Smart Meter Data Analytics

presented by

Mayur Machhi

Diksha Phatak

Aniket Chavhan

Under Guidance of

Prof. A. W. Kale

Department of Information Technology

Konkan Gyanpeeth’s College of Engineering, Karjat. 410201

University Of Mumbai

Feb 2020

Presentation Outline

- Introduction
- Aims and Objectives
- Scope of Project
- Literature Survey
- Data sets and Algorithms
- Problem Statement
- Project Methodology
- Results
- Objectives Achieved
- Conclusion

Introduction

India's energy scenario

- India is the 3rd largest producer of electricity in the world

- The world loses $89.3 billion annually through power theft while India loses US$ 10- 15 billion annually

- 240 million Indians live without access to electricity

Introduction

What are SMART METERS?

House/
Industry

Smart Meter

Server

- Track and store the amount of energy used.
- Send the collected data to the Energy Distribution company server at regular time intervals.

Aims and 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.

Scope of Project

- We will use a 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. As 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 manipulation and 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.

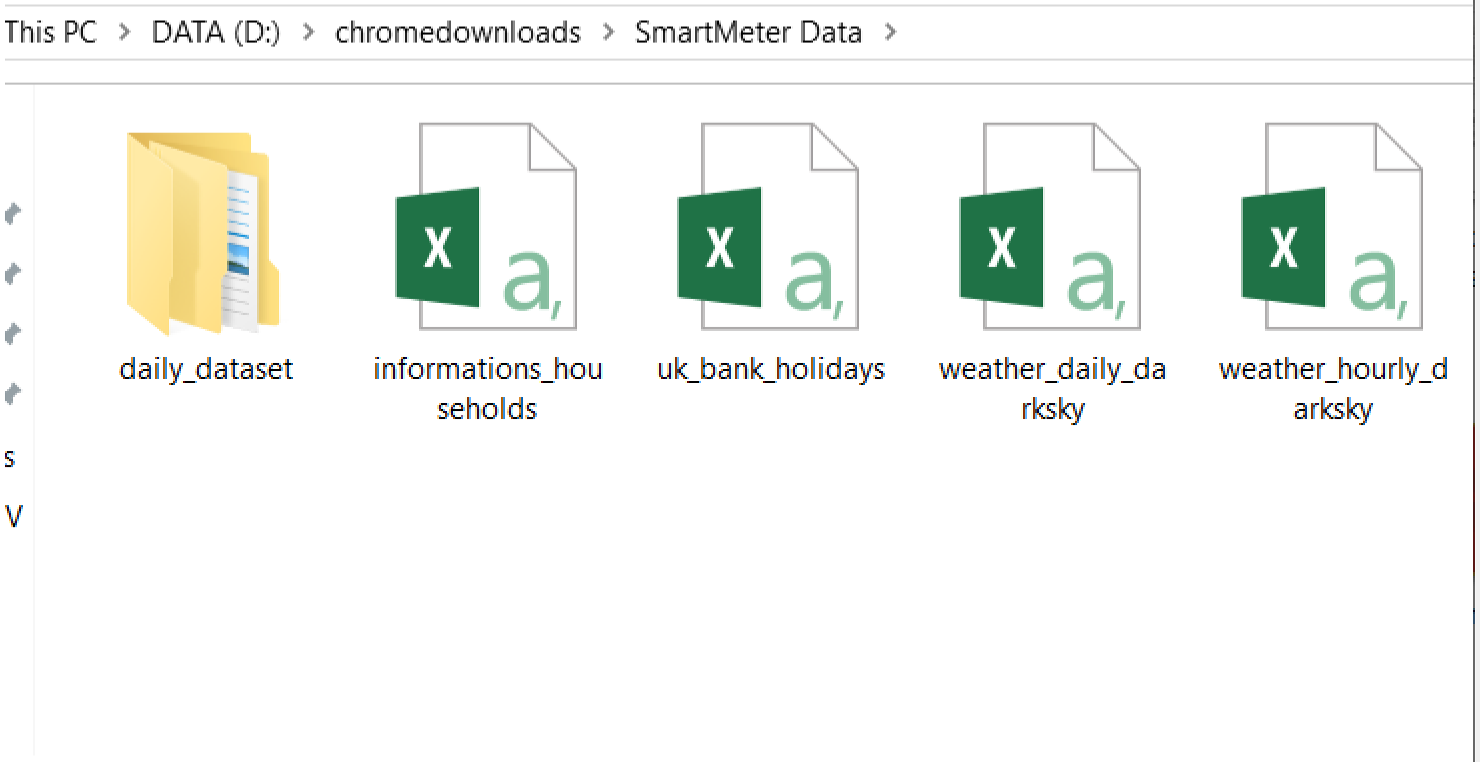
Literature Survey

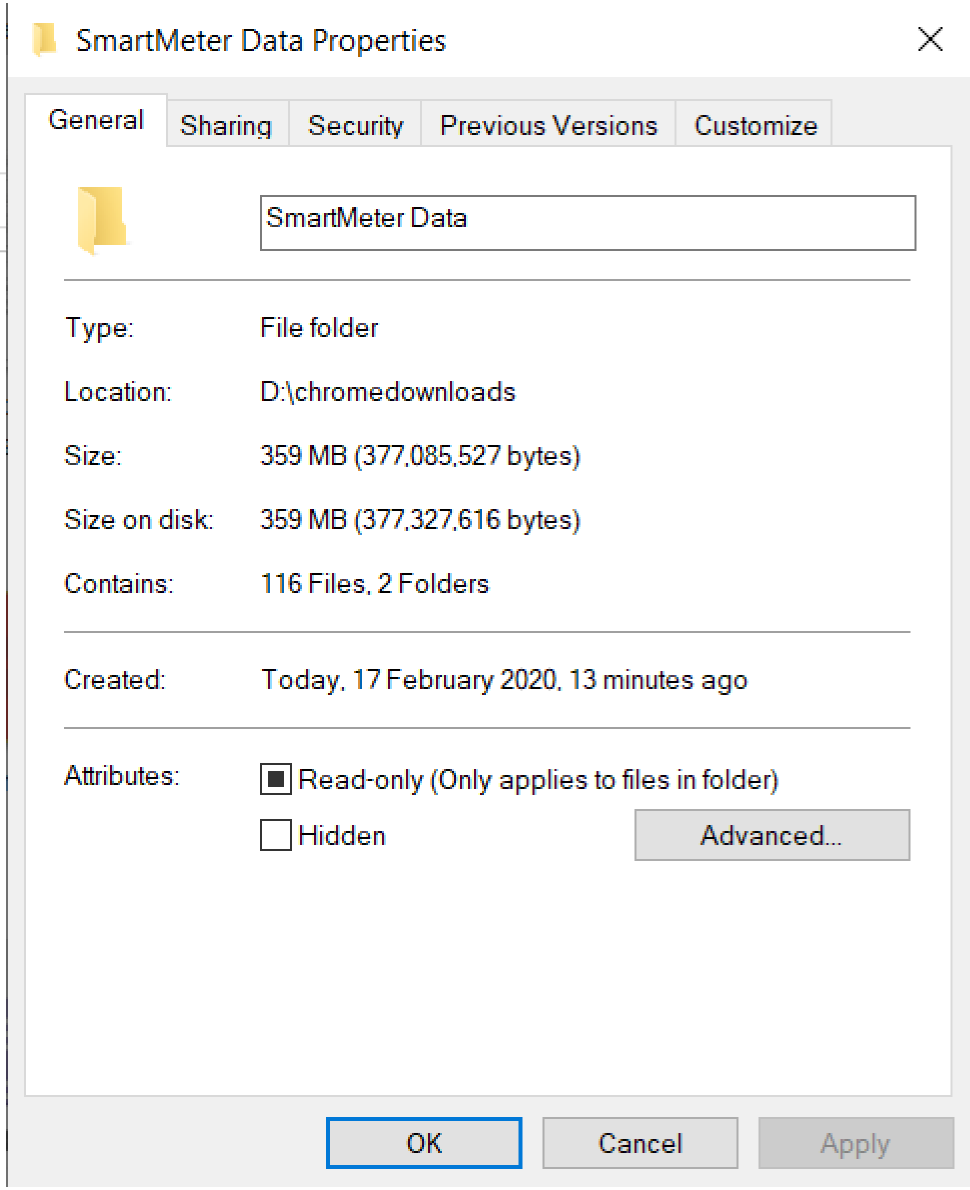
Sr. No.	Title of Paper, Author Names	Details of publication	Brief Description	Observations and Gap Identified
1	SMAS: A Smart Meter Data Analytics System Authors:A. J. Nezhad, T. K. Wijaya, M. Vasirani, and K. Aberer.	1.)2015 IEEE 31st International Conference on Data Engineering 13-17 April 2015 10.1109/ICDE.2015.7113405	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.	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.

