AI-BASED RESIDENTIAL LOAD FORECASTING

Niki Sadat Abbasian

Mina Farmanbar

A Preliminary Master's Thesis
DEPARTMENT OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE
University of Stavanger

June 19, 2022

Scientific environment

This study is carried out at the Department of Electrical Engineering and Computer Science, University of Stavanger.

Acknowledgements

I would like to express my gratitude to my primary supervisor Mrs Mina Farmanbar for excellent support and guidance to carry out this thesis in a best way. I wish to extend my thanks to Mrs Aida Mehdipour, my co-supervisor for clarifications and support throughout this study and research. I wish to acknowledge the help and support provided by the technical and support staff in the department of electrical engineering University of Stavanger during this master study. At the end I would like show my deep appreciation to my beloved family and friends specially my husband and my to beloved kids - Arman and Armita- for their support, patience and motivation to accomplish this project.

Niki Abbasian
Norway, June 2022

Abstract

The increasing levels of energy consumption worldwide is raising issues with respect to surpassing supply limits, causing severe effects on the environment, and the exhaustion of energy resources. Buildings are one of the most relevant sectors in terms of energy consumption in the world. Many researches have been carried out in the recent years with primary concentration on efficient Home or Building Management Systems. In addition, by increasing renewable energy penetration, modern power grids demand more accurate consumption predictions to provide the optimized power supply which is stochastic in nature. This study will present an analytic comparison of day-ahead load forecasting during a period of two years by applying AI based data driven models. The unit of analysis in this thesis project is based on households smart meter data in England. The collected and collated data for this study includes historical electricity consumption of 75 houses over two years of 2012 to 2014 city of London. Predictive models divided in two main forecasting groups of deterministic and probabilistic forecasting. In deterministic step, Random Forest Regression and MLP Regression employed to make a forecasting models. In the probabilistic phase,DeepAR, FFNN and Gaussian Process Estimator were employed to predict days ahead load forecasting. The models are trained based on subset of various groups of customers with registered diversified load volatility level. Daily weather data are also added as new feature in this study into subset to check model sensitivity to external factors and validate the performance of the model. The results of implemented models are evaluated by well-known error metrics as RMSE,MAE, MSE and CRPS separately for each phase of this study. The findings of this master thesis study shows that the Deep Learning methods of FNN, DeepAR and MLP compared to other utilized methods like Random Forest and Gaussian provide better data prediction reslts in terms of less deviance to real load trend, lower forecasting error and computation time. Considering probabilistic forecasting methods it is observed that DeepAR can provide better results than FFNN and Gaussian Process model. Although the computation time of FFNN was lower than others

Contents

1 Introduction **1**

1.1 Problem background 2
1.2 Motivation 5
1.3 Problem Statement and Research Questions 5
1.4 scope 6
1.5 Structure of the thesis 7
1.6 modus operandi 8

2 literature review **9**

3 Methodology **19**

3.1 data prepossessing 20
3.1.1 data integration 20
3.1.2 feature selection 21
3.1.3 data cleaning 21
3.1.4 data normalization 22
3.2 Methods and Techniques 22
3.2.1 MLP Regression 22
3.2.2 Random Forest Regression 23
3.2.3 DeepAR Estimator 24
3.2.4 Gaussian Process 25
3.2.5 Feed Forward Estimator 25
3.3 Model Evaluation Metrics 26
3.3.1 MAE 27
3.3.2 RMSE 27
3.3.3 CRPS 27

4 Dataset and Statistical Analysis **29**

4.1 Data sets description 29
4.2 data cleaning and subset selection 30
4.3 data normalization 38

5 Results and Discussion **39**

5.1 deterministic forecasting 40
5.2 probabilistic forecasting 43
5.3 Discussion 44

6 Conclusions and Future Work **49**
6.1 Conclusion 49
6.2 Future Work 50

7 REFERENCES

Chapter 1

Introduction

The increasing levels of energy consumption worldwide is raising issues with respect to surpassing supply limits, causing severe effects on the environment, and the exhaustion of energy resources. Buildings are one of the most relevant sectors in terms of energy consumption which are among the largest energy consumers in the world; as such, efficient Home or Building Management Systems are an important topic of researches these days. As new technologies have been developed, great advances have been made in buildings, turning conventional buildings into smart buildings. These smart buildings have allowed for greater supervision and control of the energy resources within the buildings, taking steps to energy management strategies to achieve significant energy savings. The forecast of energy consumption in buildings has been a very important element in these energy strategies since it allows adjusting the operation of buildings so that energy can be used more efficiently[1].

The considerable amount of energy consumption associated to the residential sector justifies and supports energy consumption modeling efforts. Moreover modern power grids face the challenge of increasing renewable energy penetration that is stochastic in nature and calls for accurate demand predictions to provide the optimized power supply. Hence, increasing the self-consumption of renewable energy through demand response in households, local communities, and micro-grids is essential and calls for high demand prediction performance at lower levels of demand aggregations to achieve optimal performance[2]. Load forecasting is a pivotal part of the power utility companies. With the evolution of smart grids in recent years, load forecasting has received more research focus than ever before[2]. To provide load-shedding free and uninterrupted power to the consumer, decision-makers in the utility sector must forecast the future demand of electricity with the least amount of error percentage. Load prediction with less percentage of error can save millions of dollars to the utility companies. There are many techniques to amicably forecast the demand of electricity. Different forecasting methods are studied in nonresidential and residential buildings for building energy consumption forecasting. Energy consumption forecasting is a critical and necessary input to planning and controlling energy usage in the building sector which accounts for 40 percent of the worlds energy use and the worlds greatest fraction of greenhouse gas emissions. However, due to the diversity and complexity of buildings as well as the random nature of weather conditions, energy consumption and loads are stochastic and difficult to predict[3].

1.1 Problem background

Smart meters are the well-known devices which have been widely utilized over the last decades. Smart meters collect energy consumption to help costumers to obtain overview on daily electricity consumption. For instance US and UK have almost 86 million smart meters installed in small and large grades, where residential consumers have taken 90 percent of installed smart meters.The device not only has economical impact but also leads to people think more green by increasing people awareness about energy consumption level. Smart meters provide costumers and suppliers with information about cost and consumption rate in real time. The device helps to make comprehensive overview on consumption habits which not only leads to deliver more accurate bills by suppliers but also caters valuable time-based pricing rate and daily consumption pattern leading not to use appliances in the peak hours for costumers. Moreover smart meter generate high quality and resolution data which can be employed by suppliers to have full control on power quality monitoring and power loss identification and other functions. Recorded data by smart meter can be deployed to predict customers future consumption by trading habits and consumption patterns which bring up the load forecasting concepts and importance [4][5]. Load forecasting refers to anticipate power requirements by costumers in the future. Load forecasting also define as a technique which is used by power supplier companies to predict the power or energy needed to balance the supply and load demand at all the times[5]. Since storing the produced power by utilities is not always possible, its vital for suppliers to estimate demand loads, this estimation or forecasting,helps power grid suppliers to moderate demand and supply in order to make the power grid more reliable and accurate and also avoid any electricity shortage in the future[52]. Generally, forecasting results can be classified into two main groups, deterministic and probabilistic forecasting. In deterministic forecasting, the result of forecast would be a certain value per each time horizon or input.In contrast, in the probabilistic forecasting, a set of probabilities related to all possible outputs will be produced instead of pinpointing one particular outcome as forecast.In probabilistic forecasting, all the outputs per a valid input are possible with different probabilities[6].

There are three categories based on different targets for load forecasting including short-term load forecasting, medium-term load forecasting and long-term load forecasting which are applied in wide range of time horizon.

- Short term load forecasting: this type of forecasting predicts a few minutes to 1 day ahead consumption, mostly focusing on adjusting supply electricity to requirements and can appropriate information for the system management of daily operations and unit commitment[1].

- Medium load forecasting: in this method, 1 day to 1 year can be estimated and its significant to apply it on different time horizon for different operations of supplier company. Medium term forecasting is utilized for the purpose of scheduling electricity supplies and unit management[1].

- Long term load forecasting: this type of forecasting refers to forecasting time periods longer than 1 year. The method is mainly used for long term planing in utility companies such as power infrastructure expansions in hiring new employees or having new investments[1].

Load forecasting can be employed in different scales form a building or community to a city or a country. There are many factors which affect load forecasting with various level of impact in different domain. For instance, weather condition can affect electricity consumption in residential more than industry.There are many parameters like weather parameters, size of the building, heating and cooling system, insulation, which should be taken into consideration when load forecasting is the problem at hand[2].

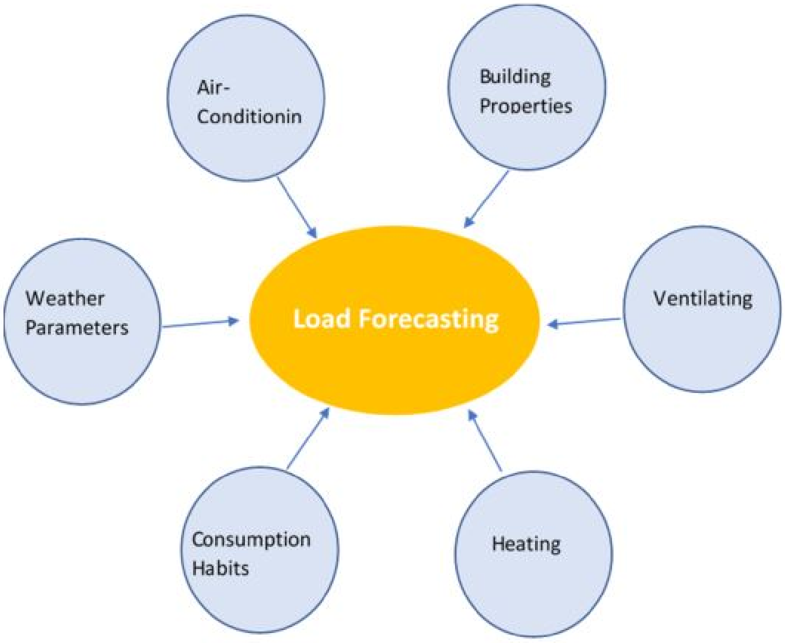


Figure 1.1: influential parameters on power consumption

The figure illustrates different parameters which influence power consumption in a building. Studies show that load forecasting for building is most challenging as oscillating factors such as weather parameters and building properties affect the consumption drastically. Since the operation of power system is relying on precise and appropriate forecasting, many researches and studies are conducted based on smart meter data[2]. Overlay, the building energy consumption approaches have been divided in two main categories of engineering and data-driven methods, while former focusing on mathematical equation to present the physical performance of building, the latter uses widely real-time and historical consumption data to predict energy consummation[3]. The engineering methods demand high level of computation capacity, expertise and tremendous amount of details which is not always available. While data-driven methods are able to build relatively precise forecasting model based on the available consumption data. Hence data-driven methods are becoming widely popular in the recent years. Data-driven approaches are grouped into statistical and AI-based techniques. Since, statistical methods utilize historical data for its analysis is really vital to apply statistical model on high amount and qualified data to achieve most accurate and effective results[6].The statistical methods such as Traditional linear, Gaussian mixture models (GMM), Conditional demand analysis (CDA), Regression models and auto-regressive moving average (ARMA) and ARIMA use time series for prediction in many applications.To overcome linearity in time series models a main limitation, Support Vector Machines (SVM), Classification and Regression

Trees (CART) have been introduced to be deployed in forecasting applications. On the other side AI- based techniques such as Artificial Neural Networks(ANNs) have been broadly employed in prediction and forecasting purposes. Similar to statistical models, ANNS models are built up on historical data. Although, ANN based models are considered as complicated model by difficulty to underlining relationship between input and output variables, they are increasingly are used as self-adaptive methods to achieve meaningful and practical patterns in the data based on training data. Layering structures, learning rate beside other parameters enable ANNS to seizes more advantages in comparison to other statistical and classical machine learning methods[7].

1.2 Motivation

The reasons to conduct a new research in this field is that: with regard to power management, load forecasting has become one of the hot topics which helps suppliers to make balance between produced and consumed power to improve power system efficiency. The availability of momentary data in household, beside importance of load forecasting open great opportunity to research and studies in this field. There are several studies which have used deterministic load forecasting for households but probabilistic forecasting has been less applied in the literature. In the recent decades emergence of more natural disasters which are not under the control of human- being causes more power shortages globally. Fluctuation of seasonal weather situation on the other side make it necessity to consider weather parameters in predicting costumers consumption patterns in the future. Recognized research gap in the literature that address all these aspects in comprehensive models has inspired me to apply different forecasting model on residential data, using both deterministic and probabilistic approaches. Further I describe and evaluate models in order to find best fitting model by lowest possible error for this type of analysis in this thesis[47].

1.3 Problem Statement and Research Questions

The main problem addressed in this research is to find the best fitting method to forecast residential power consumption based on available historical and real-time data. Following research questions are further explored and answered in this master thesis

1. Question 1: how we can build up a load forecasting model for residential data?
2. Question 2:what kind of features should be utilized to get more accurate result?
3. Question 3:how weather parameters affect prediction and which parameter has the the most significant impact?
4. Question 4:which machine learning algorithm is most suitable for selected data?
5. Question 5:how can data be selected for train and test data sets ?
6. Question 5: which time lag length provide result with minimum error rate?
7. Question 6:do the probabilistic algorithms applicable on these data/ this approach lead us to more accurate models?
8. Question 7:which error metrics or accuracy score should be deployed in order to validate outputs?

This thesis investigates these question by developing a forecasting model which can receive prepossessed and clean data in different time horizon to produce predicted value by using machine learning and deep learning algorithms.

1.4 scope

This study focuses on the three main areas which are connected together in most current problems of load forecasting for residential. We try to employ AI based techniques and tools to handle data and provide proper answers for our research questions.

- Load Forecasting is a technique to predict load demand of costumers in different time horizon from a minute a head to one year ahead. there are three different Load forecasting short-term load forecasting, medium-term load forecasting and long-term load forecasting applying in wide range of time horizon.load forecasting applying on time series data in various time lag. Time lag could be selected based on our selected approach for instance for short term forecasting lag can be day, hour or a passed minutes.
- AI based Algorithms and Techniques AI base, machine learning or deep learning techniques have been employed in different field of data analysing, consist of algorithms

and methods to establish analytical model. The techniques can be utilized for supervised learning when we deal with labeled classes with known correlation between the variables. Or unsupervised learning is used to group, unlabeled and irrelevant data and variables.

- household and weather data, smart meter is a device to record and store momentary electricity consumption (in kw)for each individual user. The data is recorded every half hour. In addition, as weather parameter playing important roll in electricity consumption, these data should be fed to the model for more accurate prediction.

