World Energy Consumption

Name:
Sohel Arsad Salam

Student ID:
001166291

Programme of Study:
MSc in Data Science

Final Report

Date Proposal Submitted: 27th May, 2022

Project Hand In Date: 16th September, 2022

Supervisor: Prof. Tony Mann, Director, Greenwich Math Center

Abstract

Today's world requires extremely efficient energy consumption. Demand is rising as a result of the industrial sector's quick advancements, making energy efficiency initiatives essential to reducing energy waste and satisfying demand. According to the study of numerous scenarios used by policymakers, at least a 50% reduction in industrial energy use is required if the increase in global temperature is to be kept to less than 2°C by the end of the 21st century. It is crucial that we include a trustworthy forecast of the future that can be used to estimate the energy usage based on numerous anticipated elements in order to remain on track with all these eventualities and to meet the desired objectives.

Accurate power demand and consumption forecasting gives businesses a competitive edge by enabling the most efficient and ideal operating system. The conventional time-series forecasting techniques cannot be utilized when working with seasonal data and outside influences. Accurate power forecasting for electrical consumption prediction is a crucial tool in energy planning for producing companies since it helps analysts comprehend and forecast market electricity demand. In a deregulated market, their power output can be modified accordingly. In order to explicitly deal with seasonality as a class of time-series forecasting models, Persistence Models (Naive Models), Seasonal AutoRegressive Integrated Moving Averages with eXogenous regressors (SARIMAX), and Univariate Long-Short Term Memory Neural Network (LSTM) are utilized.

The main goal of this dissertation is to analyze London's power's exploratory data before using several forecasting models to anticipate the following 24 hours' worth of energy consumption and daily peak demand once every day. The very first three years were utilized as the training set and the remaining years as the test set to divide the data on power use. The acquired findings demonstrated that the evaluated methods were surpassed by the machine learning techniques suggested in the most recent literature.

To directly compare models to energy values in the data, root mean squared error, or RMSE, is used in the evaluation process. RMSE has two methods for calculation. The first to illustrate the flaw in hourly prediction. second to illustrate the general performance of the models. The findings demonstrate that electricity demand can be predicted using machine learning algorithms, renewable energy deployment, preparing for peak load days, and reduction of waste from polluting reserve standby generation. The results also demonstrate the quantification of energy and cost-saving measures and the detection of anomalies in consumption trends.

Acknowledgement

Without the tremendous assistance and support of many people, my dissertation would not have been completed. This is my chance to express my gratitude to everyone who helped me along the way on this difficult trip, which has been a once-in-a-lifetime experience.

I want to start by expressing my gratitude to Prof. Tony Mann, my supervisor, for all his time and effort that he has put into my capstone project as well as for his professional guidance, support, and encouragement. Working in such proximity to him has been a priceless gift.

I thank the University of Greenwich's Data Science Faculty as well as the university for their all-around assistance.

Finally, I want to sincerely thank my friends and family, especially my mother Kauser Begum and father Abul Salam, for their love, support, and encouragement over the years. I dedicate this dissertation to them and have no words to adequately express my gratitude

Contents

List Of Figures

List of Tables

1. Introduction

The global energy demand is increasing every day, and during the past 20 years, power consumption has nearly doubled. Fossil fuels continue to be a major source of energy in many nations, and the demand for sustainable energy resources has developed as a result of the depletion of traditional energy sources like natural gas, oil, or coal. As a result, research is now more heavily focused on improving the utilization of renewable energy and renewable energy sources.A projection of future energy consumption is a vital input to many analyses of economic, energy, and environmental policies (Bhattacharyya and Timilsina, 2009).World Energy Consumption is a project which is to collect the different energy consumption data from around the globe and optimize the energy consumption. One of the rapidly expanding technological topics that combines computer science and analytics is machine learning (ML). It deals with the problem of creating algorithms that are more effective through experience-based learning in computers. Thanks to new techniques, accessible processing power, and the accessibility of internet data, machine learning (ML) is advancing.

This project explores machine learning approaches and creates data-driven models for forecasting energy consumption and performance in order to determine whether a broad and straightforward strategy based on machine learning models can produce satisfactory results in a complicated forecasting problem. Transmission Service Operators (TSOs) of the electrical grid publish energy demand predictions once every day to properly satisfy energy demands for the upcoming 24-hour period. This problem, which forecasts the anticipated maximum energy demand on an hourly basis and comprises of 24-hourly slices, is widely used and extremely significant. These forecasts are supplemented with ultra-short term (6-hour or less) forecasts to maintain grid balance and schedule supply dispatch for day-ahead bidding procedures.

High precision power demand forecasting could introduce a great collection of values for a nation, city, or even individual houses. If stakeholders choose to satisfy their energy needs from outside sources, they can change their power production to do so and cut costs, or they can purchase enough energy to meet their needs. The stakeholders may make additional profit in some specific circumstances, such as in the tendering procedures for daily energy exchange.

This documentation has 6 chapters explaining every single bit of the project, the research and other technologies involved. The first chapter is the introduction under which the motivation for this project is answered and also the scopes and limitations and target group is described. In chapter 2, all the research work is explained under literature review. Chapter 3 & 4 describe the theory behind the technologies used and the methodologies used to achieve the results.

1.1. Motivation

Every year the stakes develop higher in the battle to save the climate and battle an unnatural weather change. Presently like never before, we're mindful of the harmful impacts that our ongoing reliance on non-renewable energy sources holds for our aggregate fates. Throughout the course of recent years, the world economy has dramatically multiplied. Albeit monetary development increased living standards in many nations, it was likewise answerable for a decrease in regular assets and an expansion in ozone harming substance emanations. That's what a few conjectures recommend: with speeding up development of populace and Gross domestic product, by 2050 we will be confronted with a significant challenge of not having an adequate number of assets; this, thus, will sabotage further monetary advancement particularly in poor areas. Energy policy and the development of energy resources in a sustainable manner can be developed by economies if the relationship between energy use and economic growth is understood. Having access to a system capable of rendering the data from all over the world on different kinds of energy and providing a visualized form making it easier to comprehend the data and also suggesting a replacement sustainable energy could bring some real changes.

1.2. Scope & limitation

Some restrictions must be placed on this research as a result of the specified deadline in order for it to be completed on time:

- This project will be more focused on visualizing some very specific columns of the dataset.
- As the evaluation is done on the results of the Machine Learning approach, other techniques will not be considered.
- Big energy consumers in major cities is weather-related factors. It is possible to study and use the relationship between weather characteristics and electricity consumption as a key input in the model.
- In addition to London, the shared energy regions of Portugal and France also have high electricity consumption. The model performance may be impacted by the power transfers and correlation between these locations.
- SARIMAX models cannot accommodate dual seasonality in the data, such as both weekly and yearly patterns.

1.3. Target Audience

Government, commercial utilities companies, and policymakers must prepare extensively. Basic and more advanced energy prediction algorithms have both been utilized to obtain an accurate estimate of future global energy consumption. Machine learning approaches have been able to enhance or, in the long term, replace conventional energy forecast techniques since the development of artificial intelligence.This work is especially interesting for Economists, Policy Makers, Energy Suppliers, Governments as the energy consumption and economic growth compliments each other and also the findings of this project will help in making better energy conservation policies, proper energy regulations, growth in GDP.