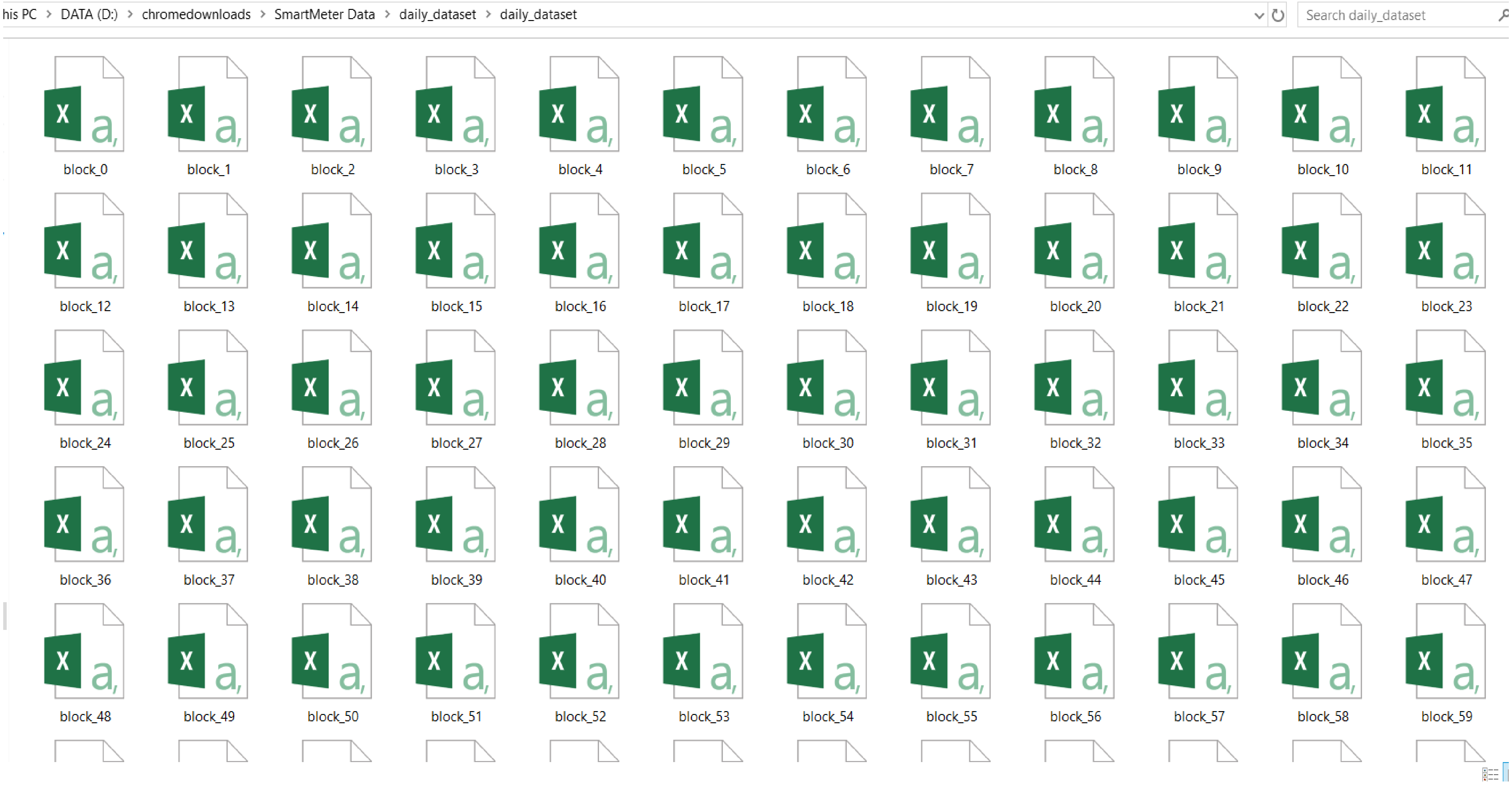
Literature Survey

Sr. No.	Title of Paper, Author Names	Details of publication	Brief Description	Observations and Gap Identified
2	Advanced Analytics for Harnessing the Power of Smart Meter Big Data Authors: R. Gerwen, S. Jaarsma and R. Wilhite	1.) 2013 IEEE International Workshop on Inteligent Energy Systems (IWIES) 14-14 Nov. 2013 10.1109/IWIES.2013.6698559	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.	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 categorized 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.

Data set


This PC > DATA (D:) > chromedownloads > SmartMeter Data >
daily_dataset
informations_households
uk_bank_holidays
weather_daily_darksky
weather_hourly_darksky


SmartMeter Data Properties
General
Sharing
Security
Previous Versions
Customize
SmartMeter Data
Type: File folder
Location: D:\chromedownloads
Size: 359 MB (377,085,527 bytes)
Size on disk: 359 MB (377,327,616 bytes)
Contains: 116 Files, 2 Folders
Created: Today, 17 February 2020, 13 minutes ago
Attributes:
Read-only (Only applies to files in folder)
Hidden
Advanced...
OK
Cancel
Apply



his PC > DATA (D:) > chromedownloads > SmartMeter Data > daily_dataset > daily_dataset
Search daily_dataset
block_0
block_1
block_2
block_3
block_4
block_5
block_6
block_7
block_8
block_9
block_10
block_11
block_12
block_13
block_14
block_15
block_16
block_17
block_18
block_19
block_20
block_21
block_22
block_23
block_24
block_25
block_26
block_27
block_28
block_29
block_30
block_31
block_32
block_33
block_34
block_35
block_36
block_37
block_38
block_39
block_40
block_41
block_42
block_43
block_44
block_45
block_46
block_47
block_48
block_49
block_50
block_51
block_52
block_53
block_54
block_55
block_56
block_57
block_58
block_59

Home Page - Selec | smda - Jupyter No | Presentations - Go | Aftermorning | Energy Consumpti | Downloads | Home Page - Selec | energy.csv - Jupyte