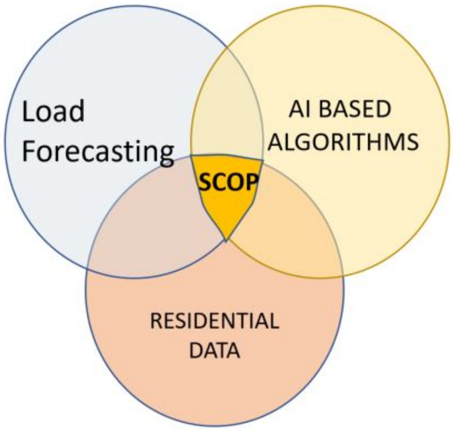


Figure 1.2: projects scope, Intersection of 3 main areas

1.5 Structure of the thesis

This thesis is structured as follows:

- literature review i) The content of this thesis starts with an explanation of the research approach in chapter1. followed by literature review including accomplished studies and researches related to load forecasting methods, analysis and AI based algorithms in chapter 2.

- Methodology Chapter 3 describes the applied methodology, and a set of techniques and approaches indicating how mentioned algorithms can be utilized to establish forecasting model. Brief explanation about Random Forest Regression,and ANN method such a MLP Regression which are applied in this thesis for deterministic forecasting model are provided in the chapter. In addition Deep learning methods and techniques which have been developed to seize probabilistic forecasting outcomes are also explained.

- experiments and methods Chapter 4 includes general description of data sets features and properties and a statistical analysis of consumption data.
- experiments and methods Chapter5 demonstrates applied methodology and selected methods on electricity consumption and weather data. Results of forecasting models are evaluated in this chapter to obtain optimum method for time series energy consumption data.
- conclusions and future works Chapter 6 includes the conclusion and discussion about proposed models and results.In addition, research opportunities and potential development of this domain are discussed in conclusion chapter.

1.6 modus operandi

At University of Stavanger, assistant professor Mina Framanbar will be the supervisor and Aida Mehdipour will be as a co-superviso

Chapter 2

literature review

This chapter presents a survey of related work are performed in this domain. This review is a broad overview to set the overall work in context. Medium and long term energy forecasts are used in capacity planning, infrastructure development, and financial budgeting. Drivers of demand on a medium and long term basis are a key difference compared to short term forecasting. Economic, demographic and climatic factors are taken into consideration including population growth, consumer income and expenditure, technological advances and efficiencies, and changing generation sources. In the short term analysis, these factors do not influence the results significantly due to their marginal changes. However they are important inputs for long forecast horizons. To overcome the incoming challenges and ensure accurate power prediction for different time horizons, sophisticated intelligent methods are elaborated.

Bouktif, S. et al (2018) have used machine learning and a long short-term memory (LSTM)-based neural network with various configurations to construct forecasting models for short to medium term aggregate load forecasting [8]. The research solves above mentioned problems by training several linear and non-linear machine learning algorithms and picking the best as baseline, choosing best features using wrapper and embedded feature selection methods and finally using genetic algorithm (GA) to find optimal time lags and number of layers for LSTM model predictive performance optimization. Using France metropolitans electricity consumption data as a case study, obtained results show that LSTM based model has shown high accuracy then machine learning model that is optimized with hyper-parameters tuning. Applying the best features, optimal lags, layers and training various LSTM configurations further improved forecasting accuracy. The result of the study concludes A LSTM model using only optimally selected time lagged features captured all the characteristics of complex time series and shows decreased Mean Absolute Error (MAE) and Root Mean Square Error

(RMSE) for medium to long range forecasting for a wider metropolitan area [8].

Fumo et al (2015) applied simple and multiple linear regression (MLR) analysis along with a quadratic regression analysis were performed on hourly and daily data from a research house [9]. In their study the time interval of the observed data showed to be a relevant factor defining the quality of the model. In their applied MLR model outdoor temperature and solar radiation offered improved coefficient of determination, but deteriorated root mean square error. Such results emphasizes the importance of using both parameters to assess and compare models. The article also conveys the authors belief that the future of residential energy forecasting is moving toward the development of individual models for each household due to the availability of data from smart meters, as well as the development of friendly and easy-to-use engineering software [9].

A time series method for medium term electricity demand was proposed by Gonzalez et all [14]. For the Spanish power system. A multi-model neural network framework to predict two components of the load time series separately, was trained on a dataset from 1975-92 to predict one month ahead values. The subsequent 10 years of monthly values comprised the test period. The economic trend and residuals time series were decomposed by means of a cubic smoothing spline and modelled using radial basis function (RBF) neural networks. Inputs to each RBF model were the previous 12 values of the corresponding time series. The trend RBF model contained 20 neurons in the hidden layer and the residuals RBF model structure contained 75 neurons. Summation of the two component forecasts constructed the final forecast. On the test period, a MAPE result of 1.89% was recorded [10].

In the study published by Baliyan et al 2015, a strategy to forecast hourly peak load using Recurrent Neural Network with Long-Short-Term-Memory architecture is presented [11]. The novelty study lies in improving the forecast accuracy by an intelligent incorporation of available domain knowledge during the forecast process. Experimentation is carried out to forecast hourly peak load in five different zones in USA [11].

The problem of modelling the electricity demand loads in Greece is addressed in another study in 2008 by Pappas et al (2008) implementing AutoRegressive Moving Average (ARMA) method [12]. In their study, the provided actual load data is deseasonilized and an (ARMA) model is fitted on the data off-line, using the Akaike Corrected Information Criterion (AICC). Finding of the article shows, the developed model fits the data in a successful manner however when the provided data includes noise or errors problems are occurred. Authors has faced similar problems when an on-line/adaptive modelling is required in their analysis. To overcome the occurred difficulties and resolve the problems in both cases simultaneous order and parameter estimation of ARMA models under the presence of noise are performed. Authors has implied the assumption that the provided data can be represented by an ARMA model. The produced results of the study indicate that the proposed method, which is based on the multi-model partitioning theory, tackles successfully the studied problem. For validation purposes the produced results are compared with three other established order selection criteria, namely AICC, Akaike's Information Criterion (AIC) and Schwarz's Bayesian Information Criterion (BIC). The developed model could be useful in the studies that concern electricity consumption and electricity prices forecasts Pappas et al (2008)[12].

XiaoshuLüa et all (2015) presents a new methodology for energy demand forecasting, which addresses the heterogeneity challenges in energy modelling of buildings [13]. In their method to improve forecast accuracy a physicalstatistical approach designed to account for building heterogeneity. The physical model provides a theoretical input to characterize the underlying physical mechanism of energy flows. Then stochastic parameters are introduced into the physical model and the statistical time series model is formulated to reflect model uncertainties and individual heterogeneity in buildings. A new method of model generalization based on a convex hull technique is further derived to parameterize the individual-level model parameters for consistent model coefficients while maintaining satisfactory modeling accuracy for heterogeneous buildings. The proposed method is applied on the data of four different sports buildings with field measurements as a case study. The results show that the proposed methodology and model can provide a considerable improvement in forecasting accuracy[13].

Lahouar, A.; Slama, J.B.H. Day-ahead (2015) in their study have applied random forest approach, characterized by immunity to parameter variations and internal cross validation for short term load prediction purposes [14]. Since several crucial tasks of power operators such as load dispatch rely on the short-term forecast, achieving similar accuracy for load prediction is intended in the study. The model is constructed following an online learning process, utilizing intelligent forecast algorithms. Such intelligent algorithms are amongst the main characteristics of smart grids, and is an efficient tool to cope with uncertainty. The study proposes a short-term load predictor, able to forecast the next 24 h of load. The inputs are refined by expert feature selection using a set of ifthen rules, in order to include the own user specifications about the country weather or market, and to generalize the forecast ability. The proposed approach is tested through a real historical set from the Tunisian Power Company, and the simulation shows accurate and satisfactory results for one day in advance, with an average error exceeding rarely 2.3%. The model is validated for regular working days and weekends, and special attention is paid to moving holidays[14].

The recent study published by Ayas Shaqour et al 2022 comprehensively investigates scale short-term load forecasting STLF of five aggregation levels (3, 10, 30, 100, and 479) based on a dataset of 479 residential dwellings in Osaka, Japan, with a sample size of (159, 47, 15, 4, and 1) per level, respectively, and investigates the underlying challenges in lower aggregation forecasting [15]. It is argued that although many of the studies have investigated both macro and micro scale short-term load forecasting (STLF), a comprehensive investigation on the effects of electrical demand aggregation size on STLF is minimal. especially with large sample sizes, where it is essential for optimal sizing of residential micro-grids, demand response markets, and virtual power plants. Five deep learning (DL) methods are utilized in their research for STLF and fine-tuned with extensive methodological sensitivity analysis and a variation of early stopping, where a detailed comparative analysis is developed. The test results reveal that a MAPE of (2.473.31%) close to country levels can be achieved on the highest aggregation, and below 10% can be sustained at 30 aggregated dwellings. Furthermore, the deep neural network (DNN) achieved the highest performance, followed by the Bi-directional Gated recurrent unit with fully connected layers (Bi-GRU-FCL), which had close to 15% faster training time and 40% fewer learnable parameters [15].

Karol Bot et al (2021) in their very recent study discusses the use of ensemble techniques in order to improve the performance of artificial neural networks models used for energy forecasting in residential houses [16]. The case study is a residential house, located in Portugal, that is equipped with PV generation and battery storage and controlled by a Home Energy Management System (HEMS). It has been shown that the ensemble forecasting results are superior to single selected models, which were already excellent. A simple procedure was proposed for selecting the models to be used in the ensemble, together with a heuristic to determine the number of models[16].

In the study published by Hambali, A.O.; Akinyemi (2017) electric power load forecast experimenting using Decision Tree Algorithms (Classification and Regression Tree) CART, (Reduced Error Pruning Tree) REPTree and Decision Stump is presented. The work revealed that, REPTree Decision Tree Technique is suitable to forecast electric load and outperformed other decision tree algorithms. The result presented by their study is expected to be useful for those are involved in Power transmission, Generation, Distribution and Marketing industry to enable them forecast electric power load and provide appropriate advising and decisions in timely manner[17, 53].

Ahmed et all in their study (2010) have tried to fill the knowledge gap they have identified on large-scale comparison studies for machine learning models of the regression or the time series for forecasting problems [53]. Hence, they have conducted a large scale comparison study on the major machine learning models for time series forecasting. Models are applied on the monthly M3 time series competition data (around a thousand time series). The models considered are multilayer perceptron, Bayesian neural networks, radial basis functions, generalized regression neural networks (also called kernel regression), K-nearest neighbor regression, CART regression trees, support vector regression, and Gaussian processes. The study reveals significant differences between the different methods. The best two methods turned out to be the multilayer perceptron and the Gaussian process regression. In addition to model comparisons, they have tested different preprocessing methods and have shown that they have different impacts on the performance [53].

Khosravani et al (2016) argue regression artificial neural networks are used to model various systems that have high dimensionality with nonlinear relations, where the system under study must have enough dataset available to train the neural network. In their study they provides a quick reference to the effects of main parameters of regression neural networks in load forecasting models. They apply and experiment various options effects on feed-foreword artificial neural network (ANN) which used to obtain regression model that predicts electrical output power (EP) of combined cycle power plant based on 4 inputs. The result of their work shows and explains the stochastic behaviour of the regression neural experiments and the effect of number of neurons of the hidden layers. Higher performance of larger training dataset size and different effect of larger number of variables as input are also presented in their findings. The reliability of the model is tested by simple statistical study performed on the error between real values and estimated values using ANN in their study [18].

In another study published by Chaturvedi (1999) it is intended to investigate the Automatic Generation Control (AGC) problem of a deregulated power system using Adaptive Neuro Fuzzy controller [19]. Three area control structure of Hydro-Thermal generation has been considered for different contracted scenarios under diverse operating conditions with nonlinearities such as Generation Rate Constraint (GRC) and Backlash. In each control area, the effects of the feasible contracts are treated as a set of new input signals in a modified traditional dynamical model. It is argued in the article that the key benefit of presented strategy is its high insensitivity to large load changes and disturbances in the presence of plant parameter discrepancy and system nonlinearities. The results of the proposed controller are evaluated with the Hybrid Particle Swarm Optimisation (HCPSO), Real Coded Genetic Algorithm (RCGA) and Artificial Neural Network (ANN) controllers to illustrate its robustness. Application of presented method leads to a flexible controller with a simple structure that is easy to realize and consequently it can be constructive for the real world power system [19].

Kuihe Yang Shijiazhuang, Lingling Zhao (2009) have adopted a short-term load forecasting model with a combined method to develop their model. The model summarizes virtues and defects of neural networks and fuzzy system, also considers that power system load has characteristics of basic load and variability load sizes [20]. Learned capability of neural networks to complete forecasting work of basic size of power load is used in the model. To correct basic load size and its variability which is affected by many factors, such as weather, data types and holidays and membership functions, fuzzy rules bases are constructed in fuzzy logic system in their study [20].

Power system is facing a transition toward a more intelligent, flexible, and interactive system with higher penetration of renewable energy generation, load forecasting, especially short-term forecasting. Other than aggregated residential load in a large scale, forecasting an electric load of a single energy user is fairly challenging due to the high volatility and uncertainty involved Kong, W (2019). Arguing the importance of load forecasting for individual electric customers in the future grid planning and operation Kong, W (2019) propose a long short-term memory (LSTM) recurrent neural network-based framework, to tackle load forecasting problem [21]. The proposed framework is tested on a publicly available set of real residential smart meter data, of which the performance is comprehensively compared to various benchmarks including the state-of-the-arts in the field of load forecasting. As a result. According to the findings of the study the proposed LSTM approach outperforms the other listed rival algorithms in the task of short-term load forecasting for individual residential households[21].

Agrawal, R et all (2018) have presented a load forecasting method with hourly resolution. The method is presented to enhance the accuracy of traditional models which are mostly restricted to electricity load data with monthly or annual granularity. The model is fundamentally cantered on Recurrent Neural Network consisting of Long-Short-Term-Memory (LSTM-RNN) cells [22]. The long term relations in a time series data of electricity load demand are taken into account using LSTM-RNN and hence results in more accurate forecasts. The proposed model is implemented on real time data of ISO New England electricity market. Precisely, publicly available data of twelve years from 2004 to 2015 have been collected to train and validate the model. Electricity demand predictions have been made for a period of five years from 2011 to 2015 on a rolling basis. The result of their study shows the proposed model is found to be highly accurate with a Mean Absolute Percentage Error (MAPE) of 6.54 within a confidence interval of 2.25%. Resulted computation time of the model is turned to be approximately 30 minutes which is favourable for offline training to forecast electricity load for a period of five years[22].

