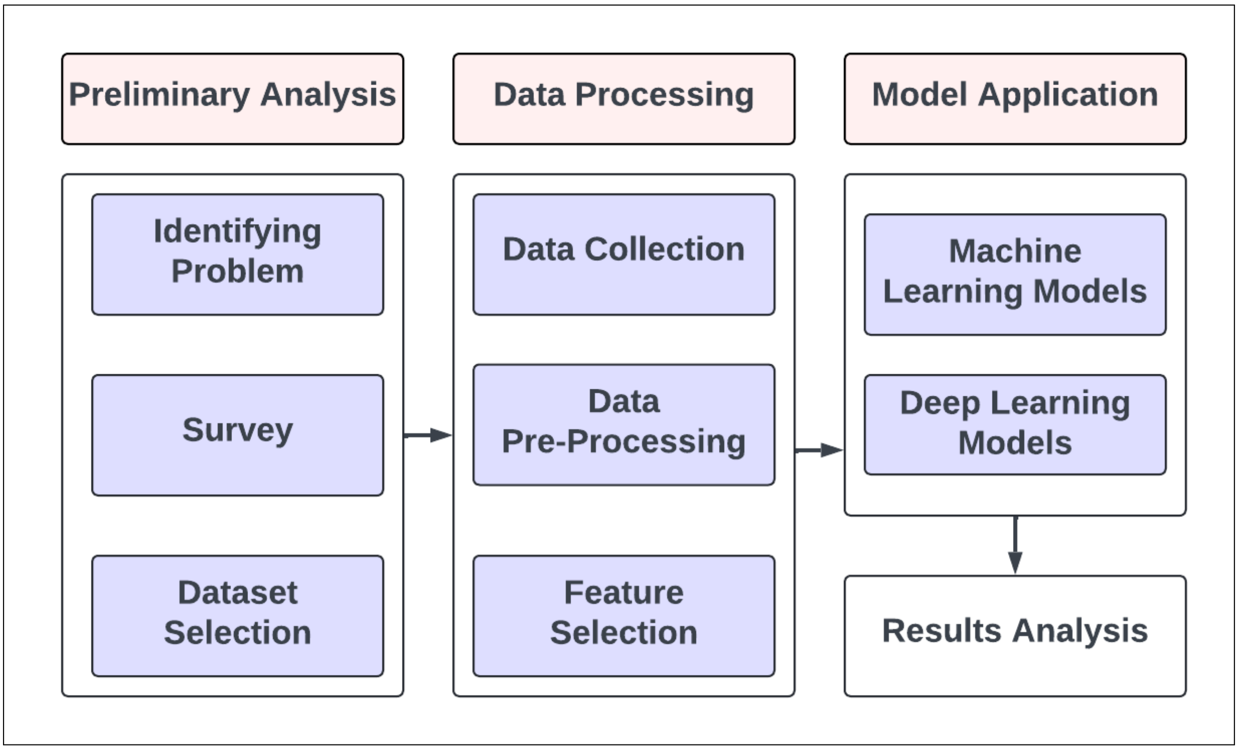

Figure 13: Implementation of ML and DL Algorithms

of each model will be evaluated using metrics like Mean Absolute Error (MAE), Root Mean Square Error (RMSE) to determine their accuracy in predicting energy consumption.

4.2 Implementation, Evaluation and Results of Machine Learning Models

Implementation is the main part of the research project.Here we have implemented Machine Learning and Deep Learning models such as Linear Regression,Random Forest,XG Boost,LSTM and RNN.The analysis has been carried out in the Jupyter notebook.After pre processing the data,it was divided into the train and test dataset and the algorithms were implmented as follws:

4.2.1 Implementation, Evaluation and Results of Linear Regression Model

Implementation :

Linear Regression (LR) is a model that helps to know the relationship between the input and the output.The inputs are nothing but return variables and output is response variable. Linear regression is used when a trend is seen and has a historical forecast data. As the applications of Linear regression models are simple it has been used in various works related to the consumption of energy. By using sklearn.linear_model library we have used Linear Regression model and have implemented it to predict the energy consumption. Later we have also cross validated the model so as to decrease the error.

