

Estimation of Power Consumption using Machine Learning

Submitted in partial fulfilment of the requirements for the award of master of Engineering degree in computer science and engineering

by

S.HEMANTHRE

40310008



DEPARTMENT OF MECHANICAL ENGINEERING

SCHOOL OF MECHANICAL ENGINEERING

SATHYABAMA

INSTITUTE OF SCIENCE AND

TECHNOLOGY (DEEMED TO BE UNIVERSITY)

Accredited with Grade "A" by NAAC

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600

119 MAY - 2022

BONAFIDECERTIFICATE

This is to certify that this Project Report is the bonafide work of **S.**

HEMANTH Reg.no.40310008 who carried out the project entitled “**ESTIMATION OF POWER CONSUMPTION FOR HOUSEHOLD APPLIANCES USING MACHINE LEARNING**” under my supervision from NOV-2020 to MAY-2022.

Internal Guide

Dr.N.SRINIVASAN.ME.,PhD

Head of the Department

Dr.S.Vigneshwari,M.E.,Ph.D

Submitted for Vivavoce Examination held on _____

Internal Examiner

External Examiner

DECLARATION

IS.HEMANTH hereby declare that the Project Report entitled “**ESTIMATION OF POWER CONSUMPTION FOR HOUSEHOLD APPLIANCES USING MACHINE LEARNING**” done by me under the guidance of **Dr.N.SRINIVASAN.ME.,PhD** ,at Sathyabama Institute of Science and Technology, is submitted in partial fulfillment of the requirements for the award of **Master of Engineering in Computer Science and Engineering**.

DATE:

PLACE:

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to Board of Management of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T.Sasikala M.E.,Ph.D,Dean**, School of computing and **Dr.S.Vigneshwari M.E.,Ph.D.and Dr.L.Lakshmanan M.E.,Ph.D.**, Head of the Department, of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr.N.SRINIVASAN.ME.,PhD** for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

ABSTRACT:

Today's youth face a lot of stress in their life as well as their work. Various NLP techniques are employed in order to handle the stress of young people. Artificial Intelligence systems are developed (AI) are developed in order to mimic human brain. A smart stress management system has been developed in order to mimic the role played by the counsellors. In the older systems of stress management, stress levels are detected and the appropriate solutions are provided. They provide user-friendly solutions but they are more time consuming and involve a lot of human power. So, a more intelligent stress management system has been developed which provides better results at a less time and with less human effort. It is a more robust system when compared to the previous one and it provides smarter solutions at a less cost.

TABLE OF CONTENT

CHAPTER NO	TITLE	PAGE NO
I.	INTRODUCTION	1
	1.1 MACHINE LEARNING	1
	1.2 Machine Learning vs. Traditional Programming	2
	1.3 APPLICATION OF MACHINE LEARNING	3
II.	LITERATURE SURVEY	5
	III. AIM	6
	3.1 OBJECTIVE	6
	3.2 AIM	6
	3.3 EXISTING SYSTEM	6
	3.4 PROPOSED SYSTEM	6
IV.	EXPERIMENTAL MATERIALS AND METHODS	8
	4.1 MODULES	8
	4.1.1 DATA COLLECTION	8
	4.1.2 DATA PRE-PROCESSING	9
	4.1.3 DATA SPLITTING	10
	4.1.4 EVALUATION MODEL	10
	4.2 PROPOSED APPROACH STEPS	10
	4.3 DATA FLOW DIAGRAM	11
	4.4 UML DIAGRAMS	13
	4.5 SEQUENCE DIAGRAM	13
	4.6 ACTIVITY DIAGRAM	14
	4.7 USE CASE DIAGRAM	15
	4.8 CLASS DIAGRAM	16
V.	RESULTS AND DISCUSSION, PERFORMANCE ANALYSIS	18
	5.1 VIRTUALIZATION	18
	5.2 LSTM RESULT	20
	5.3 FUTURE PREDICTION AND RESULTS	21
	5.4 SOURCE CODE	22
	5.5 COMPARATIVE RESULT GRAPH	45
VI.	CONCLUSION	46
VII.	REFERENCE	

LIST OF FIGURES

Fig1.1: MachineLearningBlockDiagram	2
Fig1.2: LearningPhase	3
Fig3.2: blockdiagram	14
Fig4.1: SequenceDiagram	14
Fig4.2 ActivityDiagram	15
Fig4.3 usecasediagram	17

CHAPTER

1INTRODUCTION

N

The improvement of a checking framework, which can screen every family unit electric machine, is critical to obtain essential data for building the strategy of the vitality preservation, the figure of electric vitality request to design constructing new force plants, furthermore, making better new client administrations. It is normal that the observing framework for family electric apparatuses is economical and non-meddlesome, on the grounds that the ordinary checking strategy which is set up estimating types of gear for every machine is costly and powers the client's burden. The nonintrusive observing methods the estimation of intensity load is done in the outside of a house. We have built up a non-meddlesome observing framework, which can assess the ON/OFF states of family unit electric apparatuses incorporate inverter. For the framework clients, the framework can accomplish an adequate exactness utilizing enormous edge classifiers. Consequently, the framework can gauge the force utilization of traditional electric machines, which show a consistent power utilization, in the event that we know the force utilization of every family unit electric machine. Be that as it may, an electric apparatus with an inverter circuit (inverter type machine, for example, a forced air system) doesn't show a consistent force utilization. In this way, we need to build up the non-meddlesome observing framework, which can gauge the force utilization of inverter type machines.

1.1 MACHINE LEARNING

Machine Learning is a system that can learn from example through self-improvement and without being explicitly coded by programmer. The breakthrough comes with the idea that a machine can singularly learn from the data (i.e., example) to produce accurate results.

Machine learning combines data with statistical tools to predict an output. This output is then used by corporate to make actionable insights. Machine learning is closely related to data mining and Bayesian predictive modelling. The machine receives data as input, uses an algorithm to formulate answers.

A typical machine learning tasks are to provide a recommendation. For those who have a Netflix account, all recommendations of movies or series are based on the user's historical data. Tech companies are using unsupervised learning to improve the user experience with personalizing recommendation.

Machine learning is also used for a variety of tasks like fraud detection, predictive maintenance, portfolio optimization, automate tasks and soon.

1.2 Machine Learning vs. Traditional Programming

Traditional programming differs significantly from machine learning. In traditional programming, a programmer codes all the rules in consultation with an expert in the industry for which software is being developed. Each rule is based on a logical foundation; the machine will execute an output following the logical statement. When the system grows complex, more rules need to be written. It can quickly become unsustainable to maintain.

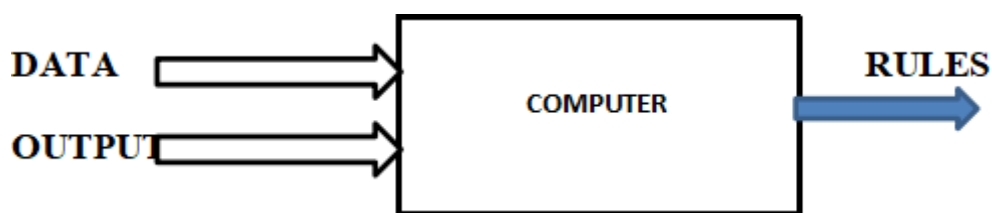


Figure 1.1: -Machine Learning Block Diagram

Machine learning is the brain where all the learning takes place. The way the machine learns is similar to the human being. Humans learn from experience. The more we know, the more easily we can predict. By analogy, when we face an unknown situation, the likelihood of success is lower than the known situation. Machines are trained the same. To make an accurate prediction, the machine sees an example. When we give the machine a similar example, it can figure out the outcome. However, like a human, if it sees a previously unseen example, the machine has difficulties to predict.

The core objective of machine learning is the learning and inference. First of all, the machine learns through the discovery of patterns. This discovery is made thanks to the data. One crucial part of the data scientist is to choose carefully which data to provide to the machine. The list of attributes used to solve a problem is called a feature vector. You can think of a feature vector as a subset of data that is used to tackle a problem.

The machine uses some fancy algorithms to simplify the reality and transform this discovery into a model. Therefore, the learning stage is used to describe the data and summarize it into a model.

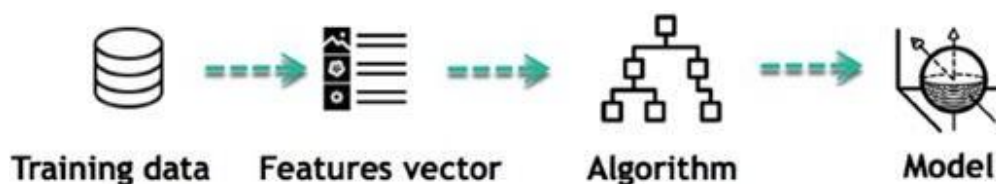


Figure 1.2:- Learning Phase

For instance, the machine is trying to understand the relationship between the wage of an individual and the likelihood of going to a fancy restaurant. It turns out the machine finds a positive relationship between wage and going to a high-end restaurant. This is the model.

1.3 APPLICATION OF MACHINE

LEARNING Augmentation

- Machine learning, which assists humans with their day-to-day tasks, personally or commercially without having complete control of the output. Such machine learning is used in different ways such as Virtual Assistant, Data analysis, software solutions. The primary user is to reduce errors due to human bias.

Automation

- Machine learning, which works entirely autonomously in any field without the need for any human intervention. For example, robots performing the essential processes in manufacturing plants.

Finance Industry

- Machine learning is growing in popularity in the finance industry. Banks are mainly using ML to find patterns in the data but also to prevent fraud.

Government organization

- The government makes use of ML to manage public safety and utilities. Take the example of China with the massive face recognition. The government uses Artificial intelligence to prevent jaywalking.