Kim, T.-Y and Cho, S.-B 2019, propose a CNN-LSTM neural network that can extract spatial and temporal features to effectively predict the housing energy consumption [23]. Based om their experiments CNN-LSTM neural network, which combines convolutional neural network (CNN) and long short-term memory (LSTM), can extract complex features of energy consumption. The CNN layer can extract the features between several variables affecting energy consumption, and the LSTM layer is appropriate for modeling temporal information of irregular trends in time series components. Their proposed CNN-LSTM method achieves permissible level of prediction performance for electric energy consumption. Also, it records the smallest value of root mean square error compared to the conventional forecasting methods for the dataset on individual household power consumption [23].

Mamun, A et al (2019) present a hybrid method integrating Genetic Algorithm (GA), which is an evolutionary algorithm, and long short-term memory (LSTM) network [24]. The study proposes a systematic method for electrical load forecasting by determining the time lags, neuron number, and batch size using GA. The authors argue inadequacy available research to increase the accuracy of the electrical load forecasting by selecting the best batch size for LSTM model. To evaluate the proposed hybrid model, the model is tested on half-hourly load data, collected from the Australian Energy Market Operator (AEMO). The experimental results show that the proposed hybrid model of GA-LSTM network surpasses the other standard models such as support vector machine (SVM), multilayer perceptron (MLP) and traditional LSTM model with the least MAE and RMSE value of 87.304 and 118.007 respectively. The proposed model shows 5.89% and 8.19% error reduction with respect to LSTM model in both MAE and RMSE respectively [24].

Jarosaw Protasiewicz and Jakub S. Sowiskiapply apply two reduction algorithms of a neural network architecture in order to improve the prediction quality of a multilayer perceptron network (MLP) [25]. The first algorithm is Optimal Brain Damage (OBD), whereas the second is Optimal Brain Surgeon (OBS). Their assumptions have been verified experimentally on the models for electricity consumption prediction using real data from the Polish electroenergetic system. Two series of tests have been carried out: the first is hourly forecast of electricity consumption for twenty four hours ahead, and the second is daily forecast of electricity consumption for a coming day considering results of performed computations. Their results show that both algorithms OBD and OBS improve the prediction quality of an MLP network. Moreover, simplification of the network has increased the speed of training process. They conclude such findings can be expanded to other time series prediction tasks [25].

A set of best practices to generate probabilistic forecasts of electricity demand using weather ensemble predictions are proposed and applied by Nicole Ludwiga,et al (2021) [26]. They use data for three weather variables (temperature, wind speed, and cloud cover), obtained from a 51-member ensemble system. For load forecasting, they investigated the efficacy of using ensemble post-processing, as opposed to using raw weather ensemble predictions from NWP systems. The research has shown how to post-process the weather ensemble predictions by accounting for spatial and temporal correlations, as well as correlations between the weather variables. It is shown in the article calibrating the weather ensemble predictions while accounting for their multivariate dependencies using a copula-based coupling approach improves the probabilistic load forecast accuracy, resulting in aCRPS that is noticeably better than a model that does not include any weather information. Nicole Ludwiga,b, Siddharth Arorac,d, and James W. Taylor 2021 (1) (PDF) Probabilistic Load Forecasting Using Post-Processed Weather Ensemble Predictions [26].

David et al 2020 propose DeepAR, a novel methodology for producing accurate probabilistic forecasts, based on training an auto-regressive recurrent network model on a large number of related time series [27]. They argue such probabilistic forecasts are crucial for example for reducing excess inventory in supply chains. It is shown in the study the proposed DeepAR model is effective at learning a global model from related time series, which can handle widely-varying scales through rescaling and velocity-based sampling. The model generates calibrated probabilistic forecasts with high accuracy, and is able to learn complex patterns such as seasonality and uncertainty growth over time from the data. The result of study shows the method works on a wide variety of datasets with little or no hyperparameter tuning, and is applicable to medium-sized datasets that contain only a few hundred time series [27].

Hilaf et al 2021 present Level Set Forecaster (LSF), as a simple yet effective general approach to transform a point estimator into a probabilistic one [28]. By recognizing the connection of algorithm to random forests (RFs) and quantile regression forests (QRFs), they prove consistency guarantees of the approach under mild assumptions on the underlying point estimator. As a by-product, they prove the first consistency results for QRFs under the CART-splitting criterion. Empirical experiments show that the approach, equipped with tree-based models as the point estimator, rivals state-of-the-art deep learning models in terms of forecasting accuracy [28].

Petrônio Silva and Frederico Gadelha Guimaraes 2017 propose a method for probabilistic forecasting based on the aggregation of seasonal Fuzzy Time-Series techniques with ensemble learning [29]. The proposed method generates dierent seasonal FTSmodels and the best ones are combined into an ensemble learning. The fore-casting procedure consists in evaluating individual models and combining their outputs into a continuous probability distribution using Kernel density estimation. The method was applied to SONDA dataset considering three seasonal indexes on solar radiation data. The best Ensemble models were those with 15minutes interval index and Entropy partitioning in their dierent parameters. The built ensemble forecasts were then compared with ARIMA and QuantileAuto-Regression models using Continuous Ranked Probability Score (CRPS)metric. The Ensemble FTS method presented a slightly larger CRPS, especially for the Epanechnikov, Tophat and Triangular kernels, which suggests a better model[29].

Summarizing reviewed literature shows load forecasting literature contains a broad range of structures and considerations. A wide variety of load forecasting modelling options are reported in the literature. Early methods focused on linear modelling techniques like multiple regression models. Some more recent studies have also used a linear approach [32]. stating that the key to their success is found in pre-processing the data. MLPs are the most well-established non-linear shallow network of choice. However, some authors extend their investigations to include other non-linear variants. The recent wave of research is focused on deep learning techniques and research in this field is ongoing. popular example of deep learning architectures for time series forecasting is the long short term memory (LSTM) recurrent neural network which is specifically designed to remember past information in such a way that can recall it at the appropriate time interval in the future to form the output prediction. With many new factors impacting the total demand, authors have moved towards modelling choices which combine forecasts from many models to create an overall prediction. Ensemble modelling has been explored in the recent studies as part of the solution to an energy forecasting competition.

Chapter 3

Methodology

The project methodology consists of 3 main steps. Including preprocessing, forecasting model development and model evaluation by error metrics which called post processing,Figure below illustrates all main steps and phase of project.

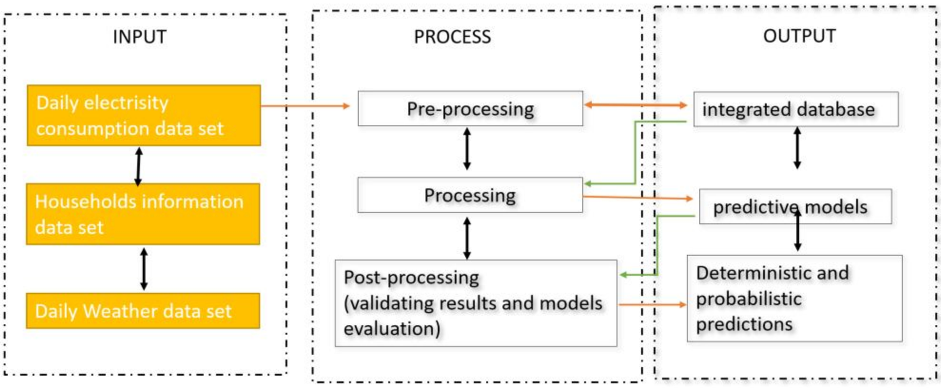


Figure 3.1: methodology steps of AI-based load forecasting on historical data

All steps and phases of the methodology are as follows:

- Inputs: Project data sources contain daily electricity consumption data household information and weather data.In this project daily min and max temperature from weather will be utilized.

- Process: first step is pre-processing data including integrating relevant data set in order to achieve coherent data set. Then visualizing data and performing statistical analysis to present data distribution and tendency. Since there are different features and some of the features are similar, a set of features should be selected which help to reduce model complexity. Selected data set should be cleaned from missing values and outliers, and finally data should be scaled to seize more accurate results. In the first step, deterministic phase, are transformed to time series data since applied models are not capable of doing such conversion automatically[31].

In the Processing step the main focus is to apply different forecasting algorithms and methods to build up an accurate and reliable forecasting model.Since in this study we are using regression model as a supervised learning, we chose to utilize random forest and MLP regression as machine learnng and deep learning methods to show a comparison between the two different approaches on the time series data. Furthermore for probabilistic forecasting, three well-known deep learning methods will be applied on data, to obtain set of output than one particular outcome per input. All achieved results of methods will be evaluated in post-processing or evaluation stage[30].

- Outputs: the outputs of this study will include data patterns and extracted knowledge from applying different models on integrated dataset which are achieved in different steps- of this analysis. The outputs are mapped to different steps of data processing as outcomes of each part while they are eventually connected internally together[31].

3.1 data prepossessing

Prepossessing is essential preliminary step in data science before persuading to knowledge discovery and data processing. Figure below shows steps of data prepossessing, consisting of data integration, feature selection, data cleansing,data transformation, data normalization,selecting train and test data, which will be explained in further parts[31].

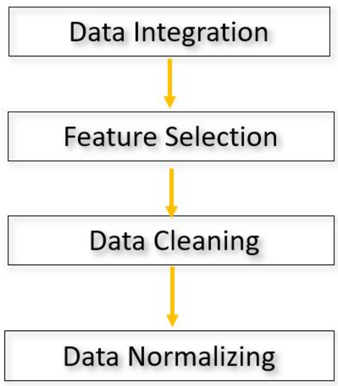


Figure 3.2: project data pre-processing steps

3.1.1 data integration

Data integration includes, gathering and combining data from different data sources which are stored using different techniques and make unified view of data. Data pre-processing is important to achieve permissible data variety and quality for data analysis. Initially, it is required to use variety of data sets based on impact of variables on final models. then required data sets should be connected together (Integration)[32].

3.1.2 feature selection

It is highly important to select optimum subset of features according to project objective, which is called feature selection. The feature selection approach aims to select a small subset of features that minimize redundancy and maximize relevance to the target. Feature selection improves learning performance, lowering computational complexity, building better generalizable models, as well as decreasing required storage capacity. Available data sources contain plenty of features and records, where performing analysis on such data sources is associated with high computational complexity. Such large data sources generally require a huge amount of installed memory and computation power in the calculation nodes. To resolve the level of complexity, normally similar features are eliminated after integration[33].

3.1.3 data cleaning

Generally substantial amount of collated data from data resources are raw data. These Raw data contains errors and missing values, not validated, in different (colloquial) formats; encoded and suspect requiring confirmation or citation. Incorrect or inconsistent data can lead to false conclusions and misdirected investments on both public and private scales. Data cleansing, data cleaning or data scrubbing is the process of detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database. Used mainly in databases, the term refers to identifying incomplete, incorrect, inaccurate, irrelevant, etc. parts of the data and then replacing, modifying, or deleting this dirty data.then it is essential in most data analyzing cases to perform data cleaning in the first step to improve the quality of existing data. Data quality is a state of completeness, validity, consistency, timeliness and accuracy that makes data appropriate for a specific use[34]. data cleaning consist of finding zero, NAN data in the data sets and replace them with appropriate value by using different method based on research purpose.

3.1.4 data normalization

data normalization refer to a act which apply to data to change data scale based on some criteria. Data normalization is practical technique especially when it is intended to implement ANNs methods as the model functions with scaled data. In the AI based modeling, normalization not only enhances algorithm functionality but also reduces computation time and helps to achieve more accurate model. data normalization can be divided in different categories such as, 1NF, 2NF and etc. One of the well known normalization methods is min-max Normalization.In this method,for each parameter or feature, minimum value should be considered as zero and maximum value should be one then the other values will be converted to a decimal between zero to one[35]. following equation illustrate min- max normalization method, as y refer to values for each feature.

$$y' = \frac{y - y_{min}}{y_{max} - y_{min}}.$$

in this research both original values and normalized values will be feed to model to show how normalization affect our results.

3.2 Methods and Techniques

this part presents brief desecration of all deployed Machine learning and AI based data driven methods.

3.2.1 MLP Regression

A multilayer artificial neuron network is an integral part of deep learning. A fully connected multi-layer neural network is called a Multilayer Perceptron (MLP).- The term "multilayer perceptron" does not refer to a single perceptron that has multiple layers. Rather, it contains many perceptrons that are organized into layers. An alternative is "multilayer perceptron network". Moreover, MLP "perceptrons" are not perceptrons in the strictest possible sense. True perceptrons are formally a special case of artificial neurons that use a threshold activation function such as the Heaviside step function. MLP perceptrons can employ arbitrary activation functions. A true perceptron performs binary classification, an MLP neuron is free to either perform classification or regression, depending upon its activation function. The term "multi-layer perceptron" later was applied without respect to nature of the nodes/layers, which can be composed of arbitrarily defined artificial neurons, and not perceptrons specifically. This interpretation avoids the loosening of the definition of "perceptron" to mean an artificial neuron in general[36].

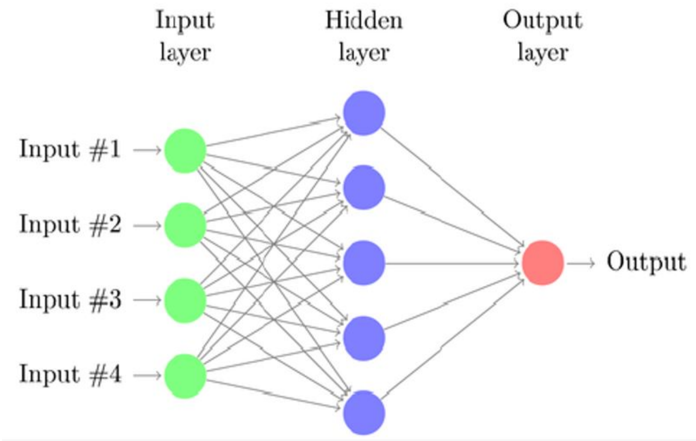


Figure 3.3: Architecture of Artificial Neural Network (ANN)

3.2.2 Random Forest Regression

Random forests (RFs) are an ensemble learning method for both classification and regression problems [4]. RF is a collection of decision trees that grow in randomly selected subspaces of the feature space. The principle of RFs is to combine a set of binary decision trees (Breiman's CART Classification And Regression Trees [5]), each of which is constructed using a bootstrap sample coming from the learning sample and a subset of features (input variables or predictors) randomly chosen at each node. Thus in contrast to the CART model building strategy, an individual tree in RF is built on a subset of learning points and on subsets of features considered at each node to split on. Moreover trees in the forest are grown to maximum size and the trimming step is skipped [36][37].