Results and Evaluation :

The Linear Regression Model achieved the accuracy of 98.80% with Mean Absolute Error of 7533 Mega Watts.

4.2.2 Implementation ,Evaluation and Results of Random Forest Model

Implementation :

Multiple decision trees are combined when combined together to increase accuracy and decrease overfitting is a Random Forest Vijayan (2022).The dataset is split into subsets and for each subset a decision tree is built and combining the final output gives us the results.By using klearn.ensemble library we have used Random Forest model and have implmented it to predict the consumption of energy. . Later we have also cross validated the model so as to decrease the error.

Results and Evaluation :

The Random Forest Model achieved the accuracy of 99.90% with Mean Absolute Error of 996 Mega Watts.

4.2.3 Implementation ,Evaluation and Results of XG Boost Model

Implementation :

It is a sequential technique in which the model goes on improving in comparision to the past one.XG Boost is also known as Extreme Gradient Boosting technique which is a combination of Gradient Boosting and Elastic Net.It adds layer L1 and L2 which prevents the overfitting of this technique.By using xgboost library we have used Random Forest model and have implemented it to predict the consumption of energy. Later we have also cross validated the model to decrease the error.

Results and Evaluation :

The XG Boost Model achieved the accuracy of 99.57% with Mean Absolute Error of 3979 Mega Watts.

4.3 Implementation,Evaluation and Results of Deep Learning Models

4.3.1 Implementation, Evaluation and Results of Long Short Term Memory (LSTM) Network Model

Implementaion :

Long Short Term Memory (LSTM) networks are special types of recurrent neural networks that address the vanishing gradient problem faced in the traditional recurrent neural networks Karim et al. (2018). The model was trained using LSTM( ) function provided Keras library from TensorFlow and compiled using the compile( ) function to predict the energy consumption. The model was further tested using the fit( ) function after passing the test data and trained model as an argument in the function. The detailed experimentation with LSTM model is document in 5.4 section.

Results and Evaluation :

The LSTM Model achieved the accuracy of 98.81% with Mean Absolute Error of 6516 Mega Watts.

4.3.2 Implementation ,Evaluation and Results of Recurrent Neural Network (RNN) Model

Implementation :

RNN is a deep learning model which functions on the principles of supervised learning Kaur and Mohta (2019). Unlike traditional neural networks, RNN uses previous information for further computation. Therefore, it is known as Recurrent as it executes similar operations for each iteration, where the outcome is dependent on previous results Kaur and Mohta (2019). To predict the energy consumption, RNN model was implemented using the Keras library from TensorFlow. The model was trained with the SimpleRNN( ) function and compiled using the compile( ) function. Once the model was trained and complied it was tested using the fit( ) function by passing the trained model with the test data. The detailed experimentation with RNN model is documented in 5.3 section.

Results and Evaluation :

The RNN Model achieved the accuracy of 99.46% with Mean Absolute Error of 5818 Mega Watts.

5 Comparision of Machine Learning and Deep Learning Models

The Machine learning and Deep learning models have been evaluated based on certain parameters. The parameters are MAE(Mean Absolute Error), MAPE(Mean Absolute Percent Error), RMSE (Root Mean Squared Error) and Accuracy.

5.1 Machine Learning

Model Name	MAE	MAPE	RMSE	Accuracy
Random Forest	996.07	0.09	1754.90	99.91
XGBoost	3979.91	0.43	6691.57	99.57
Linear	7533.68	1.20	10380.86	98.80

Table 6: Model Performance Comparison

Table 6 shows the performance of all the Machine Learning models in terms of MAE, MAPE, RMSE and Accuracy. The model with best performance, i.e. highest accuracy, is Random Forest with 99.91% Accuracy and Mean Absolute Error of 996 Mega Watts.

5.2 Deep Learning

Table 7 shows the performance of all the Machine Learning models in terms of MAE, MAPE, RMSE and Accuracy. The model with best performance i.e highest Accuracy is RNN with 99.46% Accuracy and Mean Absolute Error of 5818 Mega Watts.