2. Literature Review

This section provides an overview of the literature on energy optimization techniques and the unpredictability of energy consumption projections. Numerous research have been conducted in order to improve the energy efficiency of new or existing buildings, and two main strategies have been used in the creation of data-driven prediction systems. These data-driven prediction models can be learned from simulation-collected data using building performance simulation tools like TRNSYS and ESP-r to collect energy-related data to train models, or they can be trained through real data using smart buildings equipment to collect data to train models. For precise power forecasting and the best energy planning decisions, artificial intelligence (AI) in general and machine learning (ML) approaches, as well as a growing collection of publicly available energy consumption data, have been advocated in recent years. It helps producing businesses to cost-effectively manage energy demand. Energy resources are already running out, supply capacity is being exceeded, and there are serious environmental effects like global warming, ozone layer depletion, climate change, and others due to the world's rapidly growing energy consumption. In a mixed integer linear programming-based scheduling technique for smart home equipment conducted by Sou Kin Cheong, they considered the anticipated time frame and peak power use of the appliances. Based on a pre-existing tariff, the suggested timetable resulted in a cost savings of around 47%. The authors also showed that excellent solutions may be found with relatively little computing effort. (Cheong SK, 2011). Over the past 10 years, substantial research has been done on energy optimization. The energy efficiency of buildings has been the subject of several studies, and numerous research initiatives have been undertaken to find the most energy-efficient structures. D. Monfet and E. Arkhipova introduced a unique method for forecasting the energy requirements of commercial properties using case-based analysis. Their approach may be used to forecast energy consumption and applied to building management systems. (Monfet D, Arkhipova E, 2014). Energy consumption forecasting is essential for effective energy management and planning, according to data scientists. They are outlining a data-driven methodology that makes it possible to estimate energy use in order to make the prediction. The review demonstrates that there are numerous research gaps in the field of energy consumption prediction, including those related to long-term energy consumption, energy consumption in residential buildings, and energy consumption by building illumination. The very tiny amount of information that exists may be the cause of the paucity of research in these areas. (Amasyali K, 2018). The use of computational techniques based on historical data to aid in decision-making for a specific system is known as machine learning. These techniques are typically used to enhance performances or make precise forecasts about the future. Currently, machine learning techniques are being utilized in a variety of sectors to forecast the future using historical data, either as a replacement for traditional statistical and regression models or in conjunction with them. Energy consumption is now predicted using a variety of machine-learning methods, including but not limited to random forests, support vector machines, and neural networks. Numerous papers contend that machine learning algorithms are at the very least competitive with traditional methods. Regression is used to forecast energy consumption, which indicates that the total energy used by buildings supports the necessity for intensive study and analysis in that field. According to the article, statistical approaches are a very practical choice when creating energy models because they simply require access to historical data rather than the multiple values needed to solve engineering equations. Based on data training, Random Forest generates several decision trees. In other words, the training is what determines the outcome for a certain set of inputs. Both category and numeric output data may be used with the technique. The result from random forests is really the average of the values produced by several decision trees when dealing with numerical data (regression). The result for categorical data (classification) is the median value produced by a number of decision trees. World Energy Consumption 8 The input variables are non-linearly mapped toward a multi-dimensional space via SVMs. A linear decision surface is used to generate the output variable, which pulls information from the space and uses it to forecast new output values based on learned information. The data points are drawn against each other on an x-y graph, for instance, in the case of a model with two independent variables and one dependent variable. A line dividing the variables into two sections is the dependent variable. The model makes predictions about the dependent variable based on which section the foreseeable independent data point belongs to. ANN are machine learning models that draw inspiration from how the central nervous system of the human body functions. Human brains function in layers. The way that many relevant experiences have impacted a person's mental process will determine how they perceive a certain object or scenario. Similar to this, an artificial neural network may be created using historical data. In several layers, this neural network can learn from various historical data points. It is simpler to include anomalies in energy usage into the model when using ANNs since they not only create input-output relations but also learn from new outputs. A backpropagation neural network was used to forecast energy usage based on three input variables.The method of backpropagation involves calculating output errors and sending them back into the grid to change the weights. For the purpose of calculating energy demand, insulation, orientation angles, and transparency ratios were used as input variables.

3. Project Description

The amount of electricity needed to run all the electrical equipment in a building is known as the building electricity demand. The ventilation system, the effectiveness of the electrical equipment, and tenant behaviour can all have an impact on the electricity demand. Buildings' heating and cooling loads have to be supplied to or withdrawn from space by the heating, ventilation, and air conditioning (HVAC) system in order to provide a specified level of comfort. In recent decades, several nations throughout the world have proposed various techniques to increase building energy efficiency and estimate power demand, which have been, for example, in Europe, the greatest region in energy consumption. Conversely, the expanding need for producing energy and building new Buildings resulting from the fast increase in global population have been identified as a major source of greenhouse gas emissions. Therefore, high precision energy requirement forecasting in power demand management and energy conservation in the construction industry have gained attention to reducing the usage of fossil fuels and hazardous gasses. Due to the enormous development in the quantity of trustworthy datasets using machine learning (ML) models, the forecasting and optimization of energy usage have therefore been a long-standing concern of many academics.

Planning of energy generating, and transmission systems requires the availability of sufficiently accurate short- and long-term demand estimates. Properly. For the purpose of predicting GDP growth in general, the full real GDP (gross domestic product) can be directly connected to these estimates. From the perspective of societal advancement, electricity is crucial because it is a fundamental human necessity. It was introduced to the market as a tradable good in recent decades, at the same time that the power sector in many nations was liberalized.

3.1. Data Visualization

In simple terms, the process of converting certain information into a visual representation, for instance a map or a graph, in order to make the data easier to comprehend and extract significant results is what is known as Data Visualization. Major purpose for data visualization is to make the pattern recognition, trend recognition and outlier recognition simpler in massive datasets. Data Visualization often goes by different terms such as Information Visualization, Information Graphics, Statistical Graphics.

In Data Science, Data Visualization is one of the many processes according to which the collected, processed, and modeled data needs to be visualized to draw significant conclusions. It is also a part of DPA (Data Presentation Architecture) discipline which strives to effectively and efficiently identify, locate, modify, process, and transmit data. Almost every profession requires data visualization. For instance, professions like teachers, computer scientists, CEOs use data visualization almost everyday to display, research, and share information. Data Visualization plays a vital role when it comes to massive datasets and BigData projects.

For similar reasons, sophisticated analytics completely relies on visualization. When we build a Machine Learning algorithm or advance predictive analytics, it's important to visualize the output to keep track of the conclusions and make sure the models are working as they are meant to be. And the most important part, visually represented algorithms are much easier to understand than any numerical outputs.

3.1.1. Line Graph

A Line graph is needed to get some additional information on the visitors for the entire year. A Line Chart/Graph is very helpful to depict the progression of one or more nominal variables, straight line segments link the data points. The measurement points are organized and linked by straight line segments, kind of similar to scatter plot. Line Chart is very popular when it comes to depict a pattern in data across a time series and so the lines are frequently drawn in a chronological order.

3.1.2. Bar Graph

A Bar graph is a type of graph that uses a series of bars to display data on two axes. The x-axis divides the data into groups, with one bar for each category. It creates bars of different colors for every instance. The values for each category are shown on the y-axis. It has more applications when compared to pie charts, its versatility allows us to exhibit percentages, totals, counts, and a variety of other data. As long as you can categorize the contents of the x-axis in an acceptable fashion, it could be according to time or category. It is typically the preferable choice unless we have an unique need to use a pie chart.

3.1.3. Scatter Plot

A Scatter Plot uses dots to indicate data for different nominal values, the values for each data point are indicated by the positioning of each dot on the x and y axis. It is used to see how variables relate to each other and also to identify patterns. It is often used to identify correlation and clusters in the dataframe. Outliers can also be identified and fixed using a scatter plot.

3.2. Machine Learning

With the use of machine learning (ML), which is a form of artificial intelligence (AI), software programmes may anticipate outcomes more accurately without having to be explicitly instructed to do so. In order to forecast new output values, machine learning algorithms use past data as input. Machine learning is significant because it aids in the creation of new goods and provides businesses with a picture of trends in consumer behavior and operational business patterns. A significant portion of the operations of many of today's top businesses, like Facebook, Google, and Uber, revolve around machine learning. For many businesses, machine learning has emerged as a key competitive differentiation.

3.2.1. Persistence Model

Persistence prediction methods are used to create reference (baseline) models, compare testing, and typically used. In many circumstances, it is advantageous to create a prediction model to determine if it can perform better than a base model. Simple techniques for leveraging historical data include persistence models to anticipated data points. They are created to measure the effectiveness of feature extraction, hyper - parameter tuning, and model design against a collection of references and serve as performance benchmarks for evaluating more advanced methodologies. It may be assumed for a persistence model that the electrical load at time t + 1 is identical to the load at time t. a persistent model in which the load remains constant and equal from one day to the next, in a day-ahead prediction that used a 15-minute time window to establish time occurrences t + 1 and t would probably be unsuccessful. If it can be assumed that the electrical load at time t of day d (shortly (t, d)) would be comparable to the equivalent load at time t on the previous day d 1 or the prior identical day d 7, improved results for a persistence model could be achieved.

