8-2020

Load Forecasting Analysis Using Contextual Data and Integration With Microgrids Used For Off Grid EV Charging Stations

Ashit Neema

Utah State University

LOAD FORECASTING ANALYSIS USING CONTEXTUAL DATA AND INTEGRATION WITH MICROGRIDS USED FOR OFF GRID EV CHARGING STATIONS

by

Ashit Neema

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Computer Science

Approved:

Nicholas Flann, Ph.D. Major Professor	John Edwards, Ph.D. Committee Member
Xiaojun Qi, Ph.D. Committee Member	Richard S. Inouye, Ph.D. Vice Provost for Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2020

ABSTRACT

Load Forecasting analysis using contextual data and Integration with microgrids used for off grid EV charging stations

by

Ashit Neema, Master of Science

Utah State University, 2020

Major Professor: Nicholas Flann, Ph.D.
Department: Computer Science

Moving to a low carbon economy requires that every device run on energy made from clean and renewable sources. A smart electrical grid is needed that can efficiently manage energy production and consumption as solar and wind replace coal and natural gas, loads increase through widespread electric vehicle (EV) adoption, and storage becomes ubiquitous. Micro-grids integrate local energy generation and storage, local loads and can run in isolation or with a grid connection. Microgrids increase the adoption of solar power, improve grid resilience and expand electrical power availability in developing countries. This work considers how predicting the electrical load of a microgrid has the potential to improve the efficiency and profitability of the microgrid. The main challenge in electrical load prediction is to gather the contextual data that affects the upcoming loads. In this thesis, load or demand energy is being predicted in two distinct settings by utilizing two different machine learning models. First, in the residential sector using the ARIMA model(which gave prediction accuracy of 88%) and second, for EV charging stations using multiple linear regression model (which gave prediction accuracy of 82%). To evaluate the impact of prediction on profitability, the load predictor is integrated into a microgrid optimizer along with a solar prediction model. The results show that prediction of both energy production and demand improve the optimality of microgrid operations.

(56 pages)

PUBLIC ABSTRACT

Load Forecasting analysis using contextual data and Integration with microgrids used for off grid EV charging stations

Ashit Neema

Electricity is an essential component of the smooth working of every sector. If a successful prediction of how much electricity will be required for say the next 24 hours or 48 hours can be made, it will not only help in efficiently planning the activities and operations but also help in minimizing the cost incurred. In this thesis the same is being attempted, first, a model is created that can predict the energy consumption of households using various tools available. To achieve this, historical data of the past 5 years that has been recorded in London has been used. Secondly, a model is created that can forecast future energy requirements of EV charging stations, which will help in optimizing their working in off-grid areas. To achieve this, data that has been recorded for more than 100 charging stations across the Salt Lake area for around 4 years has been used.

The second part of this thesis involves integrating the load forecasting model with the solar energy forecasting model and microgrid optimization. Since microgrid can disconnect from the energy source and work autonomously, solar energy can be used to generate power that can be used instead of the energy that is bought from the grid. This will help in building an efficient model that maximizes the profit by making sure energy is bought at a minimum price and the same energy is being utilized when the demand is high.

To my parents, for everything they have done to get me to this day.

ACKNOWLEDGMENTS

I would like to thank my supervisor Professor Nicholas Flann for the exceptional support and for introducing me to the research. To all my co-authors thank you for the encouragement and the engaging collaborations: You have taught me a lot! Special thanks to Utah state University, Select Labs and Rocky mountain power for providing the infrastructure and resources. Last but not least, I want to thank my family and friends for their overwhelming support.

Ashit Neema

CONTENTS

LIST OF FIGURES

ACRONYMS

EV	Electric Vehicle
ARIMA	Auto Regressive Integrated Moving Average
API	Application Programming Interface
RMP	Rocky Mountain Power
RL	Reinforcement Learning
RMSE	Root mean squared error
NOAA	National oceanic and atmospheric administration

CHAPTER 1
INTRODUCTION

Electrical energy is a vital resource to drive industries. It has to be available whenever there is a demand for it. It is, therefore, imperative for the electric power utilities that the load on their systems be estimated in advance. This estimation of future load is commonly known as load forecasting [1].

The operation and planning of a power utility company require an adequate model for electric power load forecasting. Load forecasting assists an electric utility to make conclusions regarding a decision on, generating and purchasing electric power, load switching, and voltage control. Electric load forecasting is the process used for predicting electric load, given historical load and other relevant parameters.

Load forecasting also assists in deciding and planning for maintenance of the power systems. By understanding the demand, the utility can know when to carry out the maintenance and ensure that it has a minimum impact on the consumers. For example, they may decide to do maintenance in residential areas during the day when most people are at work, and demand is very low. [2].

As a part of this thesis, load forecasting has been studied for two separate application sectors which are mentioned below:

- A load forecasting study was performed in the residential sector to devise a schedule to minimize the energy consumption of households.
- A load forecasting study was performed for electrical vehicle (EV) charging stations, to devise an optimal policy for their cost-effective use in off-grid areas.

To build an accurate load forecasting model, historical values of load data, as well as other parameters affecting the load energy are required. For the study of load forecasting in the residential sector, the data used was gathered by the UK government using smart meters.

How the data was collected : To better follow energy consumption, the UK government wanted energy suppliers to install smart meters in every home in England, Wales, and Scotland. There are more than 26 million homes for the energy suppliers to get to, with the goal of every home having a smart meter by 2021. These smart meters collect the load consumption data of every household on an hourly basis. This data can be used to further study the consumption pattern of the houses and to devise a plan to minimize consumption [3].

To generate a plan or a schedule that minimizes the energy consumption of a particular household, load consumption values of several days ahead is needed. Using the data provided by the smart meters and other resources, a predictive model has been built in this work that can achieve the aforementioned result.

Time series nature of the dataset: The data that was collected by the smart meters is made up of a sequence of data points taken at successive equally spaced points in time.

Forecasting data using time-series analysis comprises the use of a model to forecast future conclusions based on known past outcomes. There are several modeling techniques available for making predictions on time series data, one of which is the ARIMA model [9].

ARIMA stands for auto-regressive integrated moving average. It's a way of modeling time series data for forecasting (i.e., for predicting future points in the series), in such a way that:

- A pattern of growth/decline in the data is accounted for ("auto-regressive")
- The rate of change of the growth/decline in the data is accounted for ("integrated")
- Noise between consecutive time points is accounted for ("moving average")

The previous studies made on ARIMA models showed that the accuracy of the model increases if additional independent parameters, which affected the dependent parameter(in this case, load consumption), are included in the list of inputs given to the model [10]. For this purpose, various studies were conducted to identify additional parameters which had an affect on load consumption. It was observed that load consumption was affected by the current weather conditions, like the outside temperature. The studies also showed that that load consumption of a given household was different for a working day than on a weekend and a holiday.

To gather the weather data for the area of London, the dark sky API [5] was utilized. The time resolution of the data was the same as that of load consumption (i.e in hourly intervals).

The weather data includes parameters like temperature, humidity, cloud coverage, dew point, precipitation, etc., but not all the parameters affected load consumption. To identify the useful parameters that had an impact on load and which can be used as an input to the ARIMA model, correlation graphs were plotted which are illustrated in Fig. 1.1 and Fig. 1.2. The figures show that temperature had an inverse relationship with energy consumption i.e. when temperature decreased the energy consumption increased, and humidity had a direct relationship with energy consumption. Based on these parameters a weather cluster was generated for every hourly value by using the technique of K-means clustering. A detailed explanation of this technique is given in Chapter 3 of this work.

The effect of holiday on load consumption was analyzed in a similar manner, and based on the results, it was identified as an independent variable that needed to be added to the list of inputs. By using the starting date, a holiday index was generated, with one being given to the date that is a national holiday and 0 to a normal working day.

The next step involved gathering the contextual data that had an impact on load consumption. This data was also provided by the smart meters in the form of ACORN values.

ACORN is a consumer classification developed by the government that segments the UK population. By analyzing demographic data, social factors, population, and consumer behavior, it provides an understanding of different groups of people. When the UK govern-

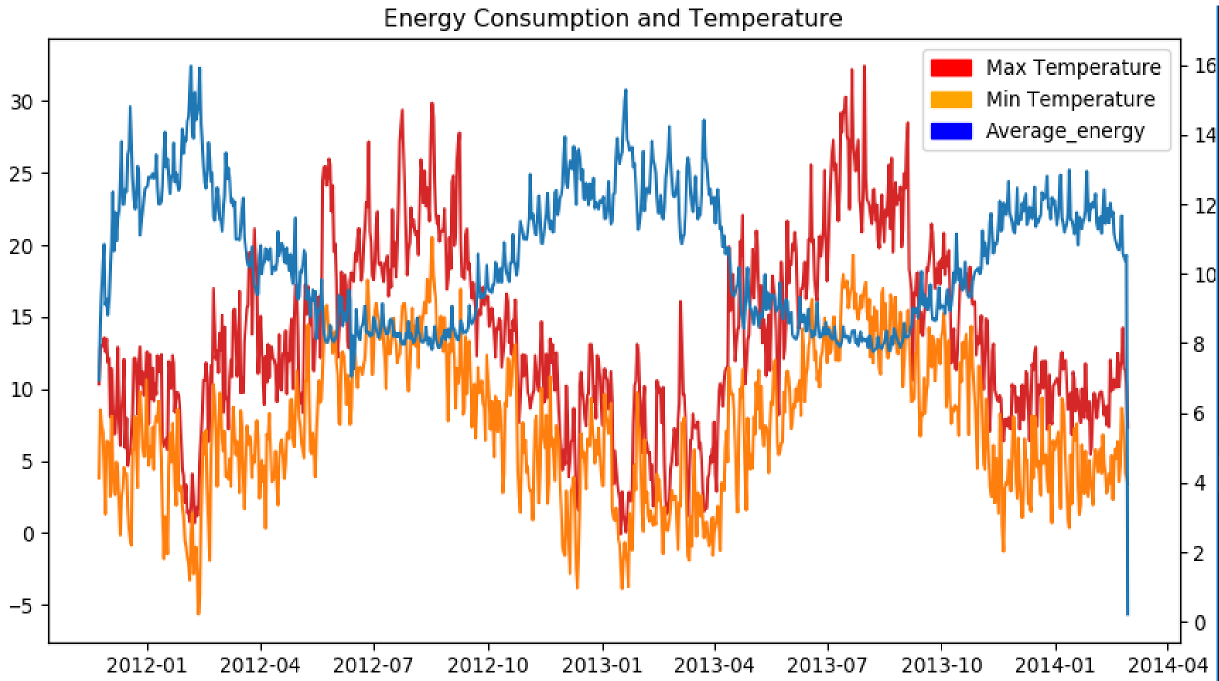

Fig. 1.1: Temperature correlation with energy consumption where left y axis denotes the temperature and right y axis denotes the average energy consumed on a day. The x axis denotes the date on which observation were made.