After individual trees in ensemble are fitted using bootstrap samples, the final decision is obtained by aggregating over the ensemble, i.e. by averaging the output for regression or by voting for classification. This procedure called bagging improves the stability and accuracy of the model, reduces variance and helps to avoid overfitting. The bias of bagged trees is the same as that of the individual trees, but the variance is reduced by reducing the correlation between the trees [49]. Breiman showed that random forests do not overfit as more trees are added, but produce a limiting value of the generalization error. The RF generalization error is estimated by an out-of-bag (OOB) error, i.e. the error for training points which are not contained in the bootstrap training sets (about one-third of the points are left out in each bootstrap training set). An OOB error estimate is almost identical to that obtained by N-fold cross-validation. The large advantage of RFs is that they can be fit in one sequence, with cross-validation being performer along the way. The training can be terminated when the OOB error stabilizes [7].

3.2.3 DeepAR Estimator

Amazon Forecast DeepAR+ is a supervised learning algorithm for forecasting scalar (one-dimensional) time series using recurrent neural networks (RNNs). Classical forecasting methods, such as autoregressive integrated moving average (ARIMA) or exponential smoothing (ETS), fit a single model to each individual time series, and then use that model to extrapolate the time series into the future. In many applications, many similar time series across a set of cross-sectional units are used. These time-series groupings demand different products, server loads, and requests for web pages. In such circumstances, it can be beneficial to train a single model jointly over all of the time series. DeepAR+ uses this approach. When the dataset contains hundreds of feature time series, the DeepAR+ algorithm outperforms the standard ARIMA and ETS methods. The trained model for generating forecasts for new time series that are similar to the ones it has been trained on can be used [27]. During training, DeepAR+ uses a training dataset and an optional testing dataset. It uses the testing dataset to evaluate the trained model. In general, the training and testing datasets should not contain the similar set of time series. A model trained on a given training set to generate forecasts for the future of the time series in the training set, and for other time series can be used. Both the training and the testing datasets consist of (preferably more than one) target time series. Optionally, they can be associated with a vector of feature time series and a vector of categorical features. Each target time series can also be associated with a number of categorical features. You can use these to encode that a time series belongs to certain groupings. Using categorical features allows the model to learn typical behavior for those groupings, which can increase accuracy. A model implements this by learning an embedding vector for each group that captures the common properties of all time series in the group. A DeepAR+ model is trained by randomly sampling several training examples from each of the time series in the training dataset. Each training example consists of a pair of adjacent context and prediction windows with fixed predefined lengths. model hyper parameters includes,context-length,prediction length,Epoch, and the other parameter related to RNN methods [38].

- context-length:the number of time points that the model reads in before making the prediction. The value for this parameter should be about the same as the Forecast Horizon. The model also receives lagged inputs from the target, so context-length can be much smaller than typical seasonality [27].
- prediction-length:or forecast horizon, this parameter controls how far in the future predictions can be made [27].
- Epoch: The maximum number of passes to go over the training data. The optimal value depends on your data size and learning rate. Smaller datasets and lower learning rates both require more epochs, to achieve good results [27].

3.2.4 Gaussian Process

builds a local time series model using Gaussian Processes (GP). Each time series has a GP with its own hyper-parameters. For the radial basis function (RBF) Kernel, the learnable hyper-parameters are the amplitude and length scale. Each time series has a GP with its own hyper-parameters. For the radial basis function (RBF) Kernel, the learnable hyper-parameters are the amplitude and length scale. The periodic kernel has those hyper-parameters with an additional learnable frequency parameter. The RBFKernel is the default, but either kernel can be used by inputting the desired KernelOutput object. The noise sigma in the model is another learnable hyper-parameter for both kernels. These parameters are fit using an Embedding of the integer time series indices (each time series has its set of hyper-parameter that is static in time). The observations are the time series values. In this model, the time features are hour of the day and day of the week [39,40].

3.2.5 Feed Forward Estimator

FeedForwardEstimator shows how to build a simple MLP model predicting the next target time-steps given the previous ones. An ANN is a system of processing units (neurons) that can be linked together in different ways and estimate various non-linear and arbitrary patterns. In a feed-forward architecture (FFNN), there is no feedback and intra-layer connections between neurons. The weights and bias of the network are estimated using a training algorithm such as the back-propagation algorithm. This algorithm measures Processes 2020, 8, 484 7 of 22 the error of output every time and feeds back this information to the network to reduce the error up to an acceptable predefined value[1]. Further, more details on back-propagation algorithms are described in [41].

common hyper parameter among all probabilistic approach forecasting is Freq that refers to time series frequency. frequency could be determined as minutes, day month and etc based on model purpose. As it was mentioned, Random Forest regression and MLP models are selected to be applied on data to achieve deterministic and particular model outputs. Next, developed models based on probabilistic methods will be assessed to find out best suit model on residential energy consumption data set[45].

3.3 Model Evaluation Metrics

To evaluate the performance of a forecasting technique, forecasting error is calculated. The lower the forecasting error, the higher the performance of the model. The forecasting error is the difference between the actual observation and the predicted value. There are many error metrics that are proposed for calculating the forecasting error and comparing the performance of time-series forecasting techniques[48]. In this study, we used three such metricsMean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Average Scaled Error (MASE) and CRPS() as a well-known metric to evaluate probabilistic forecasting[41].

If yt is the prediction value and yt is the actual value at time t and n is the number of test observations, we can define the three metrics as the following: The MAE calculates the magnitude of the errors on average and ignores whether the prediction values are higher or lower than the real values. Thus, MAE gives equal importance (weight) to all Processes 2020, 8, 484 11 of 22 individual differences. The RMSE on the other hand, penalizes large errors by calculating the squared error before averaging them. The MASE was introduced as a more applicable error metric and as an alternative to some metrics like Mean Absolute Percentage Error (MAPE) when the observation or prediction values are zero[48]. The MAPE is commonly used as a loss function in model evaluation because it can interpret the relative error. However, the problem with MAPE can occur when there are zero values in the series and there will be a division by zero. For such sequences, MASE is appropriate as it never produces infinite or unknown values. In this alternative, each actual value of the series in the MAPE formula can be replaced by the average of all actual values of that series[42].

3.3.1 MAE

The root mean square error (RMSE) has been used as a standard statistical metric to measure model performance in meteorology, air quality, and climate research studies. The mean absolute error (MAE) is another useful measure widely used in model evaluations. While they have both been used to assess model performance for many years, there is no consensus on the most appropriate metric for model errors. In the field of geosciencs, many present the RMSE as a standard metric for model errors[42].

3.3.2 RMSE

The RMSE is preferred to the MSE as it is on the same scale as the data. Historically, the RMSE and MSE have been popular, largely because of their theoretical relevance in statistical modelling. However, they are more sensitive to outliers than MAE or MdAE which has led some authors (e.g., Armstrong, 2001) to recommend against their use in forecast accuracy evaluation[42,1]

3.3.3 CRPS

Probabilistic forecasts assign a probability to every possible value in future. Yet, all probabilistic forecasts are not equally accurate, and metrics are needed to assess the respective accuracy of distinct probabilistic forecasts. Simple accuracy metrics such as MAE (Mean Absolute Error) or MAPE (Mean Absolute Percentage Error) are not directly applicable to probabilistic forecasts. The Continuous Ranked Probability Score (CRPS) generalizes the MAE to the case of probabilistic forecasts. Along with the cross entropy, the CPRS is one of the most widely used accuracy metrics where probabilistic forecasts are involved[43].

The CRPS is frequently used in order to assess the respective accuracy of two probabilistic forecasting models. In particular, this metric can be combined with a backtesting process in order to stabilize the accuracy assessment by leveraging multiple measurements over the same data set[43].

This metric notably differs from simpler metrics such as MAE because of its asymmetric expression: while the forecasts are probabilistic, the observations are deterministic. Unlike the pinball loss function, the CPRS does not focus on any specific point of the probability distribution, but considers the distribution of the forecasts as a whole[46].

if x is considered as an observation and F cumulative. distribution function associated with a probabilistic forecast, then the CRPS can be calculated as following:

in this study the result of developed models will be assessed by using mentioned metrics. in the first phase for deterministic forecasting MAE, MSE and RMSE will be applied on the results. in order to evaluate probabilistic models, CRPS and RMSE will be significant error metrics to be applied.

Chapter 4

Dataset and Statistical Analysis

4.1 Data sets description

the developed models have been applied on sub set of energy consumption data for 5567 households in London. the main data set consists of data from smart meters installed in households that involved in the UK Power Network conducted low Carbon London project November 2011 and February 2014**?**. the houses were chosen as a representative of London population. the data set contains 110 block with same number of households(50 per block) within 18 categories that called ACORN groups(from A to R).blocks are equally divided among Acorns.the data set includes complete information about social factors, population behaviors of each group and etc. Daily data set has electricity consumption information, comprising 3510433 rows and 9 columns for all households during more than 2 years from 2011-11-23 to 2014-02-28(829 days). tables below demonstrates features for different data sets.

Data Set	Features
Daily	LCLid', 'day', 'energy_median', 'energy_mean', 'energy_max', 'energy_count', 'energy_std', 'energy_sum', 'energy_min'
Info	'LCLid', 'stdorToU', 'Acorn', 'Acorn_grouped', 'file'
Acorn_details	20 categories in 200 subcategories
Weather	'temperatureMax','temperatureMaxTime', 'windBearing', 'icon', 'dewPoint','temperatureMinTime', 'cloudCover', 'windSpeed', 'pressure', 'apparentTemperatureMinTime', 'apparentTemperatureHigh', 'precipType', 'visibility', 'humidity', 'apparentTemperatureHighTime', 'apparentTemperatureLow', 'apparentTemperatureMax', 'uvIndex', 'time', 'sunsetTime', 'temperatureLow', 'temperatureMin', 'temperatureHigh', 'sunriseTime', 'temperatureHighTime', 'uvIndexTime', 'summary', 'temperatureLowTime', 'apparentTemperatureMin', 'apparentTemperatureMaxTime', 'apparentTemperatureLowTime','moonPhase'

Figure 4.1: table of data sets available features

as it is obvious in the table, there are wide range of features in each data sets in which some off them are the same or can calculated based on other features. fir instance in energy consumption data set, 'energy-mean' can be computed based on 'energy-min' and 'energy-sum'.in addition, it should be taken in consideration to select feature based on project goal. for instance for our study, the 'energy-sum' is selected as we need daily sum of energy consumption of each house in order to predict daily energy consumption in the future per households. more information about households and categories can be found in **?**. as it mentioned in Methodology section, the first step of data analysing and establishing a model based on data is data pre-processing which is divided in four steps.integrating, feature selecting, cleaning and normalizing data. as it mentioned next step before data cleaning is finding appropriates features. in this study we chose to use energy-sum ,which refer to total daily electricity consumption per household, houseid, Acorn, file, temperature-max and temperature-min which contain daily outside temperature. we finally going to utilize temperature-mean based on main features to fed the model. after integrating data sets together and selecting subset of features, the next step to be taken is refining and cleaning data.

4.2 data cleaning and subset selection

data should be cleaned and refined in order to remove noises, replacing missing and mismatch values and improving data quality in the dataset. before starting data cleaning data should be visualized.data visualization provide us with comprehensive overviews on data in term of data distribution in the groups, missing data, outliers and etc. below figure , shows daily electricity consumption distribution in all 18 groups.

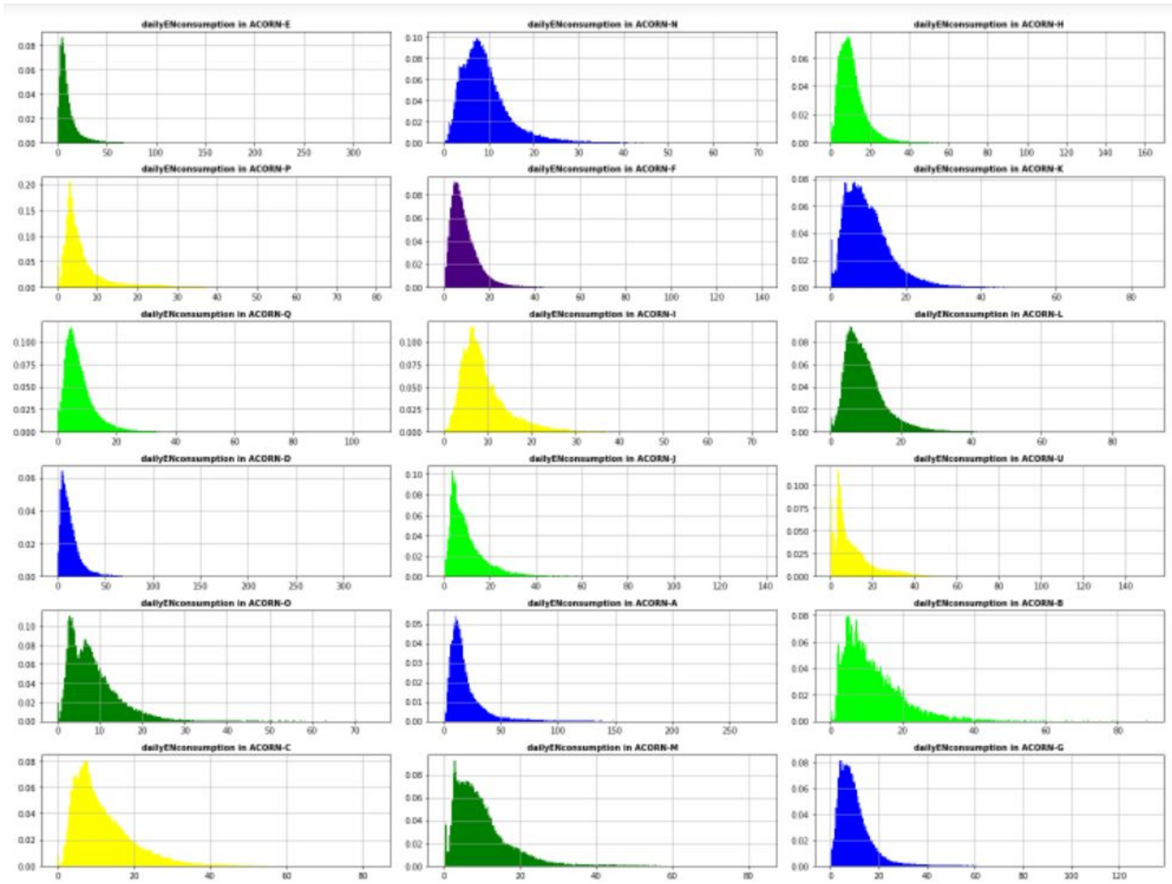
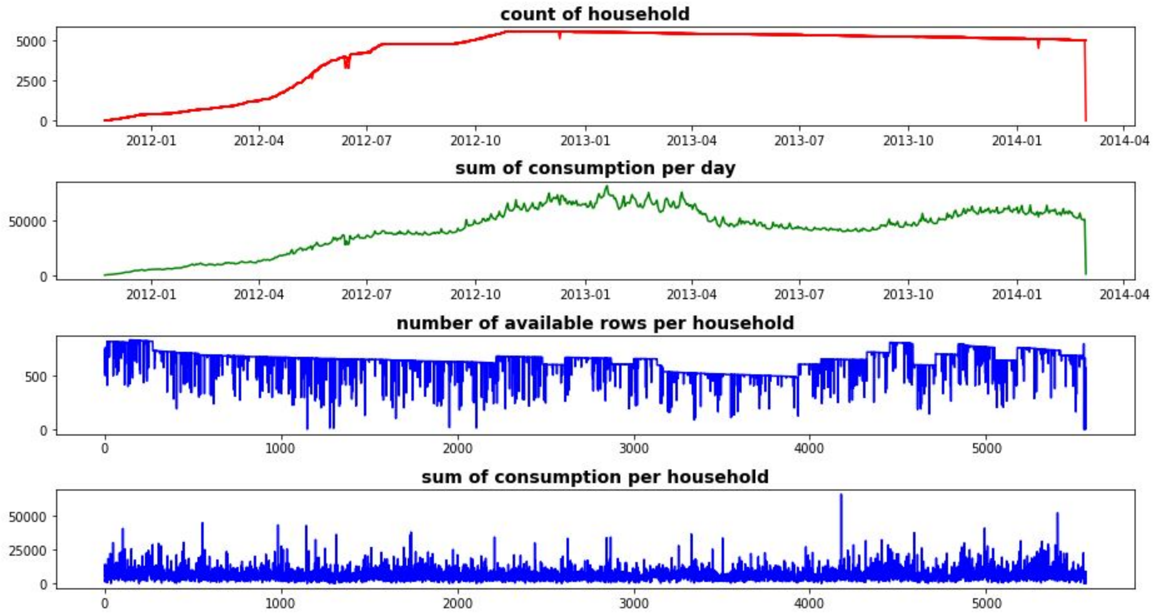