localhost:8888/edit/energy.csv

jupyter energy.csv an hour ago

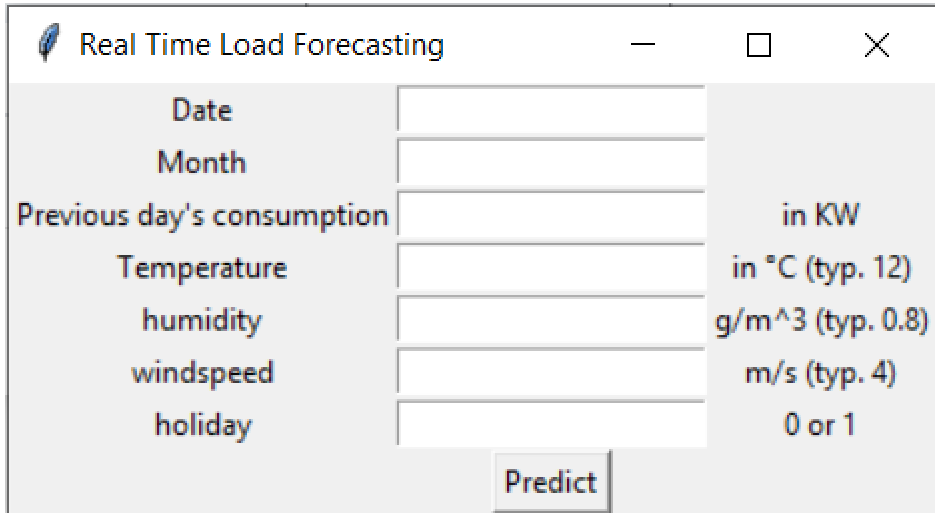
Logout

File Edit View Language current mode

```
22811,2012-03-24,MAC004451,14.8709999
22812,2012-03-25,MAC004451,11.046
22813,2012-03-26,MAC004451,15.180999999999996
22814,2012-03-27,MAC004451,13.142000000000001
22815,2012-03-28,MAC004451,13.249
22816,2012-03-29,MAC004451,12.443
22817,2012-03-30,MAC004451,13.729000000000005
22818,2012-03-31,MAC004451,18.4169999
22819,2012-04-01,MAC004451,13.511000000000005
22820,2012-04-02,MAC004451,12.997
22821,2012-04-03,MAC004451,12.117999999999995
22822,2012-04-04,MAC004451,11.380000000000004
22823,2012-04-05,MAC004451,13.4
22824,2012-04-06,MAC004451,13.853999999999996
22825,2012-04-07,MAC004451,14.951
22826,2012-04-08,MAC004451,16.131999999999998
22827,2012-04-09,MAC004451,12.652999999999999
22828,2012-04-10,MAC004451,12.2419999
22829,2012-04-11,MAC004451,13.398
22830,2012-04-12,MAC004451,14.050999999999998
22831,2012-04-13,MAC004451,13.901
22832,2012-04-14,MAC004451,14.382999999999996
22833,2012-04-15,MAC004451,12.527999999999999
22834,2012-04-16,MAC004451,14.022999900000002
22835,2012-04-17,MAC004451,13.808000000000002
22836,2012-04-18,MAC004451,14.957000000000004
22837,2012-04-19,MAC004451,14.873
22838,2012-04-20,MAC004451,16.538000000000004
22839,2012-04-21,MAC004451,13.968
22840,2012-04-22,MAC004451,14.152999999999999
22841,2012-04-23,MAC004451,13.5519999
22842,2012-04-24,MAC004451,11.874
22843,2012-04-25,MAC004451,13.182
22844,2012-04-26,MAC004451,12.463999999999995
22845,2012-04-27,MAC004451,16.488000000000003
22846,2012-04-28,MAC004451,15.588
22847,2012-04-29,MAC004451,10.514000000000003
22848,2012-04-30,MAC004451,7.121999999999999
```

ENG US 15:27 17/02/2020

GUI


Real Time Load Forecasting
Date
Month
Previous day's consumption
in KW
Temperature
in °C (typ. 12)
humidity
g/m^3 (typ. 0.8)
windspeed
m/s (typ. 4)
holiday
0 or 1
Predict

Problem Statement

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.

Project Methodology

Short-Term Load Forecasting

- In short-term load forecasting (STLF), the future load on a power system is predicted by extrapolating a predetermined relationship between the load and its influential variables namely, time and/or weather.
- Determining this relationship is a two stage process that requires (a) identifying the relationship between the load and the related variables, and (b) quantifying this relationship through the use of a suitable parameter estimation technique.
- It is well recognized that meteorological variables, such as temperature, wind speed, and cloud cover, have a very significant influence on electricity load.

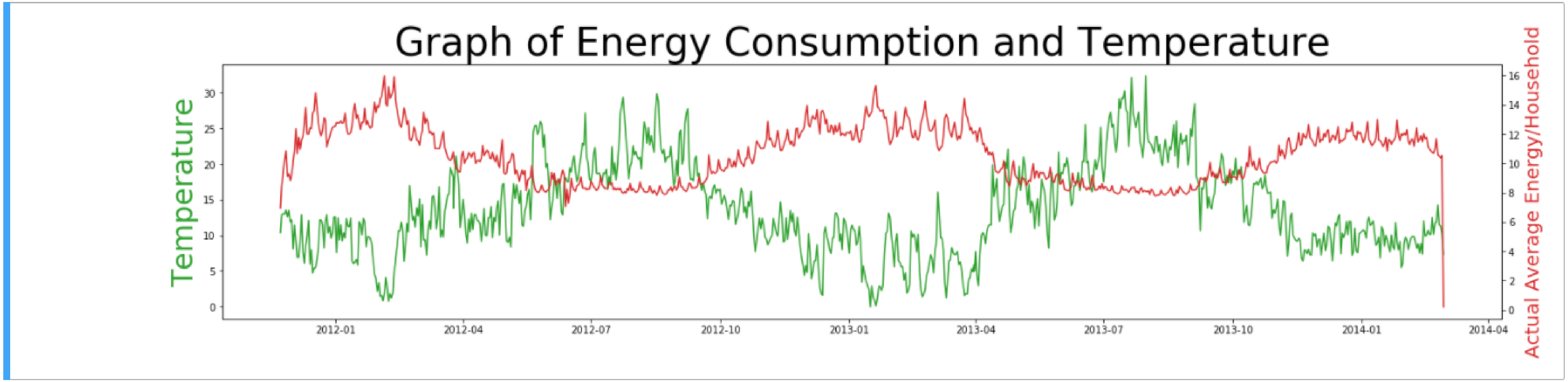
Weather-Dependent Load

- Weather contributes significantly to the dynamics of a load.
- The specific weather variables that are normally used to model weather-dependent load are:
  - Temperature
  - wind speed
  - humidity

Temperature

- The effects of temperature on load pattern are not uniform and are different from one utility to another and from one season to the next.
- A decrease in temperature below room temperature during the winter season means an increase in the heating load, but an increase in the temperature above room temperature during summer means increasing air conditioning load (increasing the cooling load).