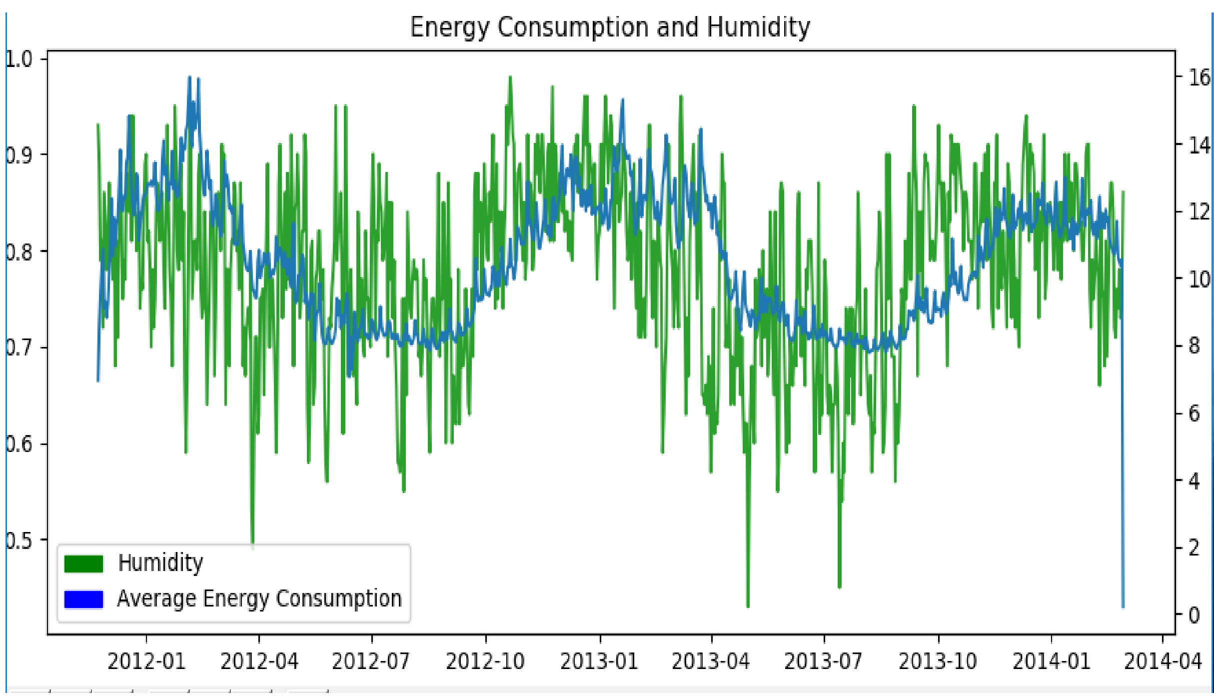

Fig. 1.2: Humidity correlation with energy consumption where y axis denotes the humidity and x axis denotes the date on which observation were made.

ment installed smart meters in households, it also classified them into Acorn groups. Acorn segments households into groups based on the number of people living in the house, their ethnicity, their occupation, their age, their financial status, etc. See [7] for more details.

The Acorn values of different households and their impact on load consumption were analyzed, and it was concluded that households belonging to affluent groups had more load consumption than those which were in financially stretched groups. Other Acorn values that affected the load consumption were, size of the house, the number of people in the house and the age group of people in the house.

After the input parameters were decided, an ARIMA model was built which gives the predicted load consumption as an output when the following input parameters were provided to it:

- Date(represented in mm/dd/yyyy format and signifies the day on which predictions are made)
- Weather cluster(0-3)
- Holiday Index(0 or 1)
- Acorn Value 1
- Acorn Value 2
- Acorn Value 3
- Acorn Value 4

A detailed explanation of how the ARIMA model was built is given in Chapter 3 of this work. Since forecasting the demand also plays a vital role in the commercial sector, the next section comprises of the load forecasting study done for **EV charging stations**.

The consistent growth and development of electric vehicles have created many new challenges for the charging stations and power systems. This is due to changes to load profiles based on EV demand, fast charging requirements, the need to keep up with the growing demand and many more. With the number of EV's gradually rising, and the charging stations working 24 hours, it was necessary to record their data continuously to meet the ever-increasing demand and also to prevent disruptions to the grid. With this idea as a foundation, RMP (Rocky mountain power) decided to set up data meters in EV charging stations located within and near Salt Lake City to better understand the characteristics of demand, and to build an efficient management system. These meters record the amount of energy consumed by the EV charging station on an hourly basis. It also records other important features like the port which was used for charging, the amount of time for which a vehicle used a charging port, and other data described later. This dataset also had a similar time series nature, since the data points were recorded every hour.

The effectiveness of such a management system is heavily dependent on the prior knowledge of future demand for energy. This knowledge can be provided by accurate time-series load forecasting techniques. The technique on which this dataset was analyzed is **multiple linear regression**.

Regression analysis is a form of predictive modeling technique that investigates the relationship between a dependent and an independent variable. This technique is used for forecasting, time series modeling and suggesting causal effects among the input and output variables. In multiple linear regression, there is more than one independent variable [13]. The regression model for this study was built using tools and libraries available in python programming language. The process of building this model is discussed in Chapter 3 of this work.

The observations made on the regression model being used for forecasting load consumption values verified that as the independent input parameters are increased the accuracy of the model also increases. To find those parameters which had maximum impact on load consumption values an analysis was performed summarised in Fig. 1.3 and Fig. 1.4.

Fig. 1.3 indicated that the frequency of people arriving at a charging station varies by the hour. As the number of people increased the energy consumption also increased. Thus the load consumption had different values at different hours. For example, most charging stations had not more than one customer at 2:00 am, thus the demand of energy was minimum at this time. Since the forecasting was achieved on an hourly basis this was an important feature that needed to be added in the input parameters.

The graph in Fig. 1.4 indicated that the energy consumption at a given hour of the day changed based on whether the day was during the work week or at a weekend. Since this affected the load consumption a weekday-index was added to the list of input parameters.

Other parameters were also added after some preprocessing was done on the dataset. Once the preprocessing was done it was observed that the load consumption at a given hour was also affected by values of load consumption at previous hours. For example, the amount of energy consumed at 2:00 pm is directly affected by the amount of energy consumed several hours before that time. Based on this observation, 24 new input parameters for the energy consumption values of the past 24 hours, were added to the system.

It was also observed that the dependent variable (i.e. load consumption) was different for a different season. For example, the load consumption for a given hour of the day increases in the summer season and less in winter, as people use electric vehicles more during the summer.

When all the parameters impacting load consumption were identified, a forecasting model using the linear regression technique was built, using the following parameters as its input:

- Date (represented in mm/dd/yyyy format and signifies the day on which predictions are made)
- Hour of the day (represented in 24 hour format and signifies the starting hour for prediction of load consumption)
- Weekday-index(0 or 1)
- Holiday-index(0 or 1)
- Season-index(0-3)
- 24 values of energy consumption of previous 24 hours

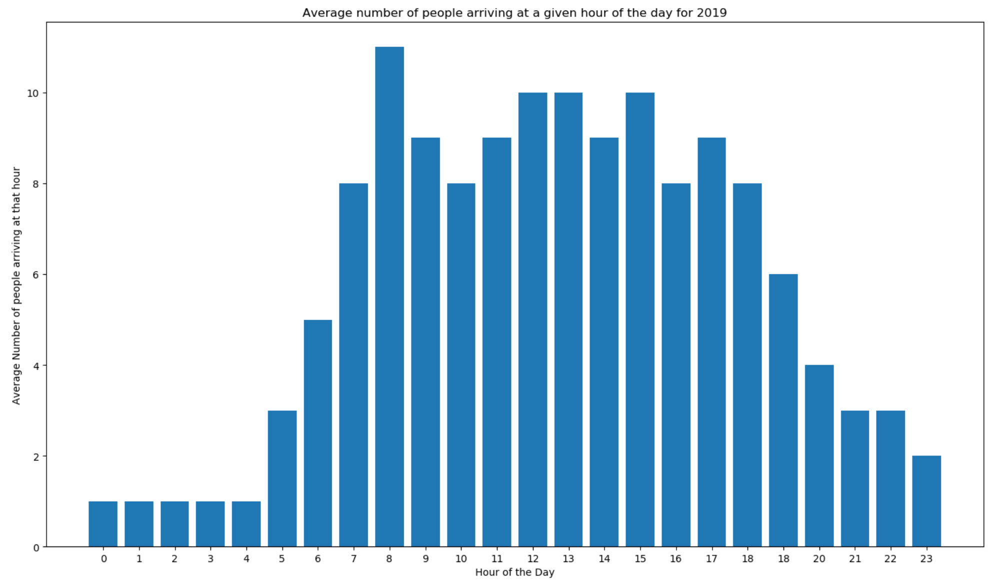

Fig. 1.3: Average Number of people arriving at a given hour of the day.

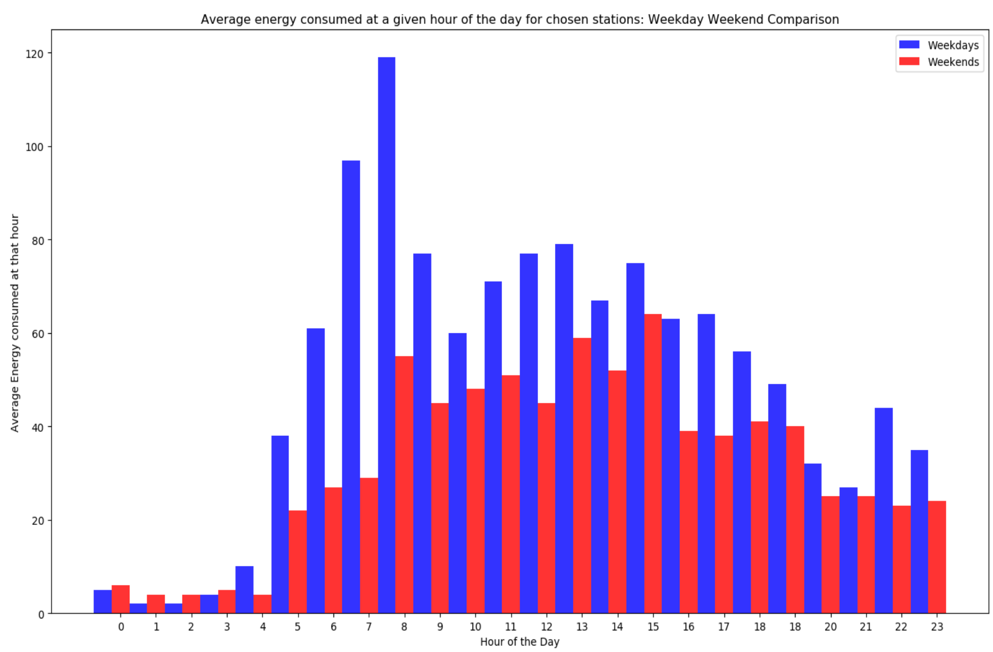

Fig. 1.4: Average energy consumed at given hour of the day: Weekday and Weekend comparison.

1.1 Problem Definition

Load forecasts are the building blocks for both short term and long term planning efforts, and to get accurate values of load consumption further into the future is a challenging task. Previous results show that the time resolution of the prediction increases and the time ahead of prediction increases, the accuracy decreases [8].

The core problem that is addressed in this work is to achieve accurate values of load consumption several days ahead of the given time in one-hour resolutions based on data sources described above.

To evaluate the usefulness of the load prediction within the larger context of EV charging station operation and planning, the predicted values are then used to devise an optimal working policy for the use of micro-grid in EV charging stations. Micro-grids combine local energy generation (often using solar panels) with battery storage and maybe on-grid or off-grid. The integration of microgrids with EV charging stations offers a new way for local entrepreneurs to build a profitable business, especially for locations that are geographically isolated and off-grid. In this situation, load prediction may be particularly useful in increasing the margins of such business operations.

The optimal policy of the micro-grid determines when and by how much to service loads, charge/discharge the battery and sell/buy electricity from the grid (if connected) such that profit is maximized. The schedule is determined by using a reinforcement Learning algorithm, called value iteration [16]. The optimizer requires accurate forecasted values of demand (load consumption) and supply (solar energy available). Once these values are known, they are integrated with the battery modeling of the grid and energy pricing of the day (if on grid).

