Smartmeter Data Analytics

Submitted in partial fulfillment of the requirements

for the degree of

Bachelor of Engineering

by

Aniket A. Chavhan

Roll No. 25

Mayur R. Machhi

Roll No. 16

Diksha M. Phatak

Roll No. 21

Under the Supervision of

Prof. Anil W. Kale

DEPARTMENT OF INFORMATION TECHNOLOGY

KONKAN GYANPEETH COLLEGE OF ENGINEERING

KARJAT-410201

September 2020

Certificate

This is to certify that the project entitled **Smartmeter Data Analytics** is a bonafide work of **Aniket A. Chavhan (25), Mayur R. Machhi (16), Diksha M. Phatak (21)** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **Undergraduate** in **DEPARTMENT OF INFORMATION TECHNOLOGY**.

Supervisor/Guide

Prof. A. W. Kale

Department of Computer

Head of Department

Prof. J. P. Patil

Department of Information Technology

Principal

Dr. M. J. Lengare

Konkan Gyanpeeth College of Engineering

Project Report Approval

This thesis / dissertation/project report entitled **Smartmeter Data Analytics** by **Aniket A. Chavhan (25), Mayur R. Machhi (16), Diksha M. Phatak (21)** is approved for the degree of **DEPARTMENT OF INFORMATION TECHNOLOGY**.

Examiners

1..

2..

Date.

Place.

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Signature

(Aniket A. Chavhan) Roll No 25

Signature

(Mayur R. Machhi) Roll No 16

Signature

(Diksha M. Phatak) Roll No 25

Date.

Abstract

With the development of smart meters, electric utilities can obtain precise and detailed data of single consumers. This has lead to the generation of various big data sets with previously unattainable information leading to efficient smart grid performance in areas such as load forecasting. Accurate load forecasting promotes effective functioning of smart grids and power system ultimately leading to proper utilization of resources and avoidance of power outages. Meter level predictions, also known as domestic level forecasting, for short period of time finds applications in load scheduling, management of demand response and storage systems. This report describes the use of an approach combined using Seasonal Autoregressive Integrated Moving Average Exogenous (SARIMAX) and Long Short Term Memory networks (LSTM) with regression models to predict daily energy consumption of customers located in London area.

Acknowledgements

Success is nourished under the combination of perfect guidance, care blessing. Acknowledgement is the best way to convey. We express deep sense of gratitude brightness to the outstanding permutations associated with success. Last few years spend in this estimated institution has molded us into condent and aspiring Engineers. We express our sense of gratitude towards our project guide Prof. A.W. Kale. It is because of his valuable guidance, analytical approach and encouragement that we could learn and work on the project. We will always cherish the great experience to work under their enthusiastic guidance.

We would also like to acknowledge with much appreciation the crucial role of the staff and Head of the Department of Information Technology Department Prof. J. P. Patil, who gave the permission to use the all required software/hardware and the necessary material to completing to the project.

We are also grateful to our principal Dr. M. J. Lengare who not only supporting us in our project but has also encouraging for every creative activity. We extend our special thanks to all teaching and non-teaching staff, friends and well-wishers who directly or indirectly contributing for the success of our maiden mission. Finally, how can we forget our parents whose loving support and faith in us remains our prime source of inspiration. Lastly, we would like to thank all those who directly and indirectly helping to complete this project.

Contents

List of Figures

Abbreviations

LSTM	**L**ong **S**hort **T**erm **M**emory Networks
AMI	**A**dvanced **M**etering **I**nterface
SARIMA	**S**easonal **A**uto**R**egressive **I**ntegrated **M**oving **A**verage
SMAS	**S**mart **M**eter Data **A**nalytics **S**ystem

Chapter 1

INTRODUCTION

1.1 Introduction

As conventional power grids turn into smart power grids, traditional electricity meters are being replaced by smart meters. The main operational advantage of smart metering systems is their ability to automatically collect fine-grained (typically hourly) electricity consumption data, which enables accurate billing without sending workers to customer premises to manually read the meters. Additionally, a new application area has emerged: Smart Meter Data Analytics. The vast amounts of data collected by smart metering systems can help understand electricity consumption patterns, thereby helping governments and utilities plan for the future and helping consumers reduce their bills. The market for smart grid data analytics is growing rapidly and is expected to reach over 4 billion USD by year 2020

1.2 Objectives

In Smart Meter Data Analytics, we will study databases and try to discover patterns and on that basis will create a data models.
Our study includes, Combining all blocks into a single data frame- keeping on relevant columns.
Use day-level energy consumption data per household to normalize data for inconsistent household count.
Explore relationships between weather conditions and energy consumptions and create clusters for the weather data.
With the help of data models we will try to find out the suitable algorithm for our dataset for data processing.

1.3 Purpose, Scope, and Applicability

Purpose, Scope and Applicability: The description of Purpose, Scope, and Applicability are given below:

1.3.1 Purpose

Since the inception of electricity deregulation and market-driven pricing throughout the world, utilities have been looking for a means to match consumption with generation. Non-smart electrical and gas meters only measure total consumption, providing no information of when the energy was consumed. Smart meters provide a way of measuring this site-specific information, allowing utility companies to charge different prices for consumption according to the time of day and the season.

1.3.2 Scope

We will use 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.
We also collected weather data for London area from the darksky to explore relationships between weather conditions and energy consumptions and create clusters for the weather data- using which we can add weather identifiers to day-level data.
We will try to discover patterns by studying these datasets and on that basis will create a data models to find out the suitable algorithm for data processing.
We will use Python and Pandas. Python provides constructs that enable clear programming on both small and large scales. And pandas is a software library written for the Python programming language for data analysis.
On that basis we will create a system that would help the consumers recognize their electric usage patterns and be able to optimize their usage.

1.3.3 Applicability

Applicability: This system will help the customers to analyse their electricity usage patterns which will help them to optimize their usage. It will also be able to predict the electricity usage for the future.

1.4 Organisation of Report

Organization of Report: Further we have Chapter wise overview of our project report. Next chapter is Literature Survey where we'll discuss the previous works that were performed by others. Then comes the Survey of Technologies in which we'll discuss about the softwares and Programming languages which can be used to build our project. Requirement and Analysis chapter brings out the softwares and hardwares that will be required for making our project. We'll see these concepts in futher chapters.

Chapter 2

LITERATURE SURVEY

SMAS: A Smart Meter Data Analytics System, 2015 IEEE 31st International Conference on Data Engineering, Authors:A. J. Nezhad, T. K. Wijaya, M. Vasirani, and K. Aberer.

Smart electricity meters are replacing conventional meters worldwide and have enabled a new application domain:smart meter data analytics. In this paper, we introduce SMAS, which demonstrates the actionable insight that consumers and utilities can obtain from smart meter data. Notably, we implemented SMAS inside a relational database management system using open source tools: PostgreSQL and the MADLib machine learning toolkit. In the proposed demonstration, conference attendees will interact with SMAS as electricity providers, consultants and consumers, and will perform various analyses on real data sets.
Observations and Gap Identified: 1. Electricity providers will identify different types of consumers and predict future consumption. 2.Electricity consultants, they will perform virtual audits to understand consumption patterns and find ways to save electricity. 3.Electricity consumers, they will see how their consumption ranks against that of their neighbours, and they will obtain advice on how to change their consumption habits to lower their bills and reduce their carbon footprint.

Advanced Analytics for Harnessing the Power of Smart Meter Big Data, 2013 IEEE International Workshop on Inteligent Energy Systems (IWIES), Authors: R. Gerwen, S. Jaarsma and R. Wilhite.