Figure 4.2: daily energy consumption distribution for all 18 groups

demonstrated histogram give excellent overview of data distribution in different groups. as it is obvious,the groups number A,E,D have higher daily consumption respectively 300,300 and 200 Kw among all groups whereas the mean consumption of them equals to 25 kw. it convey potential of having outlier in the dataset. the most of data sets have daily consumption around 10 Kw, but still outliers in all the plots are distinct. on the other hand,daily consumption for 5567 houeholds in 829 days should be equivalent to 4,615,043 records, while the recorded data rows in dataset equals to 3,510,433 rows that imply existence of almost 30% missing data in data set. moreover,considering all the group of data to feed the model not only increasing computation time and complexity but also reducing model accuracy and precision. therefor we chose to select a subset among 18 groups. as we decide to analyse daily data, availability data per day was a main consideration as well as group diversity.in order to seize that have been checked to find the group with maximum available data records and minimum outliers(829 records per households in each group). figure below, demonstrates diversified plot related to data availability and consumption tendency in all groups.



Í

Figure 4.3: data availability for house profile, total daily consumption

first plot should number of available household in each day, as it apparent, this number increasing since early 2012 but still fluctuating over 2 years. plot2, which shows daily total consumption,the consumption is lowest situation although is for winter time when the consumption have to be in highest demand. the reason absolutely is limited number of households participated in survey. to find outlier for electricity consumption, according to (https://www.ovoenergy.com/guides/energy-guides/how-much-electricity-does-a-home-use)average energy consumption for households in UK over 1 years could be around 3000 kw/h,which in our case within 2years become 6000 as a minimum and considering 3 times more than mean as a maximum criteria (20000 Kw). then the 'energy-sum' between 6000 as a lower bound and 20000 as a upper bound selected and others have been considered as outliers. in addition, record by energy-sum equal and lower than zero have been eliminated.moreover, the households with energy-sum lower than 3kw for more than 30 days have been ignored. after achieving partial cleaned data set , groups have been checked to find out group with maximum number of households and most available records,(20%) missing records per house holds is acceptable. based on our assumption and assessment three groups namely E,F,Q have been chosen in this study which are related to different Acorn categories and have maximum data. according the Acorn report, group E and D are belong to 'Affluent achiever' category, which refers to the costumers who are living in big houses in wealthy region. on the other hand, the households in group Q which is called 'difficult circumstances' are mostly single and young families, receiving benefits,living in high rise flats which are located in deprived areas. choosing households from different categories enable us to check how model works with data variety. in addition, the model will be trained on each category separately. In addition, we chose to select a limited number of household around 75 for our experiment. rather than selecting household randomly among the three selected group, household have been chose equally from each group. the 25 top households per group with maximum available data have been selected. since in the further step,model selection, data should be divided in train data and test data. we consider 15 housholds to be train data and 10 household to be test data. figure below illustrates, sum of daily energy consumption of 75 selected households.

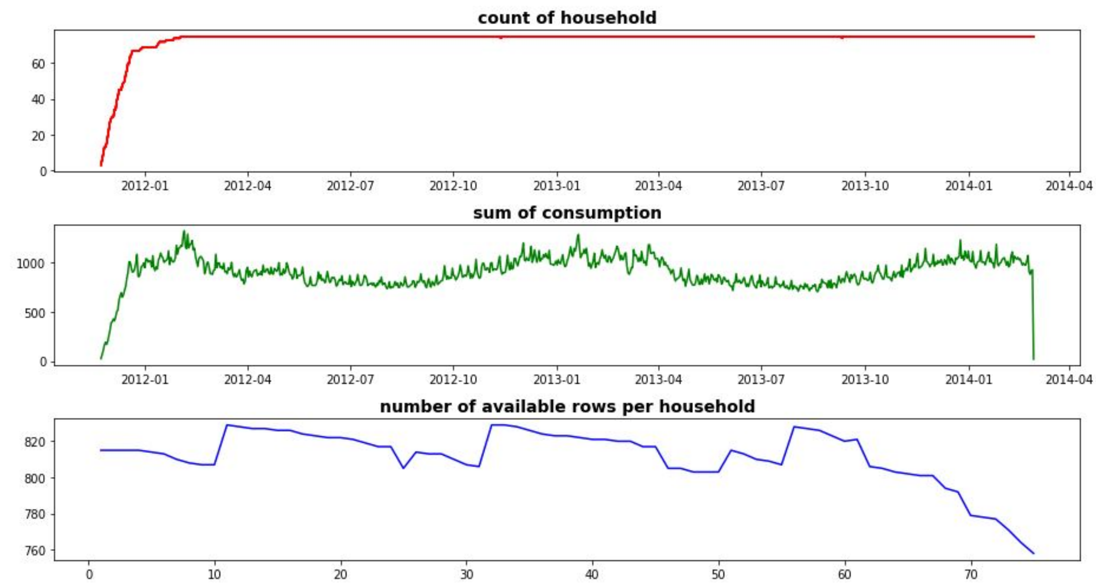


Figure 4.4: data availability for house profile, total daily consumption for selected subset

as it is shown in figure, total consumption in first 3-4 month of each year peek at more than 1000 kw.in addition,it is mean number of data records per household is around 790 records. figures below demonstrates the energy consumption of fifteen houses in the data set in three acorn groups during more than 2 years.as it is obvious, plots illustrates different amounts of daily electricity consumption per sample houses.

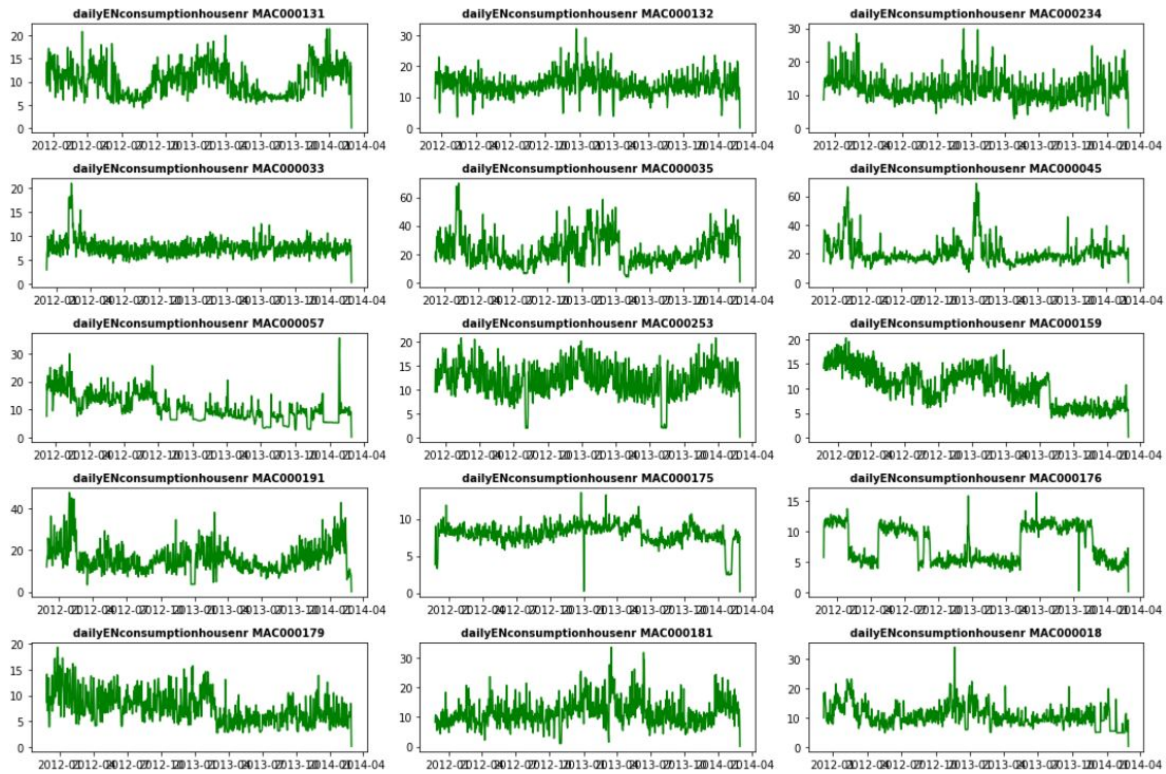


Figure 4.5: daily electricity consumption for some houses in group E

figure above, shows load readings for houses belong to group F over two years.as the figure shown the daily consumption range in this group is between 3 to 60kw.

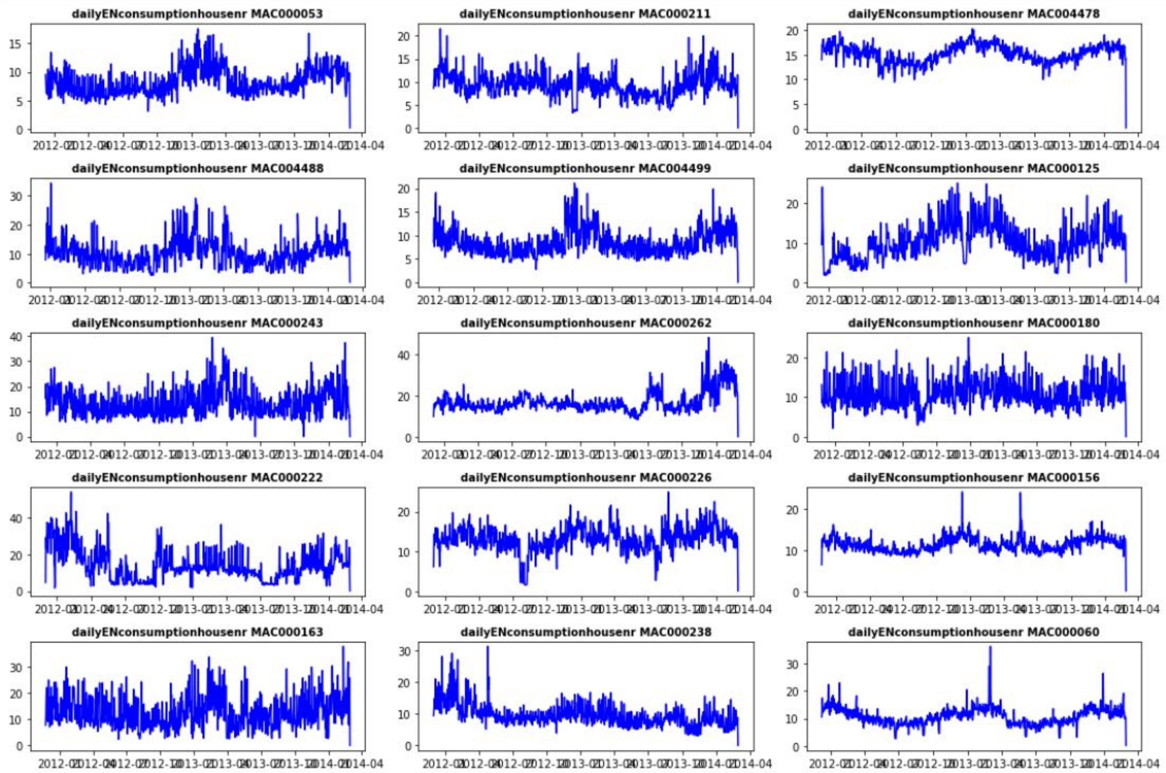


Figure 4.6: daily electricity consumption for some houses in group F

this figure, shows daily electricity consumption for houses within group E over two years.as the figure shown the consumption level varies from3 to 40kw in some houses but for most houses top level hits 20kw per day.