1.2 Organization of this work:

This work is organized as follows:

- Chapter 2 of this thesis gives explanation about the methods used in building the prediction models.

- Chapter 3 elaborates on how the models were built, how the dataset was preprocessed and used, and the techniques used to obtain the results.

- Chapter 4 inlcudes the results that were achieved as part of this thesis.

- Chapter 5 includes the integration of the results with microgid to obtain an optimal working policy for EV chargng stations.

- Chapter 6 includes the conclusions made by analyzing the results of the studies performed as a part of this work.

CHAPTER 2
LEARNING METHODS

This chapter describes about the methods that were used and analyzed as part of this work. The following sections explain about the ARIMA model that is being used for forecasting load consumption of residential sectors.

ARIMA stands for auto-regressive integrated moving average. It's a way of modeling time series data for forecasting (i.e., for predicting future points in the series) using the past values. Its a combination of auto-regressive and moving average models [9].

2.1 Autoregressive Models

In a multiple regression model, forecasting of the variable of interest is performed using a linear combination of predictors. In an autoregression model, forecasting of the variable of interest is done using a linear combination of past values of the variable. The term autoregression indicates the regression is also performed between variable against itself.

Thus, an autoregressive model of order p can be written as:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ...\phi_p y_{t-p} + \varepsilon_t \quad [9] \tag{2.1}$$

where y_t is the variable of interest, ε_t is white noise, y_{t-1} and y_{t-2} are past values of variable of interest and c is a constant. ϕ_1 and ϕ_2 are the hyper-parameters which can be changed to obtain different results. This is like a multiple regression but with lagged values of y_t as predictors.

2.2 Moving Average

Rather than using past values of the forecast variable in a regression, a moving average model uses past forecast errors in a regression-like model.

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{q-1} \quad [9] \tag{2.2}$$

where y_t is the variable of interest, ε_{t-1} and ε_{t-2} are past forecasted errors, ε_t is white noise and c is a constant. θ_1 and θ_2 are hyper-parameters. This is referred as a **MA(q) model**, a moving average model of order q.

2.3 Non Seasonal ARIMA Model

If differencing(the process of computing the differences between consecutive observations of variable of interest in order to make a non-stationary time series stationary) is combined with autoregression and a moving average model, a non-seasonal ARIMA model is obtained. ARIMA is an acronym for AutoRegressive Integrated Moving Average. The full model can be written as:

$$y_t^{'} = c + \phi_1 y_{t-1}^{'} + ... + \phi_p y_{t-p}^{'} + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad [9] \tag{2.3}$$

where $y_t^{'}$ is the differenced series (it can be differenced more than once). The "predictors" on the right hand side include both lagged values of y_t and lagged errors. The other parameters have the same meaning as in 2.2 and 2.3. We call this an ARIMA(p, d, q) model, where:

p = order of the autoregressive part;

d = degree of first differencing involved;

q = order of the moving average part

2.4 Seasonal ARIMA Model

A seasonal ARIMA model is formed by including additional seasonal terms in the ARIMA models we have seen so far. It is written as $ARIMA(p, d, q)(P, D, Q)m$ [9]

where (p, d, q) = Non seasonal part of the model

(P, D, Q) = Seasonal part of the model

m = number of observations per year

The seasonal part of the model consists of terms that are similar to the non-seasonal components of the model, but involve backshifts of the seasonal period. For example, an ARIMA$(1,1,1)(1,1,1)_4$ model is for quarterly data (m=4), and can be written as:

$$(1+\phi_1 B)(1-\Phi_1 B^4)(1-B)(1_B^4)y_t = (1+\theta_1 B)(1+\Theta_1 B^4)\varepsilon_t \quad [9] \tag{2.4}$$

The additional seasonal terms are simply multiplied by the non-seasonal terms.

The above sections defined the ARIMA model and how it can be used to predict the variable of interest(in this work, load consumption of residential sector). The upcoming sections explain, what regression is and how can this technique be used for prediction of a variable.

2.5 Regression Analysis

Regression analysis is a form of predictive modelling technique that investigates the relationship between a dependent and an independent variable. This technique is used for forecasting, time series modelling and suggesting a causal effect relationship between the variables.

Regression analysis is an important tool for modelling and analyzing data. It fits a curve to the data points, in such a manner that the differences between the distances of data points from the curve is minimized [13].

2.6 Why Regression ?

Regression analysis estimates the relationship between two or more variables. There are multiple benefits of using regression analysis. They are [13]:

- It indicates the significant relationships between dependent variable and independent variable.
- It indicates the strength of impact of multiple independent variables on a dependent variable.

Regression analysis also allows us to compare the effects of variables measured on different scales, such as the effect of price changes and the number of promotional activities. These benefits help us to eliminate and evaluate the best set of variables to be used for building predictive models.

2.7 Regression Techniques

There are various kinds of regression techniques available to make predictions. These techniques are mostly driven by three metrics (number of independent variables, type of dependent variables and shape of regression line).

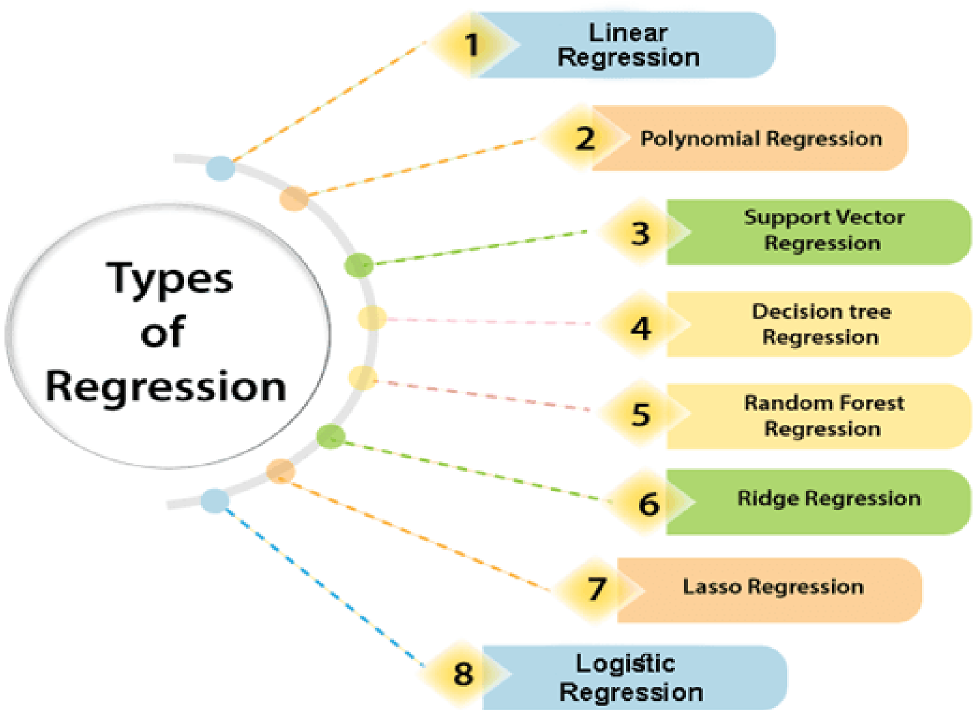

Fig. 2.1: Types of Regression.

2.8 Regression Used : Linear Regression

In this technique, the dependent variable is continuous, independent variable or variables can be continuous or discrete, and nature of regression line is linear.

Linear Regression establishes a relationship between dependent variable (Y) and one or more independent variables (X) using a best fit straight line, also known as regression line. It is represented by an equation:

$$Y = a + b * X + e \quad [13] \tag{2.5}$$

where a is intercept, b is slope of the line and e is error term. This equation can be used to predict the value of target variable based on given predictor variable(s).

The difference between simple linear regression and multiple linear regression is that multiple linear regression has more than 1 independent variables, whereas simple linear regression has only 1 independent variable. Some of the important points to keep in mind before using multiple linear regression are [13]:

- There must be linear relationship between independent and dependent variables
- Multiple regression suffers from multicollinearity, autocorrelation, heteroskedasticity (which refers to the circumstance in which the variability of a variable is unequal across the range of values of a second variable that predicts it).
- Linear Regression is very sensitive to outliers causing the regression line and forecasted values to have high error.
- Multicollinearity can increase the variance of the coefficient estimates and make the estimates very sensitive to minor changes in the model. The result is that the coefficient estimates are unstable.
- In case of multiple independent variables, we can go with forward selection, backward elimination and step wise approach for selection of most significant independent variables.

Using these methods load consumption forecasting analysis has been done in this work. The details of how the data has been utilized and how the forecasting is achieved has been given in the next chapter.

CHAPTER 3
METHODOLOGY

As mentioned before this work presents the studies made on load forecasting in two application sectors, one for the residential sector and another for EV charging stations. The analysis was done on different datasets, and by using separate machine learning models. This section presents the details about the datasets being utilized, how the pre-processing on those datasets were done, and how the models were built to predict the forecasted values of load consumption.

3.1 Dataset for Households

The load consumption data is collected by the smart meters in London which were installed in around 5,000 households by the UK government. Historical weather data has been collected from dark sky API [5]. For building the forecasting model seasonal ARIMA was used and results were generated using the matplotlib package of python.

The entire data-set is available on kaggle under "Smart meters in London", including the Acorn groups and their values. The acorn groups can better be understood by referring to Fig. 3.1

3.2 Dataset for EV Charging stations

The dataset for EV charging stations was recorded by Rocky Mountain power over 5 years. They introduced their meters in over 118 charging stations across the state of Utah. They recorded various parameters including the charging time of a vehicle, the energy consumed in that time, type of vehicle used, power outlet type used and many other miscellaneous parameters. See Fig. 3.2.

Both datasets had to be filtered to use only those parameters that affected the energy consumption of the charging station.

Fig. 3.1: Acorn Relations.

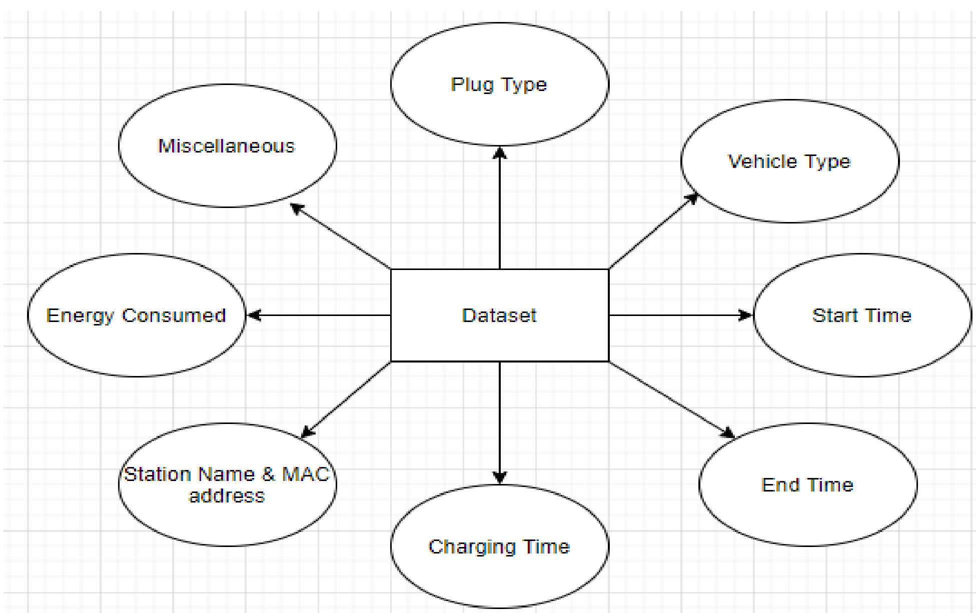

Fig. 3.2: EV Charging Station Dataset

3.3 Risk Assessment

Both the datasets used for load forecasting had a few limitations which needed to be addressed.