Smart meters are the basic building block of the smart grid. The key functionality of the smart meter is the capture and transfer of data relating to the consumption (electricity, gas) and events such as power quality and meter status. Such capability has also resulted in the generation of an unprecedented data volume, speed of collection and complexity, which has resulted in the so called big data challenge. In this paper we define a smart metering landscape and discuss different technologies available for harnessing the smart meter captured data.

Observations and Gap Identified: 1.This paper has described smart meter data as big data and presented a smart metering landscape which can be used to position diverse meter data analytics applications. 2.AMI data can be categorised as consumption and events data and at present most of the analytics is carried out on consumption data. Although the analytics techniques used are not novel, the volume and velocity and the variance of the AMI collected data have enabled applications such as load profiling ,forecasting, pricing intelligence.

Chapter 3

SURVEY OF TECHNOLOGIES

3.1 Smart Meter

A smart meter is an electronic device that records consumption of electric energy and communicates the information to the electricity supplier for monitoring and billing. Smart meters typically record energy hourly or more frequently, and report at least daily. Smart meters enable two-way communication between the meter and the central system. Such an advanced metering infrastructure (AMI) differs from automatic meter reading (AMR) in that it enables two-way communication between the meter and the supplier. Communications from the meter to the network may be wireless, or via fixed wired connections such as power line carrier (PLC). Wireless communication options in common use include cellular communications (which can be expensive), Wi-Fi (readily available), wireless ad hoc networks over Wi-Fi, wireless mesh networks, low power long range wireless (LODA), ZigBee (low power, low data rate wireless), and Wi-SUN (Smart Utility Networks).

3.2 Python

Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace. It provides constructs that enable clear programming on both small and large scales. Van Rossum led the language community until July 2018. Python is dynamically typed and garbage-collected. It supports multiple programming paradigms, including procedural, object-oriented, and functional programming. Python features a comprehensive standard library, and is referred to as "batteries included". Python interpreters are available for many operating systems. CPython, the reference implementation of Python, is open-source software and has a community-based development model. Python and CPython are managed by the non-profit Python Software Foundation.

3.3 Pandas

Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It is free software released under the three-clause BSD license. The name is derived from the term "panel data", an econometrics term for data sets that include observations over multiple time periods for the same individuals. Pandas is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language. Pandas is a NumFOCUS sponsored project. This will help ensure the success of development of pandas as a world-class open-source project, and makes it possible to donate to the project.

3.4 Machine Learning

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

3.5 Big Data

Big data is a field that treats ways to analyze, systematically extract information from, or otherwise deal with data sets that are too large or complex to be dealt with by traditional data-processing application software. Data with many cases (rows) offer greater statistical power, while data with higher complexity (more attributes or columns) may lead to a higher false discovery rate. Big data challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating, information privacy and data source. Big data was originally associated with three key concepts: volume, variety, and velocity. Other concepts later attributed with big data are veracity (i.e., how much noise is in the data) and value.
Current usage of the term big data tends to refer to the use of predictive analytics, user behavior analytics, or certain other advanced data analytics methods that extract value from data, and seldom to a particular size of data set. "There is little doubt that the quantities of data now available are indeed large, but that's not the most relevant characteristic of this new data ecosystem." Analysis of data sets can find new correlations to "spot business trends, prevent diseases, combat crime and so on." Scientists, business executives, practitioners of medicine, advertising and governments alike regularly meet difficulties with large data-sets in areas including Internet search, fintech, urban informatics, and business informatics. Scientists encounter limitations in e-Science work, including meteorology, genomics, connectomics, complex physics simulations, biology and environmental research.

3.6 LSTM

Long Short Term Memory networks – usually just called "LSTMs" – are a special kind of RNN, capable of learning long-term dependencies. They were introduced by Hochreiter

Schmidhuber (1997), and were refined and popularized by many people in following work.1 They work tremendously well on a large variety of problems, and are now widely used.

LSTMs are explicitly designed to avoid the long-term dependency problem. Remembering information for long periods of time is practically their default behavior, not something they struggle to learn!

3.7 SARIMA

SARIMA is an Ideology that captures autocorrelation in the series by modelling it directly.It is a series which needs to be differentiated in order to be made stationary is an integrated (I) series.Lags of the stationarized series are called ˆaautoregressiveˆa that refers to (AR) terms

Lags of the forecast errors are called moving average which refers to (MA) terms.It is basically used for forecasting Arima is a Generalized random walk models which is fine-tuned

to eliminate all residual autocorrelation. It is a Generalized exponential smoothing model that can incorporate long-term trends and seasonality. The Stationarized regression model uses lags of the dependent variables and/or lags of the forecast errors as regressors.Here the

forecasting model of time series can be stationarized by using transformations like differencing, logging and deflating. By this we can say that a time series is ˆaStationaryˆa if all the

Statistical properties like mean, variance, autocorrelation etc. are constant in time.

Chapter 4

REQUIREMENTS AND ANALYSIS

4.1 Problem Definition

Study datasets, try to discover patterns, and on that basis will create a data models to find out the suitable algorithm for our dataset for data processing, and thus, predicting the future readings.

4.2 Requirements Specification

Data server (Such as: Core i5, RAM 4 GB, 120 GB Hard Drive)

Operating System (Such as: Windows/ Linux Operating System)

Smart meter datasets

Pandas Library

Programming languages (Python, Java, Hadoop)

Neural networks (LSMT, ARIMA, SARIMA)