Model Name	MAE	MAPE	RMSE	Accuracy
RNN	5818.85	0.54	8019.67	99.46
LSTM	6516.73	1.20	7818.21	98.81

Table 7: Model Performance Comparison

Below is a bar graph 14 that shows accuracy of all the models:

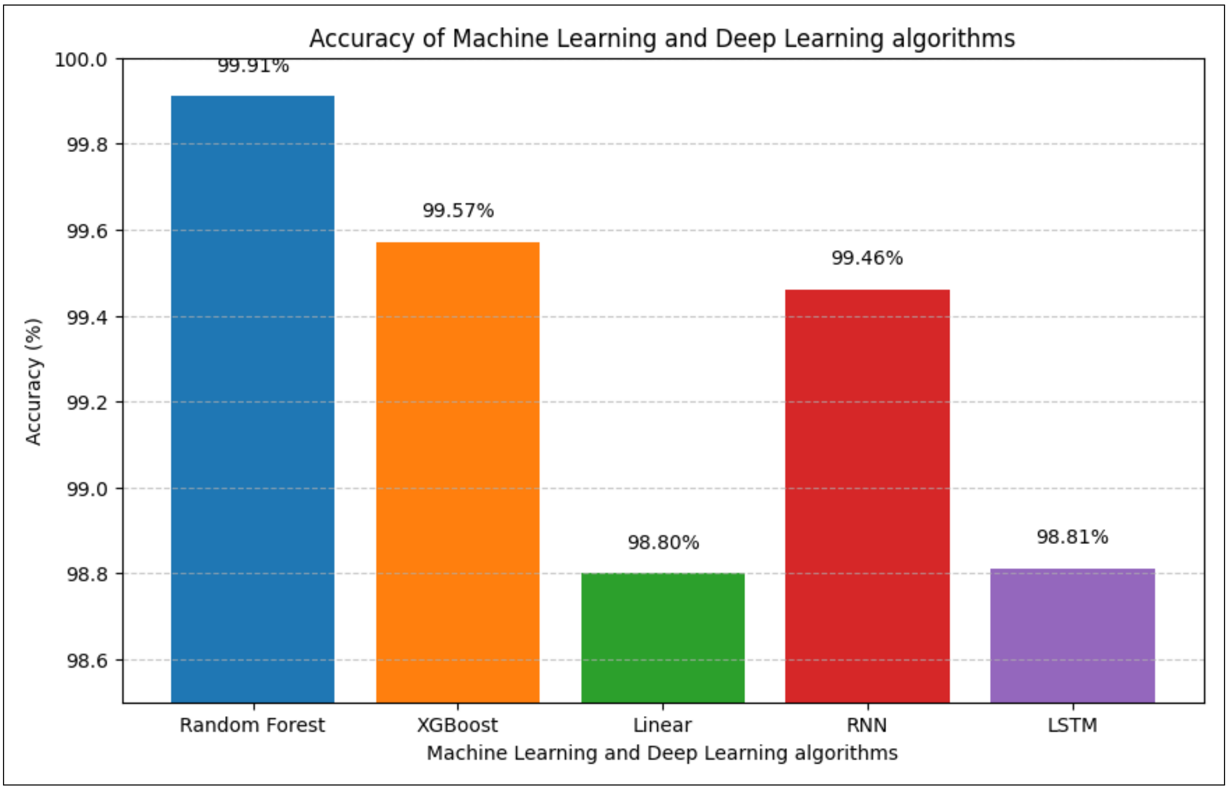

Figure 14: Accuracy of Machine Learning and Deep Learning algorithms

Below is a bar graph 15 that shows Mean absolute error of all the models:

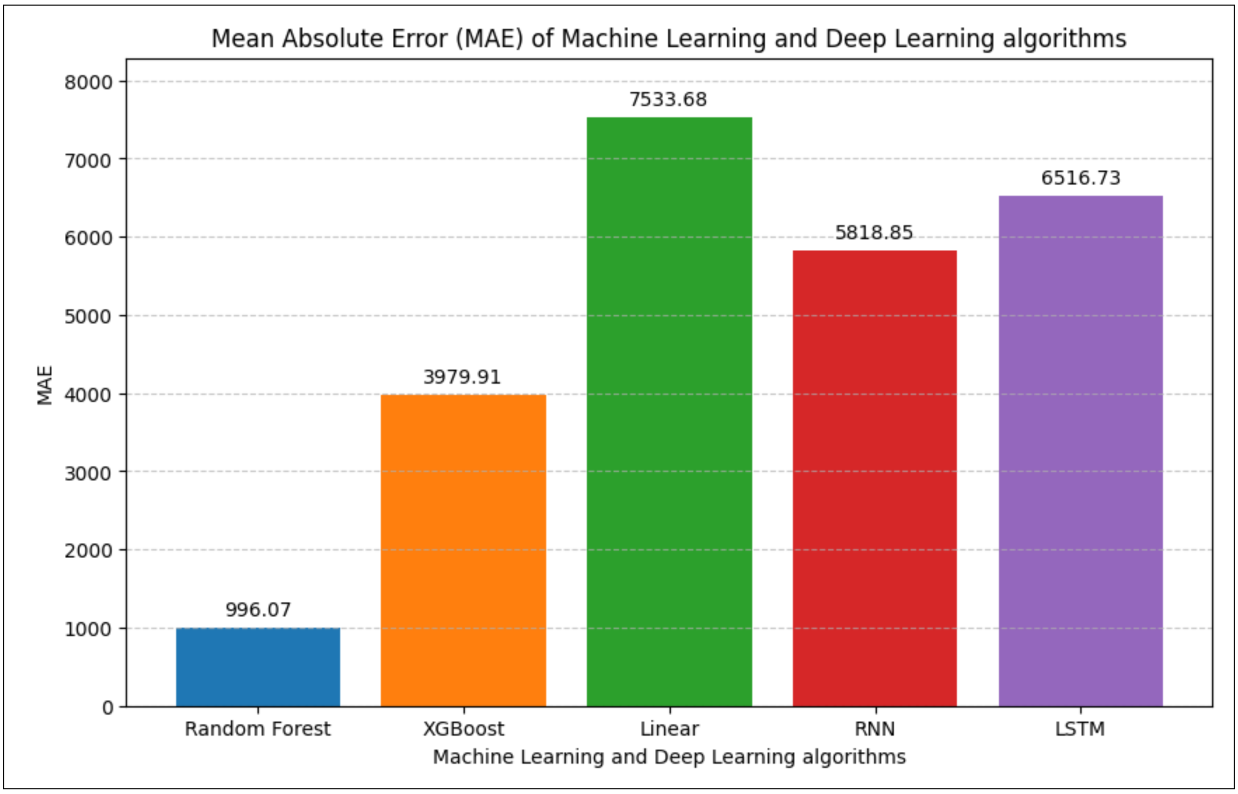

Figure 15: MAE of Machine Learning and Deep Learning algorithms

5.3 Experiment with RNN

Model Name	MAE	MAPE	RMSE	ACCURACY
Simple RNN	11732.28	2.2313	18078.34	97.7687
RNN with 3 layers	5744.52	0.4888	8026.06	99.5112

Table 8: Evaluation Metrics for RNN Models

Table 8 compares two RNN models: **Simple RNN** and **RNN with 3 layers**, showing that RNN with 3 layers significantly outperforms Simple RNN in all key metrics. It has lower errors (MAE: 5,818.85 vs. 13,314.74, RMSE: 8,019.67 vs. 20,149.56), a lower percentage error (MAPE: 0. 54% vs. 2. 50%) and higher precision (99. 46% vs. 97. 50%). The complex RNN Model architecture consists of three SimpleRNN layers. Layer normalization is applied after each of the first two RNN layers to stabilize training and dropout layers are included to prevent overfitting by randomly deactivating half of the units during training. The final SimpleRNN layer reduces the sequence to a single output, which is passed through a Dense layer to produce the final prediction.