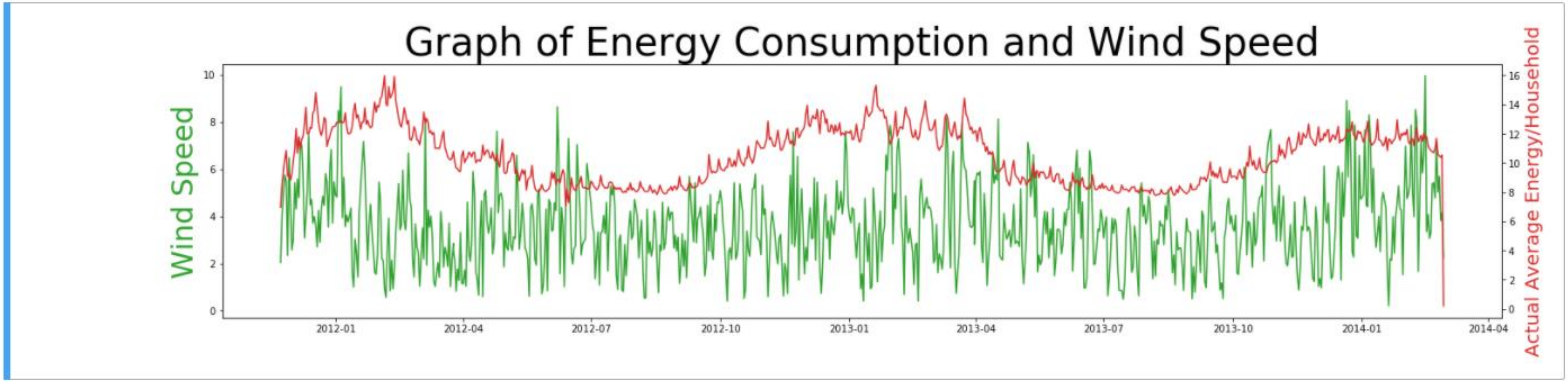
1. Energy Consumption and Temperature



Windspeed

- Wind effects are especially prevalent during winter and are a direct consequence of the cooling power of the wind.
- The cooling effect of the wind depends on the wind speed and the dry bulb temperature.
- The heat loss from a building is proportional to the product of the square root of the wind speed and the temperature deviation from the comfort level of approximately 18°C.
- This effect is relatively small in postwinter seasons.

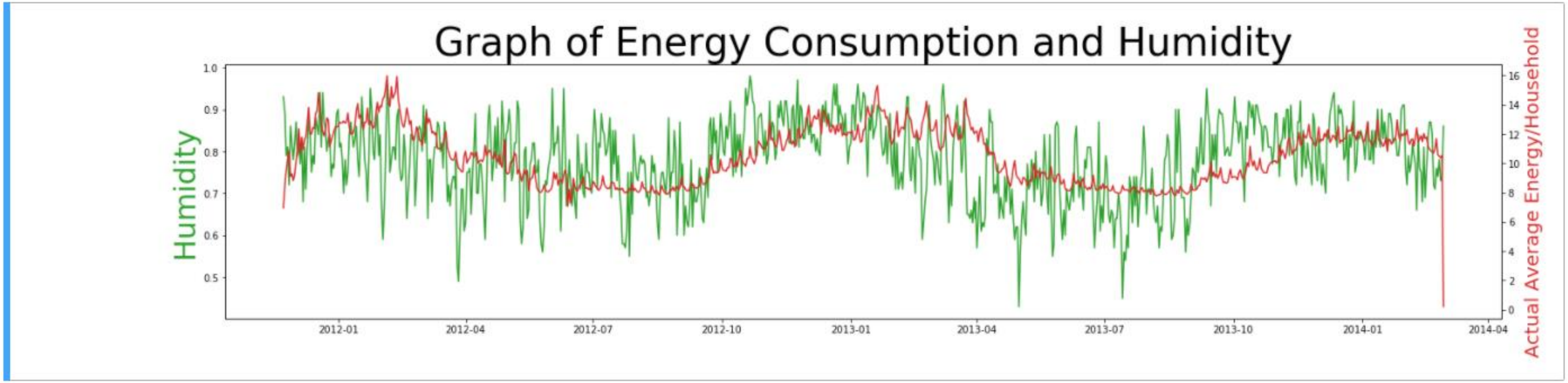
Energy Consumption and Wind Speed



Humidity

- A weather variable that greatly influences air conditioning and other related cooling loads in summer is the level of humidity in the atmosphere.
- The effects of high humidity are generally noticeable only when the temperature is quite high, usually above room temperature.

Energy Consumption and Humidity



Long Short Term Memory networks(LSTM)

- Long Short Term Memory networks usually just called LSTM 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.
- They work tremendously well on a large variety of problems, and are now widely used.

Seasonal Autoregressive Integrated Moving Average (SARIMA)

- Seasonal Autoregressive Integrated Moving Average, SARIMA or Seasonal ARIMA, is an extension of ARIMA that explicitly supports univariate time series data with a seasonal component.
- It adds three new hyperparameters to specify the autoregression (AR), differencing (I) and moving average (MA) for the seasonal component of the series, as well as an additional parameter for the period of the seasonality.

Performance matrics

- The accuracy of forecasting is determined by the performance metrics.
- Essentially, it expresses the error between predication and real load observation.
- Root Mean Square Error (RMSE):Root Mean Square Error (RMSE) is the standard deviation of the residuals (prediction errors).
- The MAE is the most intuitive of the metrics since we are just looking at the absolute difference between the data and the model’s predictions.
- The mean absolute percentage error (MAPE) is the percentage equivalent of MAE. The equation looks just like that of MAE, but with adjustments to convert everything into percentages.