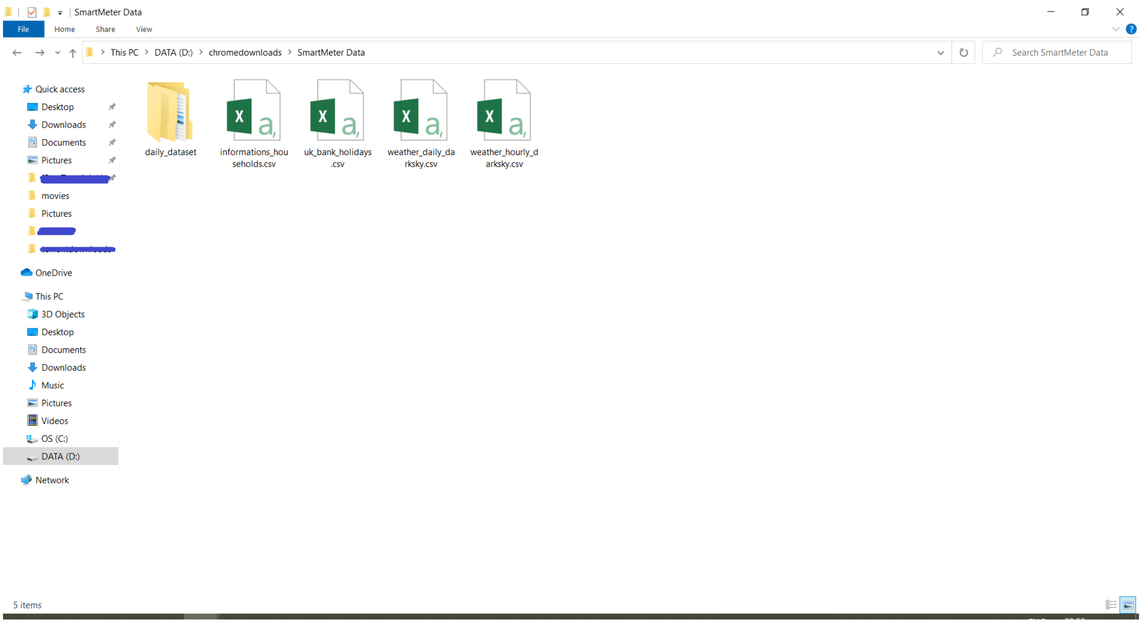

FIGURE 4.1: Datasets

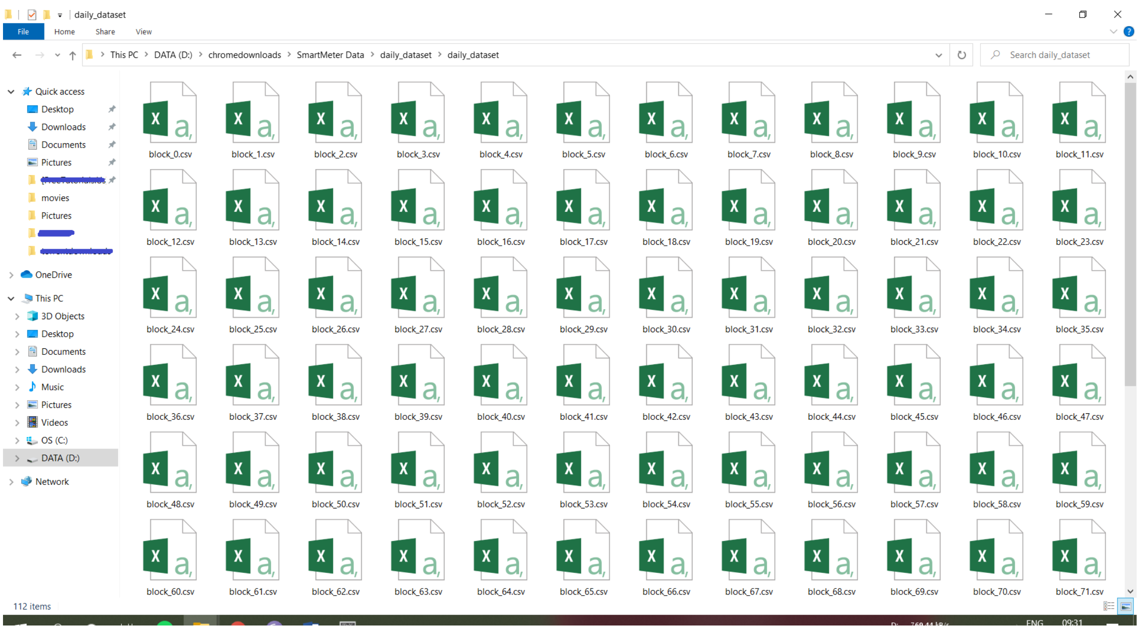

FIGURE 4.2: Datasets

4.3 Planning and Scheduling

August - Revise Python. Collect a pre-existing dataset of readings of Smart Meters. Learn Cloud Computing Technologies.
September - Learn requirements for data analysis.
October - Process and plot graphs and other charts of the data of smart meters. Learn predictive analytics.
November - Learn and implement various data analysis algorithms required for predictive analytics.
December - Implement algorithms that will take various data related decisions and display them on the dashboard.
January - Interface the system with cloud and smart meters.
February - Test and provide a complete data and predictive analytics of the electricity meter data on the dashboard.
March - Test and provide a complete data and predictive analytics of the electricity meter data on the dashboard using which energy loss can be prevented.

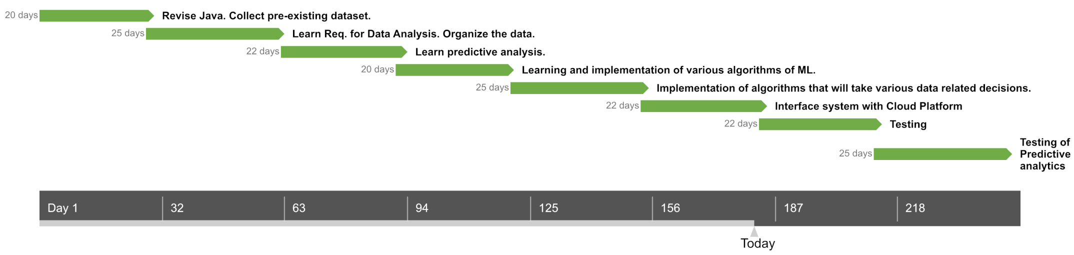

FIGURE 4.3: Gantt Chart

4.4 Software and Hardware Requirements

Hardware Requirement: Data server (Such as: Core i5, RAM 4 GB, 120 GB Hard Drive)

Software Requirements: Windows or Linux Operating System, Pandas Library, Smart meter data sheets, Programming languages (Python, Java, Hadoop), Neural networks (LSMT, ARIMA, SARIMA)

4.5 Preliminary Product Description

This project will contain a GUI which will take different inputs from the user such as Date, Previous Day's consumption, Temperature, Humidity, etc and after you click Predict, it will give you the predicted value.

4.6 Conceptual Models

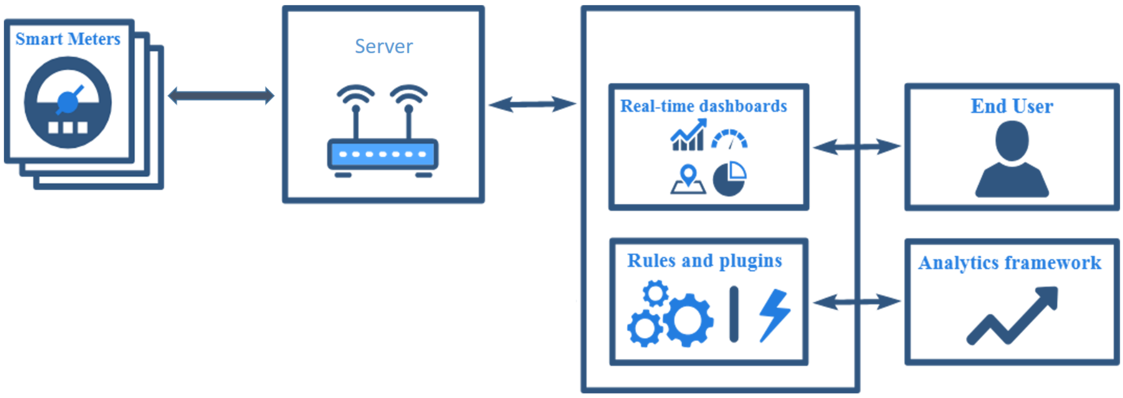

FIGURE 4.4: Proposed Block Diagram

Chapter 5

SYSTEM DESIGN

Describes desired features and operations in detail, including screen layouts, business rules, process diagrams, pseudo code and other documentation. Elements of Project Development

5.1 User interface design

Here the user will enter the data like the day and month of which the value is to be predicted. Also some additional data such as Temperature, Humidity, Windspeed are needed to be entered because these parameters directly relate to the energy consumption. On the day of a Holiday, the energy consumption tends to increase due to more usage so we need to specify that in binary(0=false/1=true). Thus, predicting the energy consumption of a particular day.