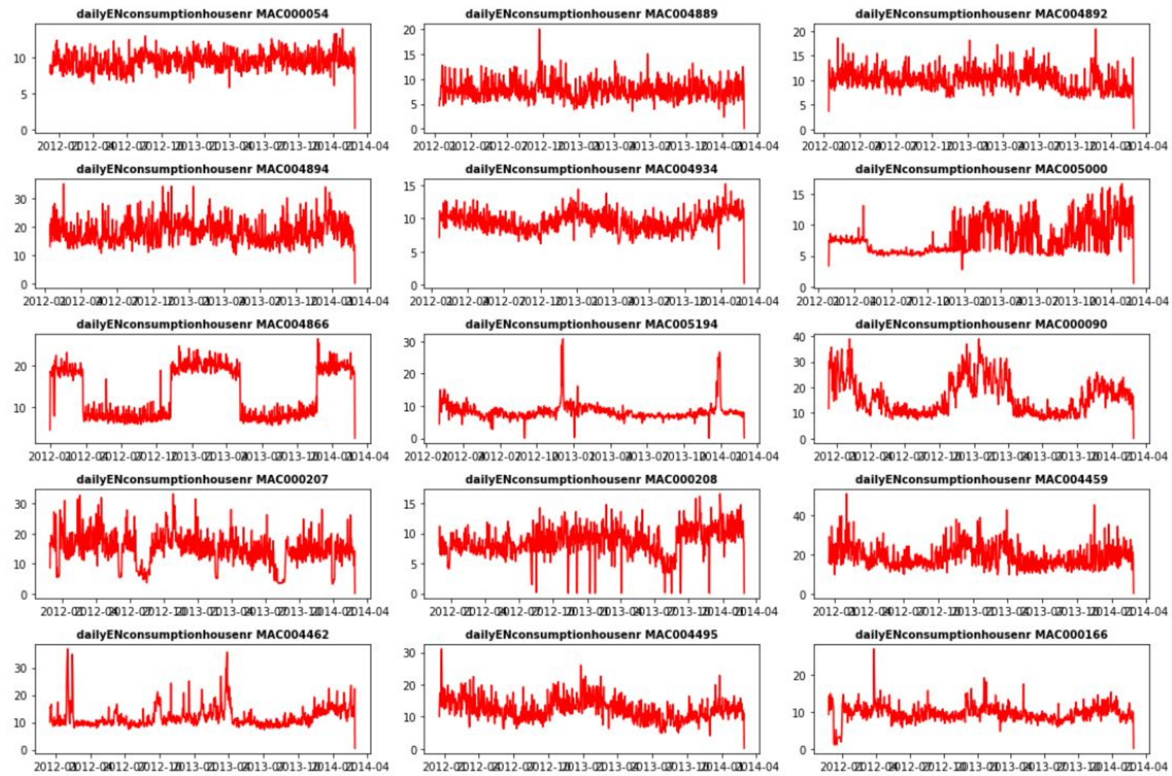


Figure 4.7: energy consumption for houses in group Q

as it mentioned,Outlier detection is interesting and important task in data preprocessing, Outliers will affect results of analytical model and leads to inaccurate outcomes. on the other hand, load volatility is considered as an influential factor in residential load forecasting. load volatility refers to deviation from mean of energy consumption which leads to high level of fluctuation in house consumption patterns as it is affected by different factors. for instance, in this case of analyzing daily data, seasonal factor like outside temperature can be important. as the figure shows, consumption raises significantly in winter and partiality in autumn. another effective factor could be day of week, as consumption increases in the weekend and holiday since people spend more time at home.however, high deviation of consumption results in complexity in load profile, and complication to build an accurate model. in the below figures, the boxplots depicts load volatility for houses in selected groups over 2 years. the box plot is good tool to obtain information on variation of electricity consumption and median values.

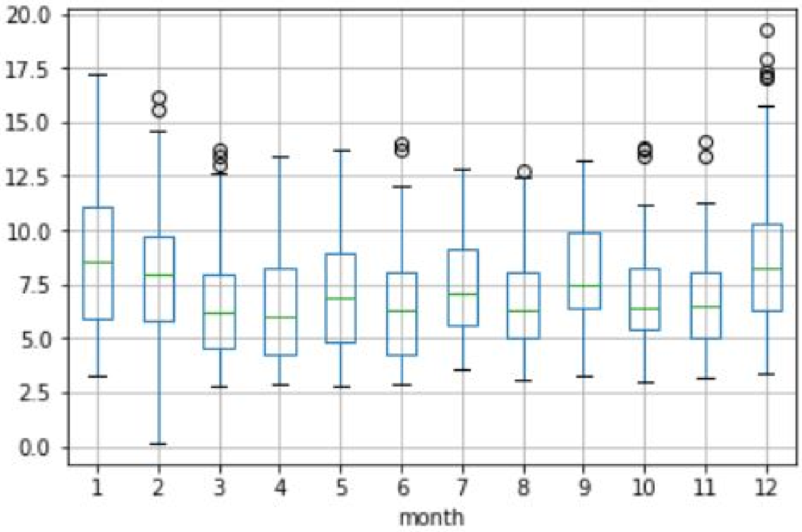


Figure 4.8: daily energy consumption for house number 15 in group E over 12 month of year

on the other hand, box plot can provide us with overview that energy consumption of one house changes in different month of a year. figure below shows electricity consumption of house number 15 from group E over time.

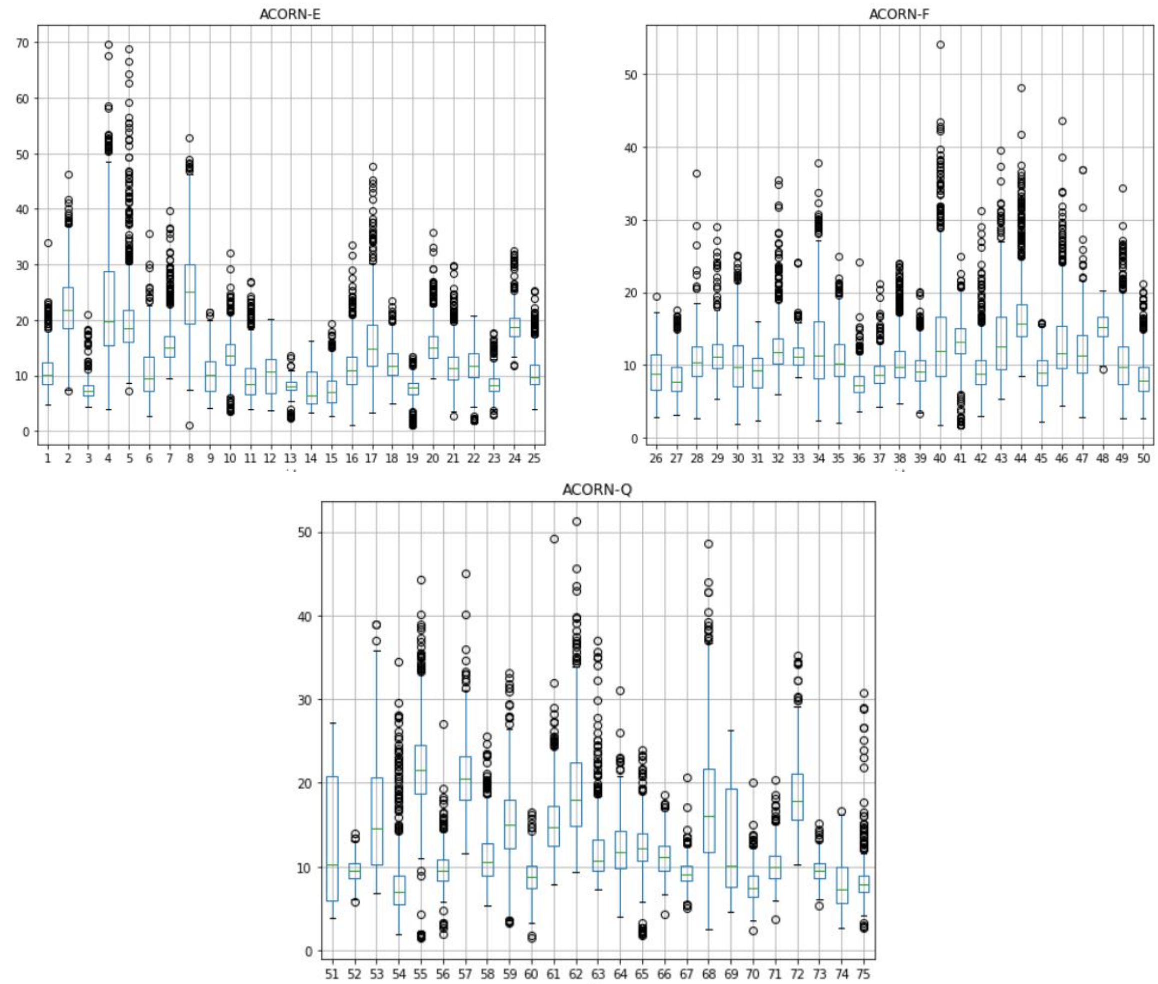


Figure 4.9: energy consumption for houses in group Q

box plot is a method which can illustrate locality of data in the quantile. the values lower than minimum or larger than maximum are represented by bobbles in the plot. then the bobble indicates value out of quantil range. for instance in the above pictire for house number 15 in group B,maximum and minimum range in November is 16,3 respectively, by using the equation Q3+1.5*(Q3-Q1) and replacing Q3= 8,25 and Q1=5 in the equation, then we have 8.25+1.5*(8.25-5) which equals to 13.1,which considered as outlier and shown by bobble in the plot. above boxplot demonstrate load volatility for each particular group in each group. the similar comprehension can be obtained by checking other boxplot, for instance,houses 3,13 in group E and 50 in group F had experienced small changes over years,which make prediction of future load forecasting easier while in some other houses (51 and 69 in the group Q) there are major changes that result in difficulty forecasting. (Vapnik, V.N. Statistical Learning Theory; John Wiley Sons, Inc.: New York, NY, USA, 1998) after detecting outlier in our data set, now the time is for removing or replacing these noisy data. we decided to replace than remove outliers as we would avoid to decrease the data set size. outlier should be replaced by mean of corresponds day in belonging group. in addition after joining weather and consumption data sets, some Nan VALUES appear in new data set which need to be replace by using fillna method in python.

table below provide description of three group of acorn which have been selected for rest of study over test period.as it is shown in the table, group A has highest mean and standard deviation values of daily and monthly electricity consumption in comparison to other groups.the group Q costumers consumed slightly lower electricity than households in group E.the lowest daily and monthly electricity consumption has been registered for group F by daily mean and standard deviation 11 and 4.6 respectively.(aida)final dataset consist of 75 houses with 829 rows per house, equals to 62,175 rows.

Group	blocks	number of houses	Mean and STD of daily consumption	mean and std of monthly consumption	Total consumption for 2 years
ACORN-E	[block_22, block_36, block_35, block_29	25	[12.82, 6.48]	[385, 75.3]	252924.58299
ACORN-F	[block_52, block_54, block_46, block_44	25	[11.21, 5.09]	[330,54, 57.4]	221035.37500
ACORN-Q	[block_96, block_98, block_99, block_104	25	[13.5, 6.12]	[405, 69.89]	265017.47499

Figure 4.10: Descriptive statistical description of selected subset

in the next section data will be normalized and serialized to have same value range in time series format.

4.3 data normalization

as it was explained in methodology chapter,data normalization is practical technique to scale data especially when we come to using ANNs methods as they working finely with scaled data. in the AI based modeling, normalization not only enhance algorithm functionality but also reduce computation time and help to achieve more accurate model. data normalization can be divided in different categories such as, 1NF, 2NF and etc. one of the well known normalization method is min-max Normalization. in this case, data have been normalized by using min-max normalization in range 0 to 1 to be fed to MLP deterministic forecasting model and project probabilistic models. in forecasting regression, the model will be applied on original value of data. moreover, data should be converted to time series format in different horizon or lag. in deterministic forecasting step. for this purpose, we have defined function named 'series $_{t}osupervise()'whichreceiveddata, numberofvariable, desiredinputandoutputLaglengthandreturntimeseriesd$

Chapter 5

Results and Discussion

As discussed in the methodology chapter, data forecasting process is divided into two main steps of deterministic forecasting and probabilistic forecasting. In this research deterministic forecasting models were implemented by using supervised learning and regression methods. Further in the process model tries to learn itself to find mapping function between input and output [50]. Such mapping function is used to forecast output from new input. Data Random forest and MLP regression methods have been applied on training and test data of the deterministic forecasting phase of this thesis.To perform such experiments, scikit-learn libraries in python were utilized. The computation environment has been jupyter Note book for deterministic and Google Colab for probabilistic forecasting model on a pc, Asus AMD Ryzen 7 3700U with Randon Vega Mobile Gfx,2.30 GHz with 16 GB RAM. The data set used for this analysis, comprises daily consumption of 75 house profiles, which are split in two different subsets, test sets with 15 houses and train set with 60 houses. The houses data for test and train set were not selected randomly from all data set in this analysis. Instead the houses in data base are categorized in three size groups and equivalent number of house from each group(3 groups in total) are randomly selected. This way of clustered sampling from population, helps to evaluate model performance with diversified consumption level over years.Statistical specifications and various house profiles are used in the data sampling. In this study it was also evaluated how machine learning models are affected with multitude and diversity of input variables as well as training size. The number of variables and train size could be considered as an influential parameters in model performance. For instance, if the model is trained by simple and inappropriate variables length and training size, it will not be able to learn the type of relationship among input and output properly, which stems to inaccurate results. On the other hand feeding model by many variable and enormous training data increase model complexity and computation time. In the processing, the models would be fed with different input length which called load Lag or time horizon which were produced from previous time steps in pre-processing step, to assess model accuracy and performance.

5.1 deterministic forecasting

As it explained, in the Deterministic forecasting, the result of forecast is a certain value per each time horizon, and a explicit input. Among wide range of methods and techniques which are applicable for deterministic forecasting, in this study, we chose to employ two machine learning and deep learning algorithms of Random Forest Regression and MLP regression respecteivly. As mentioned earlier in the methodology chapter, Random forests (RFs) are an ensemble learning method for both classification and regression problems. RF is a collection of decision trees that grow in randomly selected sub spaces of the feature space. The principle of RFs is to combine a set of binary decision trees [7],each of which is constructed using a bootstrap sample coming from the learning sample and a subset of features (input variables or predictors) randomly chosen at each node. In order to implement forecasting model based on Random Forest algorithm, RandomforestRegressor() from Python, Scikit-learn was utilized.the function receives different parameter as a input argument's such as,number of estimator or trees,maximum deep of forest and etc and build a model based on parameters and training data set. Then selecting appropriate parameters was considered as a important step. as it mentioned,MLP is an integral part of deep learning. A fully connected multi-layer neural network. it contains many perceptrons that are organized into layers. MLP perceptrons can employ arbitrary activation functions[25]. A perceptron performs binary classification, an MLP neuron is free to either perform classification or regression, depending upon its activation function. In this project, MLP regression from sickitlearn, neural netwwork library has been utilized to build deterministic forecasting model. MLP in sickit libarary dosnot have activation function in output layer,and uses squre error as a loss function. Meanwhile mlp is able to function with single and multiple targets. MLP has different parameters such as 'hidden-layer-size' the number of neurons in the hidden layers,'activation function' for the hidden layer, learning-rate to control updating weight and etc [36]. Since selecting suitable hyper parameters to implement predictive models is paramount, in this project, Gridesearch from sickit-learn was utilized to find best values. Output of Gridserach provides optimal combination of hyper parameters based on predefined range of values corresponding to each parameters.it can be utilized to obtain optimal model rather than trying different values for each parameters. This method uses cross-validation techniques to train and test model in order to achieve most appropriate hyper parameters set. although, Gridsearch is practical methods to find hyper parameters set, it demand high computation time,as the number of input parameters increase the computaion time inccreasing significantly [51]. table below shows the input parameters to Gridsearch and optimal output parameters set for both randomforest and mlp regression.

methods	Input parameters	Output optimal set
Random Forest Regression	n_estimators = [np.linspace(start = 200, stop =600, num = 3)] max_depth = [np.linspace(10, 20, num = 5)] min_samples_split = [2, 5] min_samples_leaf = [1, 2] bootstrap = [True,False]	n_estimators: 400 min_samples_split: 5 min_samples_leaf:1 max_depth: 10 bootstrap: True
MLP Regression	hidden_layer_sizes: [(1,),(50,)] batch_size = [100, 500] Max_itr:[1000,20000] Activation : ["identity", "logistic", "tanh", “relu"]	batch_size = 200 max_iter=10000 hidden_layer_sizes = 9 Activation=relu

Figure 5.1: table of optimal combination of hyper parameters providing by Gridsearch Method

as it shown in the table, the Gridsearch, could find the best models parameters based on training data and primarily range of desired parameters. Based on the results for random forest, 400 trees with using bootstrap technique is appropriate selection. Load Lag lentgh is another issue that should be taken into consideration. We considered different number of load lags from 1 to 12, for instance, load consumption of previous 1 to 3 days, load consumption of previous1 to 6 days, load consumption of previous 1 to 9 days and finally load consumption of previous 1 to 12 days. Then four different load Lags were fed into model and the results were evaluated for both forecasting models by random forest or MLP. Table below demonstrates evaluation of erorr metrics on training data for different load Lag sizes or input variable for two regression models. Model hyper parameters were selected based on Gridsearch techniques are indicated in the table....