Healthcare industry

- Healthcare was one of the first industries to use machine learning with image detection.

Marketing

- Broad use of AI is done in marketing thanks to abundant access to data. Before the age of mass data, researchers developed advanced mathematical tools like Bayesian analysis to estimate the value of a customer. With the boom of data, the marketing department relies on AI to optimize the customer relationship and marketing campaign.

CHAPTER

2 LITERATURE SURVEY

Y

Santos, R. et.al., discussed about Software-Defined Networking (SDN) the controller is the only entity with a comprehensive perspective of the network, and it serves as the brain, managing traffic based on its global knowledge of the network.

As a result, an attacker attempts to steer malicious traffic to the controller, potentially paralyzing the entire network. Deep Learning techniques are utilised in this study to safeguard the controller by implementing high-security measures, which is critical for the network's ongoing availability and connection. To identify and prevent intrusion threats, Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU) are proposed. We test our model on a freshly available dataset (In SDN dataset).

S. Wilson Prakash. et.al., discussed about the Open Flow switches in the SDN are used to deliver the Dynamic Server Load Balancing algorithm (DServ-LB). It is a load balancing method based on packets. Requests are sent to the webserver with the most server resources based on the server resource. When compared to the present load, the suggested DServ-LB enhances overall network performance and efficiently leverages server resources algorithms for balancing. A Flow-Based Dynamic Load Model was used in the second investigation. In cloud networking using SDN, the balancing (FB-LB) method is introduced. The suggested method manages flows without knowing anything about them beforehand size. Small fluxes are handled by ECMP, whereas big volumes are handled by FB-LB algorithm.

Dong, S., et.al., discussed about for decades, the Distributed Denial of Service (DDoS) assault has severely harmed network availability, and there is still no viable protection mechanism against it. However, the growing Software Defined Networking (SDN) technology offers a fresh method to examine DDoS defences. We offer two strategies for detecting DDoS attacks in SDN in this research. To identify a DDoS assault, one way uses the degree of the DDoS attack. To detect the DDoS

assault, the other technique employs an updated K-Nearest Neighbors (KNN) algorithm based on Machine Learning (ML).

CHAPTER 3

AIM AND SCOPE OF THE PRESENT INVESTIGATION

3.1 OBJECTIVE

The main objective of this study is to come to evaluate useful regulations to the machine learning community that permits them the fundamental realization to use and build energy estimation methods for machine learning algorithms.

3.2 AIM

To estimate energy consumption in general and machine learning applications in particular. Our goal is to provide useful guidelines to the machine learning community giving them the fundamental knowledge to use and build specific energy estimation methods for machine learning algorithms.

3.3 EXISTING METHOD

We compare with several regression algorithms to estimate the power consumption of an air conditioner as follows: multi-layered perceptron's (MLP), radial basis function networks (RBFN) and support vector regression (SVR). MLP is a conventional nonlinear regression method. So it is a good basis to compare with the other methods. In RBFN, each basis function, which is able to have different parameters, acts like a hidden unit of MLP. SVR has an advantage in high dimensional space using kernel functions. Therefore, we expect that RBFN and SVR have better results than MLP.

Disadvantages

- We only did with non-intrusive monitoring system but not for monitoring system.

3.4 PROPOSED METHOD

In this we try to improve the accuracy while compared to the previous one and here we try to find the sufficient accuracy for monitoring System which can estimate the power consumption of electric appliances, to actual fields.

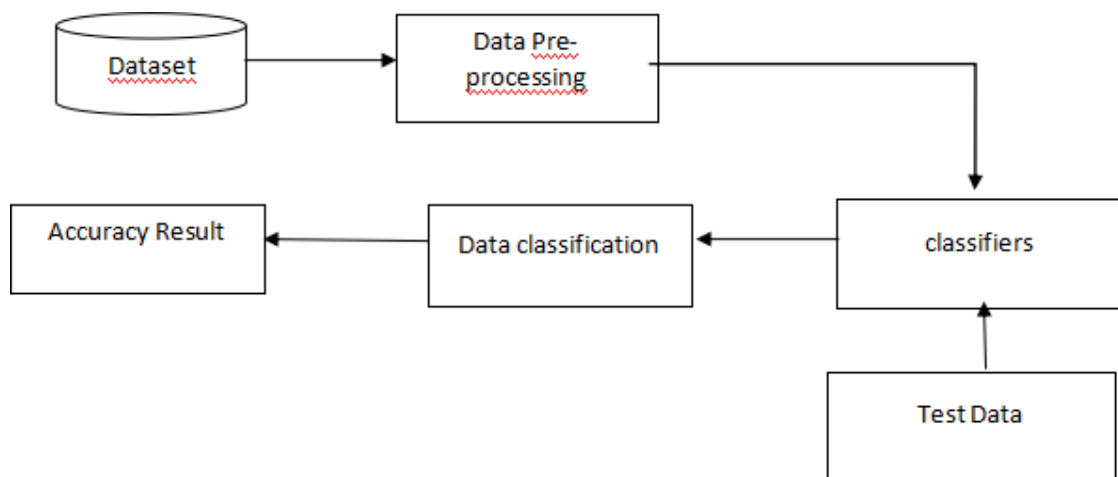


Figure 3.1:-BLOCK DIAGRAM

Advantages

- In this we find the accuracy for monitoring system.

CHAPTER 4

EXPERIMENTAL MATERIALS AND METHODS; ALGORITHM USED

4.1 MODULES

1. DATA COLLECTION
2. DATA PRE-PROCESSING
3. DATA SPLITTING
4. EVALUATION MODEL

4.1.1 DATA COLLECTION

Data used in this paper is a set of meter data having several appliances.

Dataset is a multivariate time series dataset that describes the electric consumption for a household.

Global_active_power, global_reactive_power, voltage, global_intensity, ST, ET etc. are the attributes used.

The dataset is split randomly into two parts: training set, test set.

Mean Absolute Errors (MAE), Root Mean Squared Errors (RMSE) are calculated for the datasets.

The algorithm which gives the lowest error is considered to be accurate.

ML problems start with data preferably, lots of data (examples or observations) for which you already know the target answer. Data for which you already know the target answer is called *labelled data*.

Table 1: Description of dataset

*	Attributes	UNITS	Remarks
1	Date	dd/mm/yy yy	
2	Time	hh/mm/ss	The time is given in hours, minutes, and seconds (hh/mm/ss), where the hour values range from 0 to 23, and minute values range from 1 to 60
3	Global Active Power (GAP)	Kilowatts	Total average active power of the household for each minute.
4	Global Reactive Power (GRP)	Kilowatts	Total average reactive power of the household for each minute
5	Voltage (V)	Volts	Total average voltage for each minute.
6	Global Intensity (GI)	Amperes	Total average current intensity for each minute
7	Sub-metering 1 (S1)	Watt-hours	Active energy related to kitchen, including dishwasher, oven and microwave.
8	Sub-metering 2 (S2)	Watt-hours	Active energy related to laundry room, including washing machine, tumble-drier and refrigerator
9	Sub-metering 3 (S3)	Watt-hours	Active energy related to electric water-heater and air-conditioner.

4.1.2 DATAPRE-PROCESSING

Organize your selected data by formatting, cleaning and sampling from

it. Three common data pre-processing steps are:

- **Formatting:** The data you have selected may not be in a format that is suitable for you to work with. The data may be in a relational database and you would like it in a flat file, or the data may be in a proprietary file format and you would like it in a relational database or a text file.
- **Cleaning:** Cleaning data is the removal or fixing of missing data. There may be data instances that are incomplete and do not carry the data you believe you need to add to solve the problem. These instances may need to be removed. Additionally, there may be sensitive information in some of the attributes and these attributes may need to be anonymized or removed from the data entirely.
- **Sampling:** There may be far more selected data available than you need to work with. More data can result in much longer running times for algorithms and larger computational and memory requirements. You can take a smaller representative sample of the selected data that may be much faster for exploring and prototyping solutions before considering the whole dataset.

4.1.3 DATA SPLITTING

Then we have pre-processed the data we have remove the missing values and we have split the data into train and test we have given the ratio 80:30 We have given 80% to train the data and 20% to test the data.

4.1.4 EVALUATION MODEL

Model Evaluation is an integral part of the model development process. It helps to find the best model that represents our data and how well the chosen model will work in the future. Evaluating model performance with the data used for training is not acceptable in data science because it can easily generate over optimistic and over fitted models. There are two methods of evaluating models in data science, Hold-Out and Cross-Validation. To avoid over fitting, both methods use a test set (not seen by the model) to evaluate model performance.

Performance of each classification model is estimated based on its averaged. The result will be in the visualized form. Representation of classified data in the form of graphs.

Accuracy is defined as the percentage of correct predictions for the test data. It can be calculated easily by dividing the number of correct predictions by the number of total predictions.

4.2 Proposed Approach Steps

The prediction method will employ 3 machine learning algorithms which are LSTM, SVM and LR. Feature attributes for this prediction will use electrical power data consisting of power factor, voltage and current, in which the demand would be the targeted output.