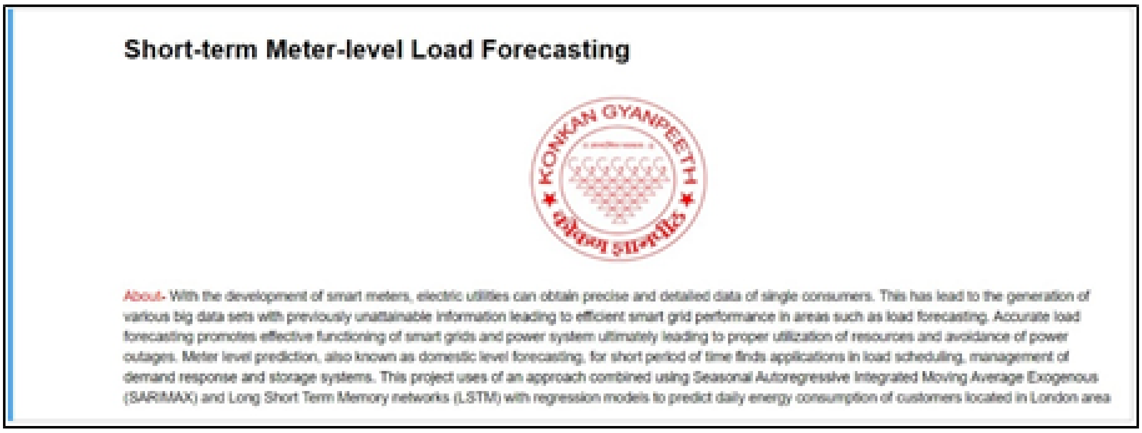

FIGURE 5.1: Introduction in GUI

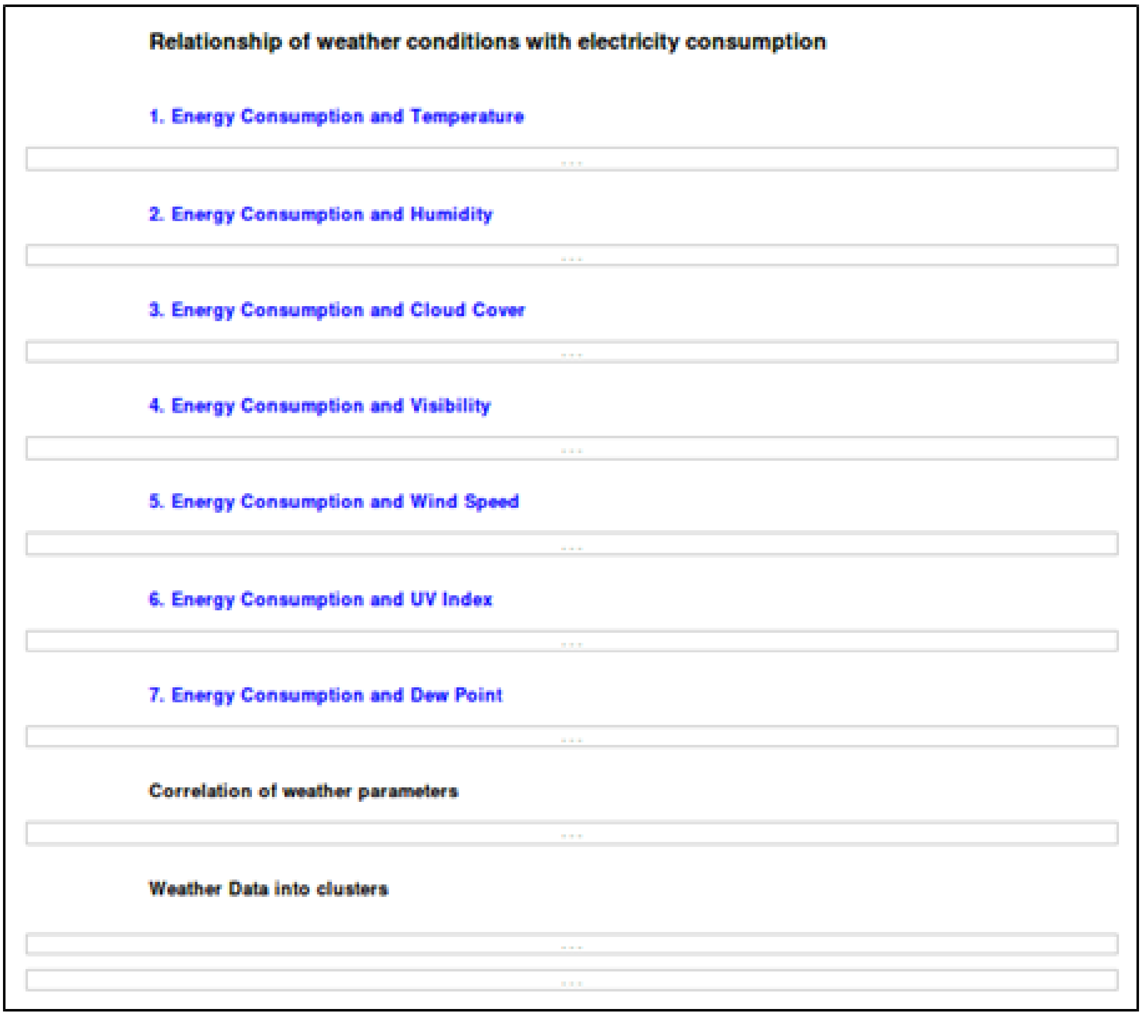

FIGURE 5.2: Weather condition analysis

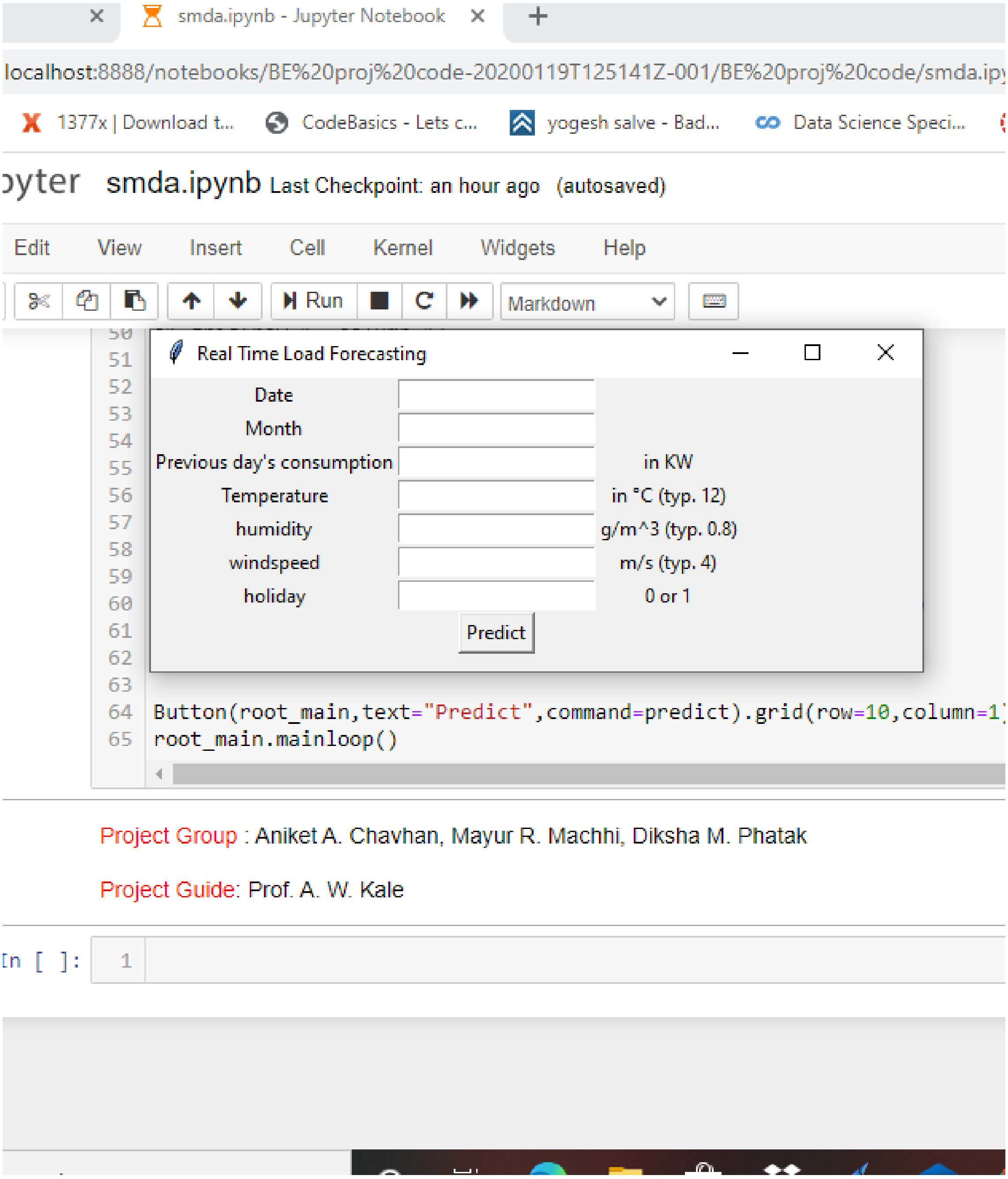

Figure 5.3: Real Time Demonstration GUI

5.2 Flowchart

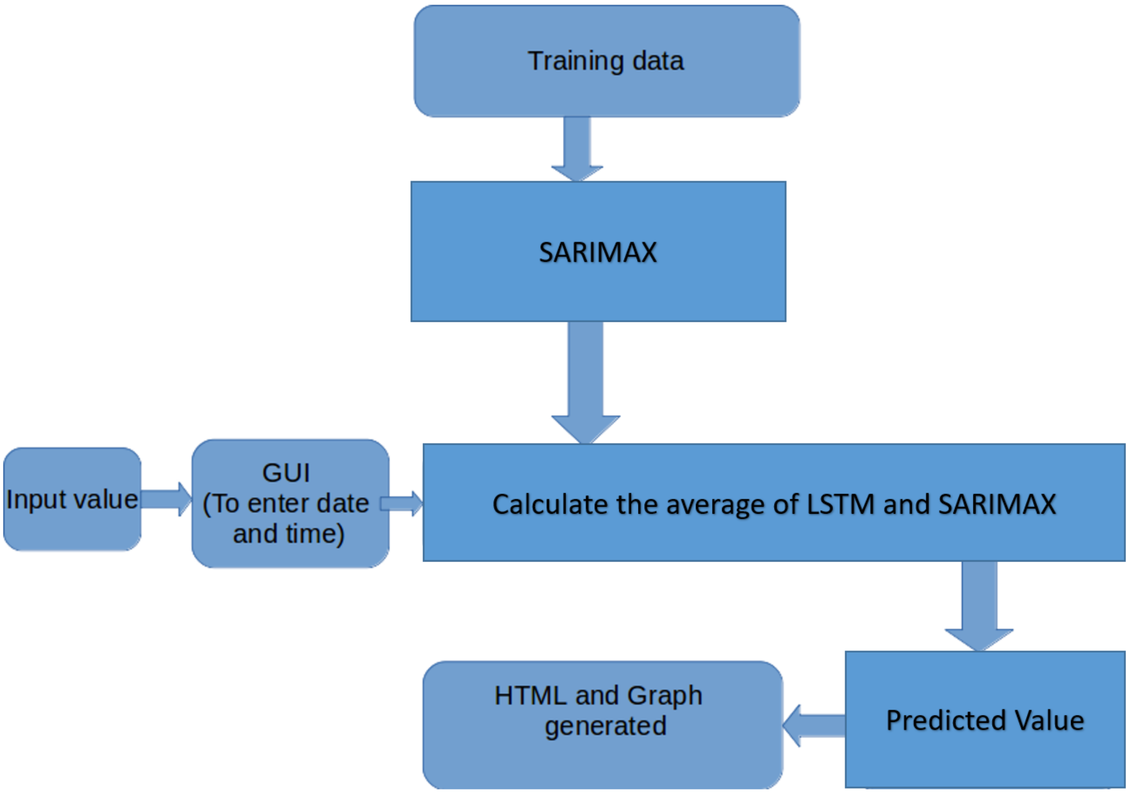

FIGURE 5.4: Flowchart

Chapter 6

RESULTS AND DISCUSSIONS

6.1 Result

The interface that we have created seems to be working by predicting the energy consumption in kW for the dates entered by the user. It also takes into consideration different parameters such as Temperature, Humidity, Windspeed because these parameters directly relate to the energy consumption.