Let yd(t) represent the electrical load on a typical home at time instance t on day d. Next, we have a 1-day ahead persistence model.

$$\tilde{y}d\ PM = yd{-}1\ (t).$$

Additionally, the mathematical definition of an n-day ahead persistence model is as follows:

$$\tilde{y}d\ PM = 1\ N \sum yi{-}1\ (t)$$

In other words, it uses an average of the N preceding days' load simultaneously. One can further refine an N-day persistence model by just taking into account the N prior same days. enhanced the aforementioned model since the presence of residents has a strong correlation with power load. within a home. Considering that we're looking for a day-ahead estimate for time t, and that d equals matched a Monday, and the typical load at that hour on the most recent N preceding It is necessary to produce Mondays. We shall call to this as a copy-last-days persistence model.

3.2.2. ARIMA Model

One of the most popular time series models is the ARIMA (autoregressive integrated moving average) model, which incorporates the well-known Box & Jenkins approach, GM (Grey model), and statistical features into the modeling process. When predicting energy usage in other models that can be employed include artificial neural network models and multiple regression models for economies. The ARMA process, which deals with utilizing just the previous values from the time series and the present value, combines a solely autoregressive (AR) and moving average (MA) approach. A random process's previous values and past values to forecast future values. We can apply this approach to stationary data even though most real-world situations do not include stationary data. The method used to handle non-stationary data is referred to as differencing, and it involves replacing a number with its difference from earlier values. The components of an ARIMA model and its mathematical representation are described below.

A weighted linear sum of the previous p values plus white noise can be used to numerically depict an autoregressive process of order p, called AR(p).

$$y_t = \varnothing_1 y_{t-1} + \varnothing_2 y_{t-2} + \cdots \varnothing_p y_{t-p} + z_t \,.$$

Where $\varnothing_1, \varnothing_2, \cdots \varnothing_p$, represent the AR order's coefficients while z_t represent the error term with 0 mean.

- ➔ p order of AR term
- ➔ d required differencing for data stationary
- ➔ q order of MA term

3.2.3. SARIMAX Model

Autoregressive Integrated Moving Average (ARIMA) models, which operate on stationary and linear data, are used to handle non-stationary data. When dealing directly with seasonality in data, the Seasonal ARIMA (SARIMA), a generalized variant of the ARIMA, uses seasonal AR, MA, and terms that differ in the model. The seasonal ARIMA additionally allows for the input of external variables, allowing the user to enter the effects of such factors on the model. The term "exogenous variable" refers to variables that may affect a model but are unaffected by it. The weather falls under this definition.

A SARIMAX (p, d, q) (P,D,Q)s can be mathematically represented as:

- ➔ p order of seasonal AR term
- ➔ d required differencing for data stationary
- ➔ q order of MA term
- ➔ P order of seasonal AR term
- ➔ D required differencing for data stationary
- ➔ Q order of MA term
- ➔ S number of periods

$$y_t = \beta_0 + \beta_1 X_{1.t} + \beta_2 X_{2.t} + \cdots + \beta_k X_{k.t}$$

$$+ \frac{(1-\Theta_1 B^s - \Theta_2 B^{2s} \cdots - \Theta_Q B^{Qs})(1-\Theta_1 B - \Theta_2 B^2 \cdots - \Theta_q B^q)}{(1-\phi_1 B^s - \phi_2 B^{2s} \cdots - \phi_p B^{Ps})(1-\phi_1 B - \phi_2 B^2 \cdots - \phi_p B^p)} Z_t,$$

3.2.4. LSTM Model

A particular kind of RNN called an LSTM neural network, developed in 1997 by Hochreiter and Schmidhuber, may be more suitable for modelling long-range dependencies. In the LSTM architecture, memory blocks are used in place of hidden units. Negative sigmoidal memory block's memory cells are modulated by gates that are applied several times. Memory cells share the same gates in order to reduce the parameters. These gates control whether the model retains the values at the gates or discards them, allowing the network to take use of long-range temporal contexts.

Computing a mapping sequence on the output $y = (y1, y2, \cdots, yT)$,

let $x = (x1, x2, \cdots, xT)$, be the input. Using the following equation, unit activations can be determined:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f)$$

$$c_t = f_t c_{t-1} + i_t\, tanh\,(W_{xc}x_t + W_{hc}h_{t-1} + b_c)$$

$$f_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_0)$$

$$i_t = o_t\, tanh\,(c_t)$$

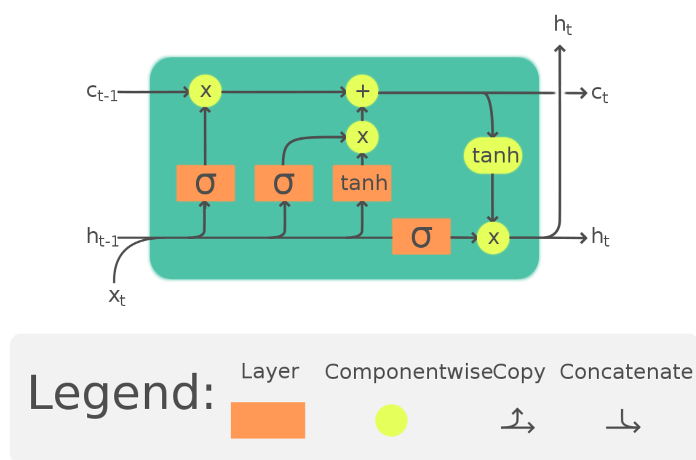

Figure 1: LSTM cell

Where,

- σ represents the logistic sigmoid function.
- f,o,i,c, represents, forget gate, output gate,input gate, and cell activation vector, respectively.
- W represents weight matrices.
- tanh represents output activation function.

4. Data Analysis

The ability of traditional statistical models to predict vs contemporary applications of neural networks for the practical goal of anticipating short-term energy demand.

In this chapter, the forecasting skills of traditional statistical models and cutting-edge neural network implementations were evaluated using a practical job of anticipating short-term energy consumption.

The project's primary inquiry is:

What supervised machine learning problem formulation and forecasting model results in the lowest MAE when computing power is limited?

The following timeseries forecasting methods were used to answer this question:

- ➔ Persistence Model(Naive)
- ➔ ARIMA- Autoregressive Integrated Moving Average
- ➔ SARIMAX- Seasonal Autoregressive Integrated Moving Average with eXogenous Parameters.
- ➔ LSTM- Long Short-Term Neural Memory

4.1. Dataset Preprocessing

The supply and demand equilibrium can be accurately predicted using this data. We propose a test case for modeling electrical demand prediction based on analytical data in recognition of the advancements in the previous research and the significance of household energy usage forecast. This is accompanied by modifications to home consumption patterns which are also impacted by advancements in energy efficiency. Information regarding London's hourly electricity production and weather from 2015 to 2019 was taken from Kaggle.

	day	LCLid	energy_sum
0	2012-10-12	MAC000002	7.098000000000001
1	2012-10-13	MAC000002	11.087000000000002
2	2012-10-14	MAC000002	13.222999999999999
3	2012-10-15	MAC000002	10.257
4	2012-10-16	MAC000002	9.769
5	2012-10-17	MAC000002	10.885000000000002
6	2012-10-18	MAC000002	10.751

Figure 2: Energy Dataset Sample

The dataset includes data on the hourly electrical load as well as projected TSO load and energy pricing data for upcoming data points. We concentrate on forecasting electrical consumption more accurately than the data's existing forecast. Mean Absolute Percentage Error, or MAPE, is the measure we are comparing. To accomplish this goal, the preprocessing dataset's "total load real" and "total load forecast" features were taken out and used as the models' input variables. The method used to create and purify the dataset is described in this section. The "format data" function is used to rename the columns, shorten the text identification for times, and convert to a Datetime index. The "interpolate nans" function is then used to fill in the missing values with a linear interpolation technique.

It's crucial not to alter the data's structure when working with NAN values. Dropping values or altering the amount of observations made each day can cause this. The number of daily observations per day must coincide with the days before and after, or the temporal character of the data will not be accurately represented if the missing values are filled in with the series mean value. Since there were only 36 NAN values in the entire dataset, which had a length of 35064, we utilised a linear interpolation function without altering the distribution's structure.

```
[ ] energy.describe()
```

	energy_sum	LCLid	avg_energy
count	829.000000	829.000000	829.000000
mean	43535.325676	4234.539204	10.491862
std	20550.594031	1789.994799	1.902513
min	90.385000	13.000000	0.211766
25%	34665.436002	4084.000000	8.676955
50%	46641.160997	5138.000000	10.516983
75%	59755.616996	5369.000000	12.000690
max	84156.135002	5541.000000	15.964434

Figure 3: Day-Level Energy Consumption

We needed the right number of values per 24-hour period because we were working with sequence data and needed to check for duplicate timestamps. If not, the data may eventually become skewed and a source of error for the model. In this instance, a day consists of 24 readings and has 24 hours. As a result, we can determine how many data points we should have throughout a specific time period. In particular, the years 2015, 2017, 2018, and 2019 are non-leap years for the five years cleared in this example, although 2016 is, so we must take that into account.