STEPS for Proposed Approach-

Step 1 Normality testing of

dataset

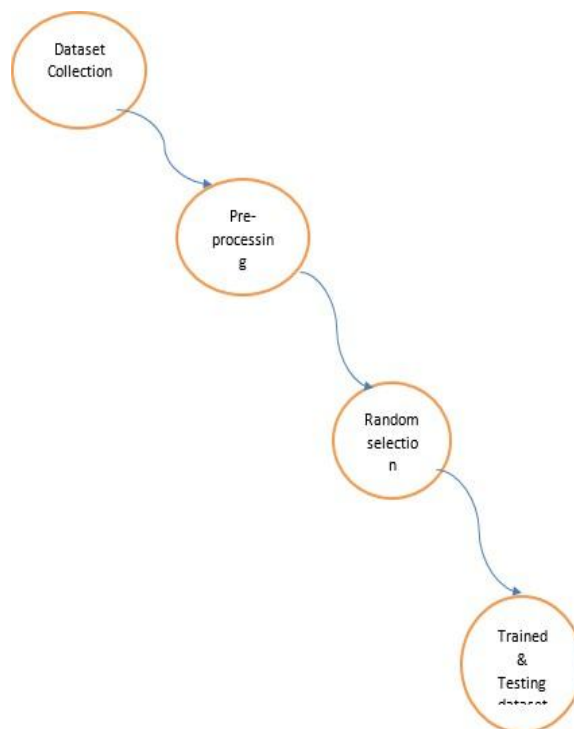
Step 2 Data pre-processing

Step 3 Model development (training)

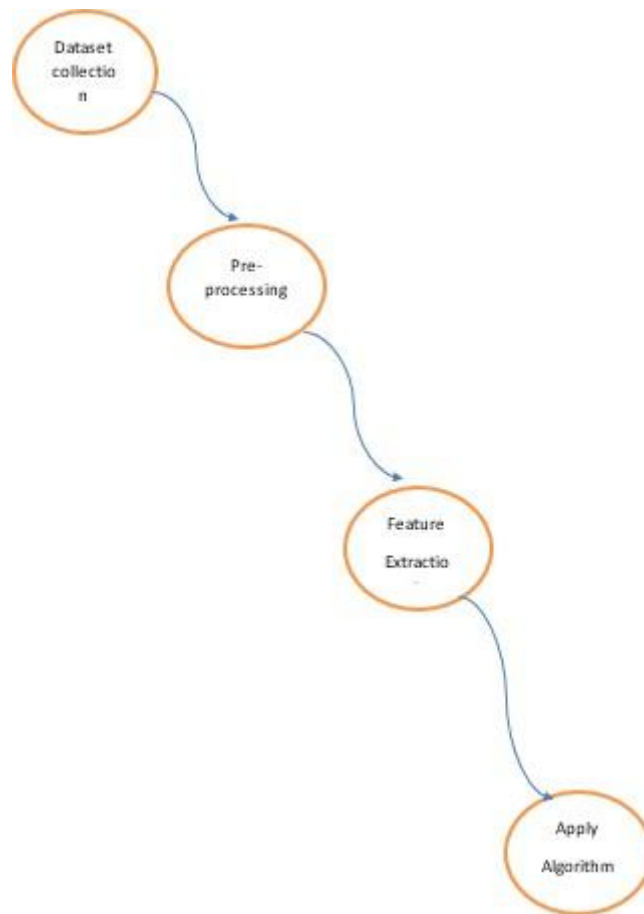
Step4Modevaluation(test)

4.3 DATAFLOWDIAGRAM

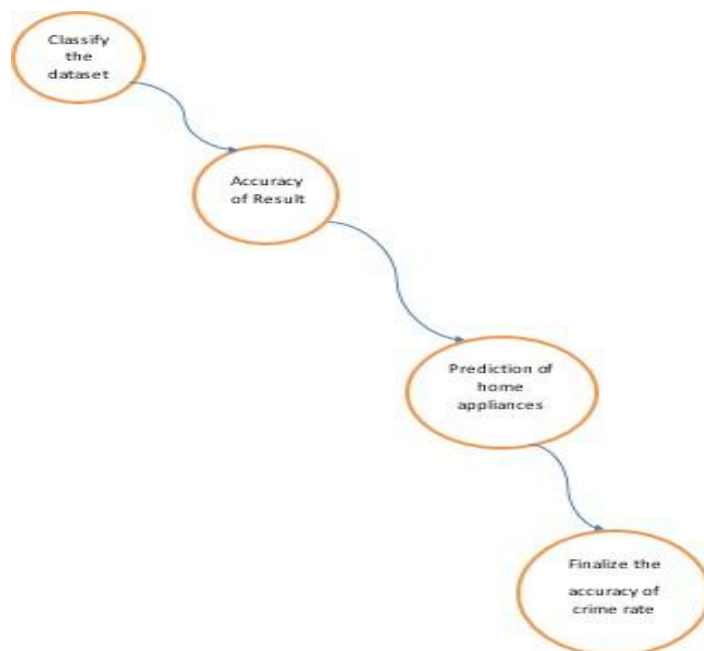
LEVEL 0



LEVEL 1



LEVEL 2



4.4 UML DIAGRAMS

The Unified Modeling Language (UML) is used to specify, visualize, modify, construct and document the artifacts of an object-oriented software intensive system under development. UML offers a standard way to visualize a system's architectural blueprints, including elements such as:

- actors
- business processes
- (logical) components
- activities
- programming language statements
- database schemas, and
- Reusable software components.

UML combines best techniques from data modelling (entity relationship diagrams), business modelling (work flows), object modelling, and component modelling. It can be used with all processes, throughout the software development life cycle, and across different implementation technologies. UML has synthesized the notations of the Booch method, the Object-modelling technique (OMT) and Object-oriented software engineering (OOSE) by fusing them into a single, common and widely usable modelling language. UML aims to be a standard modelling language which can model concurrent and distributed systems.

4.5 Sequence Diagram

Sequence Diagrams represent the objects participating in the interaction horizontally and time vertically. A Use Case is a kind of behavioural classifier that represents

a declaration of an offered behaviour. Each use case specifies some behaviour, possibly including variants that the subject can perform in collaboration with one or more actors. Use cases define the offered behaviour of the subject without

reference to its internal structure. These behaviours, involving interactions between the actor

and the subject, may result in changes to the state of the subject and communications with its environment. A use case can include possible variations of its basic behaviour, including exceptional behaviour and error handling.

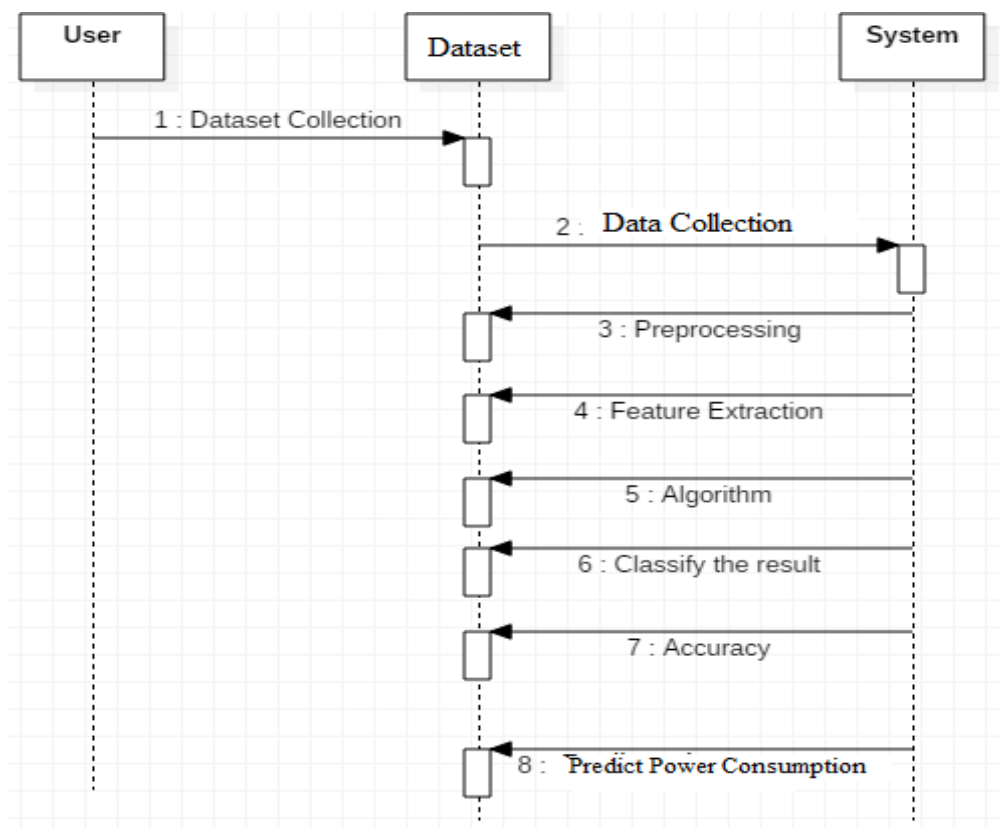


Figure 4.1 SEQUENCE DIAGRAM

4.6 Activity Diagrams

Activity diagrams are graphical representations of Workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.