- Household Load dataset:
  - The dataset had several inconsistencies due to the fact that, the meters were not installed in all the houses on a single day. For example, on day one, data was recorded for 13 houses and on day two it was recorded for 25 houses. Thus, the number of households (for which energy data was collected) varied across different days. This prevented the use of the entire dataset. Only those observations for which the data was consistent were chosen.
  - The weather data collected was from a single location and the houses were spread across multiple locations. Some houses were located around 300 miles from the weather station.Thus it was observed that some chosen parameters did not have the expected impact on the load consumption.
  - The multiple weather parameters that affected load consumption had to be clustered together so that they could be used as an input parameter.
  - The acorn values were not given in a numerical format. The houses were divided into acorn groups that had different features. It was not possible to use them in the form they were presented.
- EV charging station load dataset:
  - The assumption that the vehicles started to charge as soon as they arrived did not hold, as data showed that people had to wait for a free port to charge their vehicles. Date-time calculations were done to get consistent values.
  - The data was collected only in the area of Utah. When the forecasting model was integrated with the optimizer for micro-grid, assumptions had to be made to enable transfer of the results from one location to another.

  - From the year 2014 to 2017 the recording of data was in its initial stages therefore the data was not consistent. For example, a lot of missing values were observed in the dataset due to meters being sent for maintenance. Thus the data after that period was used, which reduced the training set.

3.4 Data Preparation

The dataset chosen for load forecasting studies performed in this work was available in the forms of multiple files. These files had some data that was not required for this study and so the relevant data had to be extracted and the irrelevant data discarded. This section explains how the pre-processing was done on the dataset to obtain the input parameters.

- Household data: Initially the dataset had 7 files, from which the useful data was to be extracted. The process of getting the input parameters was done by performing the following steps:
  - The data for the years 2013 and 2014 were separated to get consistent data.
  - The weather data and load consumption data were combined based on the household id and the date of observation. This step was performed using the pandas library of python.
  - The weather parameters that had an impact on load consumption were clustered together and by using kmeans [14] an index was generated based on their values. The indexes were 0, 1, 2 and 3.
  - The holiday index was added by comparing the date of observation with the list of national holidays in the UK.
  - Acorn values were generated from the data by using feature vectorization in python i.e. a numerical value was given to the acorn feature based on its impact on the load consumption.

- EV charging station Data: Initially the data file was in the form of a single csv file which had 37 columns, but most of them were not useful as they didn't have any impact on energy consumption. In order to prepare the data in a way which could be used, following steps were followed:
  - The data was separated for the year 2018 and 2019.
  - The stations were grouped together and their data separated.
  - According to the start date weekend index column was generated using the Date Time Library of Python.
  - using the holiday list obtained from the web holiday index column was generated.
  - Season index column was then produces using the date column.
  - 24 new columns were generated based on the previous 24 hour value of energy consumption from the historical data. This was done using the pandas library of Python.

3.5 Forecasting Plan

Forecasting is about predicting the future as accurately as possible, given all of the information available, including historical data and knowledge of any future events that might impact the forecasts.

To plan a day ahead for the efficient working of EV charging stations, an accurate forecast of the load consumption, solar energy available and optimal policy of the micro-grid is essential. Planning is a response to these forecasts. It involves determining the appropriate actions that are required to make the forecasts match with the goals.

As the time resolution of the forecasting is increased, the accuracy decreases. Therefore to maintain the accuracy several experiments were conducted by increasing the time resolution of prediction slowly. The hyper-parameters were also changed multiple times to maintain the accuracy of the forecasts.

3.6 Forecasting Process

This section explains how the forecasting models were built for both the studies performed in this work and the scenarios in which the results were observed.

- Household load forecasting process: The data-set of 2 years, from 2013 to 2014 was divided into training and testing sets, with a 70 to 30 percent ratio i.e. the first 70 percent of the data was converted into training set and the remaining data was converted into testing set. The model was fit using the training set and predictions were made on both training as well as test data-sets.

  The weather parameters which correlated with energy consumption of households were combined to form a weather cluster index. The method used to achieve this result was kmeans clustering using sklearn in python. A cluster refers to a collection of data points aggregated together because of certain similarities. We defined a target number k, which refers to the number of centroids we need in the data-set. A centroid is the location in the data space representing the center of the cluster. Every data point is allocated to each of the clusters through reducing the in-cluster sum of squares [14]. The result was that we got 4 weather clusters indexed 0,1,2 and 3.

  The Acorn values in the data-set were given in a form that could not be used directly as an input feature, thus feature vectorization technique was used to convert them into features that can be fed to the model. For this purpose, 4 Acorn values were selected from the file and merged into the data-set of the corresponding household. The model was built in python using the SARIMAX package of the statsmodels library.

  This model was used to analyze the results in three different ways:

  - First the average energy of all the households for a particular day was given as an input feature to the model.
  - Second, the input features corresponding only to a single household was fed to the model. This process was conducted for 20 houses.

  - Lastly, the input features of individual households belonging to different acorn groups, was fed to the model along with the acorn values of that group.

- EV charging station load forecasting process:

  The division of dataset into training and testing sets was done using the same 70-30 rule, in a time ordered manner by using the data from the year 2018 and 2019. The input parameters were identified in a different manner than in the previous dataset.

  The parameters used were starting date, start time, charging time, energy consumed, and MAC address of the station. Based on these values the hourly requirement of energy was computed through sampling for a given station.

  On analysis of the data , it was observed that the energy consumed is different for :

  - Weekday and Weekend.
  - Different seasons, example summer and winter.
  - Ordinary day and a holiday.

  Based on the above findings these parameters were also added as an index to list of input parameters in the model. The model was built using the Keras library of python and linear regression was done by using neural networks which are explained in the following section.

  Neural Networks:

  Neural networks are a means of doing machine learning, in which a computer learns to perform some task by analyzing training examples. They are multi-layer networks of neurons (simple processing nodes that are densely interconnected) that are used to classify things, make predictions, etc. Neural networks are organized into layers of nodes. An individual node might be connected to several nodes in the layer beneath it, from which it receives data, and several nodes in the layer above it, to which it sends data [11].

The first layer is the input layer and it has the same number of nodes as the number of inputs. The last layer is the output layer, which gives the result of the network. The layers in the middle are called hidden layers, and their number can vary according to the program. To each of its incoming connections, a node will assign a number known as a “weight” and each node has its own ”bias”. Each node applies the weight and bias to the incoming value, and passes it through an activation function that is associated with the layers in which the node is placed. The result is forwarded to the next layer. After every iteration the weights and biases of the nodes are adjusted by the network to give the correct output and this is how the network learns.

The model that was built using Keras had 4 layers, one input, one output, and 2 hidden layers. Each layer had ”relu” as its activation function. The optimizer used was ”adam” and the loss function used was RMSE(Root mean squared error). The model was trained on 10000 epochs(number of cycles). The training time for one iteration was about 22 minutes.

This model was used to analyze the results in two different ways:

– The load consumption forecasting model was applied to the data of different charging stations, which were picked based on their locations. The forecasted values were compared with the actual values of the consumption to get the accuracy of the model.

– Predictions were made on the load consumption of a charging station for two months of different seasons and the results were analyzed.

For better understanding of the process for the load forecasting model Fig. 3.3 can be referred:

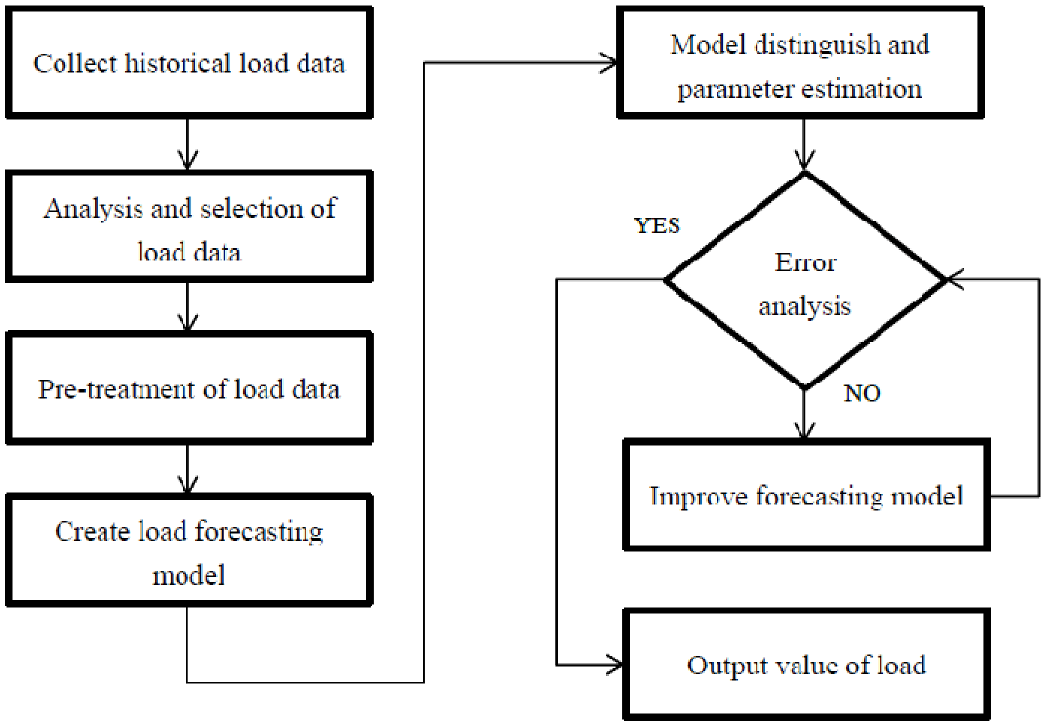

Fig. 3.3: Flow chart of the complete load forecasting process.

This chapter explained how the ARIMA and regression models were built in this work and what tools were used to do the same. The next chapter illustrates the results that were achieved after using these models to predict the load consumption data.

CHAPTER 4

RESULTS

This chapter presents the results of different load forecasting studies done in this work. Sections 5.1, 5.2 and 5.3 present the results of the study done in the residential sector whereas sections 5.4 and 5.5 presents the result of the study conducted for EV charging stations. These results are used to validate the models that have been used to predict the load consumption and check the accuracy with which the prediction is made. The different scenarios in which the models are tested are also explained.

4.1 Average energy consumption of all houses combined:

The dataset used in this section has the data from the London data store, that contains the energy consumption readings for a sample of 5,567 London Households that took part in the UK Power Networks led Low Carbon London project between November 2011 and February 2014 [3].

In this result prediction is being made for energy demand in the future, therefore only the sum of energy, i.e. total energy use per day for a given household was considered. This data was then combined with the weather data from the dark sky API [5] along with the holiday index. The resulting dataset was used to train the ARIMA model and the results can be seen in Fig. 4.1.

In Fig. 4.1, the actual values are plotted against the predicted values using matplotlib library of python. The unit of measurement for energy consumption is KW/hr. The prediction accuracy came out to be 92 percent when data for all the households was used as a single entity.