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fa161dc37d0>
```

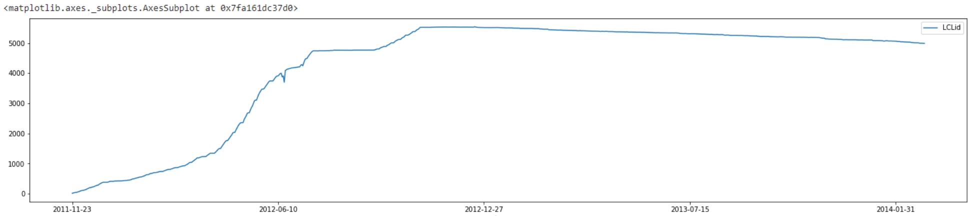

Figure 4: House Count Graph

4.2. Analysis of Energy Load data

By plotting an annual line, a grouped by month line, and a grouped by days of the week-mean line, we assess the energy load data as a whole in this section. Study of the stationary properties of hourly segments by repeating this part with each of the hourly slices.

The goal was to comprehend the overall structure of the energy demand data. We need to consider the following issues:

- ➔ What are the level and noise variances, and are there any seasonal or trending elements in the data?
- ➔ Is the data still? Can we make it stationary if not?
- ➔ What connection exists between the days of the year and the months?
- ➔ What variations exist between the actual load and the ACORN predictions for London?

The present short-term forecasts accurately reflect the actual distribution. This is seen by their different levels and inter - quartile bands (28723 MWh for the amount of projections and 28810 MWh for the amount of demand, respectively). We were also interested in the baseload when it came to meeting energy demand. The baseload is the minimum demand over a period of time (usually weekly). Our global baseload in this instance is the bare minimum over the four years of data (baseload is 18000 MWh).

Following are the Visualized data of different factors of weather affecting energy consumption, factors like: Temperature, Humidity, Cloud Cover, Dew Point:

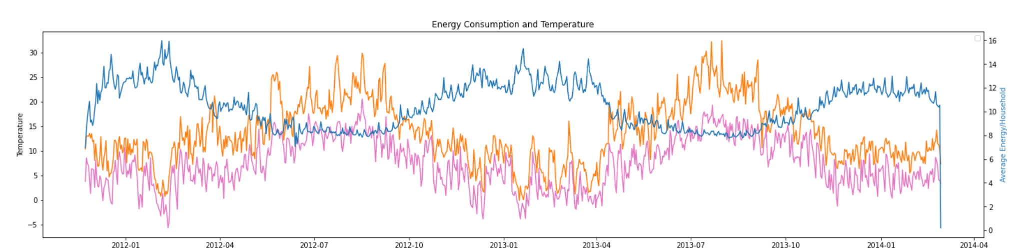

Figure 5: Energy Consumption and Temperature

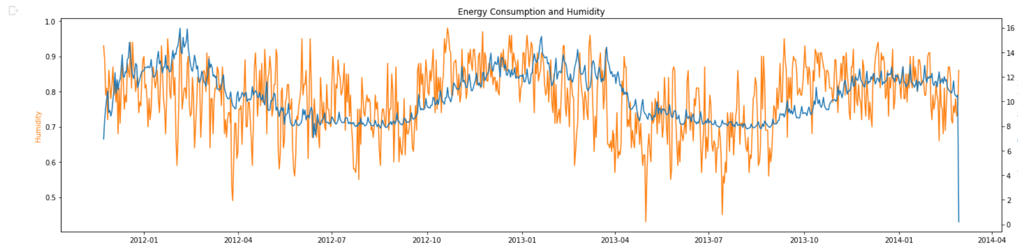

Figure 6: Energy Consumption and Humidity

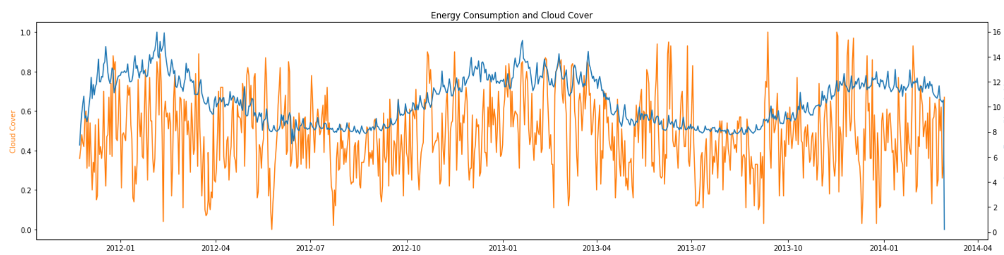

Figure 7: Energy Consumption and Cloud Cover

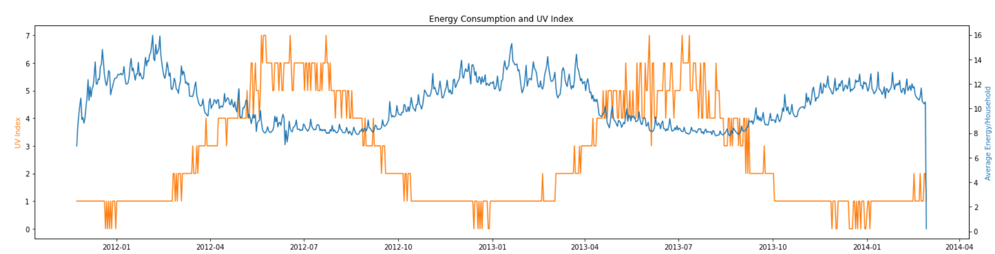

Figure 8: Energy Consumption and Dew point

4.3. Machine Learning Models

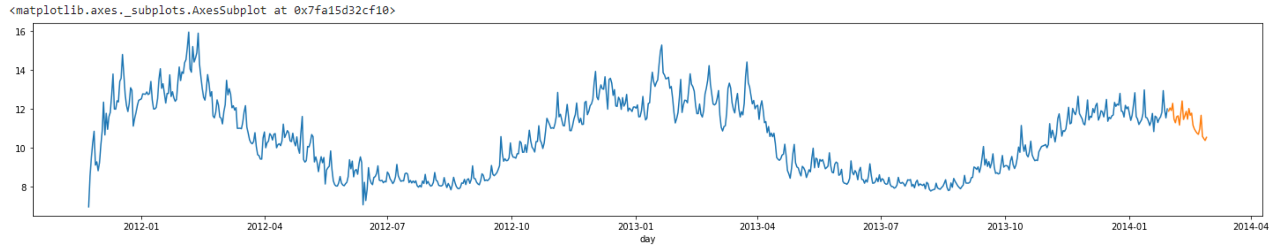

Figure 9: Dataframe split 70-30

4.3.1. Persistence Models

Simple methods called persistence models (also known as naive models) can predict future data points by referencing historical data. They are created to measure the performance of feature engineering, hyperparameter tuning, and model design against a collection of references and serve as performance benchmarks when evaluating more complicated methods. Data is available for the period 2015-01-01 through 2019-08-25. For ease of use, we fixed the sample from 2015-01-01 to 2018-12-31, a span of precisely 4 years, for training and assessing models.

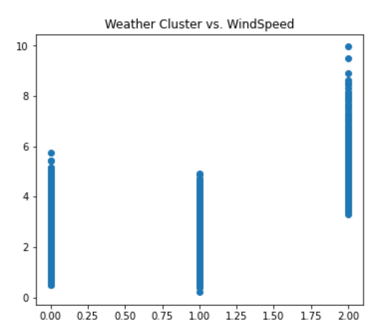

Figure 10:Cluster vs WindSpeed

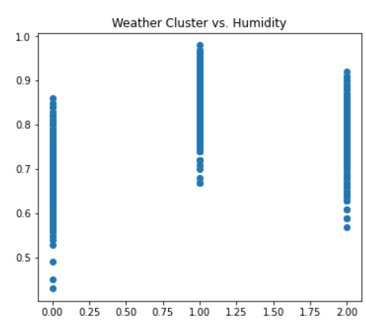

Figure 11:Cluster vs Humidity

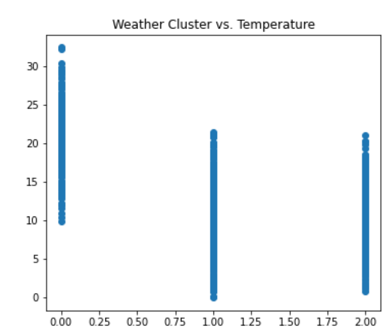

Figure 12: Cluster vs Temperature

	avg_energy	temperatureMax	dewPoint	cloudCover	windSpeed	pressure	visibility	humidity	uvIndex	moonPhase
avg_energy	1.000000	-0.846965	-0.755901	0.241779	0.149624	-0.028851	-0.246404	0.361237	-0.733171	-0.031716
temperatureMax	-0.846965	1.000000	0.865038	-0.333409	-0.153602	0.118933	0.259108	-0.404899	0.696497	0.003636
dewPoint	-0.755901	0.865038	1.000000	-0.025207	-0.092212	-0.028121	0.042633	0.055514	0.486692	-0.008239