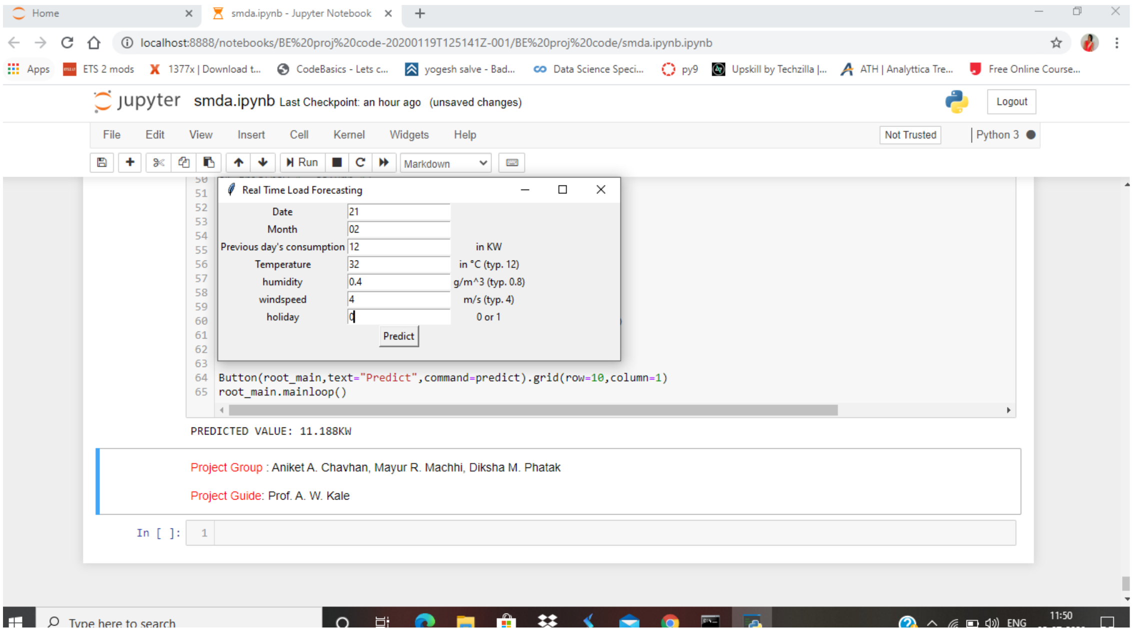

Figure 6.1: Prediction No. 1

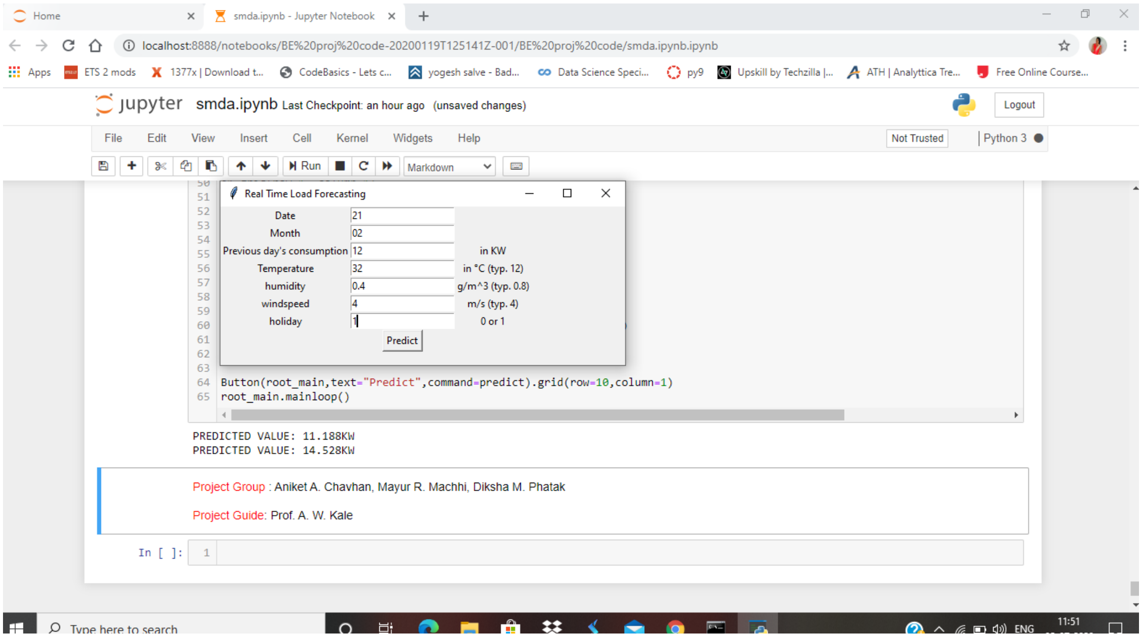

FIGURE 6.2: Prediction No. 2

Also the holidays tend to increase the energy consumption due to more people being at home rather than work or schools and thus increasing usage of electrical appliances.

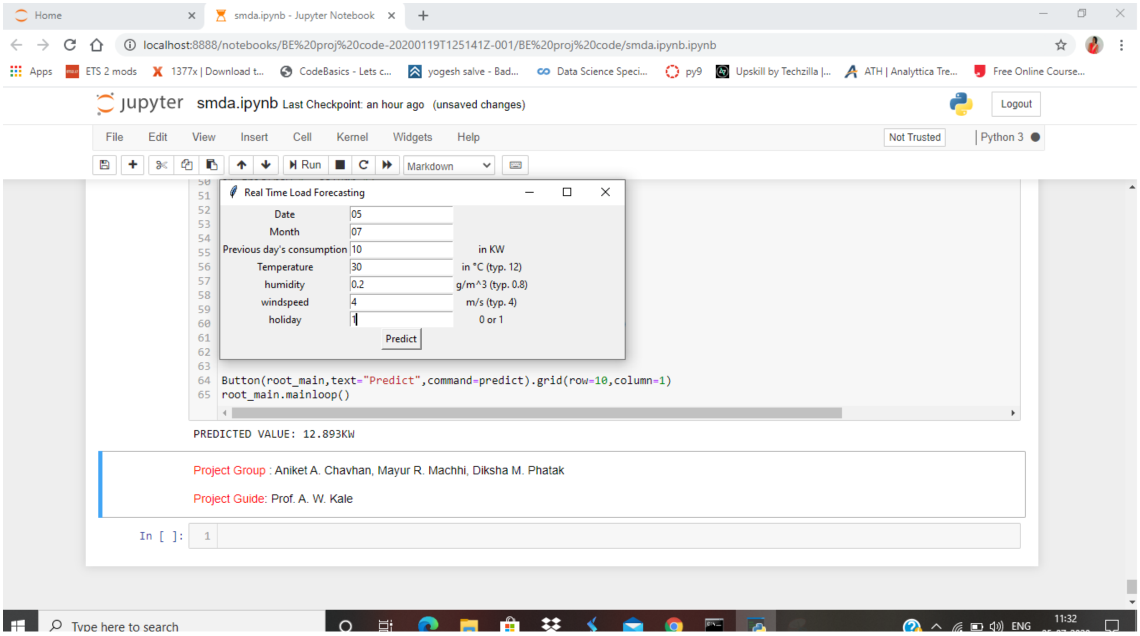

FIGURE 6.3: Prediction No. 3

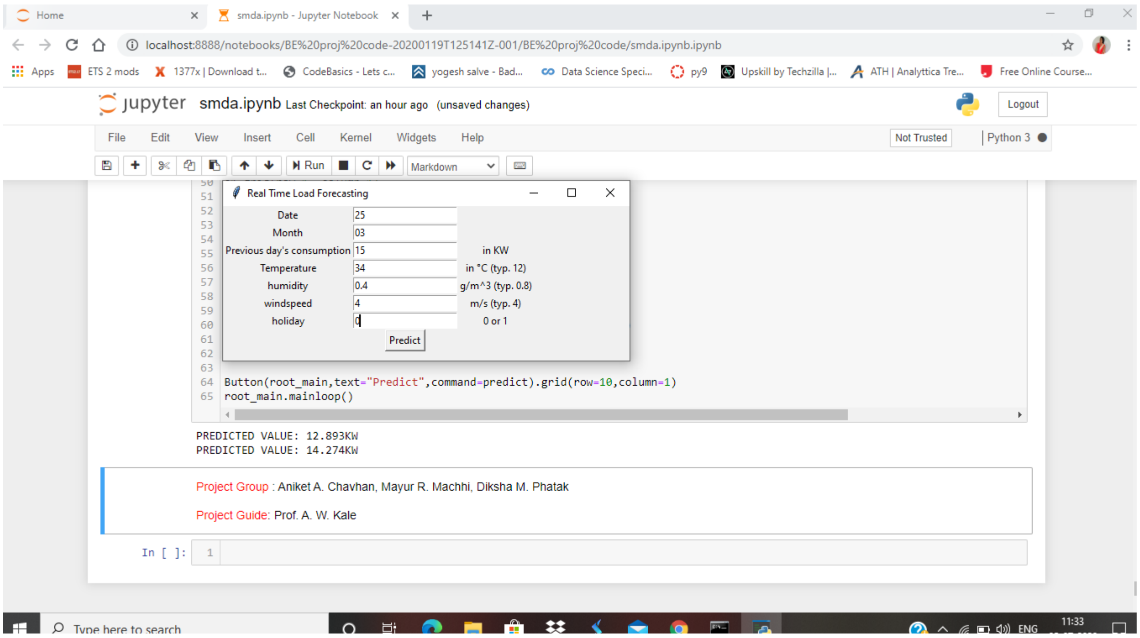

FIGURE 6.4: Prediction No. 4

Further is the graph of the actual consumption and predicted consumption vs time. We can see fairly accurate predictions done.