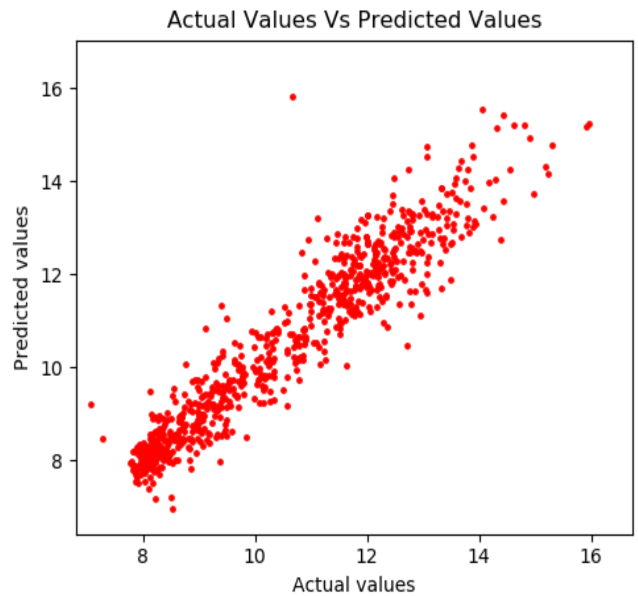

Fig. 4.1: Actual vs Predicted values for average energy consumption of all the houses combined over three years.

4.2 Average energy consumption for a single household:

This section presents the result of the energy consumption forecast when the ARIMA model is applied to the data for a single household. This includes the individual aspects of the household, like the number of people living in the house, the financial status of the family, etc. The accuracy of the model can be deduced by referring to Fig. 4.2.

The prediction has been made for daily resolutions i.e. average energy consumption of a day for a single household is being predicted in this analysis. From Fig. 4.2 it can be seen that a single household energy consumption forecast is more scattered that average of all houses combined. This is because of the outliers in the dataset, which emerged due to the unpredictable nature of people living in the house.

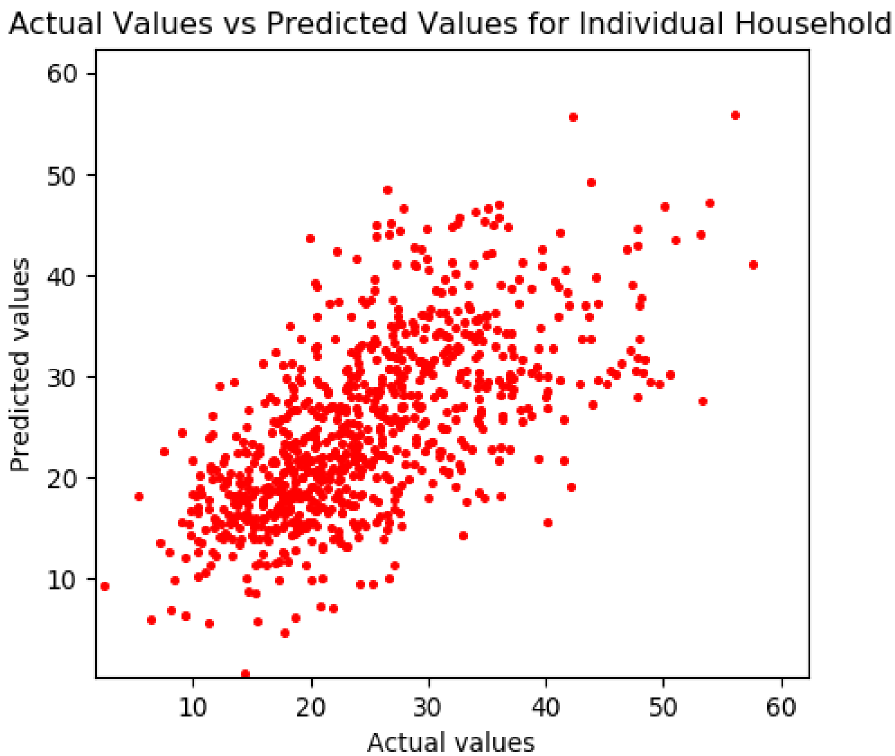

Fig. 4.2: Actual vs Predicted values for average energy consumption of a single house combined over three years.

4.3 Average energy consumption of houses with Acorn values as a part of input features:

This section presents the analysis done on the energy consumption forecast when the Acorn values were included in the input features for particular households. These values had a large impact on the energy consumption of a household. Acorn values signified the contextual or behavioral aspect of a particular household.

To obtain the results three houses belonging to different communities were observed and their data was recorded and merged with their respective Acorn values. The consumption of the three houses can be seen in Fig. 4.3

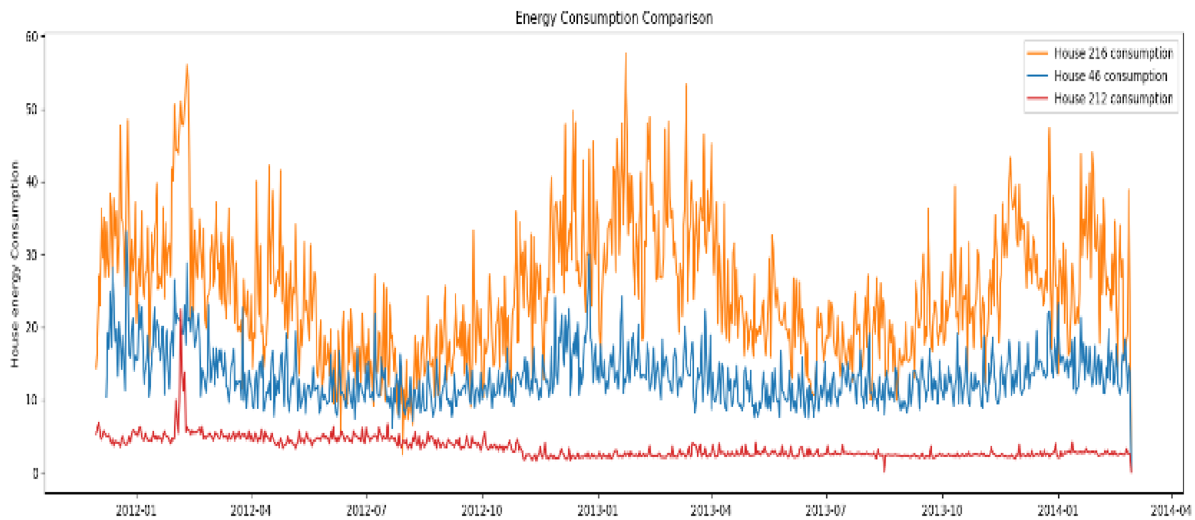

Fig. 4.3: Energy consumption analysis of three houses of different groups

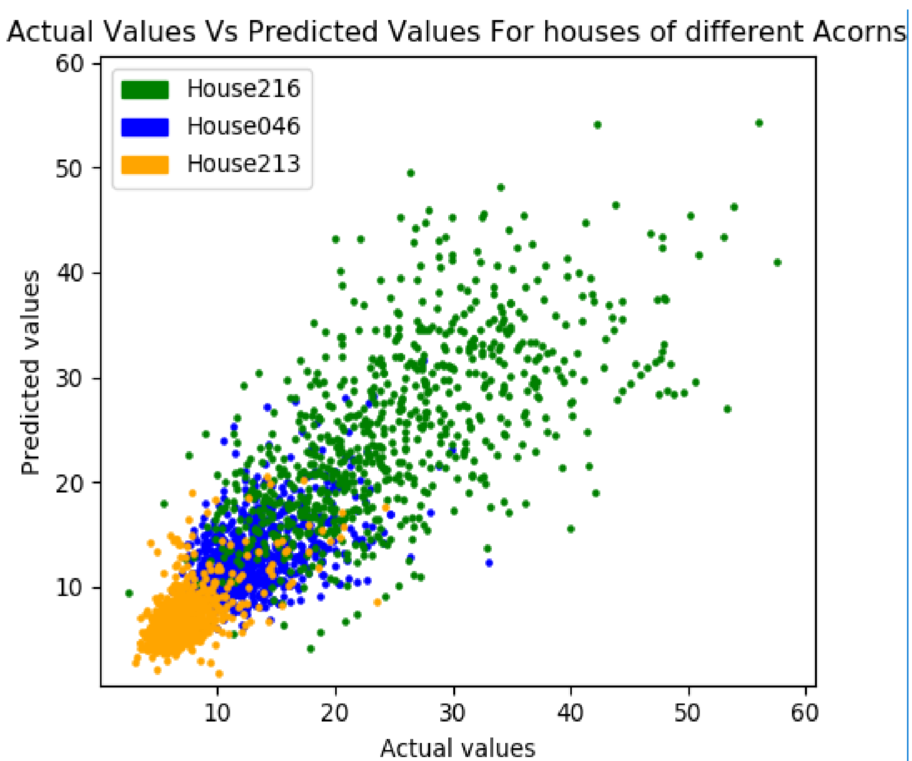

Fig. 4.4: Actual vs Predicted values for average energy consumption of houses when their acorn values are also included in the input features

It was observed from Fig. 4.3 that households belonging to different acorn groups had different energy consumption profiles. This analysis was used to provide the data to the ARIMA model. This is the most vital result as it includes the contextual data as an input feature and it can be referred to in Fig. 4.4. The green dots denotes the energy consumption of house belonging to the comfortable community and that is the reason it has more energy consumption. The Acorn communities can be seen in Fig. 3.1. The orange dots denote the energy consumption of the house belonging to the financially stretched community and thus it has less energy consumption.

4.4 Actual vs Predicted values of load consumption for EV charging stations:

The dataset used to obtain the result in this section contains the data of all the stations in which the meters were installed, i.e. a total of 118 stations. The data is being recorded since 2014, but only the years 2018 and 2019 have been taken into the training set as the values were much more consistent after that time.

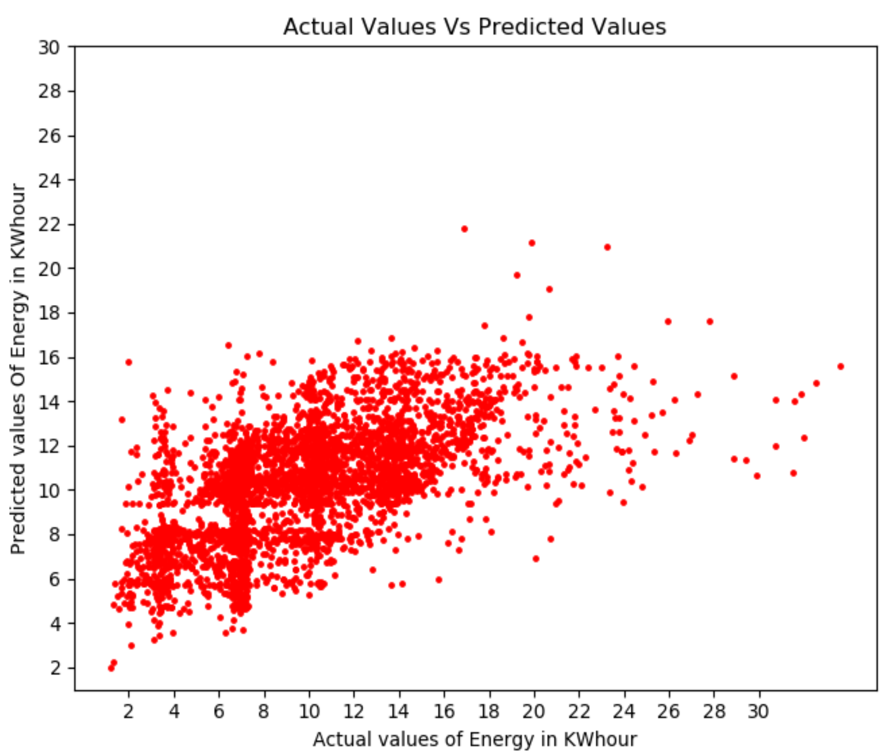

Fig. 4.5: Actual vs Predicted values for EV charging station load consumption

The selected input parameters were fed to a linear regression model. The output of the model was compared to the actual values of the load consumption available from the historical data. The result of the analysis can be seen in Fig. 4.5. The predictions made by the model had a time resolution of 48 hours i.e. it could predict 48 values of load consumption for 48 hours ahead of the given time.