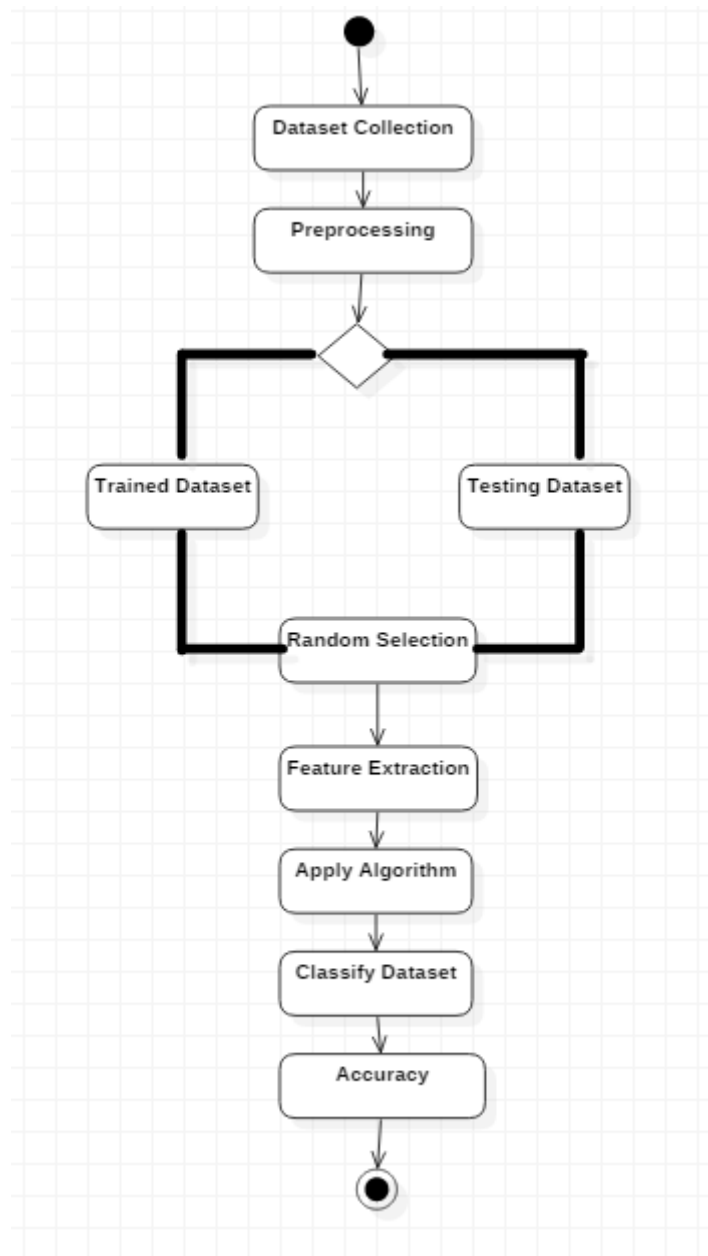


Figure4.2ActivityDiagram

4.6 Usecasediagram

UML is a standard language for specifying, visualizing, constructing, and documenting the artifacts of software systems.

UML was created by Object Management Group (OMG) and UML 1.0 specification draft was proposed to the OMG in January 1997.

OMG is continuously putting effort to make a truly industry standard. UML stands for **U**nified **M**odeling **L**anguage.

UML is a pictorial language used to make software blueprints

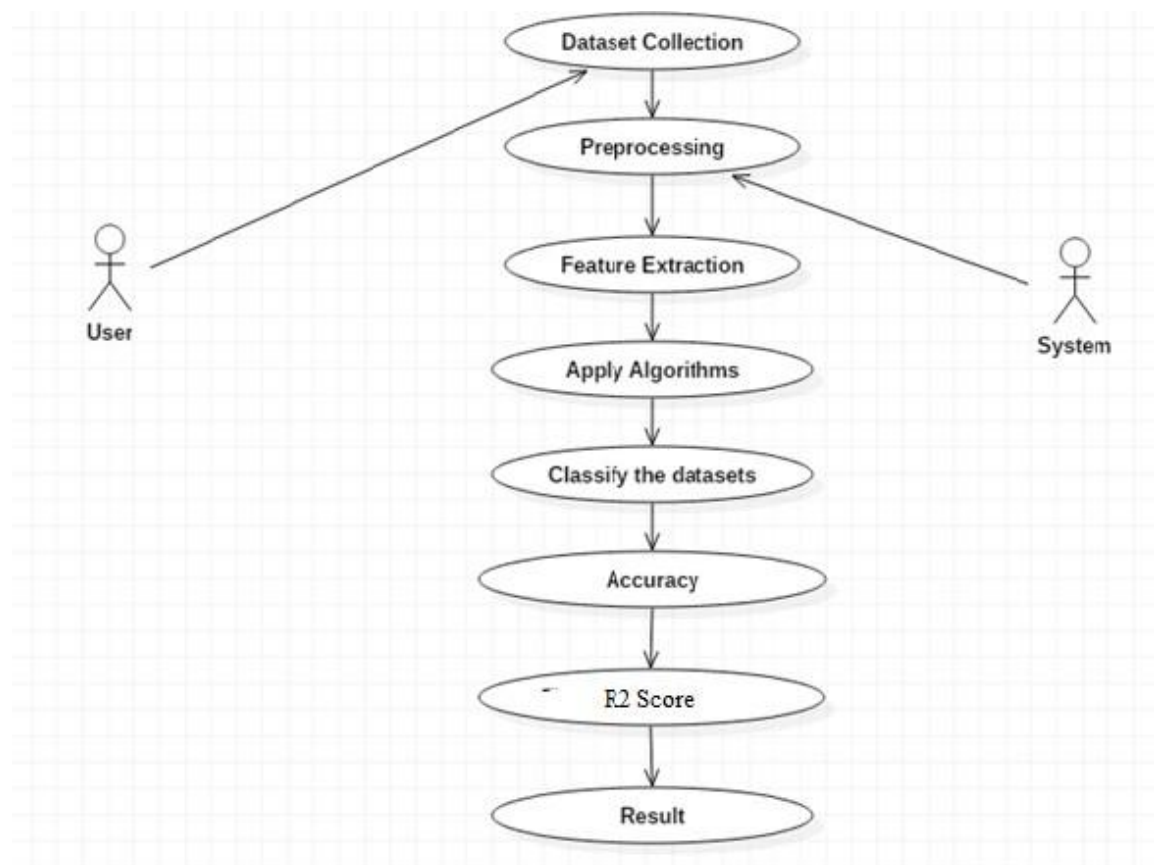


Figure 4.3 UML Use Case Diagram

4.7 Class Diagram

The class diagram is the main building block of object-oriented modeling. It is used for general conceptual modeling of the system of the application, and for detailed modeling translating the models into programming code. Class diagrams can also be used for data modeling. The classes in a class diagram represent both the main elements, interactions in the application, and the classes to be programmed.

In the diagram, classes are represented with boxes that contain three compartments:

The top compartment contains the name of the class. It is printed in bold and centered, and the first letter is capitalized.

The middle compartment contains the attributes of the class. They are left-aligned and the first letter is lowercase.

The bottom compartment contains the operations the class can execute. They are also left-aligned and the first letter is lowercase.



Figure 4.4 USE CASE DIAGRAM

4.8 TESTING

Software testing is an investigation conducted to provide stakeholders with information about the quality of the product or service under test. Software Testing also provides an objective, independent view of the software to allow the business to appreciate and understand the risks at implementation of the software. Test techniques include, but are not limited to, the process of executing a program or application with the intent of finding software bugs.

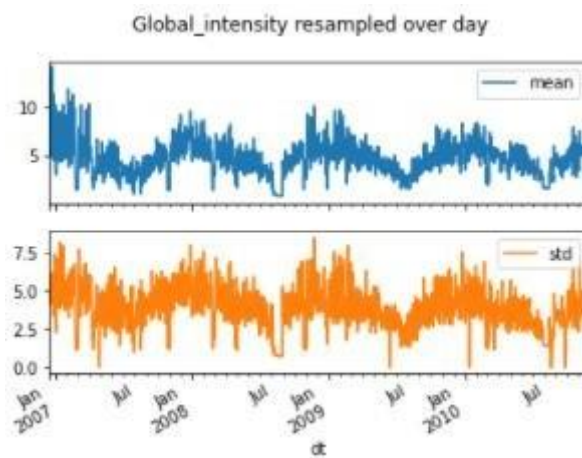
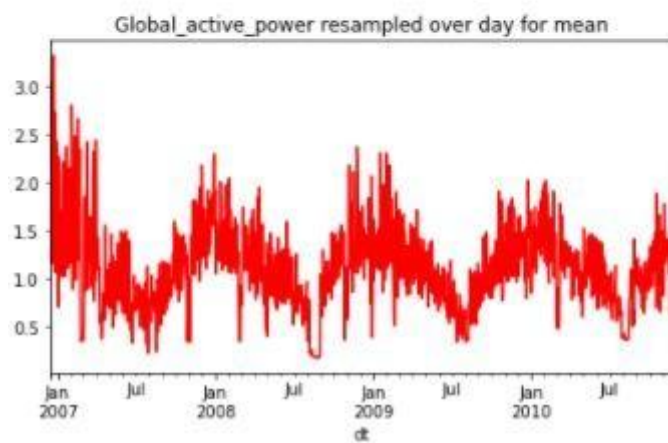
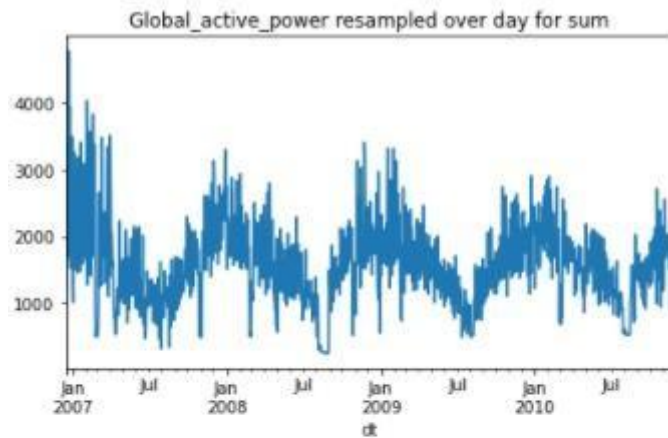
Software Testing can also be stated as the process of validating and verifying that a software program/application/product:

- Meets the business and technical requirements that guided its design and Development.
- Works as expected and can be implemented with the same characteristics.