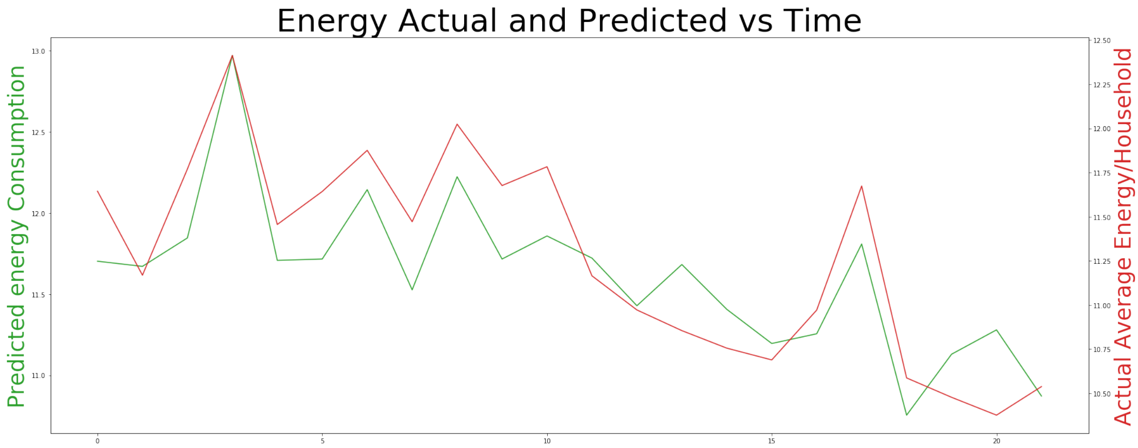

FIGURE 6.5: Energy Consumption Actual and Predicted

6.2 Discussions

The electric demand in most every electric utility system displays variations that have an obvious tie to season and weather. Demand is higher in summer and winter than at other times of the year. During periods of extremely hot weather, or very cold weather, demand is higher than when weather conditions are milder. The tie between weather and demand seems clear to most utility engineers, operators, and planners. "Weather sensitive demands" cause it - appliances and usage patterns that are (at least partly) influenced by weather conditions. During hotter weather, air conditioners have to run more often, thereby using more electric power. Similarly, during really cold spells, more electric power is used by residential, commercial, and industrial heating systems than during periods of mild weather. This chapter address three highly intertwined analytical applications involving weather and electric demand forecasting:

1. Understanding of consumer needs and load patterns can come from analysis of weather sensitive demand. Study of weather and its impact on load can help utility planners and marketers better understand the variability of demand. It can provide indications of the composition of consumer demand - the most obvious example

being how the market penetrations of appliances like air conditioning and chillers change over time.

2. Weather normalization of historical load readings involves adjusting load readings taken at different times (and therefore under different weather conditions) so they approximate the readings that would have been taken under identical weather conditions. Weather varies from year to year and month to month. As a result, demand readings meant to represent annual peak readings, for example, all taken at different times, are not directly comparable. These demand readings must be adjusted to a standard set of weather conditions if forecasters are to make a valid comparison of those readings to identify trends in consumer growth and usage patterns.

3. Demand forecasts done to standard weather design criteria involve producing forecasts of future demand that are adjusted to represent expected demand under some standard set of weather conditions. Generally, the forecasts will be produced to represent demands under the same "standard conditions" that represent "extreme enough" weather and demands for planning.

6.2.1 Weather Analysis

We can see that energy and temperature have an inverse relationship-we can see the peaks in one appearing with troughs in the other. This confirms the business intuition that during low temperature, it is likely that the energy consumption through heaters etc. increases as in Fig 6.6. Humidity and the average consumption of energy seems to have the same trend as in Fig 6.7. The cloud cover value seems to be following the same pattern as the energy consumption in Fig 6.8. The visibility factor does not seem to affect energy consumption at all- since visibility is most likely an outdoors factor, it is unlikely that it's increase or decrease affects energy consumption within a household in Fig 6.9. Like visibility, wind speed seems to be an outdoors factor which does not affect in the energy consumption as such as in Fig 6.10. The UV index has an inverse relationship with energy consumption as shown in Fig 6.11. Dew Point- is a function of humidity and temperature therefore it displays similar relation to energy consumption as in Fig 6.12.