As the load consumption of a vehicle can vary depending on the model, the outliers which had a difference of 4KW/hr from the actual value were not considered as errors. The resulting model gave an accuracy of 82 percent. To generate the new result, the outliers were eliminated and new graph was generated. It was also observed that the number of outliers in the result were 642 and the total number of observations were 4234. The outcome of this process is illustrated in Fig. 4.6

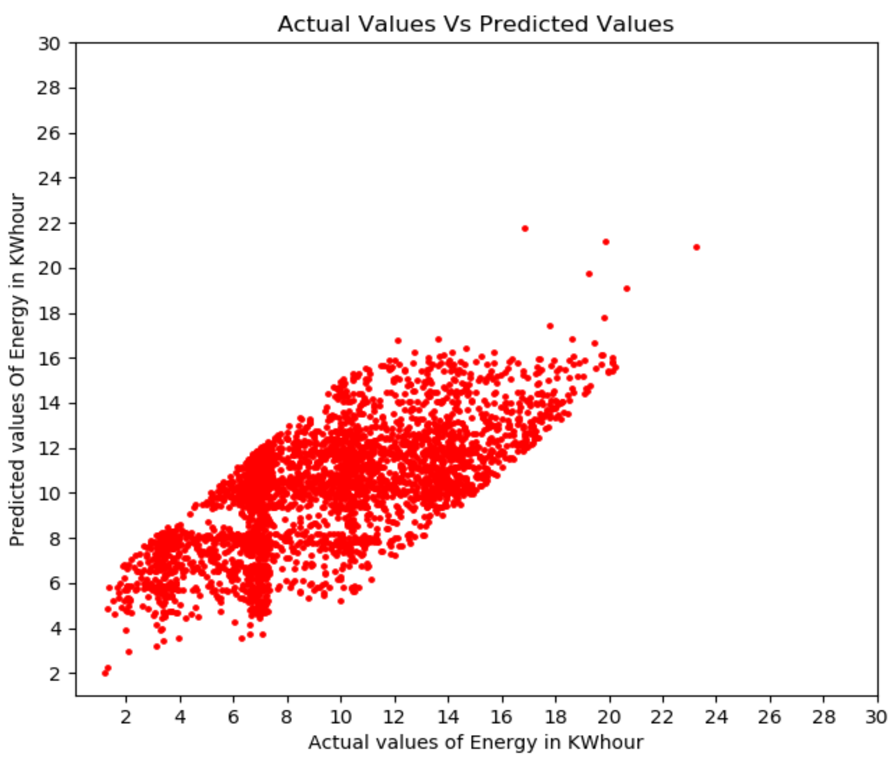

Fig. 4.6: Actual vs Predicted values for EV charging station load consumption after eliminating the outliers

4.5 Load consumption prediction of EV charging stations for different seasons:

This section presents the results when the prediction was being made for an entire month. The months chosen for this analysis are February (i.e end of winter) and the month of May (i.e summer month). The dataset used to obtain the result contains the data of 10 chosen stations which were picked based on their locations.

To make predictions for a month, each day of the month was divided into 4 quarters and then the regression model was used to predict the data for that quarter. When each quarter of each day of the month was combined, it produced the predicted data for the month. These values were then plotted on the bar graph using the matplotlib package of python.

The prediction results for the month of February can be seen in Fig. 4.7 and the same for the month of May can be seen in Fig. 4.8.

To show the variation in the load consumption values for the two chosen months Fig. 4.9 was plotted, from which it can be inferred that the EV charging station had more load consumption in the summer months than in winter as the graph clearly shows that more energy is consumed in May than in February.

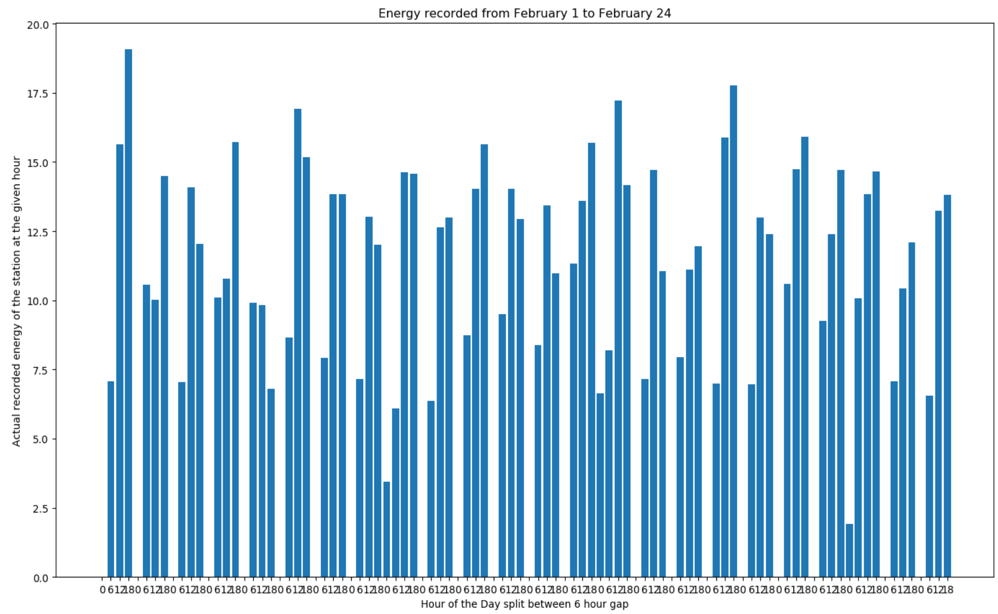

Fig. 4.7: Predicted values for EV charging station load consumption, for the month of February.

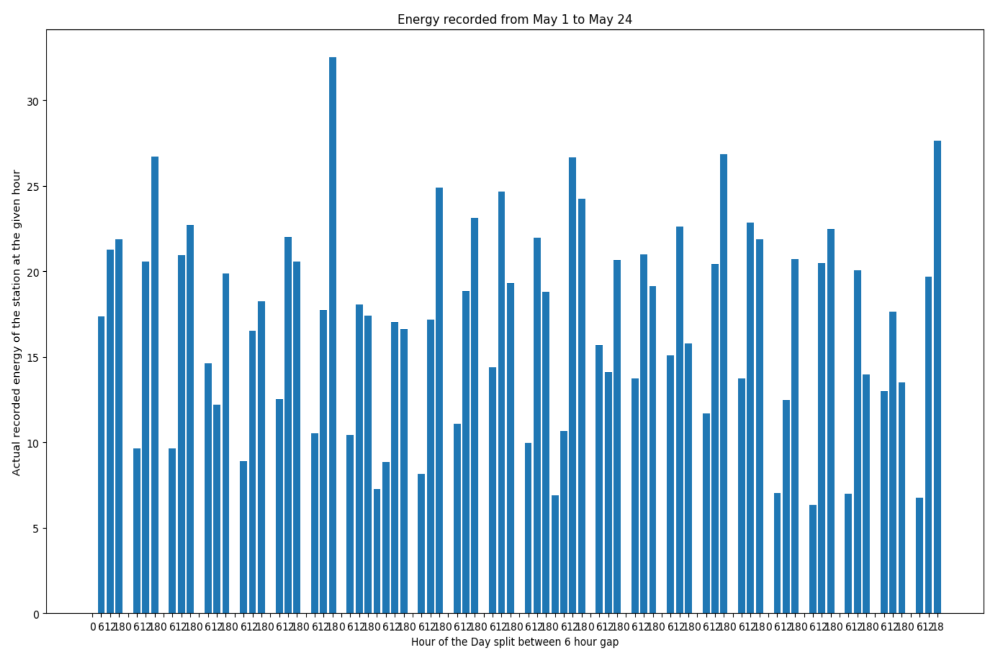

Fig. 4.8: Predicted values for EV charging station load consumption for the month of May.

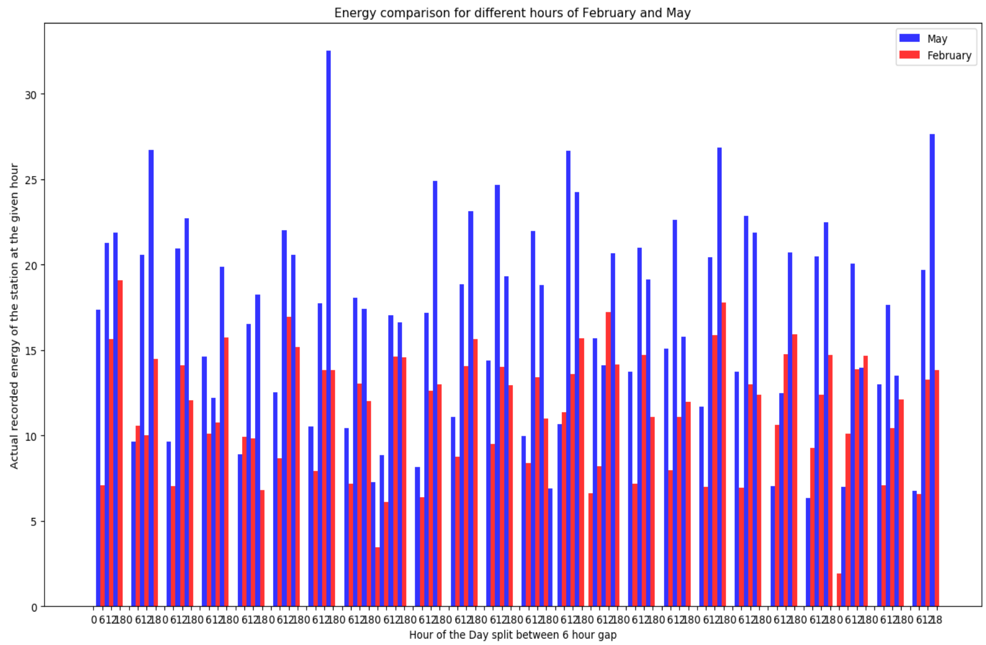

Fig. 4.9: Comparison of Predicted values for EV charging station load consumption for the month of February(end of winter) and May(Summer).The y axis denotes the predicted energy consumption of a EV charging station and x axis denotes the hour of the day for which prediction is made. Each day was divided into 4 quarters of 6 hours each, and thus each day has 4 predicted values of energy consumption.

4.6 Sample Input given to the optimizer after forecasting values 48 hours ahead:

One of the input given to the optimizer had 48 values of load consumption, forecasted for the next 48 hours. The starting hour changed for testing different scenarios. Two sample inputs with different starting hours are illustrated in Fig. 4.10 and Fig. 4.11. From the figures it can be seen that the regression model also gives load consumption values as 0 KW hour, when predicting for the time when there are no vehicles being charged at the stations.

This chapter showed the results obtained for different scenarios of residential sector and EV charging stations. The next chapter deals with integrating the results with microgrid.