Model	RMSE	MAE	MSE
Random Forest Regressor	1.25	0.96	2.1
MLP Regressor	0.083	0.069	0.07

Input size 3

Model	RMSE	MAE	MSE
Random Forest Regressor	1.19	0.91	2.08
MLP Regressor	0.082	0.061	0.069

Input size 6

Model	RMSE	MAE	MSE
Random Forest Regressor	1.101	0.803	2.04
MLP Regressor	0.061	0.057	0.067

Input size 9

Model	RMSE	MAE	MSE
Random Forest Regressor	1.26	1.04	2.36
MLP Regressor	0.85	0.076	0.071

Input size 12

Figure 5.2: Error analysis with respect to the multitude of input variables or Lag size

The figure report obtained error by applying models on different Lag size. As the number of Lag increased the error decreased by Lag length 9 (number of input days per one output), then the model accuracy increased. This behavior implies to the fact that selecting number of appropriate Lag size can be crucial in deterministic time series forecasting. It is concluded that MLP model compared to Random Forest models reached the better performance by considering MSE, MAE and RMSE values in both data subsets (with temperature without temperature data).the RMSE values for MLP and Random Forest are 1.101 and 0.061 respectively for Lag size 9. In the single variable subset the result in temperature included subset for Lag size 18 are 0.889 and 0.039 for both Random Forest and MLP techniques. In fact, the error metrics values for all two models decreased when weather feature added to subset. The results show that the models performances increase by adding more relevant input features.

as it was noted weather parameters are other features which influence considerably load forecasting models. Here, we chose to utilize outside temperature as a weather parameter which can be fed to the predictive model beside energy consumption values. Weather dataset contains daily min- temperature and max-temperature, which should be converted to time series format. We calculated daily mean-temperature based on mentioned features. Table below illustrates evaluation of error metrics based on daily temperatures and total energy consumption for both deterministic models.

Model	RMSE	MAE	MSE
Random Forest Regressor	0.98	0.84	1.8
MLP Regressor	0.043	0.054	0.004

Input size 6

Model	RMSE	MAE	MSE
Random Forest Regressor	0.94	0.81	1.91
MLP Regressor	0.034	0.042	0.0031

Input size 12

Model	RMSE	MAE	MSE
Random Forest Regressor	0.884	0.78	1.83
MLP Regressor	0.039	0.046	0.0036

Input size 18

Model	RMSE	MAE	MSE
Random Forest Regressor	00.987	0.97	1.98
MLP Regressor	0.0396	0.076	0.0051

Input size 24

Figure 5.3: Error analysis with respect to the number of input variables or Lag size with temperature

After applying model on training and test data , models were evaluated with well-know error metrics for regression, including MSE,MAE and RMSE.Above tables, show evaluation of calculated error for all models.As presented in th figure, random forest hits maximum error rate in comparison to the model MLP model in both data subsets (with and without temperature).

5.2 probabilistic forecasting

As it was described, Probabilistic forecasting devotes a probability of possible value and produce set of predicted values rather than one single value. Since, probabilistic forecasts are not precises and accurate, the metrics are required to verify accuracy of distinct probabilistic forecasts [50]. In this thesis, three well-known supervised deep learning methods are applied on train data to obtain set of forecasts per input. DeepAR+ as a ANNs algorithm provided by Amazon for forecasting time series using RNNs is the first method. When the dataset contains hundreds of feature time series, the DeepAR+ algorithm outperforms the standard ARIMA and ETS methods, then the performance of model can be improved by increasing number of input variables [27].

Another deployed algorithm to build a predictive model is FeedForwardEstimator which implemented based on simple MLP model predicting,and the next target time-steps given the previous ones. In a feed-forward architecture (FFNN), there is no feedback and intra-layer connections between neurons. The weights and bias of the network are estimated using a training algorithm such as the back-propagation algorithm.[41]. The last method to apply on the residential data, builds a local time series model using Gaussian Processes (GP). Each time series has a GP with its own hyper-parameters [1]. table below depict information about important hyper parameter which were tuned to fined appropriate parameters for three selected algorithms DeepAR algorithm.

DeepAR	simple_feedforward	Gaussian Process
freq="D" context_length=45 prediction_length=30 num_layers = 2, num_cells = 40 distr_output=StudentTOut put() dropout_rate=0.01 epochs=50	freq="D", context_length=150, prediction_length=30, distr_output=StudentTOutput() epochs=50 learning_rate=1e-3, num_batches_per_epoch=100 patience=10	freq="D" prediction_length=30 cardinality = 10 epochs=100

Figure 5.4: optimal set of hyper parameters for 3 models

common hyper parameter among all probabilistic approach forecasting is Freq that refers to time series frequency.

5.3 Discussion

the best deterministic and probabilistic models were selected based on minimum corresponding error rates. In the probabilistic forecasting when the models had slight differences in error rate we considered computation time to select best model.Table below provide information about best hyper parameters, size of training data, computation time. The selected probabilistic models have been applied on the train and test data in this study. The result of each model evaluated by error metrics to figure out the accuracy of the models. Unlike to previous step, the data set contains total daily consumption of all groups divided in train and test set. Since Probabilistic model is able to convert data to time series Lag, it is not necessary to convert data in preprocessing step of this phase.Those models are able to produce multiple times series Lag based on input data. +Models were fed with train and test data one time with and one time without temperature, besides mentioned optimal hyper parameters. Following tables presents the evaluation of forecasting models, resulted by utilizing predefined metrics for 30 days ahead predictions. The provided information imply that DeepAr model slightly outperforms compared to FFNN and Guassian Process in daily load forecasting.

Model	RMSE	MAPE	CRPS
DeepAR	0.056	0.063	0.09
simple_feedforward	0.05	0.06	0.101
Gaussian Process Estimator	0.04	0.05	0.13

Figure 5.5: Error analysis with respect to the number of input variables or Lag size with temperature

Model	RMSE	MAPE	CRPS
DeepAR	0.073	0.09	0.1603
simple_feedforward	0.080	0.1	0.1865
Gaussian Process Estimator	0.089	0.102	0.21

Figure 5.6: Error analysis with respect to the number of input variables or Lag size with temperature

As the figures show DeepAr model compared to FFNN models reached the better performance by considering CRPS and MAPE values.However, DeepAr and FFNN find better prediction sets in comparison to Gaussian Process in both subsets(with temperature and without temperature data)the CRPS values for DeepAR, FFNN and Gaussian Process are 0.09,0.101 and 0.13 respectively. In fact, the error metrics values for all three models decreased when weather feature were added to subset. The results show that the models performances increase by adding more relevant input features. As it was commented, in probabilistic forecasting, uncertainty of the results are higher. Therefore CRPS metrics was implemented to compare models accuracy. The CRPS values for DeepAr, FFNN and Gaussian Process for consumption data are 0.1603,0.1865 and 0.21 respectively. Figure below show time series curve of electricity consumption over 2 years and histogram related to distribution of total daily consumption. As results show in term of computation time, FFNN outperform better than 2 other models with shorter computation time. While DeepAR has a higher model complexity than two others.

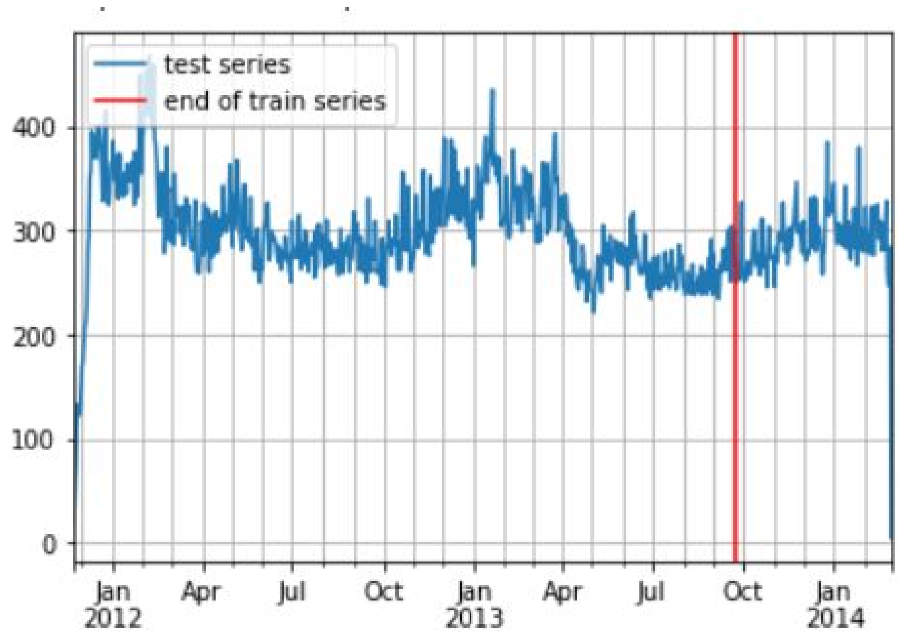


Figure 5.7: time series cure depicts total daily consumption of houses over two years

figures below demonstrates probabilistic prediction electricity consumption for three model separately,

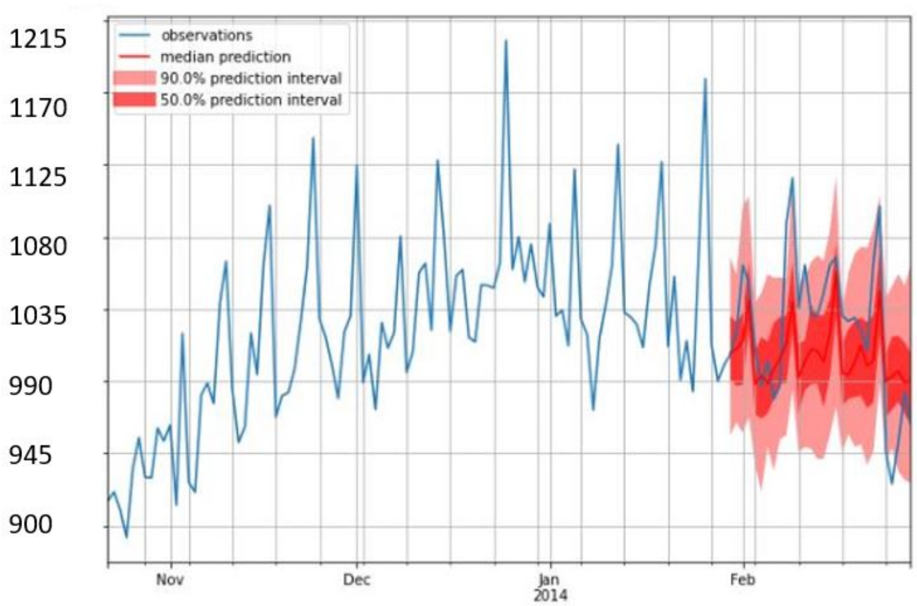


Figure 5.8: Real total daily consumption of houses in 3 group versus predictions by DeepAr model over two years

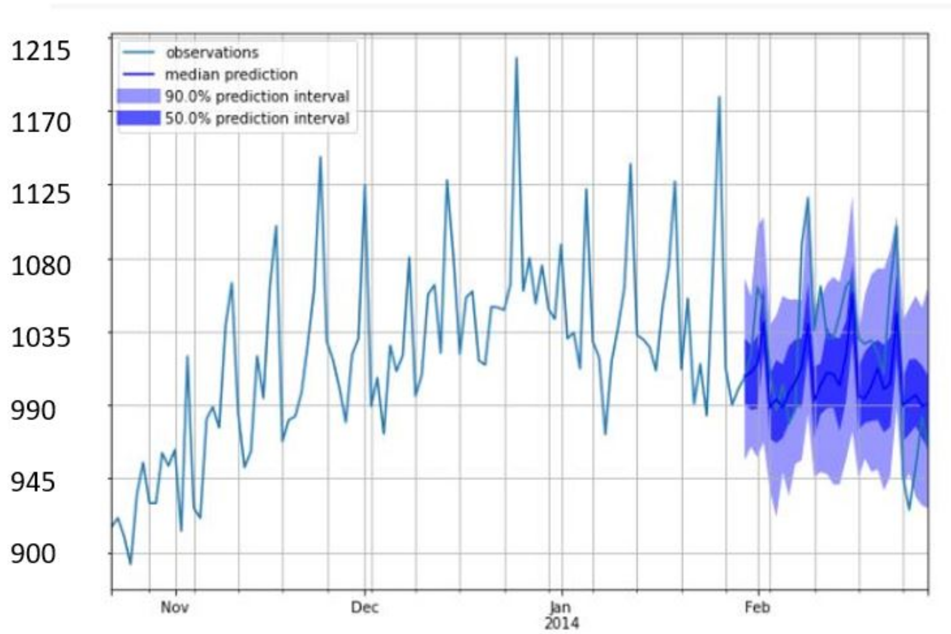


Figure 5.9: Real total daily consumption of houses in 3 group versus predictions by FeedForward model over two years

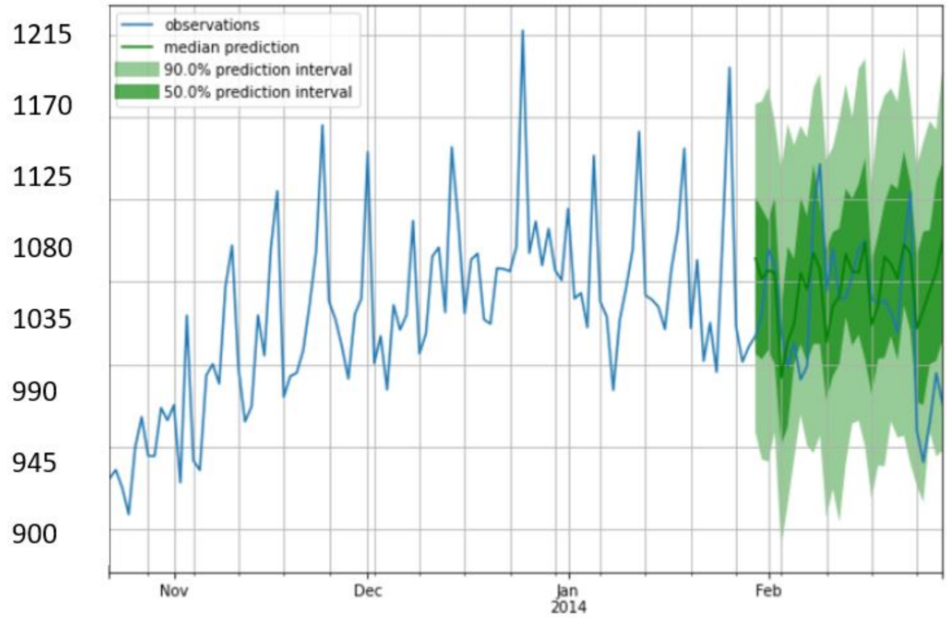


Figure 5.10: Real total daily consumption of houses in 3 group versus predictions by Gaussian process model over two years

As it is illustrated in figures, DeepAR and FFNN curves have proper match and steady daily prediction for three days ahead forecasting in their quantile. Although, the average line of predictions per points are slightly far from input spots, the forecasting curves follow similar trend as real data time series curve. On the other hand, forecasting vectors for Gaussian Process covered wider range than previous algorithms and the average line is not exactly in the same trend to real data.