5.4 Experiment with LSTM

Model Name	MAE	MAPE	RMSE	ACCURACY
Simple LSTM	25003.99	3.6583	47848.85	96.3417
Simple LSTM with SELU activation	11951.88	1.5534	19208.34	98.4466
Stacked LSTM with 0.5 Dropout	6516.73	1.2015	7818.22	98.7985
Simple LSTM with 0.5 dropout & GC	10928.21	2.6946	13514.86	97.3054

Table 9: Model Performance Comparison

Table 9 compares the performance of various LSTM models, showing that the **Stacked LSTM with 0.5 Dropout** achieves the best results, with the lowest errors (MAE: 4766.43, RMSE: 5812.44) and the highest accuracy (98.81%). The **Simple LSTM with SELU activation** also performs well, significantly reducing errors compared to the basic **Simple LSTM** model.

5.5 Discussion

The research project highlights on the models performed under machine learning and deep learning algorithms for prediction of consumption of energy.Based on the outcomes ,random forest model ,a machine learning model had the highest accuracy of 99.91% with lowest mean absolute error 996.01MW,whereas RNN a deep learning model had accuracy of 99.46% and mean absolute error of 5818.85 Mw.Whereas if we compare the best model amongst all in terms of less error score and highest accuracy it is concluded that Random forest a machine learning algorithm has outperformed the Deep learning Model.

6 Conclusion and Future Work

The research project primarily was focused on two main problem things a) accuracy and b)least error score using the machine learning and deep learning algorithms.It was seen that Random forest amongst both machine learning and deep learning algorithm performed well.It has good accuracy and least error score.

Using real time data can give us more prominent results in case of accuracy,using real time data of newly introduced electrical grid companies under the PJM grid can give us an overview of entire US and not limit to certain states.Going ahead a hybrid model can be designed combining the best suited approach including both Machine Learning and Deep Learning algorithms.Due to time constraints only two deep learning algorithms were implemented,in future many other algorithms can be implemented to get results. Other factors such as the weather conditions and economic factors can also be taken into consideration for predicting the energy usage.

Acknowledgement

I wish to thank Dr. Catherine Mulwa for her contribution and support during the span of thesis project and her help throughout this tenure. I am also grateful to the lecturers and staff of National College of Ireland for their help and support in my Masters in Data Analytics.

References

Abraham, A. V., Sasidharan, P., Tejas, S. S., Manohara, M., Muthu, R. and Naidu, R. C. (2022). Predicting energy consumption using lstm and cnn deep learning algorithm, *2022 7th International Conference on Environment Friendly Energies and Applications (EFEA)*, pp. 1–6.

Karim, F., Majumdar, S., Darabi, H. and Chen, S. (2018). Lstm fully convolutional networks for time series classification, *IEEE Access* **6**: 1662–1669.

Kaur, M. and Mohta, A. (2019). A review of deep learning with recurrent neural network, *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, pp. 460–465.

Olanrewaju, O. A. and Mbohwa, C. (2017). Predicting world energy consumption: Comparison of ann and regression analysis, *2017 IEEE International Electron Devices Meeting (IEDM)*, pp. 1–5.

Smpokos, G., Elshatshat, M. A., Lioumpas, A. and Iliopoulos, I. (2018). On the energy consumption forecasting of data centers based on weather conditions: Remote sensing and machine learning approach, *2018 11th International Symposium on Communication Systems, Networks Digital Signal Processing (CSNDSP)*, pp. 1–6.

Swan, L. G. and Ugursal, V. I. (2009). Modeling of end-use energy consumption in the residential sector: A review of modeling techniques, *Renewable and Sustainable Energy Reviews* **13**(8): 1819–1835.
URL: *https://www.sciencedirect.com/science/article/pii/S1364032108001949*

Vijayan, P. (2022). Energy consumption prediction in low energy buildings using machine learning and artificial intelligence for energy efficiency, *2022 8th International Youth Conference on Energy (IYCE)*, pp. 1–6.

Wu, Z. and Chu, W. (2021). Sampling strategy analysis of machine learning models for energy consumption prediction, *2021 IEEE 9th International Conference on Smart Energy Grid Engineering (SEGE)*, pp. 77–81.

Yan, G., Han, T., Zhang, W. and Zhao, S. (2019). Short-term load forecasting of smart grid based on load spatial-temporal distribution, *2019 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia)*, pp. 781–785.