cloudCover	0.241779	-0.333409	-0.025207	1.000000	0.170235	-0.101079	-0.330177	0.480056	-0.248695	-0.062126
windSpeed	0.149624	-0.153602	-0.092212	0.170235	1.000000	-0.344354	0.281088	-0.042391	-0.152634	-0.023273
pressure	-0.028851	0.118933	-0.028121	-0.101079	-0.344354	1.000000	-0.012508	-0.250941	0.100774	0.038462
visibility	-0.246404	0.259108	0.042633	-0.330177	0.281088	-0.012508	1.000000	-0.578130	0.240485	0.062813
humidity	0.361237	-0.404899	0.055514	0.480056	-0.042391	-0.250941	-0.578130	1.000000	-0.533919	-0.013997
uvIndex	-0.733171	0.696497	0.486692	-0.248695	-0.152634	0.100774	0.240485	-0.533919	1.000000	0.012833
moonPhase	-0.031716	0.003636	-0.008239	-0.062126	-0.023273	0.038462	0.062813	-0.013997	0.012833	1.000000

Figure 13: Correlation Data

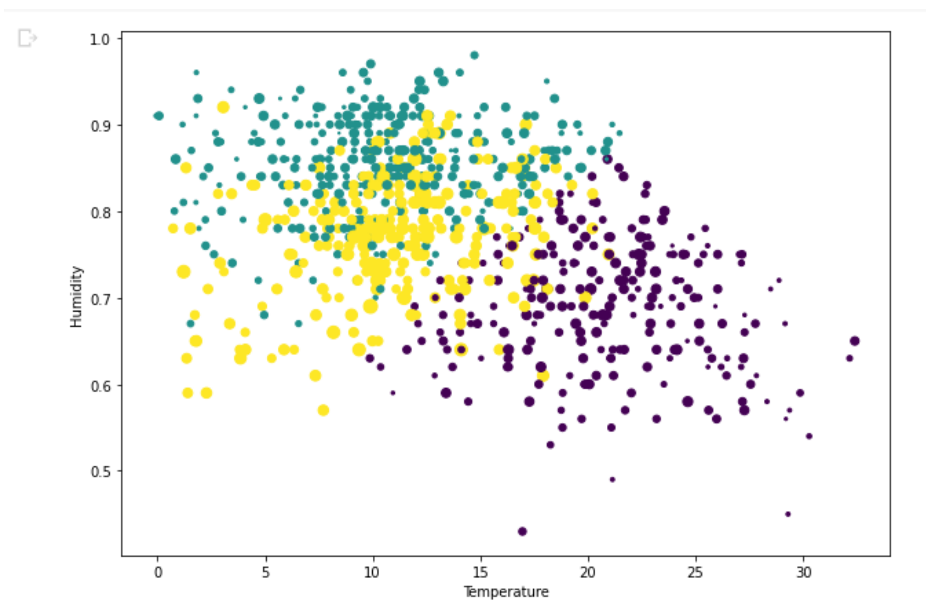

Figure 14: Scatter Plot Temperature vs Humidity

To directly compare models to energy values in the data, root mean squared error, or RMSE, is used in the evaluation process. RMSE has two methods for calculation. The first to illustrate the flaw in hourly prediction (i.e. one error per-hourly slice). Second, to depict the overall effectiveness of the model (one value). Walk-forward forecasting is used to create predictions. Walk forward generates predictions by iteratively going over the samples and making predictions along the way. The test value is reused and added to the end of the training set after a forecast has been formed.

4.3.2. ARIMA Model

Autoregressive Integrated Moving Average is referred to as ARIMA. In contrast to the model above, it predicts t using a linear mix of prior time steps and moving averages. Only stationary time series are used in ARIMA. The load data can be made stationary to produce analysis findings, as was discussed in the section on data analysis.

We used the "statsmodels.api" ARIMA model, which accepts the following arguments:

- ➔ p denotes lag order.
- ➔ d denotes degree of differencing.
- ➔ q denotes order of moving average.

	avg_energy	weather_cluster	holiday_ind
day			
2014-01-30	11.886982	1	0

Figure 15: Head of new Dataframe

Above figure demonstrates that the test's p-value is significantly below the cutoff of 0.05, and as a result, the null hypothesis is rejected and a stationary dataset is assumed. As a result, the differencing value's default model parameter is 0. We looked into the difference between a parameter of 24 and 168, which correspond to the day before and the week before, respectively.

4.3.2.1. Dickey Fuller's Test:

Test Statistics	-1.87279
p-Value	0.344966
Lags Used	21.00000
No. of Observations Used	776.00000
Data type	float64

Table 1: Before Differencing

Test Statistics	-6.715004e+00
p-Value	3.600554e-09
Lags Used	2.000000e+01
No. of Observations Used	7.760000e+02
Data type	float64

Table 2: After Differencing

4.3.3. Auto-correlation and Partial Auto-correlation

ACF- Explains the connections between lagging (shifted) autoregressive features that are both direct and indirect. Taking into account the interrelationships between characteristics, in this case, t 1, t 2, t 3, etc., in the linkages between t, t 1, t 2, t 3, etc.

PACF- Exclusively discusses the connections between lagging (shifted) and autoregressive aspects that are direct.