Performance matrics formulas and comparions

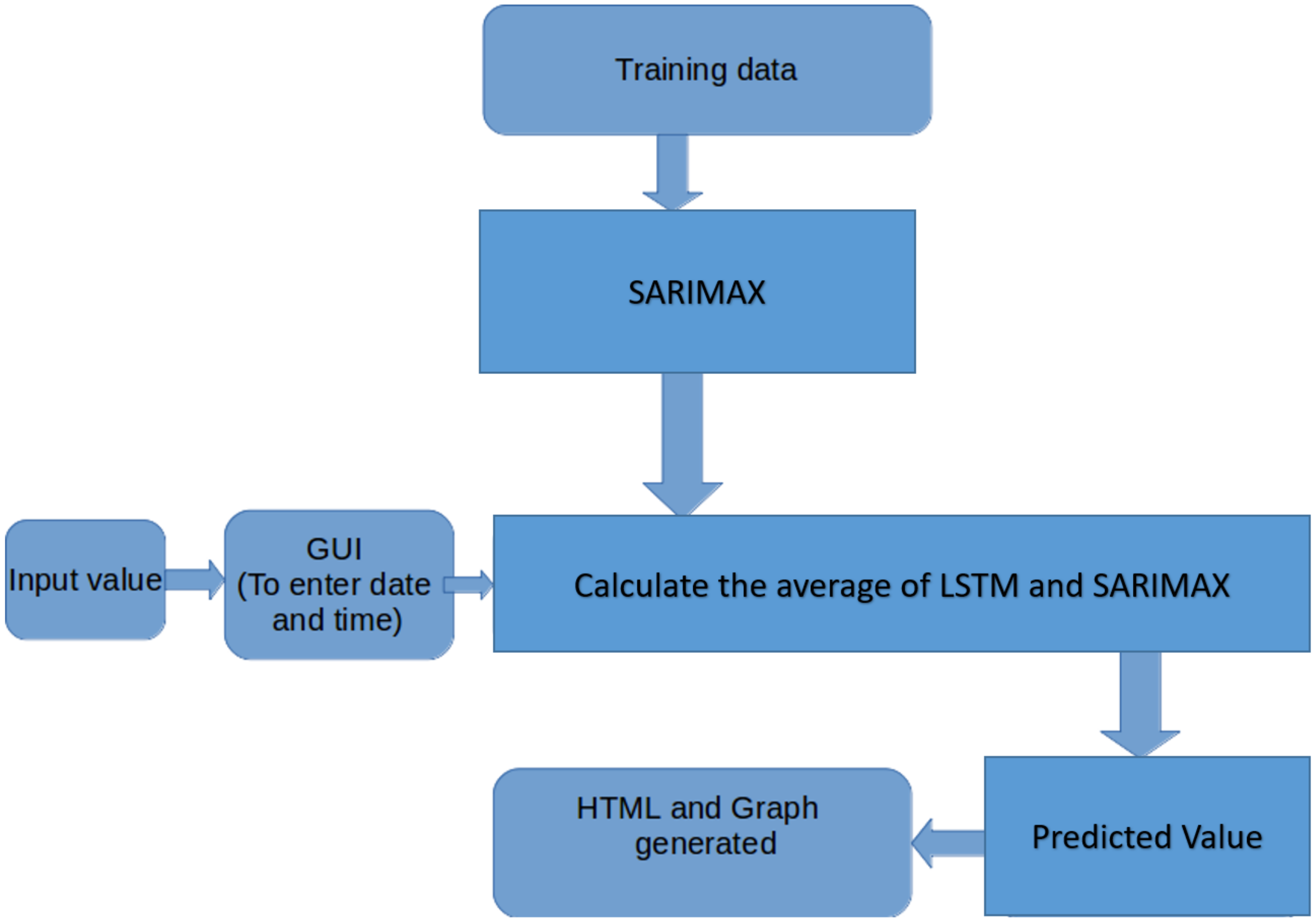
$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(p_i - a_i)^2}{n}}$$

$$MAE = \overset{\text{Divide by the total number of data points}}{\boxed{\frac{1}{n}}} \underset{\text{Sum of}}{\sum} \underbrace{\left|\overset{\text{Actual output value}}{\boxed{y}} - \overset{\text{Predicted output value}}{\boxed{\widehat{y}}}\right|}_{\text{The absolute value of the residual}}$$

$$MAPE = \frac{\overset{\text{Multiplying by 100\% converts to percentage}}{\boxed{100\%}}}{n} \sum \left|\frac{\overbrace{y - \widehat{y}}^{\text{The residual}}}{\underset{\text{Each residual is scaled against the actual value}}{\boxed{y}}}\right|$$

Performance Metrics	Comparison of Various Algorithms		
	LSTM	SARIMAX	HYBRID
RMSE	0.432	0.773	0.408
MAE	0.366	0.607	0.327
MAPE	3.277	5.523	2.959