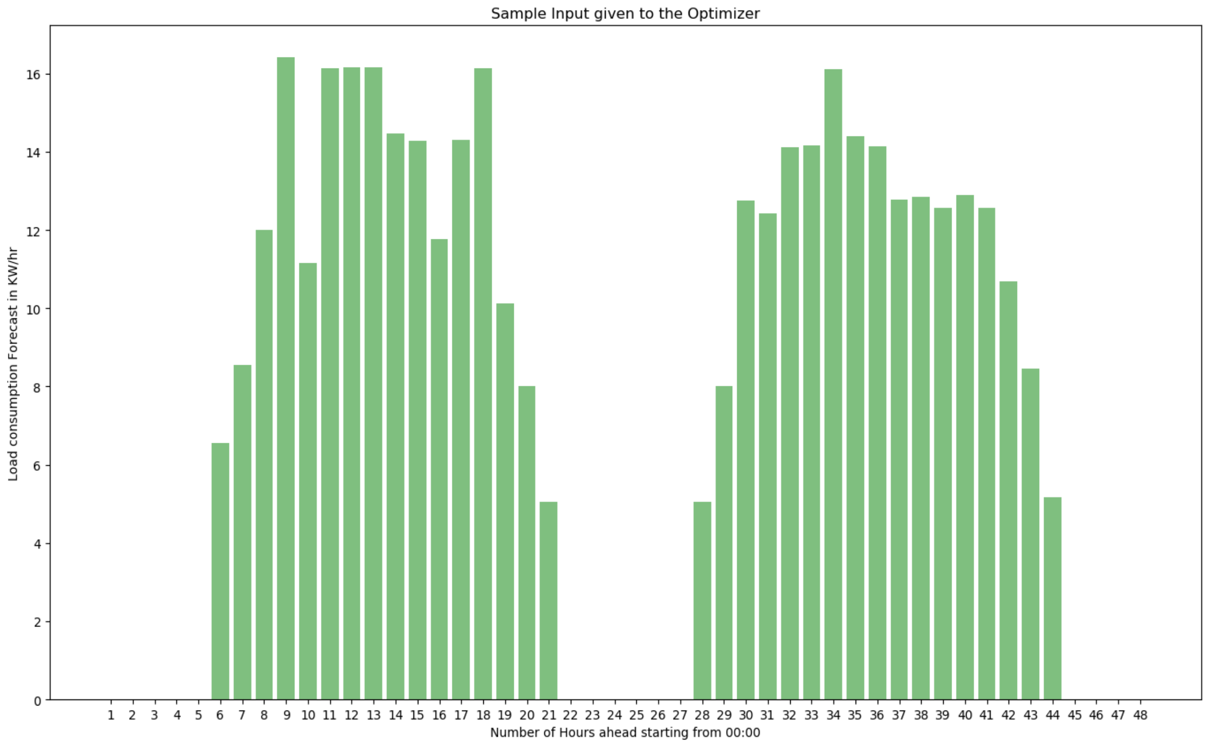

Fig. 4.10: Forecasting values 48 hours ahead of the the starting time 00:00

Fig. 4.11: Forecasting values 48 hours ahead of the the starting time 12:00 pm

CHAPTER 5

INTEGRATION WITH MICRO-GRIDS IN EV CHARGING STATIONS

Microgrids are self-contained electric grids that can operate even in off-grid areas, independent of the central power grid. In order to run the experiments for finding the optimal working policy of the microgrid, accurate forecasted values of demand (load consumption) and supply (solar energy available) were required. The forecasted values of demand were obtained from the studies, demonstrated in Chapter 3 and Chapter 4. This Chapter illustrates how the optimal policy for microgrid was generated and how the the forecasted load consumption values affect that policy.

5.1 Working of optimizer

The optimizer is using a reinforcement learning algorithm called value iteration to determine the optimal policy for the working of the grid. The data it requires for that purpose includes:

- Forecasted load values for next 48 hours,
- Forecasted solar energy values for the next 48 hours,
- energy pricing policy, and a
- battery cost model.

The optimizer applies the value iteration algorithm on the given inputs to generate a sequence of optimal actions. The actions can be to buy energy or sell energy based on future values of supply and demand. All the inputs are fed to the optimizer in one hour resolution, i.e. for every hour an action is given by the optimizer.

This action needs to be applied to the power management system so as to maximize the use of solar energy available to us to meet the load requirements while also minimizing the cost incurred.

The experiments were run on a microgrid model of an EV charging station with a solar panel of 100 KiloWatts, where the source of energy is only solar power. The configuration and data folow is illustrated in Figure 5.1

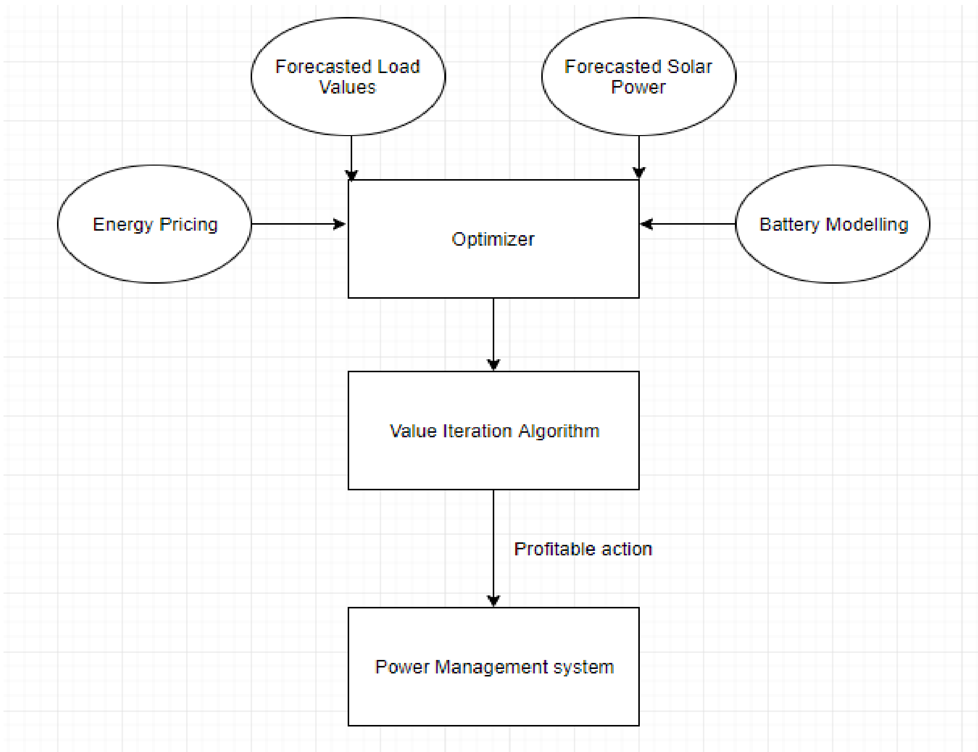

Fig. 5.1: Working of the Microgrid optimization algorithm.

5.2 Working of Solar Energy Predictor

Similar to the load consumption prediction problem, solar energy production is also dependent on multiple factors. The dataset used for this purpose is from NOAA and numerical weather prediction data from which historical values of solar irradiance was computed for one hour resolutions. Solar irradiance is defined as the amount of energy received from the sun for a one-meter square area of the Earth's surface. The measure of solar irradiance is highly dependent on sun position, diurnal cycle and weather parameters like cloud cover, relative humidity, temperature, dew point, precipitation, etc. Solar irradiance is directly proportional to solar energy produced. Hence predicting solar irradiance is equivalent to predict solar energy.

Solar irradiance is referred to as GHI (Global Horizontal Irradiance) and is measured in Watts/meter square. The predictor can predict values for 10 days ahead but the studies have shown that the first 48 hours is the most reliable. The predictor bases its outputs on the following parameters:

- Meteorological parameters:
  - Temperature
  - Humidity
  - Precipitation
- Sun position parameters:
  - Zenith angle
  - Azimuthal angle
- Diurnal cycle parameters:
  - Year
  - Month
  - Day
  - Hour

In order to generate results of solar energy 48 hours ahead of the given input following algorithms were studied:

- Support vector regression
- Xg-boost method
- Cat-boost regression

The forecasted values of solar energy which were supplied as an input to the optimizer, were obtained by using the Xg-boost method which gave a prediction accuracy of 85%.

5.3 Optimizer Results

In order to easily interpret the results produced by the optimizer, a specific format was used for the graphs. The format is illustrated in Fig. 5.2. The result figures consist of five different graphs, aligned vertically on the same time scale. The top graph among the five represents the input values of supply and demand given to the system. The supply is the predicted solar power represented by orange bars. The demand is the forecasted load values for that timestamps represented by gray bars. the unit on the y-axis is KW since it illustrates power. The second graph below represents the state of charge of the battery for each time step for two days in blue. The unit used here is KWh for the y-axis since energy is measured. The third graph describes the policy determined by the optimization algorithm as a charge/discharge sequences from the battery for the maximum profit. Maroon colour denotes charging the battery at a power value and sea green denotes the discharge power of the battery. The fourth graph shows the action of buying and selling electricity power to/from the grid. The final (bottom) graph shows the actual cost or profit which can be made by following the action given in fourth graph.

The optimizer determines an optimal policy based on the time of day electricity price. After observing Fig. 5.2 it was concluded that the optimizer uses the stored electricity or solar power to either fulfill the incoming demand or sell this energy back to the grid when the price of electricity is highest. On the other hand, it buys the electricity from the grid to either support the incoming demand or store this energy in the battery when the price of electricity is low. It was also observed that after applying the action sequences produced by the optimizer, the system can make a total profit of 5 dollars in 2 days.

The result shown in Fig. 5.2 has been plotted for the battery size of 20KWh. When the battery size was increased to 50KWh, it was observed that the optimizer takes an action to buy more electricity from the grid in the off-peak hours and then sells it back in peak hours. The result of which is illustrated in Fig. 5.3. It can be observed that as the battery size of the system increases the profit also increases, as the battery can now store more energy. The new configuration produced a profit of 40 dollars in 2 days.