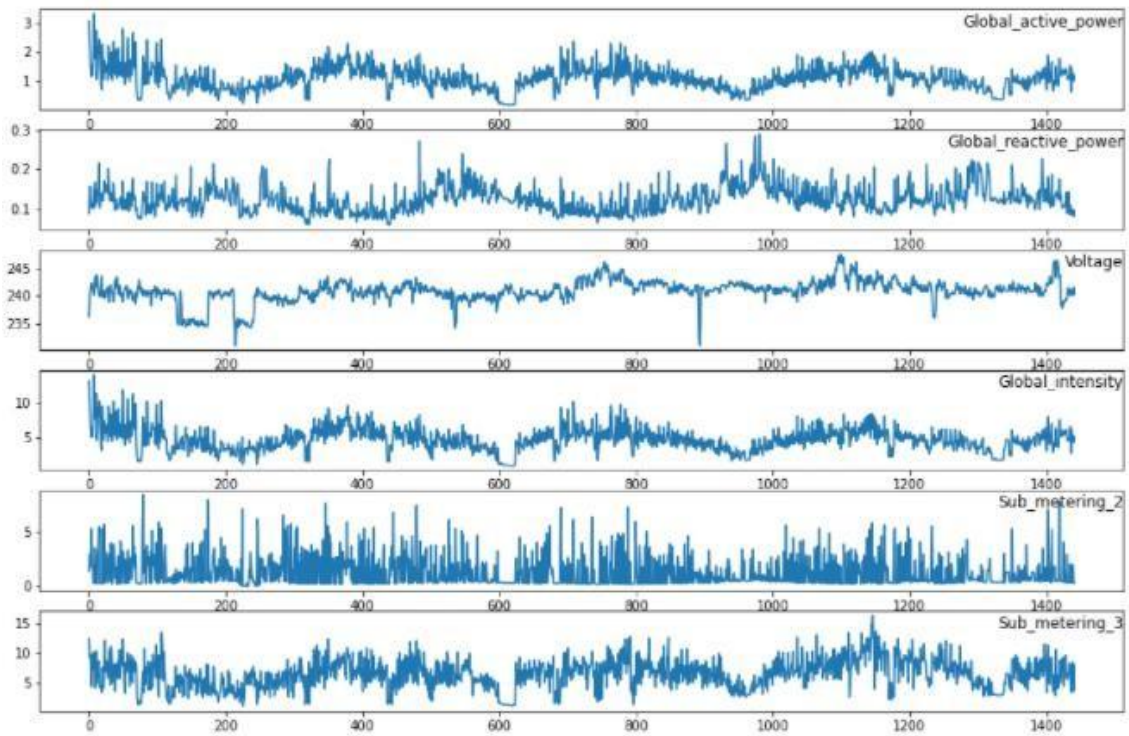
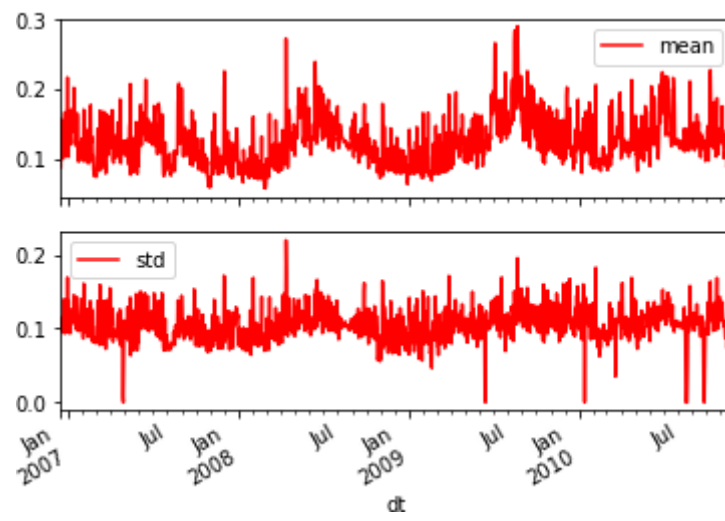
CHAPTER

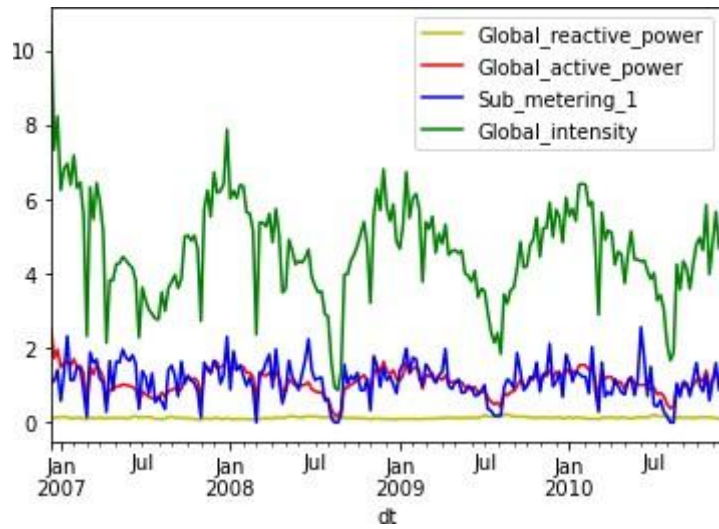
5 RESULTS AND DISCUSSION

5.1 VISUALIZATION

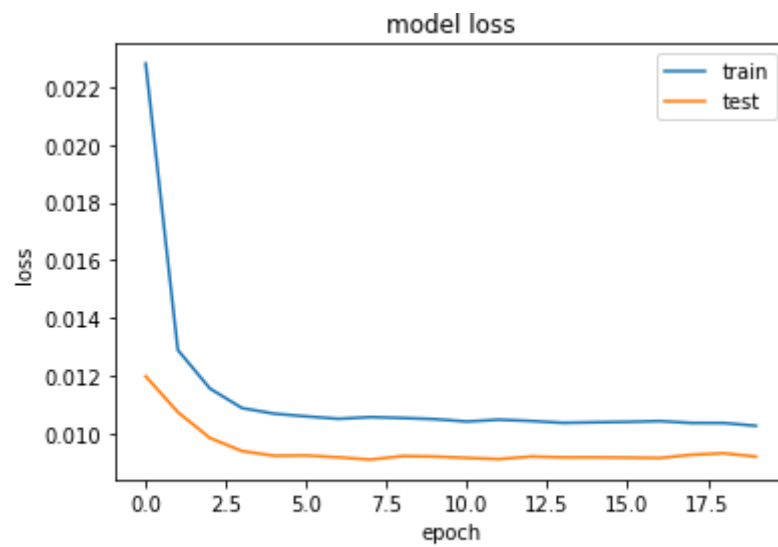


Global_reactive_power resampled over day



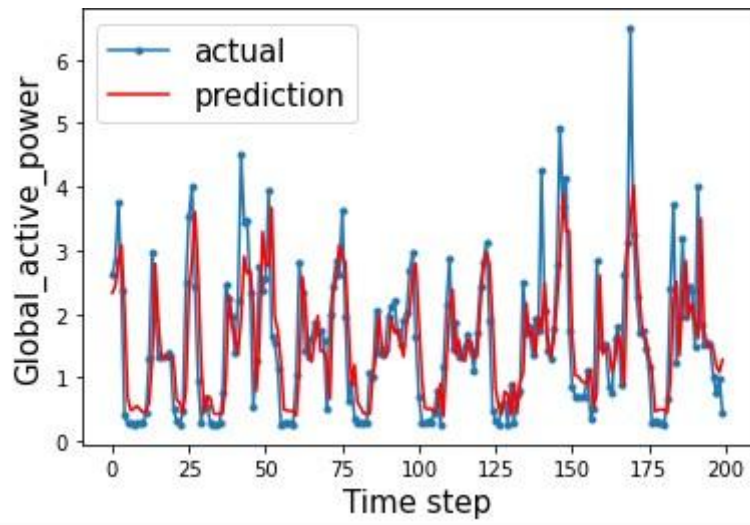


5.2 LSTMRESULT



Test RMSE: 0.617

5.3 FuturePredictionResult



5.4 SOURCECODE:

```
{
  "cells": [
    {
      "cell_type":
      "code", "execution_c
ount": 1, "metadata": {
        "_cell_guid": "7b27c994-a143-422a-a8cc-
51e6a1a688d3", "_uuid": "4e690c4a961b9475a93164720c7aeaa
b99fe7061"
      },
      "outputs":
      [], "source":
      ["import sys\n",
        "import numpy as np\n",
        "#linear algebra\n",
        "import pandas as pd\n",
        "#data processing, CSV file I/O (e.g. pd.read_csv),
        "data manipulation as in SQL\n",
        "import matplotlib.pyplot as plt\n",
        "# this is used for the plot the graph
        "\n",
        "import seaborn as sns\n",
        "# used for plot interactive graph. \n",
        "from sklearn.model_selection import train_test_split\n",
        "# to split the data into two parts
        "\n",
        "from sklearn.preprocessing import MinMaxScaler\n",
        "from sklearn import metrics\n",
        "# for the check the error and accuracy of the model\n",
        "from sklearn.metrics import
        mean_squared_error, r2_score\n",
        "import keras\n",
        "from keras.layers import
        Dense\n",
        "from keras.models import Sequential\n",
        "from tensorflow.keras.utils import
        to_categorical\n",
        "# from keras.utils import to_categori
        "cal\n",
        "from keras.optimizers import
        gradient_descent_v2\n",
        "# from keras.optimizers import
```

```
SGD\n",  
"from keras.callbacks import  
EarlyStopping\n", "from keras.utils  
import np_utils\n",
```

```

"importitertools\n",
"fromkeras.layersimportLSTM\n",
"fromkeras.layers.convolutionalimportConv1D\n",
"fromkeras.layers.convolutional import
MaxPooling1D\n", "fromkeras.layersimport Dropout"
]
},
{
"cell_type":
"code", "execution_c
ount": 2, "metadata": {
"_cell_guid": "3b7223da-4278-4fa5-add9-
8e238facc74a", "_uuid": "7ef91d05da654a56aaccb166dfa30d5
0cae2fb86"
},
"outputs": [],
"source": [
"df=pd.read_csv('household_power_consumption.txt', sep=';', \n",
"                parse_dates={'dt': ['Date', 'Time']}, infer_datetime_format=True,
\n",
"                low_memory=False, na_values=['nan', '?'], index_col='dt')"
]
},
{
"cell_type":
"code", "execution_co
unt": 27, "metadata": {
"_cell_guid": "e4be7beb-bed9-45ae-8f7b-
5fed896f26e6", "_uuid": "bbcddeb8ff813f822725d7c73ceeee9e
497a4f06"
},
"outputs": [
{