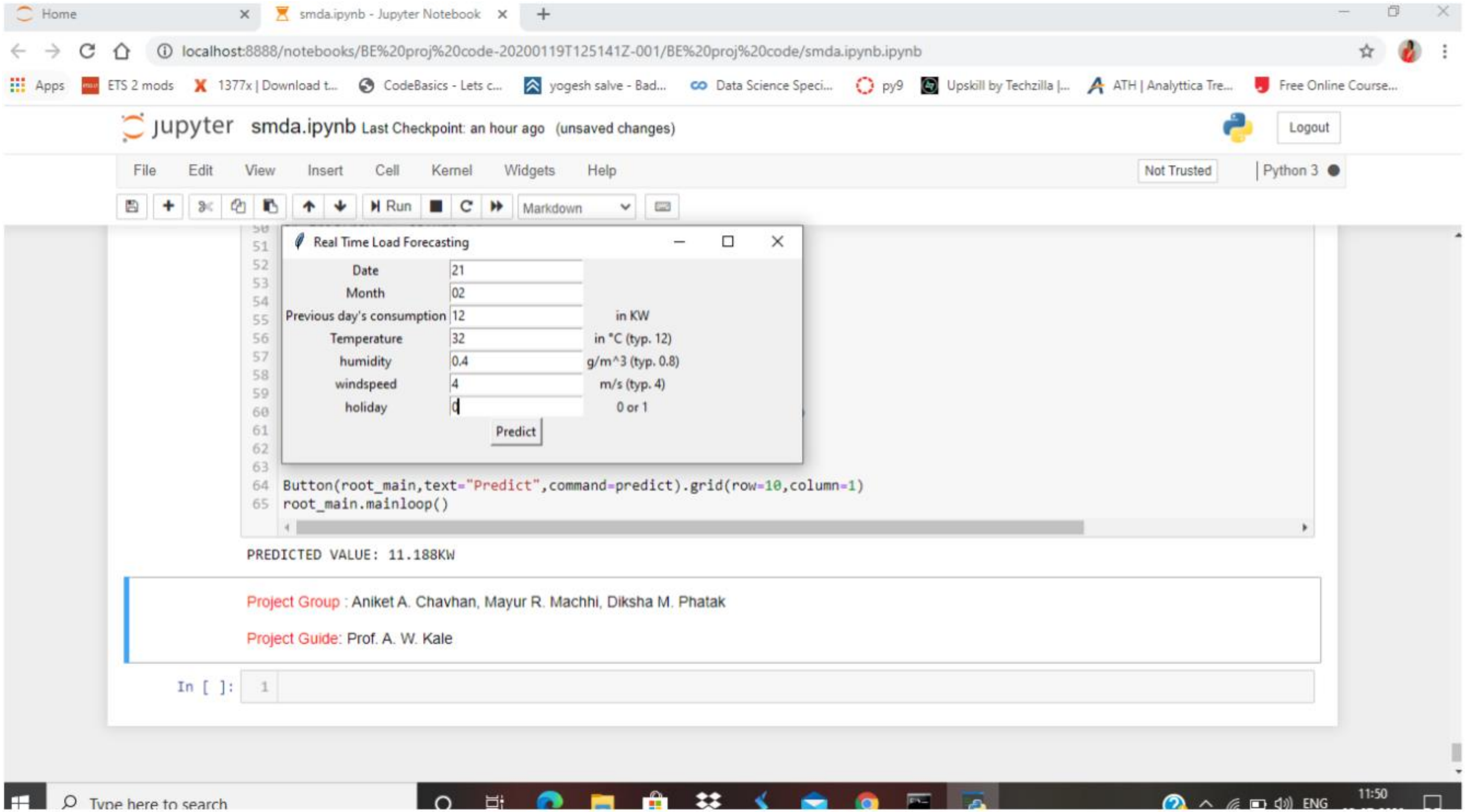
Flow chart


Training data
SARIMAX
Input value
GUI
(To enter date
and time)
Calculate the average of LSTM and SARIMAX
HTML and Graph
generated
Predicted Value

Resources Required for Project

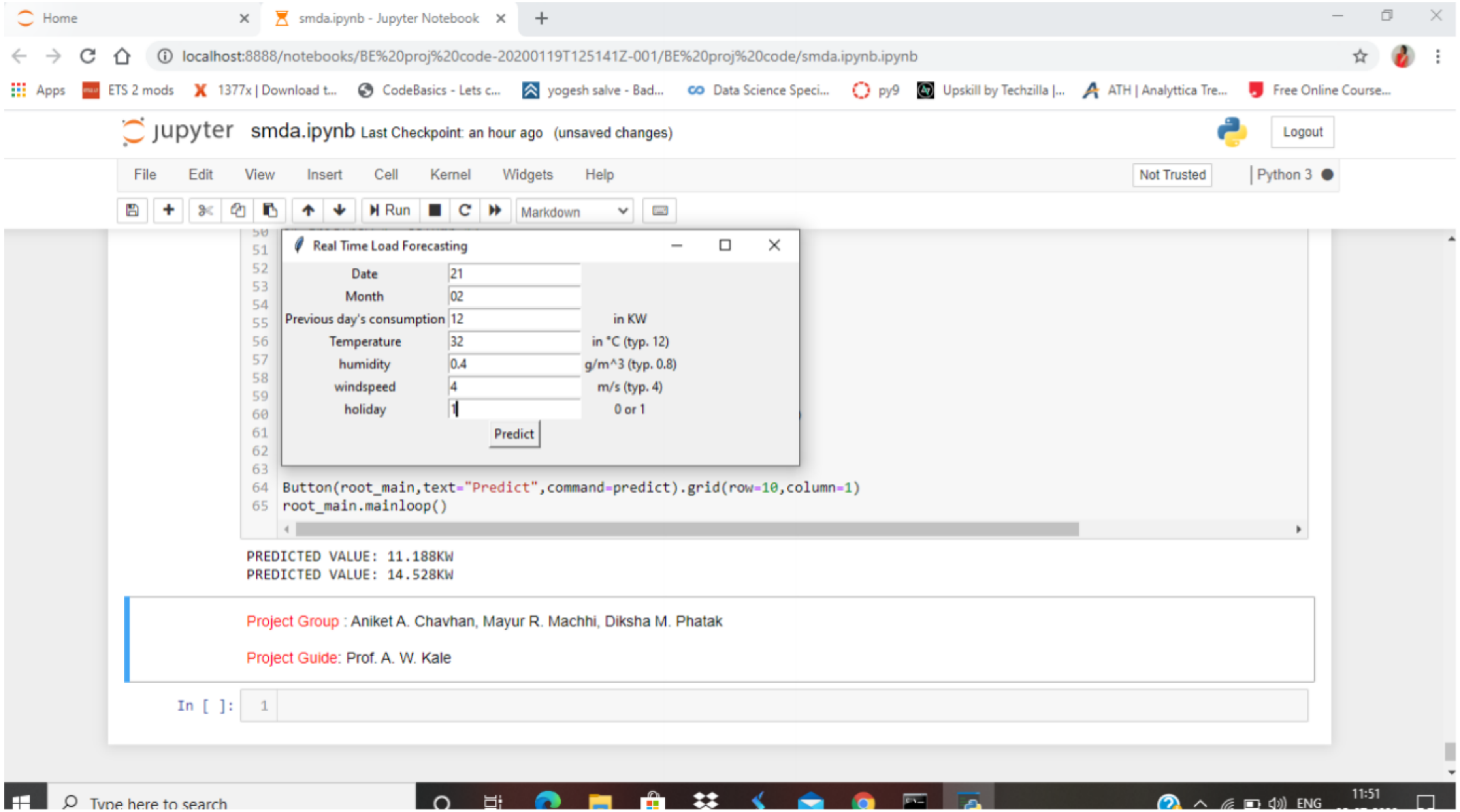
- Python programming language
- Linux based OS/ Windows 10
- Pandas
- Smart meters data sets