Chapter 6

Conclusions and Future Work

6.1 Conclusion

This study has presented an analytic comparison of day-ahead load forecasting during a period of two years by applying AI based data driven models. The project analyzes electricity consumption of houses by applying two different forecasting approaches, deterministic and probabilistic. In the deterministic step the two methods of Random Forest Regression and MLP Regression was employed to make a predictive models. In the probabilistic phase,DeepAR, FFNN and Gaussian Process Estimator were employed to predict future days load forecasting. The result of models were assessed by well-known error metrics as RMSE,MAE, MSE and CRPS separately for each phase. The models were trained on UK residential smart meter historical data. The subset were selected from 3 various groups of costumer with registered diversified load volatility level and daily consumption. In addition daily weather(temperature) data was added to subset to check model sensitivity and correctness. Moreover, in both deterministic and probabilistic forecasting models, different time series Lag were fed to models to find more accurate model. The main achievements and findings of this study are as follows:

- The Deep Learning methods of FNN, DeepAR,MLP compared to other utilized methods like Random Forest and Gaussian can provide better better results considering evaluation criteria of following similar trend to real load consumption, lower forecasting error and computation time.

- The forecasting models were able to predict 30 days ahead forecasting by acceptable errors although only energy consumption and mean temperature were fed to models.

- With regards to deterministic forecasting MLP outperformed Random Forest model based on forecast consumption error, while temperature was included in the data for both models

- In the probabilistic forecasting, DeepAR could provide better results than FFNN and Gaussian Process models. Although the computation time of FFNN was lower than others.

- It turned out that increasing number of input variable and training houses improve the accuracy of forecasts in employed methods on chosen data subsets but increase model complexity and computation time.

- In probabilistic forecast adding weather data could improve model performance and deduct error values for all models. It was figured out in the study that the seasons and months with lower outside temperature could have higher load violations that makes predicting more difficult

- It is also concluded in this study that Hyper parameters tuning has crucial role in model performance and achieving more accurate model.

6.2 Future Work

This study creates basis for future studies in household load forecasting models by expanding the number variables into model to improve the results. In such circumstances the models can be trained with more weather features to check out model outputs and accuracy. In the further work in short term load forecasting, clustering algorithms could be deployed to divide data set in meaningful clusters. Further achieved clusters could be fed into model for for further analysis. Such approach can help to make better comparison based on data diversity, as well as, finding new group of consumers with similar consumption pattern. Increasing power load balance flexibility by utilizing probabilistic forecasting model in the case study can also be applied in future works. Developing on line AI based probabilistic load forecasting by using real time data from houses is also another suggested expansion to this thesis work.

REFERENCES

[1] Aida Mehdipour Pirbazari, Mina Farmanbar, Antorweep Chakravorty and Chunming Rong ,short-Term Load Forecasting Using Smart Meter Data:A Generalization Analysis,MDPI,2020,

[2] Farmanbar, M.; Parham, K.; Arild, Ø.; Rong, C,A Widespread Review of Smart Grids Towards Smart Cities}, Energies, 2019

[3] Pirbazari, A.M.; Chakravorty, A.; Rong, C, Evaluating Feature Selection Methods for Short-Term Load Forecasting}, In Proceedings of the 2019 IEEE International Conference on Big Data and Smart Computing (BigComp),February,2019, 1-8

[4] Mocanu, E.; Nguyen, P.H.; Gibescu, M.; Kling, W.L, Comparison of machine learning methods for estimating, energy consumption in buildings, In Proceedings of the 2014 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), JULY, 2014,1-6

[5] Tahreem Anwar, Bhaskar Sharma, Koushik Chakraborty and Himanshu Sirohia Jayoti Vidyapeeth, introduction to Load Forecasting}, International Journal of Pure and Applied Mathematics},2018

[6] Zhao, H.; Magoulès, F, A review on the prediction of building energy consumption, Renew. Sustain. Energy, 2012, 16, 3586–3592

[7] Grzegorz Dudek, Short-Term Load Forecasting Using Random Forests, Intelligent Systems book, January, 2015, 821-828

[8] Bouktif, S.; Fiaz, A.; Ouni, A.; Serhani, M, Optimal Deep Learning LSTM Model for Electric Load Forecasting using Feature Selection and Genetic Algorithm: Comparison with Machine Learning Approaches, Energies, November, 2016

[9] Kong.w, Regression analysis for prediction of residential energy consumption, Renew. Sustain. Energy Rev, 2015, 332–343

[10] Eva Gonzalez-Romera, Miguel A Jaramillo-Moran, and Diego Carmona-Fernandez, IEEE Transactions on power systems, 2006, 332–343

[11] Baliyan, A.; Gaurav, K.; Mishra, S.K., A Review of Short Term Load Forecasting using Artificial Neural Network Models, Procedia Computer, 2015, 121–125

[12] Pappas, S.S.; Ekonomou, L.; Karamousantas, D.C.; Chatzarakis, G.E.; Katsikas, S.K.; Liatsis, P, Electricity demand loads modeling using AutoRegressive Moving Average (ARMA) models}, Energy , 2008, 1353–1360

[13] XiaoshuLüaTaoLuaCharles J.KibertbMarttiViljanena, Journal of Applied Energy}, April,2015, 261-275

[14] Lahouar, A.; Slama, J.B.H., Day-ahead load forecast using random forest and expert input selection, Energy Convers. Manag, April, 2015, 1040–1051

[15] Ayas Shaqour a , Tetsushi Ono b , Aya Hagishima a , Hooman Farzaneh, Electrical demand aggregation effects on the performance of deep learning-based short-term load forecasting of a residential building, journal of Energy AI, 2022

[16] Karol Bot, Samira Santos, Inoussa Laouali, Antonio Ruano and Maria da Graça Ruano, Design of Ensemble Forecasting Models for Home Energy Management Systems, MDPI, 2021

[17] Hambali, A.O.; Akinyemi, M.; JYusuf, N , Electric Power Load Forecast Using Decision Tree Algorithms, Comput. Inf. Syst. Dev. Inform. Allied Res,2017, 29-42

[18] Khosravani, H.; Castilla, M.; Berenguel, M.; Ruano, A.; Ferreira, P , A Comparison of Energy Consumption Prediction Models Based on Neural Networks of a Bioclimatic Building, energies,2016

[19] Chaturvedi, D.K., Satsangi, P.S. and Kalra, P.K, Load Frequency Control A Generalized Neural Network Approach, International Journal of Electrical Power & Energy Systems, 1999, 405-415

[20] Kuihe Yang Shijiazhuang, Lingling Zhao, Application of Mamdani Fuzzy System Amendment on Load Forecasting Model, Symposium on Photonics and Optoelectronics, 2009, 405-415

[21] Kong, W.; Dong, Z.Y.; Jia, Y.; Hill, D.J.; Xu, Y.; Zhang, Y, Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network, IEEE Trans. Smart Grid, 2019, 841–851

[22] Agrawal, R.; Muchahary, F.; Tripathi, M, Long Term Load Forecasting with Hourly Predictions Based on Long-Short-Term-Memory Networks, Texas Power and Energy Conference (TPEC), College Station, 2018

[23] Kim, T.-Y.; Cho, S.-B, Predicting residential energy consumption using CNN-LSTM neural networks, Energy,2019, 72–81

[24] Mamun, A.; Hoq, M.; Hossain, E.; Bayindir, R. A, Hybrid Deep Learning Model with Evolutionary Algorithm for Short-Term Load Forecasting, International Conference on Renewable Energy Research and Applications (ICRERA),2019,

[25] Jarosław Protasiewicz; Jakub S. Sowiński, Improvement of MLP models for forecasting electrical energy consumption using OBD and OBS methods, International Conference on Industrial Technology (ICIT), November, 2015

[26] Probabilistic_Load_Forecasting_Using_Post-Processed_Weather_Ensemble_Predictions}, Available from:https://www.researchgate.net/publication/348248021_Probabilistic_Load_Forecasting_UsinPost-Processed_Weather_Ensemble_Predictions,2022

[27] David Salinas, Valentin Flunkert, Jan Gasthaus, Tim Januschowski, DeepAR: Probabilistic forecasting with autoregressive recurrent networks, International Journal of Forecasting, 2020, 1181-1191

[28] Hilaf Hasson, Bernie Wang, Jan Gasthaus, Probabilistic Forecasting: A Level-Set Approach, NeurIPS Computer Science ,2021

[29] Petrônio Silva and Frederico Gadelha Guimaraes, Probabilistic Forecasting with Seasonal Ensemble Fuzzy Time-Series, Academia.edu, 2017

[30] Prentice Hall, Big Data Fundamentals: Concepts, Drivers & Techniques, By Thomas Erl, Wajid Khattak, Paul Buhler, Part of the The Prentice Hall Service Technology Series from Thomas Erl series,2016

[31] Big Data Now: 2012 Edition, by O'Reilly Media, Inc. October 2012, http://oreilly.com/catalog/errata.csp?isbn=9781449356712, 2012

[32] Saurabh , Inderveer Chana, A Survey of Clustering Techniques for Big Data Analysis, Department of Computer Science and Engineering Thapar University Patiala,India, 2012

[33] Charu C. Aggarwal, Chandan K. Reddy, Data Clustering: Algorithms and Applications, Chapman and Hall/CRC ,Reference - 652 Pages - 102 B/W Illustrations ,ISBN 9781466558212 - CAT# K15510, 2013

[34] Cleaning data: Wikipedia,https://en.wikipedia.org/wiki/Data cleansing, 2015

[35] Rohit Sharma, normalization-in-data-mining:https://www.upgrad.com/blog/normalization-in-data-mining, 2020

[36] Jaroslaw Protasiewicz, Improvement of MLP models for forecasting electrical energy consumption using OBD and OBS methods, International Conference on Industrial Technology, 2015

[37] simplilearn, An Overview on Multilayer Perceptron (MLP):https://www.simplilearn.com/tutorials/deep-learning-tutorial/multilayer-perceptr, 2022

[38] by Amazon, DeepAR Forecasting Algorithm, www://docs.aws.amazon.com/sagemaker/latest/dg/deepar.html

[39] Oscar Knagg , An intuitive guide to Gaussian processes, Toward Datascience, 2019

[40] Gaussian Process:https://ts.gluon.ai/v0.8.x/_modules/gluonts/model/gp_forecaster/gaussian_process.html

[41] Géron, A., Hands on Machine Learning with Scikit-Learn and Tensorflow, O'Reilly Media: Sevastopol, 2019

[42] Tianfeng Chai, Root mean square error (RMSE) or mean absolute error (MAE), National Oceanic and Atmospheric Administration, 2014

[43] SmartMeter Energy Consumption Data in London Households—London Datastore, https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-househ lds (accessed on 10 February 2020)

[43] Available online: https://acorn.caci.co.uk/downloads/Acorn-User-guide.pdf, https://acorn.caci.co.uk/downloads/Acorn-User-guide.pdf(accessed on 23 February 2022)}

[45] J. M. Hernandez-Lobato and R. Adams, Probabilistic backpropa-gation for scalable learning of bayesian neural networks, the 32nd International Conference on Machine Learning, 2015, 1861–1869

[46] Lingxue Zhu,Nikolay Laptev, Deep and Confident Prediction for Time Series at Uber,

arXiv:1709.01907v1 [stat.ML], 2017

[47] D.W. van der Meer, J. Widén, J. Munkhammar, Review on probabilistic forecasting of photovoltaic power production and electricity consumption, Renewable and Sustainable Energy Reviews, 2018, 1484-1512

[48] Asare-Bediako, B.; Kling, W.L.; Ribeiro, P.F, Day-ahead residential load forecasting with artificial neural networks using smart meter data, IEEE Grenoble Conference, Grenoble, 2013, 1-6

[49] uiz-Abellón, M.; Gabaldón, A.; Guillamón, A, Load Forecasting for a Campus University Using Ensemble Methods Based on Regression Trees, Energies, 2018

[50] Alexandrov, A., Benidis, K., Bohlke-Schneider, M., Flunkert, V., GluonTS: probabilistic time series models in Python, ICML time series workshop, 2019

[51] Grid Search Explained – Python Sklearn Examples,Data Analytic,https://vitalflux.com/grid-search-explained-python-sklearn-examples/, 2020

[52] Pérez-Lombard, L.; Ortiz, J.; Pout, C, A review on buildings energy consumption information, Energy Build., 2008, 394-398

[53] Ahmed, N.K.; Atiya, A.F.; Gayar, N.E.; El-Shishiny, H, An Empirical Comparison of Machine Learning Models for Time Series Forecasting, Comput. Inf. Syst. Dev. Inform. Allied Res,2010, 594–621