```

"data":{

```

"text/html":
["<div>\n",
"<scoped>\n",
" .dataframe tbody tr th:only-of-type
{\n"," vertical-align:middle;\n",
" }\n",
"\n",
" .dataframe tbody tr th
{\n"," vertical-align:top;\n",
" }\n",
"\n",
" .dataframe thead th
{\n"," text-align:right;\n",
"
    }\n","</st
yle>\n",
"<table border=\"1\"
class=\"dataframe\">\n","<thead>\n",
" <tr style=\"text-align:
right;\">\n"," <th></th>\n",
" <th>Global_active_power</th>\n",
"
    <th>Global_reactive_power</th>\n",
" <th>Voltage</th>\n",
"
    <th>Global_intensity</th>\n","
    <th>Sub_metering_1</th>\n","
    <th>Sub_metering_2</th>\n","
    <th>Sub_metering_3</th>\n","</tr
>\n",
"<tr>\n",
"
    <th>dt</th>\n","<th

```

></th>\n",

" <th></th>\n",

" <th></th>\n",

```

"    <th></th>\n",
"    <th></th>\n",
"    <th></th>\n",
"    <th></th>\n",
"  </tr>\n",
" </thead>\n",
" <tbody>\n",
"   <tr>\n",
"     <th>2006-12-1617:24:00</th>\n",
"     <td>4.216</td>\n",
"     <td>0.418</td>\n",
"     <td>234.84</td>\n",
"     <td>18.4</td>\n",
"     <td>0.0</td>\n",
"     <td>1.0</td>\n",
"     <td>17.0</td>\n",
"   </tr>\n",
"   <tr>\n",
"     <th>2006-12-1617:25:00</th>\n",
"     <td>5.360</td>\n",
"     <td>0.436</td>\n",
"     <td>233.63</td>\n",
"     <td>23.0</td>\n",
"     <td>0.0</td>\n",
"     <td>1.0</td>\n",
"     <td>16.0</td>\n",
"   </tr>\n",
"   <tr>\n",
"     <th>2006-12-1617:26:00</th>\n",
"     <td>5.374</td>\n",
"     <td>0.498</td>\n",
"     <td>233.29</td>\n",

```

```

"
    <td>23.0</td>\n","
    <td>0.0</td>\n","
    <td>2.0</td>\n","

    <td>17.0</td>\n","
    </tr>\n",
" <tr>\n",
"   <th>2006-12-1617:27:00</th>\n",
"   <td>5.388</td>\n","
    <td>0.502</td>\n","
    <td>233.74</td>\n",
"   <td>23.0</td>\n",
"   <td>0.0</td>\n","
    <td>1.0</td>\n",

    <td>17.0</td>\n","
    </tr>\n",
" <tr>\n",
"   <th>2006-12-1617:28:00</th>\n",
"   <td>3.666</td>\n","
    <td>0.528</td>\n","
    <td>235.68</td>\n",
"   <td>15.8</td>\n",
"   <td>0.0</td>\n",
    <td>1.0</td>\n",

    <td>17.0</td>\n","
    </tr>\n",
"</tbody>\n","</table>
le>\n","</div>"
],

```

```
"text/plain":[
"          Global_active_powerGlobal_reactive_powerVoltage\\n","dt
\\n",
```

```

"2006-12-1617:24:00      4.216      0.418  234.84  \n",
"2006-12-1617:25:00      5.360      0.436  233.63  \n",
"2006-12-1617:26:00      5.374      0.498  233.29  \n",
"2006-12-1617:27:00      5.388      0.502  233.74  \n",
"2006-12-1617:28:00      3.666      0.528  235.68  \n",
"\n",
"
      Global_intensitySub_metering_1Sub_metering_2\\n","d
t
      \n",
"2006-12-1617:24:00      18.4      0.0      1.0  \n",
"2006-12-1617:25:00      23.0      0.0      1.0  \n",
"2006-12-1617:26:00      23.0      0.0      2.0  \n",
"2006-12-1617:27:00      23.0      0.0      1.0  \n",
"2006-12-1617:28:00      15.8      0.0      1.0  \n",
"\n",
"      Sub_metering_3\n",
"dt      \n",
"2006-12-1617:24:00      17.0\n",
"2006-12-1617:25:00      16.0\n",
"2006-12-1617:26:00      17.0\n",
"2006-12-1617:27:00      17.0\n",
"2006-12-1617:28:00      17.0"
]
},
"execution_count":
27,"metadata":{},
"output_type":"execute_result"
}
],
"source":[
"df.head()"
]

```

},

```

{
  "cell_type":
  "code", "execution_c
ount": 4, "metadata": {
    "_cell_guid": "76a8cf19-7cfd-4b46-b508-
200ead2cafc3", "_uuid": "11872f6f01ded0e63e9a32e82acc2808
60c5a15d"
  },
  "outputs": [
    {
      "name":
      "stdout", "output_type":
      "stream", "text": [
        "<class 'pandas.core.frame.DataFrame'>\n",
        "DatetimeIndex: 2075259 entries, 2006-12-16 17:24:00 to 2010-11-
2621:02:00\n",
        "Data columns (total 7
columns):\n", "#Column  Dtype\n",
        "-----\n",
        "0Global_active_power
                                float64\n", "
1Global_reactive_powerfloat64\n", "2V
oltage                                float64\n", "
3Global_intensity
                                float64\n", "
4Sub_metering_1
                                float64\n", "
5Sub_metering_2
                                float64\n", "
6Sub_metering_3
                                float64\n", "
dtypes:float64(7)\n",
        "memoryusage:126.7MB\n"
      ]
    }
  ]
}

```

```
]
}
],
"source":[
"df.info()"
```

```

]
},
{
  "cell_type":
  "code", "execution_c
ount": 5, "metadata": {
    "_cell_guid": "a26765e3-2ac8-441a-a3d8-
ff32ef5c80a0", "_uuid": "cfbb517b90839c772001575bb5636053
db10159f"
  },
  "outputs": [
    {
      "data":
      {"text/plain":
      [
        "Global_active_power
                                float64\n", "Globa
l_reactive_power
                                float64\n", "Volta
ge
                                float64\n", "Globa
l_intensity
                                float64\n", "Sub_
metering_1
                                float64\n", "Sub_
metering_2
                                float64\n", "Sub_
metering_3
                                float64\n", "dtype:
object"
      ]
    },

```

```
"execution_count":  
5,"metadata":{},  
"output_type":"execute_result"  
}  
],  
"source":  
["df.dtypes  
"  
]  
},
```

```

{
  "cell_type":
  "code", "execution_c
ount": 6, "metadata": {
    "_cell_guid": "5eddbf54-4dd0-49d1-af1f-
4d6a97275247", "_uuid": "a0e9aa33b5f6231b67cfce24445fbd7
95190fd42"
  },
  "outputs": [
    {
      "data":
      {"text/plain":
      ["(2075259,7)"
      ]
      },
      "execution_count":
      6, "metadata": {},
      "output_type": "execute_result"
    }
  ],
  "source":
  ["df.shape
  "
  ]
},
{
  "cell_type":
  "code", "execution_c
ount": 7, "metadata": {
    "_cell_guid": "234d0be9-5117-40f6-93d8-
8c4f425f1926", "_uuid": "e4c89c0f6b3e00baa511d2b611079d6
a00648c83"
  },

```

```
"outputs":[  
  {
```

```

"data":{
  "text/html":
    ["<div>\n",
     "<scoped>\n",
     "  .dataframe tbody tr th:only-of-type
     {\n","  vertical-align:middle;\n",
     "  }\n",
     "\n",
     "  .dataframe tbody tr th
     {\n","  vertical-align:top;\n",
     "  }\n",
     "\n",
     "  .dataframe thead th
     {\n","  text-align:right;\n",
     "
     } \n","</st
yle>\n",
     "<table border=\"1\"
     class=\"dataframe\">\n","<thead>\n",
     "  <tr style=\"text-align:
     right;\">\n","    <th></th>\n",
     "    <th>Global_active_power</th>\n",
     "
     <th>Global_reactive_power</th>\n",
     "    <th>Voltage</th>\n",
     "
     <th>Global_intensity</th>\n","
     <th>Sub_metering_1</th>\n","
     <th>Sub_metering_2</th>\n","
     <th>Sub_metering_3</th>\n","</tr
     >\n",
     "
     </thead>\n","

```

```
<tbody>\n","<
tr>\n",
"  <th>count</th>\n",
```

```

"    <td>2.049280e+06</td>\n",
"    <td>2.049280e+06</td>\n",
"    <td>2.049280e+06</td>\n",
"    <td>2.049280e+06</td>\n",
"    <td>2.049280e+06</td>\n",
"    <td>2.049280e+06</td>\n",
"    <td>2.049280e+06</td>\n",
"  </tr>\n",
" <tr>\n",
"   <th>mean</th>\n",
"   <td>1.091615e+00</td>\n",
"   <td>1.237145e-01</td>\n",
"   <td>2.408399e+02</td>\n",
"   <td>4.627759e+00</td>\n",
"   <td>1.121923e+00</td>\n",
"   <td>1.298520e+00</td>\n",
"   <td>6.458447e+00</td>\n",
" </tr>\n",
" <tr>\n",
"   <th>std</th>\n",
"   <td>1.057294e+00</td>\n",
"   <td>1.127220e-01</td>\n",
"   <td>3.239987e+00</td>\n",
"   <td>4.444396e+00</td>\n",
"   <td>6.153031e+00</td>\n",
"   <td>5.822026e+00</td>\n",
"   <td>8.437154e+00</td>\n",
" </tr>\n",
" <tr>\n",
"   <th>min</th>\n",
"   <td>7.600000e-02</td>\n",
"   <td>0.000000e+00</td>\n",