p (AR): The number of lags beyond which there is no discernible association is the best way to define the autoregressive hyperparameter value p. The ACF interprets this as the location where plot values fall outside the significance band (light blue horizontal band).

q (MR): The relationship between the lag feature and the feature is stated as the direct relationship between the moving average hyperparameter value q.

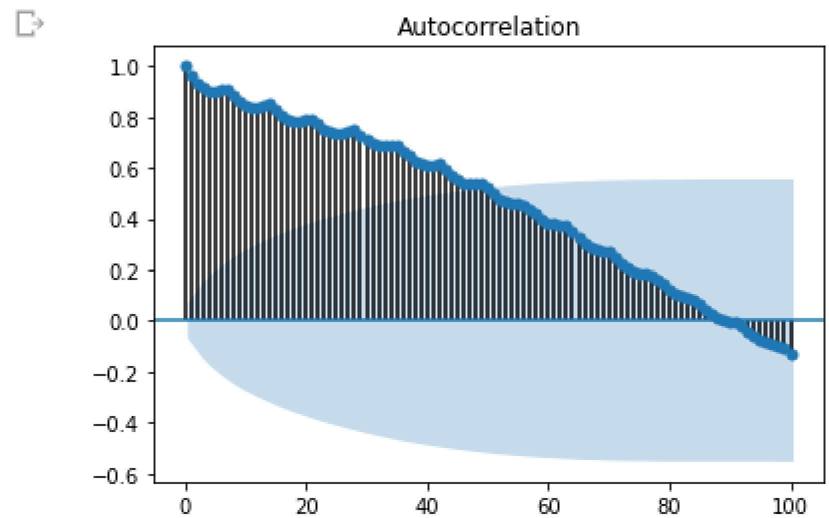

Figure 16: Autocorrelation

The autocorrelation plot seen in the following image reveals strong positive correlations in the initial nine lag. However, we are attempting to anticipate the following 24 hours in this scenario and employing a demand prediction for the following 24 hours cannot be fully understood with a lag of less than 24 hours.

The charts clearly show the autoregressive features' cyclical behavior. The ARIMA model uses The presumption that there is no association past the selected lag point (p). Looking at the bottom right plot, we can observe that this moment happens at roughly 4000 delays.

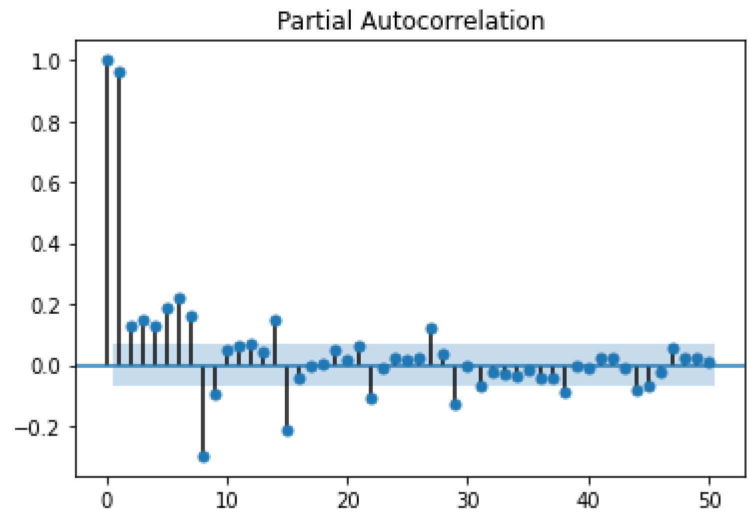

Figure 17: Partial Autocorrelation

Looking at picture 4-18 of the partial autocorrelation plot demonstrates that there is no discernible partial autocorrelation after 24 delays. In light of this, we will look into the moving average values for delays 2, 3, and 24.

4.3.4. SARIMAX Model

The SARIMAX model has a wider calculable state space due to its complexity and utilisation of mode characteristics. The documentation states that while the SARIMA model has more recent implementations, the ARIMA model is kept to a minimum. The SAIRMAX algorithm may or may not be quicker. This section compares the results of a test against an ARIMA.

The SARIMAX model may have a quicker implementation, as was already suggested. Functionally, the model also provides a second layer of seasonality-related hyperparameters, P/D/Q/m.

- ➔ P denotes seasonal autoregressive order.
- ➔ D denotes seasonal difference.
- ➔ Q denotes moving average.
- ➔ m denotes the number of time steps for a single seasonal period.

We may rephrase the forecasting issue so that each m periods is a season thanks to the new attributes. We may define P, D, and Q as functions of the season inside the season. Based on our understanding of the issue, we picked the seasonal parameters.

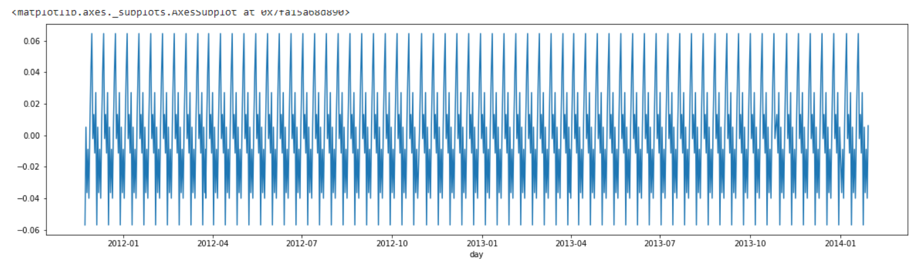

Figure 18: Seasonal Decomposition

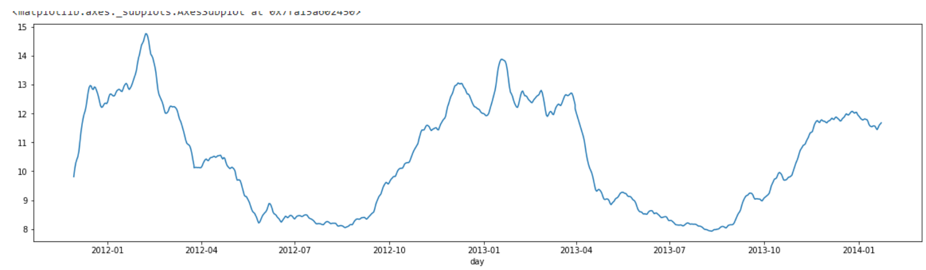

Figure 19: Decomposition Trend

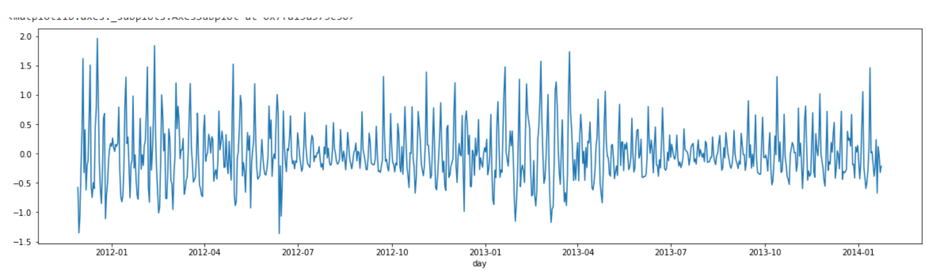

Figure 20: Decomposition Residual trend

Dependent Variable	avg_energy	No. of Observations	798
Model	SARIMAX(7, 1, 1)x(1, 1, [], 12)	Log Likelihood	-649.345
Date	Mon, 12 Sep 2022	AIC	1324.689
Time	13:32:07	BIC	1385.343
Sample	0	HQIC	1348.011

Table 3: SARIMAX Results

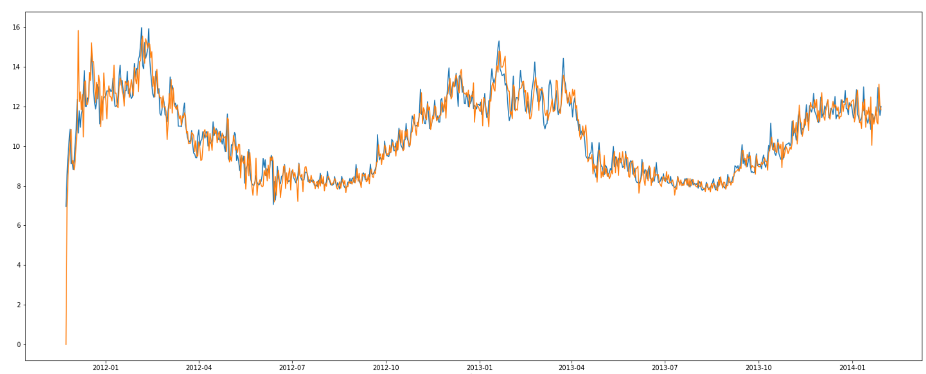

Figure 21: Model Fit Visualized

Days	avg_energy	weather_cluster	holiday_ind	predicted
2014-02-23	11.673756	2	0	11.563777
2014-02-24	10.586235	2	0	10.723157
2014-02-25	10.476498	2	0	11.463700
2014-02-26	10.375366	2	0	11.880441
2014-02-27	10.537250	2	0	11.498228

Table 4: Prediction using SARIMAX

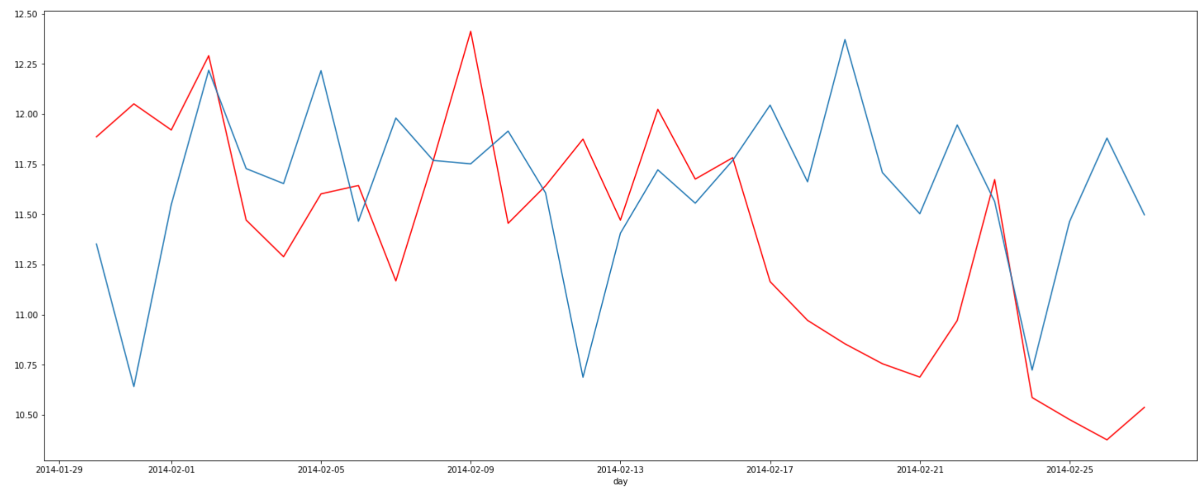