Results



Prediction No. 1

Results



Prediction No. 2

Comparison

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

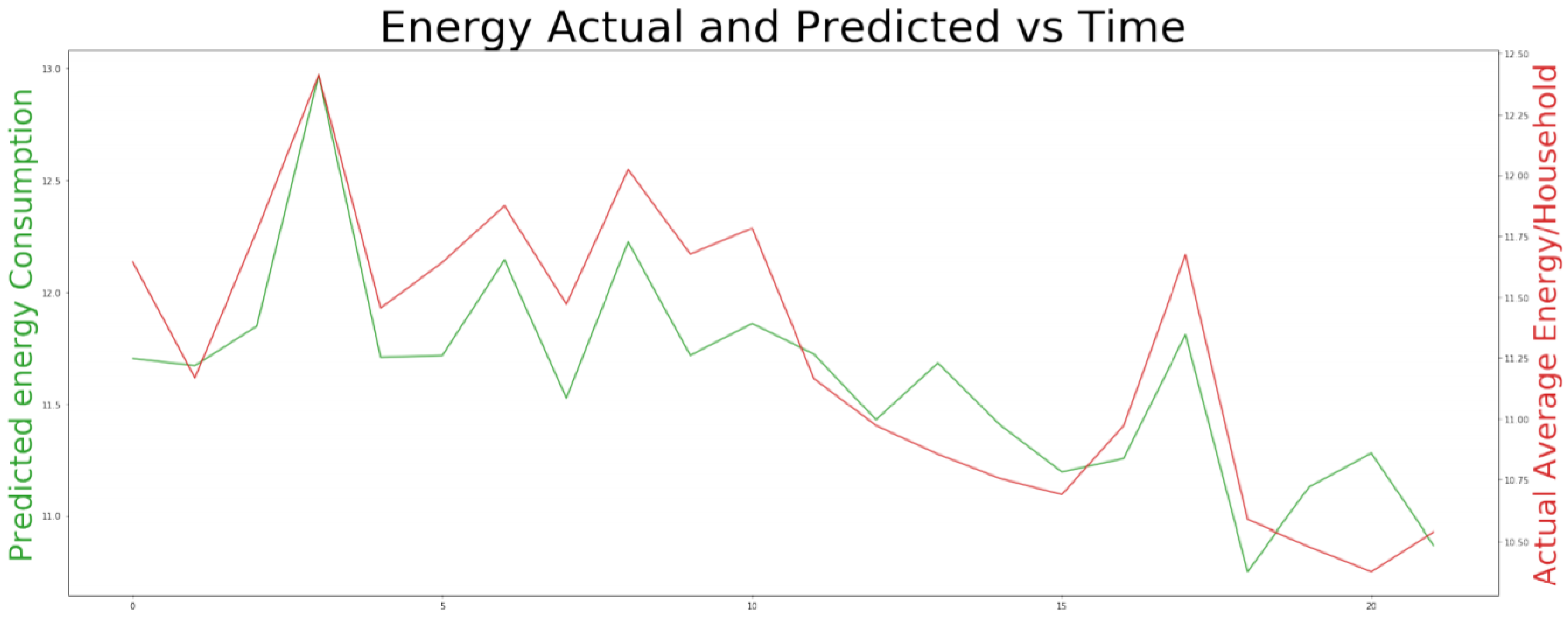


FIGURE 6.5: Energy Consumption Actual and Predicted

Objectives achieved

- We studied databases and discovered patterns.
- Our study includes, Combining all blocks into a single data frame- keeping on relevant columns.
- Used a day-level energy consumption data per household to normalize data for inconsistent household count.
- Explored relationships between weather conditions and energy consumptions and created clusters for the weather data.
- With the help of data models we found out the suitable algorithm for our dataset for data processing.

Conclusion

- 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 accuracy that can be achieved as seen in the results.
- Thus combining traditional time series algorithm with learning model using machine learning model proves to provide better results.

Thank You