```

```

"    <td>2.232000e+02</td>\n",
"    <td>2.000000e-01</td>\n",
"    <td>0.000000e+00</td>\n",
"    <td>0.000000e+00</td>\n",
"    <td>0.000000e+00</td>\n",
"  </tr>\n",
"  <tr>\n",
"    <th>25%</th>\n",
"    <td>3.080000e-01</td>\n",
"    <td>4.800000e-02</td>\n",
"    <td>2.389900e+02</td>\n",
"    <td>1.400000e+00</td>\n",
"    <td>0.000000e+00</td>\n",
"    <td>0.000000e+00</td>\n",
"    <td>0.000000e+00</td>\n",
"  </tr>\n",
"  <tr>\n",
"    <th>50%</th>\n",
"    <td>6.020000e-01</td>\n",
"    <td>1.000000e-01</td>\n",
"    <td>2.410100e+02</td>\n",
"    <td>2.600000e+00</td>\n",
"    <td>0.000000e+00</td>\n",
"    <td>0.000000e+00</td>\n",
"    <td>1.000000e+00</td>\n",
"  </tr>\n",
"  <tr>\n",
"    <th>75%</th>\n",
"    <td>1.528000e+00</td>\n",
"    <td>1.940000e-01</td>\n",
"    <td>2.428900e+02</td>\n",
"    <td>6.400000e+00</td>\n",

```

```

"
<td>0.000000e+00</td>\n","
<td>1.000000e+00</td>\n","
<td>1.700000e+01</td>\n","</tr
>\n",
"<tr>\n",
"  <th>max</th>\n",
"
<td>1.112200e+01</td>\n","
<td>1.390000e+00</td>\n","
<td>2.541500e+02</td>\n","
<td>4.840000e+01</td>\n","
<td>8.800000e+01</td>\n","
<td>8.000000e+01</td>\n","
<td>3.100000e+01</td>\n","</tr
>\n",
"</tbody>\n","</tab
le>\n","</div>"
],
"text/plain":[
"  Global_active_powerGlobal_reactive_power    Voltage\\n",
"count      2.049280e+06      2.049280e+062.049280e+06  \n",
"mean       1.091615e+00      1.237145e-012.408399e+02  \n",
"std        1.057294e+00      1.127220e-013.239987e+00  \n",
"min        7.600000e-02      0.000000e+002.232000e+02  \n",
"25%        3.080000e-01      4.800000e-022.389900e+02  \n",
"50%        6.020000e-01      1.000000e-012.410100e+02  \n",
"75%        1.528000e+00      1.940000e-012.428900e+02  \n",
"max        1.112200e+01      1.390000e+002.541500e+02  \n",
"\n",
"\n",
\n",

```

" Global_intensitySub_metering_1Sub_metering_2Sub_metering_3"count 2.049280e+

```

    "mean      4.627759e+00   1.121923e+00   1.298520e+00   6.458447e+00
\n",
    "std  4.444396e+00  6.153031e+00  5.822026e+00   8.437154e+00\n", "min
    2.000000e-01   0.000000e+00   0.000000e+00   0.000000e+00\n", "25%
    1.400000e+00   0.000000e+00   0.000000e+00   0.000000e+00
\n",
    "50%      2.600000e+00  0.000000e+00   0.000000e+00   1.000000e+00"75%
\n",
    6.400000e+00   0.000000e+00   1.000000e+00   1.700000e+01
\n",
    "max      4.840000e+01   8.800000e+01   8.000000e+01   3.100000e+01"
]
},
"execution_count":
7, "metadata": {},
"output_type": "execute_result"
}
],
"source": [
    "df.describe()"
]
},
{
    "cell_type":
"code", "execution_c
ount": 8, "metadata": {
    "_cell_guid": "25d6f9c5-fd15-4ba7-b938-
a3e927c50f1c", "_uuid": "1d8ddd56738f436c1713c451328693e
dafa1c3b4"
},
"outputs": [
    {
        "data":
        {"text/plain":

```

[

```

"Index(['Global_active_power', 'Global_reactive_power',
'Voltage',\n",
      'Global_intensity','Sub_metering_1','Sub_metering_2',\n",
"
      'Sub_metering_3'],\n",
"      dtype='object')"
```

```

]
},
"execution_count":
8,"metadata":{},
"output_type":"execute_result"
}
],
"source":
["df.columns
"
]
},
{
"cell_type":
"markdown","metadata":
{
"_cell_guid": "938ffef5-1d60-45ab-9e20-
1377c25ef3ec","_uuid":"2e248c3bc354cb0d672d173577e81d
815d1b9a3c"
},
"source":[
"##Dealingwithmissingvalues'nan'withateststatistic"
]
},
{
"cell_type":
"code","execution_c

```

```
ount": 9,"metadata":{
  "_cell_guid": "81cdc1e5-ca60-4239-ace7-
2f814ac871f5","_uuid":"efb409f0ea173f8d4589cc8c8f6aed97e
771e873"
},
```

```

"outputs":[
{
  "data":
  {"text/plain":
  [
    "[0, 1,2,3,4, 5,6]"
  ]
  },
  "execution_count":
  9,"metadata":{},
  "output_type":"execute_result"
},
]
"source":[
  "## finding all columns that have
  nan:\n","\n",
  "dropping_list_all=[]\n",
  "forjinrange(0,7):\n",
  "  ifnotdf.iloc[:,j].notnull().all():\n",
  "    dropping_list_all.append(j)
    \n","
    #print(df.iloc[:,j].unique())\n","dro
    ping_list_all"
  ]
},
{
  "cell_type":
  "code","execution_co
  unt": 10,"metadata":{
    "_cell_guid": "77b8bc3e-c837-4976-ab9f-
    bc441d251d8b","_uuid":"cd3966b66a5b8682efffa059632
    536ae7e5820e4"
  },

```

```
"outputs":[],  
"source":[
```

```

"# filling nan with mean in any
columns\n","\n",
"forj in range(0,7):    \n",
"    df.iloc[:,j]=df.iloc[:,j].fillna(df.iloc[:,j].mean())"
]
},
{
"cell_type":
"code","execution_co
unt": 11,"metadata":{
"_cell_guid": "03367835-8a7b-48d7-becb-
4f1e10f18b28","_uuid":"39f6858342a7496f73fb1d5bc1d184dd
bb75b773"
},
"outputs":[
{
"data":
{"text/plain":
[
"Global_active_power
0\n","Globa
l_reactive_power
0\n","Volta
ge
0\n","Globa
l_intensity
0\n","Sub_
metering_1
0\n",
"Sub_metering_2
0\n",
"Sub_metering_3
0\n",
"dtype:int64"
]

```

```
},  
  "execution_count":  
    11,"metadata":{},  
  "output_type":"execute_result"  
}  
],
```

```

"source":[
  "# another sanity check to make sure that there are not more any
  nan\n","df.isnull().sum()"
],
{
  "cell_type":
  "markdown","metadata":
  {
    "_cell_guid": "2cc83d0f-626b-4048-8ffd-
    51010e800810","_uuid":"64b0092f2b4702dcbebb56f868f8725
    2ea6b87eb"
  },
  "source":[
    "#Datavisualization"
  ],
},
{
  "cell_type":
  "code","execution_co
  unt": 12,"metadata":{
    "_cell_guid": "dd48cb9c-3970-4614-95d5-
    0eb668b00994","_uuid":"ab6c5bdbe295c16d37f45be52b7244
    6a1cb35806"
  },
  "outputs":[
    {,,,,
    "<FigureSize432x288with1Axes>"
  ],
},
  "metadata":
  {"needs_background":"light
  "

```

```
},  
  "output_type": "display_data"  
},
```

```

{
  "name":
  "stdout","output_type":
  "stream","text":[
    "TestRMSE:0.615\n"
  ]
}
],
"source":[
  "\n",
  "model=Sequential()\n",
  "model.add(LSTM(100,
input_shape=(train_X.shape[1],
train_X.shape[2])))\n",
  "model.add(Dropout(0.2))\n",
  "# model.add(LSTM(70))\n",
  "# model.add(Dropout(0.3))\n",
  "model.add(Dense(1))\n","model.compile(loss='mean_squared_er
ror', optimizer='adam')\n","\n",
  "\n",
  "\n",
  "#fitnetwork\n",
  "history = model.fit(train_X, train_y, epochs=20,
batch_size=70,validation_data=(test_X,test_y),verbose=2,shuffle=
False)\n",
  "\n",
  "# summarize history for
loss\n","plt.plot(history.history['loss'])\n"
,"plt.plot(history.history['val_loss'])\n","pl
t.title('model
loss')\n","plt.ylabel('loss')\n",
  "plt.xlabel('epoch')\n",
  "plt.legend(['train','test'],loc='upperright')\n","
plt.show()\n",