Figure 22: Predicted energy visualization

4.3.5. LSTM Model

In this part, we implement an LSTM model as a straightforward univariate strategy using "Keras," a Python-based deep learning API that runs on top of the TensorFlow machine learning framework. The input and output of the model are arranged as a univariate sequence for each hour of the day.

The prediction algorithm predicts the demand for the upcoming 24 hours. Each hour of the day is represented by 24 projections made by the model. The following are some advantages of the approach:

- ➔ In contrast to the autocorrelation between h0, h1, h2, h3, etc., we take use of greater direct (partial) autocorrelations between h0, h1,... h23 of today, the day previous, and so on.
- ➔ Stronger (partial) autocorrelations between hours of today, yesterday, and so forth are taken into consideration as opposed to the autocorrelation between hours of days.
- ➔ We can capture the impact of seasonality by training the algorithm on fewer samples.

The following figure illustrates the data structure for the univariate scenario. In this situation, we are making predictions based on historical data from the same hour by hour. In this way, every hour develops into its own dataset. These many sets can be combined into a single data block with the following shape:

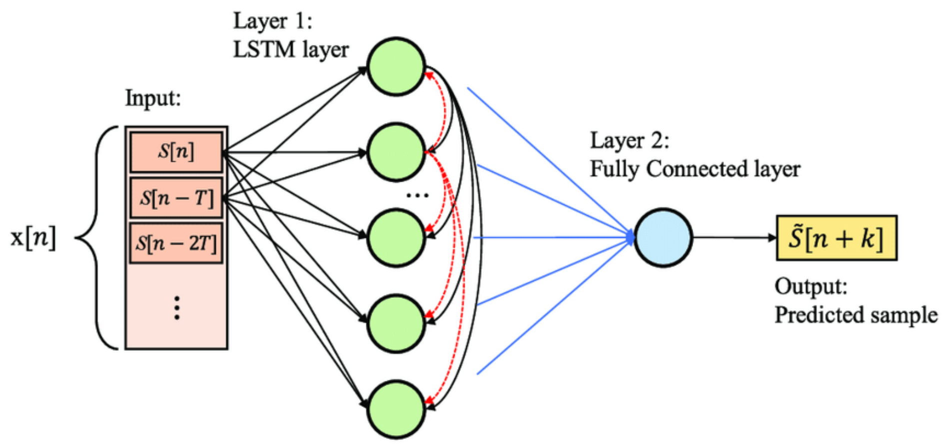

Figure 23: Data Structure LSTM problem

	var1(t-7)	var1(t-6)	var1(t-5)	var1(t-4)	var1(t-3)	var1(t-2)	var1(t-1)	var1(t)
7	6.952693	8.536480	9.499782	10.267707	10.850805	9.103382	9.274873	8.813513
8	8.536480	9.499782	10.267707	10.850805	9.103382	9.274873	8.813513	9.227707
9	9.499782	10.267707	10.850805	9.103382	9.274873	8.813513	9.227707	10.145910

Figure 24: Series to supervised visualization

4.3.6. Normalization

We normalized the values and made paired windows of X representing rows of historical data and Y representing the electricity load on the target day in order to construct the LSTM model. The model is trained on a little quantity of data from 2015 to 2018 in order to construct a cross-validation testbench. Data translation into the [0, 1] (or any other range) or simple data transformation onto the unit sphere are both examples of normalization in machine learning. Normalization and standardization are helpful for some machine learning methods.

4.3.7. Models Evaluation

After the model had been trained using the training set of data, it had to be tested using the test set in order to determine the model's correctness. Timeseries-specific cross-validation would be helpful for assessing the generalizability of the model to yet-to-be-observed data during the manufacturing phase. Additionally, there may be a number of domain-specific features or some fundamental features that have a significant impact on the performance of the model. A few examples of these features include hourly weather conditions, holidays and special days, information about energy-intensive factories, and data on the sun's rise and set.

For each hourly forecast, the MAE is calculated to assess the model's performance. To compare model runs, we may calculate the entire model MAPE. We calculate the mean of the total number of mistakes for each hour.