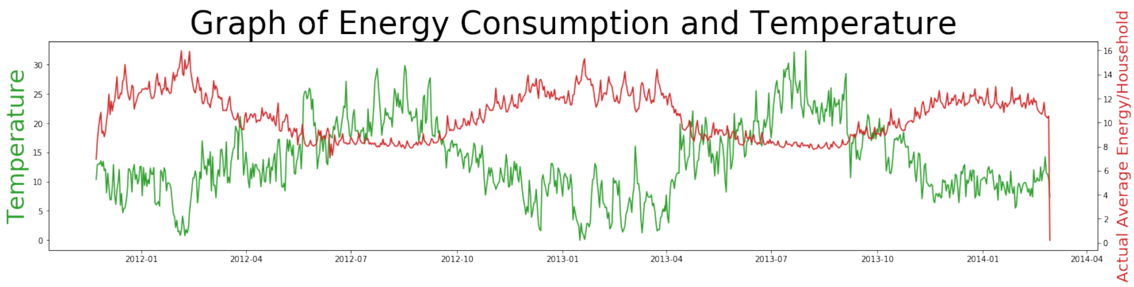

FIGURE 6.6: Energy Consumption and Temperature

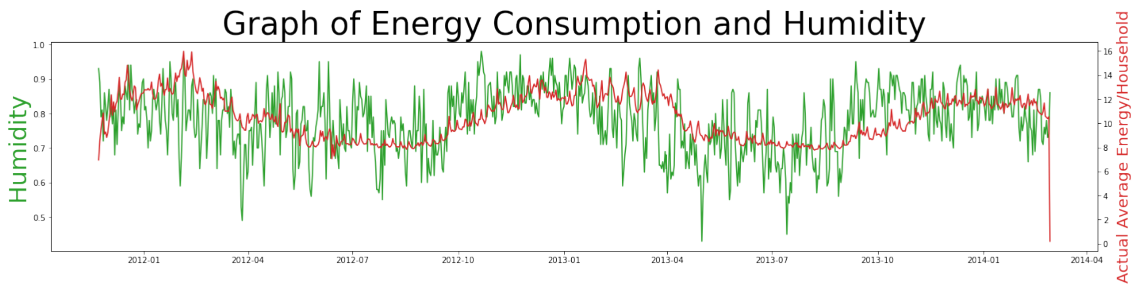

FIGURE 6.7: Energy Consumption and Humidity

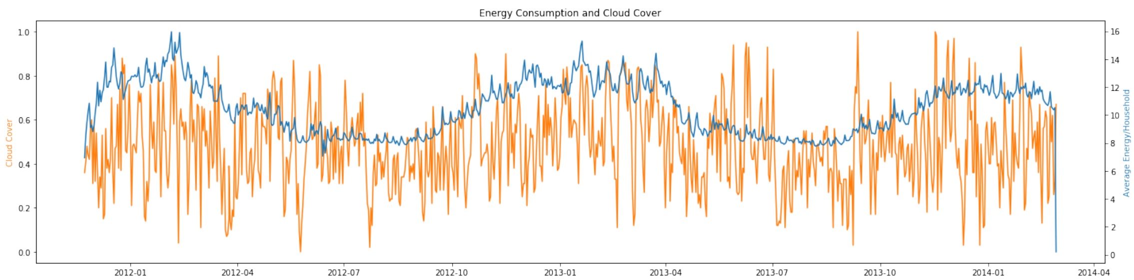

FIGURE 6.8: Energy Consumption and Cloud Cover

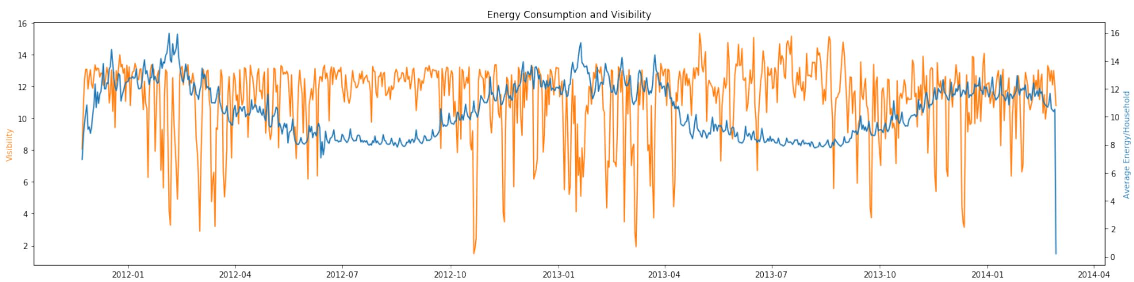

FIGURE 6.9: Energy Consumption and Visibility

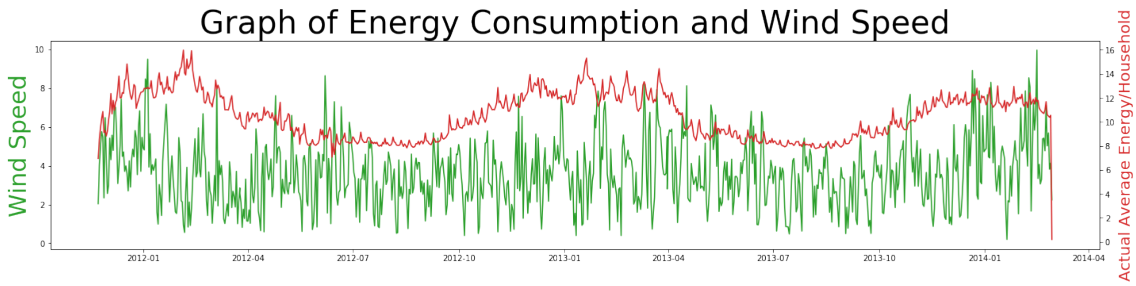

FIGURE 6.10: Energy Consumption and Wind Speed

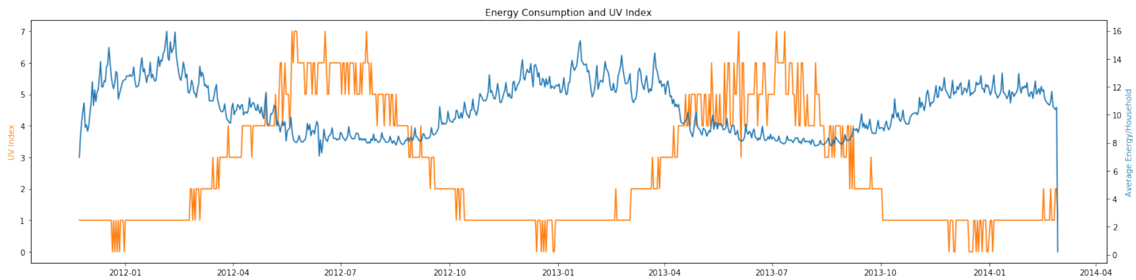

FIGURE 6.11: Energy Consumption and UV Index

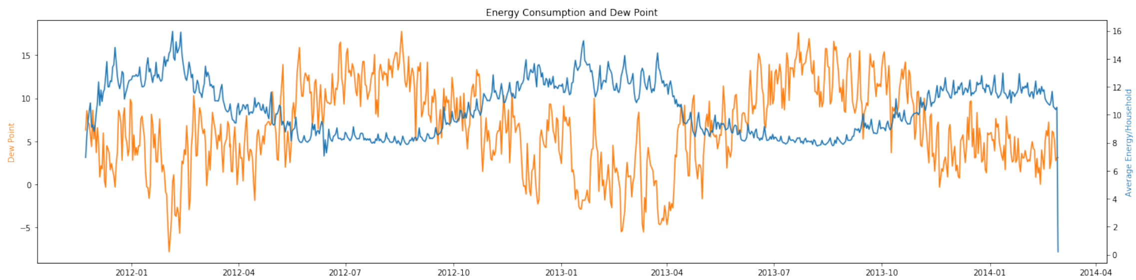

FIGURE 6.12: Energy Consumption and Dew Point

Chapter 7

CONCLUSIONS

7.1 Conclusions

The main feature highlighted in this report is the use of time series forcasting and deep learning model for short term meter level load forcasting which proved to be a more accurate approach to predict the energy values. The report discusses the detail implementation of the algorithm used. The key advantage of using the algorithm mentioned in this paper is the accuracy that can be achieved as seen in the results. Thus combining traditional time series algorithm with state-of-art deep learning model using machine learning model proves to provide better results.

7.1.1 Limitations of the System

The data we have collected across households are inconsistent, therefore we will be using energy per household as the target to predict rather than energy alone. This is an optional step as we can also predict for energy sum as whole for each household. However there are quite a lot of unique households for which we have to repeat the exercise and our ultimate goal is to predict overall consumption forecast and not at household level. This also means that since household level is removed, we are not looking into the details which is available at household level.

Bibliography

[1] Yi Wang, Qixin Chen, Tao Hong, Chongqing Kang, *Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges* IEEE TRANS. SMART GRID, March 2018.

[2] R. R. Mohassel, A. Fung, F. Mohammadi, and K. Raahemifar, *A survey on advanced metering infrastructure,* International Journal of Electrical Power Energy Systems, vol. 63, pp. 473â484, 2014.

[3] How many smart meters are installed in the united states, and who has them? https://www.eia.gov/tools/faqs/faq.php?id=108t=3, accessed July 31, 2017.

[4] Number of electric smart meters deployed in the u.s. from 2008 to 2016, https://www.statista.com/statistics/499704/number-of-smart-meters-in-the-united-states/, accessed July 31, 2017.

[5] Department for Business Energy and Industrial Strategy, *âSmart Meters, Quarterly Report to end December 2016, Great Britain,* Tech. Rep., 2017.

[6] Chen, T. and Y.-C. Wang, *Long-term load forecasting by a collaborative fuzzy-neural approach.* International Journal of Electrical Power Energy Systems, 2012. 43(1): p. 454-464.

[7] Srivastava, A.K., A.S. Pandey, and D. Singh.*Short-term load forecasting methods: A review.* in 2016 International Conference on Emerging 11Trends in Electrical Electronics Sustainable Energy Systems (ICETEESES). 2016.

[8] A. J. Nezhad, T. K. Wijaya, M. Vasirani, and K. Aberer. SmartD: Smart Meter Data Analytics Dashboard, in Proc. Int. Conf. on Future Energy Systems (ACM e-Energy), 213-214, 2014.

[9] R. Gerwen, S. Jaarsma and R. Wilhite, Smart Metering, Distributed Generation., pp 1-9, June 2006

[10] https://en.wikipedia.org/wiki/Long_short-term_memory Available as on 2nd April 2019

[11] https://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average Available as on 2nd April 2019

[12] https://en.wikipedia.org/wiki/Regression Available as on 2nd April 2019

[13] Elham M. Eskandarnia, Sameem Abdul Kareem, Hesham M. Al-Ammal, *A Review of Smart Meter Load Forecasting Techniques: Scale and Horizon* International Conference on Information Society and Smart Cities ICS2018, Cambridge, UK

[14] K. S. K. Weranga, Sisil Kumarawadu, D. P. Chandima, Smart Metering Design and Applications SPRINGER BRIEFS IN APPLIED SCIENCES AND TECHNOLOGY

[15] S. A. Soliman, Ahmad M Alkandari, Electrical load forecasting : modeling and model construction

[16] Omkar Nibandhe and Abhishek Korpe, Smart Meter Data Analytic Using Hadoop 2014 Hadoop Summit

[17] H. Lee Willis, Spatial Electric Load Forecasting

[18] Kuster, C., Y. Rezgui, and M. Mourshed, âElectrical load forecasting models: A critical systematic reviewâ Sustainable Cities and Society, 2017. 35: p. 257-270.

[19] Mohamed Akram Zaytar, Chaker El Amrani, Sequence to Sequence Weather Forecasting with Long Short-Term Memory Recurrent Neural Networks, International Journal of Computer Applications 143(11):7-11, June 2016