```

```

"\n",
"#makeaprediction\n",
"yhat=model.predict(test_X)\n",
"test_X = test_X.reshape((test_X.shape[0],
7))\n", "#invert scalingforforecast\n",
"inv_yhat = np.concatenate((yhat, test_X[:, -6:]),
axis=1)\n", "inv_yhat =
scaler.inverse_transform(inv_yhat)\n", "inv_yhat=inv_yhat[
:,0]\n",
"#invertscalingforactual\n",
"test_y=test_y.reshape((len(test_y),1))\n",
"inv_y = np.concatenate((test_y, test_X[:, -6:]),
axis=1)\n", "inv_y=scaler.inverse_transform(inv_y)\n",
"inv_y =
inv_y[:,0]\n", "#calcula
teRMSE\n",
"rmse = np.sqrt(mean_squared_error(inv_y,
inv_yhat))\n", "print('TestRMSE:%.3f' % rmse)"
]
},
{
"cell_type":
"markdown", "metadata":
{
"_cell_guid": "82163b05-3842-4108-aee9-
ca3e4b88d924", "_uuid": "57a7d528a6f621368582f2440bd3b30
b0675dd88"
},
"source":[
"###Notethatinordertoimprovethemodel,onehastoadjustepochsandbatch_size."
]
},
{

```

```
"cell_type":  
"code","execution_co  
unt":22,
```

```

"metadata":{
  "_uuid":"78449473770f61669fa5bd8acedb0b662f48ebac"
},
"outputs":[
  {
    "data":{,,,,,,,,

    "text/plain":[
      "<FigureSize432x288with1Axes>"
    ]
  },
  "metadata":
  {
    "needs_background":"light
  "
  },
  "output_type":"display_data"
}
],
"source":[
  "aa=[xforxinrange(200)]\n",
  "plt.plot(aa, inv_y[:200], marker='.',
  label=\"actual\")\n", "plt.plot(aa, inv_yhat[:200], 'r',
  label=\"prediction\")\n", "plt.ylabel('Global_active_pow
  er', size=15)\n", "plt.xlabel('Time step',
  size=15)\n", "plt.legend(fontsize=15)\n",
  "plt.show()"
]
"text/plain":[
  "<FigureSize432x288with1Axes>"
]
},
"metadata": {

```

```

    "needs_background": "light"
  },
  "output_type": "display_data"
}
],
"source": [
  "aa=[xforx inrange(200)]\n",
  "plt.plot(aa, inv_y[:200], marker='.',",
  label="\nactual\n"),\n", "plt.plot(aa, inv_yhat[:200], 'r',",
  label="\nprediction\n"),\n", "plt.ylabel('Global_active_power",
  ', size=15)\n", "plt.xlabel('Time step',",
  size=15)\n", "plt.legend(fontsize=15)\n",
  "plt.show()"
]
},
{
  "cell_type":
  "code", "execution_coun
t": null, "metadata": {
  "_cell_guid": "9e12fbc0-d0ba-4cce-a202-
a2656cb2b1e0", "_uuid": "07a28ca8f9c4964fb7a63d49baabdafc29
aaf760"
},
  "outputs": [],
  "source": []
}
],
"metadata":
{"kernel_spec": {

```

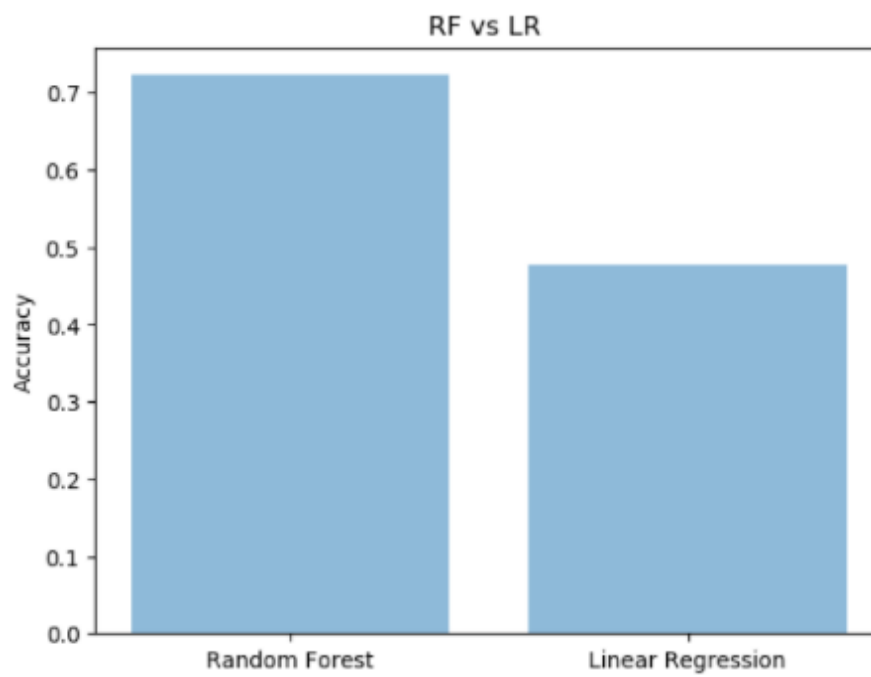
```
"display_name": "Python3(ipykernel)",
```

```

"language":
"python","name":"py
thon3"
},
"language_info":
{"codemirror_mode":
{"name":
"ipython","version":
3
},
"file_extension":
".py","mimetype": "text/x-
python","name":
"python","nbconvert_exporter":
"python","pygments_lexer":
"ipython3","version":"3.9.7"
}
},
"nbformat":4,
"nbformat_minor":1
}

```

5.5 ComparativeResultGraph



CHAPTER
6 CONCLUSION
N

In this project we compare several regression algorithms to estimate the power consumption of an air conditioner as follows: multi-layered perceptron's (MLP), radial basis function networks (RBFN) and support vector regression (SVR). MLP is a conventional nonlinear regression method. We find the sufficient accuracy for monitoring System which can estimate the power consumption of electric appliances, to actual fields.

References:

- [1] C. M. Bishop.: Newral Networks for Pattern Recognition. Oxford University Press,1995.
- [2] W. Hart.: Non-intrusive appliance load monitoring. Proceedings of the IEEE, vol.80,no.12,1992.
- [3] J.Moody et al.: Fast learning in networks of locallytuned processing units. NeuralComputation1 (2), ,692498,2001. 281-294,1989.
- [4] H.Murata et al.: Applying Kernel Based Subspace classification to a Non-IntrusiveMonitoringSystemforHousehold ElectricAppliances.ICANN2001
- [5] T.Onodaetal.:ApplyingSupportVectorMachinesandBoostingtoaNon-IntrusiveMonitoringSystemforHouseholdElectricApplianceswithInverters.NC'2000,2000.
- [6] D. E. Rumelhart et al.: Learning Internal Rep resent ations by Error Propagation.ParallelDistributedProcessing:ExplorationsintheMicrostructureofCognition,Vol.I:Foundations,318-362,BradfordBooks/MITPress,Cambridge,MA.,1986.
- [7] V.N.Vapnik.:TheNatureofStatisticalLearningTheonJ.Springer,1995.
- [8] G. Mulligan and M. Cahill , “Using geographically weighted regression to explorelocal crime patterns,” Social Sci. Comput. Rev., vol. 25, no. 2, pp. 174–193, May2007,doi: 10.1177/0894439307298925.
- [9] A. Babakura, M. N. Sulaiman, and M. A. Yusuf “Risk terrain modeling: Brokeringcriminological theory and GIS methods for crime forecasting,” Justice Quart., vol. 28,no.2,pp.360–381,Apr.2011, doi:10.1080/07418825.2010.486037.
- [10] A. Almehmadi, Z. Joudaki, and R. Jalali, “Language usage on Twitter predictscrime rates,” in Proc. 10th Int. Conf. Secur. Inf. Netw., Oct. 2017, pp. 307–310, doi:10.1145/3136825.3136854.
- [11] V. K. Borooah and N. Ireland, “Deprivation, violence, and conflict: An analysis ofNaxalite activity in the districts of India,” Int. J. Conf. Violence, vol. 2, no. 2, pp. 317–333,2008,doi:10.4119/UNIBI/ijcv.42.

[12] J. M. Caplan, L. W. Kennedy, and J. Miller, "Improved method of classification algorithms for crime prediction," in Proc. Int. Symp. Biometrics Secur. Technol. (ISBAST), Aug. 2014, pp. 250–255, doi:10.1109/ISBAST.2014.7013130.

[13] J. M. Caplan, L. W. Kennedy, and J. Miller, "Risk terrain modeling: Brokering criminological theory and GIS methods for crime forecasting," Justice Quart., vol. 28, no. 2, pp. 360–381, Apr. 2011, doi:10.1080/07418825.2010.486037.

.

[14] Z. Li, T. Zhang, Z. Yuan, Z. Wu, and Z. Du, "Spatio-temporal pattern analysis and prediction for urban crime," in Proc. 6th Int. Conf. Adv. Cloud Big Data (CBD), Aug. 2018, pp. 177–182, doi: 10.1109/CBD.2018.00040.

[15] A. Almaw and K. Kadam, "Survey paper on crime prediction using ensemble approach," Int. J. Pure Appl. Math., vol. 118, no. 8, pp. 133–139, 2018. [Online]. Available: <https://internal-pdf://107.93.182.66/18.pdf%0Ahttp://www.ijpam.eu>

[16] T. B. Hyde, H. Dentz, S. A. Wang, H. E. Burchett, S. Mounier-Jack, and C. F. Mantel, "The impact of new vaccine introduction on immunization and health systems: A review of the published literature," Vaccine, vol. 30, no. 45, pp. 6347–6358, 2015, doi: 10.1016/j.vaccine.2012.08.029

[17] S. Yadav, M. Timbadia, A. Yadav, R. Vishwakarma, and N. Yadav, "Crime pattern detection, analysis & prediction," in Proc. Int. Conf. Electron., Commun. Aerosp. Technol. (ICECA), Apr. 2017, pp. 225–230, doi:10.1109/ICECA.2017.8203676.