```
Epoch 1/50
11/11 - 3s - loss: 0.6446 - 3s/epoch - 269ms/step
Epoch 2/50
11/11 - 0s - loss: 0.5263 - 35ms/epoch - 3ms/step
Epoch 3/50
11/11 - 0s - loss: 0.3993 - 43ms/epoch - 4ms/step
Epoch 4/50
11/11 - 0s - loss: 0.2553 - 39ms/epoch - 4ms/step
Epoch 5/50
11/11 - 0s - loss: 0.1253 - 42ms/epoch - 4ms/step
Epoch 6/50
11/11 - 0s - loss: 0.1184 - 34ms/epoch - 3ms/step
Epoch 7/50
11/11 - 0s - loss: 0.1069 - 34ms/epoch - 3ms/step
Epoch 8/50
11/11 - 0s - loss: 0.0962 - 36ms/epoch - 3ms/step
Epoch 9/50
11/11 - 0s - loss: 0.0883 - 36ms/epoch - 3ms/step
Epoch 10/50
11/11 - 0s - loss: 0.0813 - 35ms/epoch - 3ms/step
Epoch 11/50
11/11 - 0s - loss: 0.0748 - 43ms/epoch - 4ms/step
Epoch 12/50
11/11 - 0s - loss: 0.0679 - 37ms/epoch - 3ms/step
Epoch 13/50
11/11 - 0s - loss: 0.0611 - 36ms/epoch - 3ms/step
Epoch 14/50
11/11 - 0s - loss: 0.0545 - 35ms/epoch - 3ms/step
Epoch 15/50
11/11 - 0s - loss: 0.0481 - 34ms/epoch - 3ms/step
```

Figure 25: Iteration of 50 Epochs

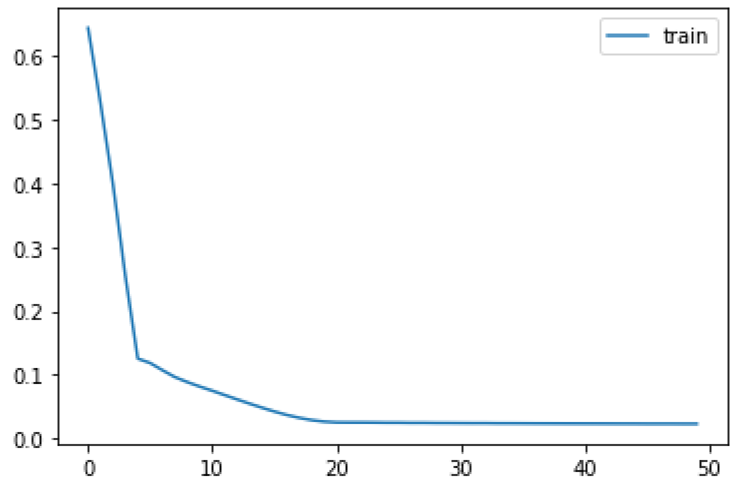

Figure 26: Training dataset trend

5. Conclusion

In this research, future load electricity demands are predicted using the hourly electrical load use. As a result, it's possible that conventional methods can't provide an accurate forecast of future values. In the energy dataset for London, the values for the hourly electrical load are given for the period from 01/01/2015 to 31/12/2018. We outlined the significance of demand forecasting and related literature in chapter 1. We began with exploratory data analysis to investigate the features of the dataset, and the second half of chapter 4 contains descriptive data. In order to better understand the trend and seasonality functions, we replaced null values with mean values during the data cleaning process, extracted redundant features, and aggregated hourly load values daily level. Four machine learning techniques are applied to the dataset in section 4 of this chapter using the Python programming language.

The Persistence models (naive models), which can be thought of as rudimentary strategies, were designed to forecast future data points using past data. The first three years (2015–2017) served as the normative training set, with the fourth and final year (2018) serving as the testing set. The issue of estimating the next day's 24-hour slices of anticipated energy demand was solved using the three persistence models, namely the prior hour-by-hour persistence, moving average (3 days) persistence, and same-day previous day persistence.

The first three hours' worth of results from our persistence models are provided. We can see that, on average, the previous-year model performed better at predicting the power consumption for the next day. This could be a sign of recurring seasonal patterns in the data. The early morning period from six to ten in the morning is difficult to anticipate using any approach. As the day goes on, the forecast grows better and better.

The second and third techniques are termed SARIMAX and ARIMA, respectively. SARIMAX is a family of time series models that automatically handles seasonality in data, and ARIMA is a typical time-series modelling methodology. A streamlined version of the walkforward validation set was put into practice.

Model	Test Set	RSME	Computation Time
ARIMA	Condensed	4125.87	21:45
SARIMAX	Condensed	2562.95	02:05

Table 5: Outcomes of ARIMA & SARIMAX model

Even though SARIMAX might not have been quicker, the forecast was noticeably superior in terms of the RMSE on the first walk forward validation step and how the forecast appeared on the plot.

With the last technique, we created a straightforward "Keras" univariate LSTM model. In this scenario, each hour turns into its own dataset in order to make hour-by-hour predictions utilising historical data from the same hour. These several sets were consolidated into a single data block. On some hours and some weeks, the model performs noticeably well.

With the final method, we produced a simple "Keras" univariate LSTM model. To produce hour-by-hour forecasts using past data from the same hour, in this situation, each hour becomes its own dataset. One data block was created from these many sets. The model operates remarkably well on some hours and on some weeks.

6. Reference

Bhattacharyya, S., Timilsina G., 2009. Energy Demand Models for Policy Formulation: A Comparative Study of Energy Demand Models World Bank Policy Research Working Paper 4866. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1368072

Cheong SK, 2011. Scheduling smart home appliances using mixed integer linear programming. In: 50th IEEE conference on decision and control and European control conference, IEEE. Available at: https://doi.org/10.1109/cdc.2011.6161081

D. Monfet, E. Arkhipova, 2014. Development of an energy prediction tool for commercial buildings using case-based reasoning, Energy Build. Available at: https://www.sciencedirect.com/science/article/pii/S037877881400499X

Amasyali K, 2018. A review of data-driven building energy consumption prediction studies, Renew Sustain Energy Rev. Available at: https://www.sciencedirect.com/science/article/pii/S1364032117306093?via%3Dihub

7. Bibliography

Ahmad, M. W., Mourshed, M., & Rezgui, Y. (2017). Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. Energy and Buildings, 147, 77–89. https://doi.org/10.1016/j.enbuild.2017.04.038

Amasyali, K., & El-Gohary, N. (2019). Predicting Energy Consumption of Office Buildings: A Hybrid Machine Learning-Based Approach. In Advances in Informatics and Computing in Civil and Construction Engineering (pp. 695–700). Springer International Publishing. https://doi.org/10.1007/978-3-030-00220-6_83

Chatfield, C., Time-series Forecasting. Chapman & Hall/CRC, 2001

Citroen, N., Ouassaid, M., & Maaroufi, M. (2015). Long term electricity demand forecasting using autoregressive integrated moving average model: Case study of Morocco. In Proceedings of 2015 International Conference on Electrical and Information Technologies, ICEIT 2015 (pp. 59–64). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/EITech.2015.7162950

Deng, H., Fannon, D., & Eckelman, M. J. (2018). Predictive modeling for US commercial building energy use: A comparison of existing statistical and machine learning algorithms using CBECS microdata. Energy and Buildings, 163, 34–43. https://doi.org/10.1016/j.enbuild.2017.12.031

Divina, F., Torres, M. G., Vela, F. A. G., & Noguera, J. L. V. (2019). A comparative study of time series forecasting methods for short term electric energy consumption prediction in smart buildings. Energies, 12(10). https://doi.org/10.3390/en12101934

Foucquier, A., Robert, S., Suard, F., Stéphan, L., & Jay, A. (2013). State of the art in building modelling and energy performances prediction: A review. Renewable and Sustainable Energy Reviews. https://doi.org/10.1016/j.rser.2013.03.004

Friedrich, L., Armstrong, P., & Afshari, A. (2014). Mid-term forecasting of urban electricity load to isolate air-conditioning impact. Energy and Buildings, 80, 72–80. https://doi.org/10.1016/j.enbuild.2014.05.011

Hahn, H., Meyer-Nieberg, S., & Pickl, S. (2009). Electric load forecasting methods: Tools for decision making. European Journal of Operational Research, 199(3), 902–907. https://doi.org/10.1016/j.ejor.2009.01.062

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hor, C. L., Watson, S. J., & Majithia, S. (2006). Daily load forecasting and maximum demand estimation using ARIMA and GARCH. In 2006 9th International Conference on Probabilistic Methods Applied to Power Systems, PMAPS. https://doi.org/10.1109/PMAPS.2006.360237

Jota, P. R. S., Silva, V. R. B., & Jota, F. G. (2011). Building load management using cluster and statistical analyses. International Journal of Electrical Power and Energy Systems, 33(8), 1498– 1505. https://doi.org/10.1016/j.ijepes.2011.06.034

Kyriakides, E., & Polycarpou, M. (2006). Short term electric load forecasting: A tutorial. Studies in Computational Intelligence, 35, 391–418. https://doi.org/10.1007/978-3-540-36122-0_16

Li, X., Bowers, C. P., & Schnier, T. (2010). Classification of energy consumption in buildings with outlier detection. IEEE Transactions on Industrial Electronics, 57(11), 3639–3644. https://doi.org/10.1109/TIE.2009.2027926

Li, X., & Wen, J. (2014). Review of building energy modeling for control and operation. Renewable and Sustainable Energy Reviews. Elsevier Ltd. https://doi.org/10.1016/j.rser.2014.05.056

Lusis, P., Khalilpour, K. R., Andrew, L., & Liebman, A. (2017). Short-term residential load forecasting: Impact of calendar effects and forecast granularity. Applied Energy, 205, 654–669. https://doi.org/10.1016/j.apenergy.2017.07.114

Notton, G. and Voyant, C. (2018). Forecasting of Intermit-tent Solar Energy Resource. In Advances in Renewable Energies and Power Technologies, 77-114.

Papaioannou, G., Dikaiakos, C., Dramountanis, A., & Papaioannou, P. (2016). Analysis and Modeling for Short- to Medium-Term Load Forecasting Using a Hybrid Manifold Learning Principal Component Model and Comparison with Classical Statistical Models (SARIMAX, Exponential Smoothing) and Artificial Intelligence Models (ANN, SVM): The Case of Greek Electricity Market. Energies, 9(8), 635. https://doi.org/10.3390/en9080635

Rodrigues, F., Cardeira, C., & Calado, J. M. F. (2014). The daily and hourly energy consumption and load forecasting using artificial neural network method: A case study using a set of 93 households in Portugal. In Energy Procedia (Vol. 62, pp. 220–229). Elsevier Ltd.https://doi.org/10.1016/j.egypro.2014.12.383

Taşpinar, F., Çelebi, N., & Tutkun, N. (2013). Forecasting of daily natural gas consumption on regional basis in Turkey using various computational methods. Energy and Buildings, 56, 23–31. https://doi.org/10.1016/j.enbuild.2012.10.023

Wang, Z., Wang, Y., Zeng, R., Srinivasan, R. S., & Ahrentzen, S. (2018). Random Forest based hourly building energy prediction. Energy and Buildings, 171, 11–25. https://doi.org/10.1016/j.enbuild.2018.04.008

Zhang, F., Deb, C., Lee, S. E., Yang, J., & Shah, K. W. (2016). Time series forecasting for building energy consumption using weighted Support Vector Regression with differential evolution optimization technique. Energy and Buildings, 126, 94–103. https://doi.org/10.1016/j.enbuild.2016.05.028

Zhao, H. X., & Magoulès, F. (2012, August). A review on the prediction of building energy consumption. Renewable and Sustainable Energy Reviews. https://doi.org/10.1016/j.rser.2012.02.049

Zhang, P. G. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159–175. https://doi.org/10.1016/S0925-2312(01)00702-0