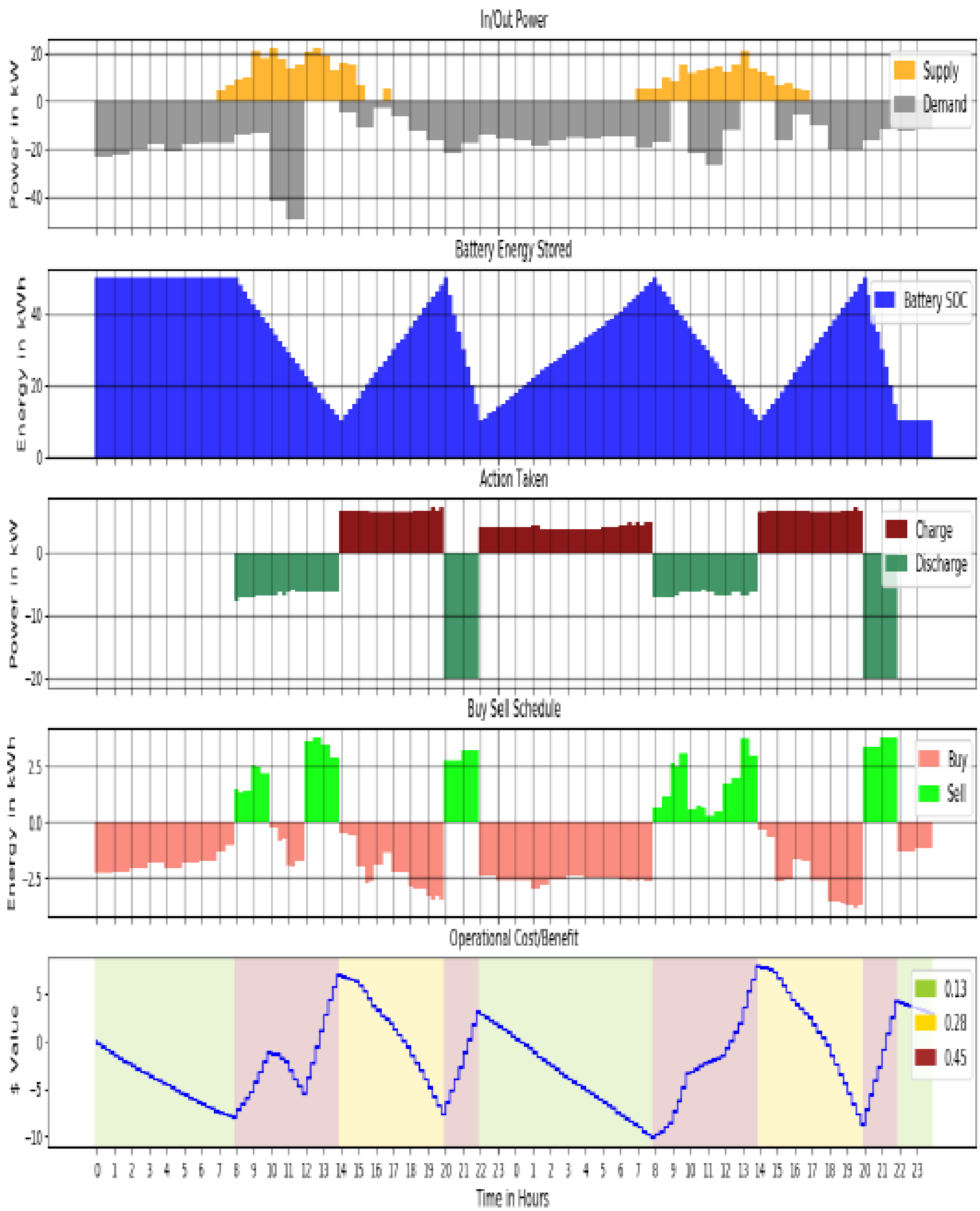

Fig. 5.2: Optimizer result for battery size 20KWh, solar capacity 100KW and maximum load 40KW from 02/01/2020 to 02/02/2020.

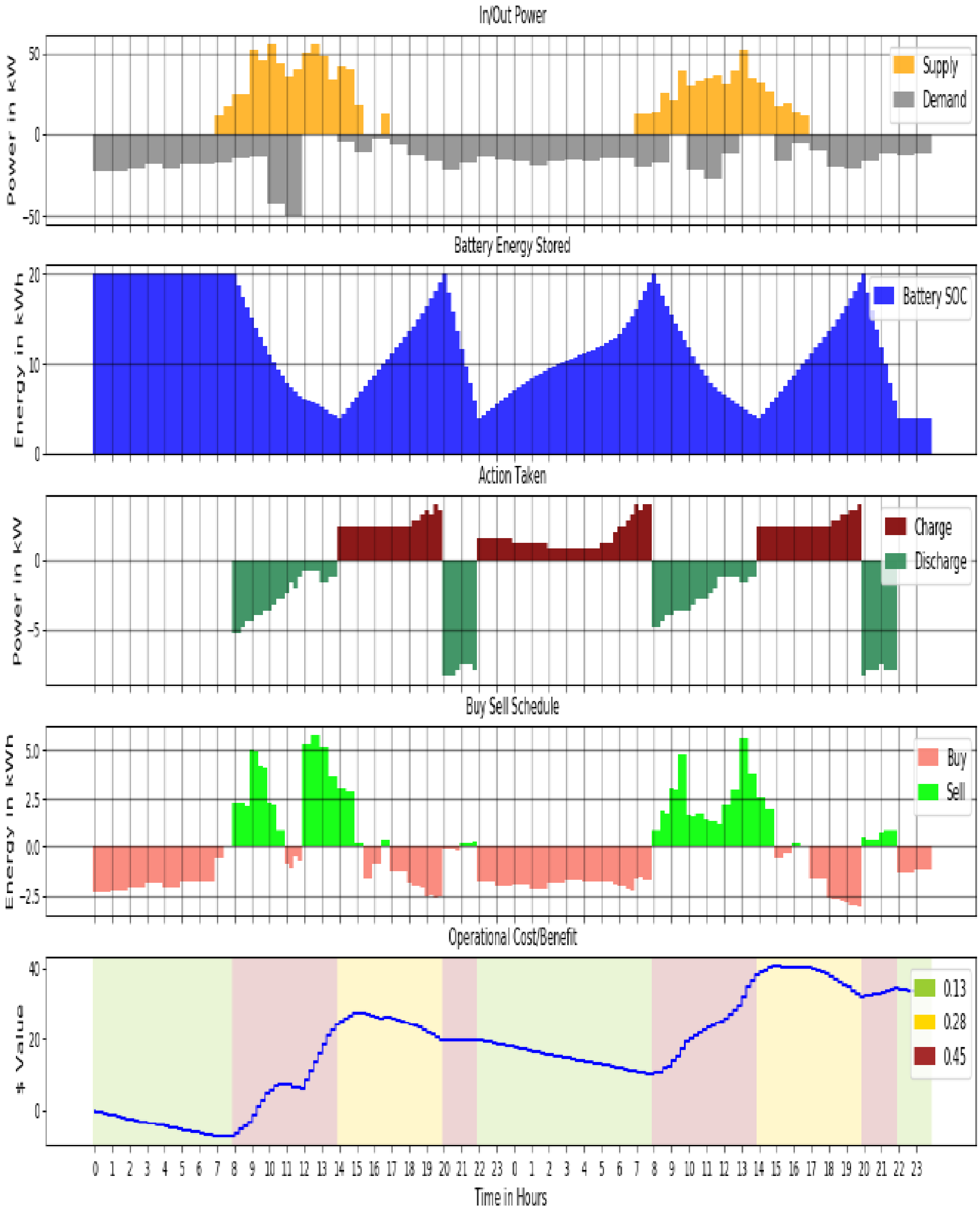

Fig. 5.3: Optimizer result for battery size 50KWh, solar capacity 100KW and maximum load 40KW from 02/01/2020 to 02/02/2020.

CHAPTER 6

Concluding Remarks

Electricity demand forecasting represents an important component in the planning and management of a power system. This paper presents two ways in which this can be achieved, by using the ARIMA model for the residential sector and multiple linear regression for the EV charging stations. The forecasting period is set to 48 hours, with a one-hour resolution i.e. the models can predict 48 values of load consumption for the next 48 hours. The ARIMA model gave a prediction accuracy of 88% and the regression model gave a prediction accuracy of 82% for load consumption. The key results can be seen in Fig. 4.6 and Fig. 4.4

For the residential sector, the load consumption is highly affected by the information of the people living in the house. After observing the results from Fig. 4.4 it was concluded that houses belonging to rich communities had more load consumption than those belonging to financially stretched communities. It was also observed that, as the number of people living in the house increases, the energy consumption also increased.

For EV charging stations, the analysis showed that the energy consumption is maximum between the time of 12:00 pm TO 7:00 pm of a given day. Load consumption is also affected by the season as summer months reported more energy demand than winter as illustrated in Fig. 4.9.

The forecasted values of the regression model for EV charging stations are integrated with the solar energy forecast and battery configuration to build an optimal policy for the working of micro-grid. The optimal policy is determined by using value iteration algorithm and the process is illustrated in Fig. 5.1. Several experiments with different battery models have been reported in this work. The results show that with prior knowledge of supply and demand energy, a decision can be made by the optimizer on whether to buy or sell energy by also taking into account the energy pricing at that time of the day, in a way that the cost incurred is minimum.

The results obtained from the optimizer in Fig. 5.2 and Fig. 5.3 show that, the battery size has a large impact on the profit of the system. When the battery size was increased from 20KWh to 50KWh , the profit of the system increased form 5 dollars to 40 dollars for 2 days.

6.1 Future Work:

The load prediciton of EV charging stations can be improved by also considering fast charging of vehicles and multiple parallel ports for a station as this will also impact the load consumption of a station.

The time resolution of the forecasting models can be increased to predict further into the future. These values can be given to the optimizer to study how it impacts the optimal policy generated. The idea is that the more information the optimizer has about the future values the more profitable decisions it can make.

REFERENCES

[1] Anwar, Tahreem Sharma, Bhaskar Chakraborty, Koushik Sirohia, Himanshu. (2018). Introduction to Load Forecasting. International Journal of Pure and Applied Mathematics. 119. 1527-1538.

[2] http://engineering.electrical-equipment.org/electrical-distribution/electric-load-forecasting-advantages-challenges.html

[3] https://www.kaggle.com/jeanmidev/smart-meters-in-london

[4] E. S. Xydas, C. E. Marmaras, L. M. Cipcigan, A. S. Hassan and N. Jenkins, "Electric vehicle load forecasting using data mining methods," IET Hybrid and Electric Vehicles Conference 2013 (HEVC 2013), London, 2013, pp. 1-6.

[5] https://darksky.net/dev

[6] F. Xie, M. Huang, W. Zhang and J. Li, "Research on electric vehicle charging station load forecasting," 2011 International Conference on Advanced Power System Automation and Protection, Beijing, 2011, pp. 2055-2060.

[7] https://acorn.caci.co.uk

[8] K. Goswami, A. Ganguly and A. K. Sil, "Day Ahead Forecasting and Peak Load Management using Multivariate Auto Regression Technique," 2018 IEEE Applied Signal Processing Conference (ASPCON), Kolkata, India, 2018, pp. 279-282.

[9] https://otexts.com/fpp2/arima.html

[10] C. Hor, S. J. Watson and S. Majithia, "Daily Load Forecasting and Maximum Demand Estimation using ARIMA and GARCH," 2006 International Conference on Probabilistic Methods Applied to Power Systems, Stockholm, 2006, pp. 1-6.

[11] https://towardsdatascience.com/understanding-neural-networks-19020b758230

[12] V. Dehalwar, A. Kalam, M. L. Kolhe and A. Zayegh, "Electricity load forecasting for Urban area using weather forecast information," 2016 IEEE International Conference on Power and Renewable Energy (ICPRE), Shanghai, 2016, pp. 355-359.

[13] https://www.sciencedirect.com/topics/social-sciences/regression-model

[14] https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1

[15] Wang, Jian, Kui Hua Wu, Feng Wang, Zhi Hui Li, Qing Song Niu, and Zhi Zhen Liu. "Electric Vehicle Charging Station Load Forecasting and Impact of the Load Curve." Applied Mechanics and Materials 229–231 (November 2012): 853–58.

[16] https://towardsdatascience.com/reinforcement-learning-demystified-solving-mdps-with-dynamic-programming-b52c8093c919

[17] Zhu, J.; Yang, Z.; Guo, Y.; Zhang, J.; Yang, H. Short-Term Load Forecasting for Electric Vehicle Charging Stations Based on Deep Learning Approaches. Appl. Sci. 2019, 9, 1723.

[18] Chokkalingam, B.; Padmanaban, S.; Siano, P.; Krishnamoorthy, R.; Selvaraj, R. Real-Time Forecasting of EV Charging Station Scheduling for Smart Energy Systems. Energies 2017, 10, 377.

[19] N. S. C. D. M. P. Robert Blaga, Andreea Sabadus, "A current perspective on the accuracy of incoming solar energy forecasting," in Progress in energy and combustion science, Oct. 2018.

[20] S. K. M.-L. N. C. P. F. M. A. F. Cyril Voyant, Gilles Nottan, "Machine learning methods for solar radiation forecasting: A review" 2017

[21] Nicholas Flann, Pratyusha Sai Kamarouthu, "Solar Irradiance prediction using XG Boost with the numerical weather forecast" (May 2020)