

**A
MINI PROJECT
on
Data analysis on Energy Conservation**

Submitted By

S.Yadavendra.Reddy (160116737120)

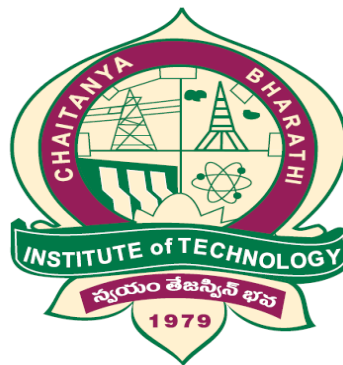
B.sanjay (160116737321)

Under the guidance of

Mrs.B.VeeraJyothi

Asst. Professor,

Dept. of IT, CBIT.



DEPARTMENT OF INFORMATION TECHNOLOGY

CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY

(Affiliated to Osmania University; Accredited by NBA (AICTE), ISO Certified 9001:2015)

Website: www.cbit.ac.in, www.cbitworld.com

HYDERABAD – 500 075

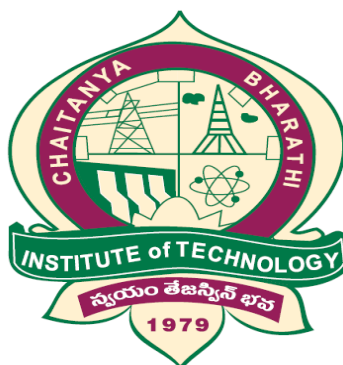
2018-19

CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY

DEPARTMENT OF INFORMATION TECHNOLOGY

(Affiliated to Osmania University)

GANDIPET, HYDERABAD – 500 075



CERTIFICATE

This is to certify that the project work titled “**DataAnalysis for energy conservation**” submitted to **CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY**, in partial fulfilment of the requirements for the award of the completion of 5th semester of B.E. in Information Technology, during the academic year 2018-19, is a record of original work done by **S. Yadavendra Reddy (160116737120)** and **B.Sanjay (160116737321)** during the period of study in Dept. of IT, CBIT, HYDERABAD, under our supervision and guidance.

PROJECT GUIDE

Mrs. B VeeraJyothi

Asst. Professor,

Dept. of IT, CBIT.

HEAD OF THE DEPARTMENT

Dr. Suresh Pabboju

Head of the department

Information Technology

C.B.I.T., HYDERABAD

DECLARATION

This is to certify that the work reported in the present report titled “**Data Analysis for Energy Conservation**” is a record of work done by us in the Department of Information Technology, Chaitanya Bharathi Institute of Technology, Hyderabad.

No part of the report is copied from books / journals / internet and wherever the portion is taken, the same has been duly referred. The reported results are based on the project work done entirely by us and not copied from any other source.

S.Yadavendra.Reddy(160116737120)

B.Sanjay (160116737321)

ACKNOWLEDGEMENT

It is our privilege to acknowledge with deep sense of gratitude and devotion for keen personal interest and invaluable guidance rendered by our Project Guide **Mrs. B.VeeraJyothi**, Assistant Professor, Department of Information, Chaitanya Bharathi Institute of Technology.

We take the opportunity to express our thanks to **Dr. Suresh Pabboju**, Professor & Head of IT Department, CBIT for his valuable suggestions and moral support.

We are grateful to our Principal **Dr. P Ravinder Reddy**, Chaitanya Bharati Institute of Technology, for his cooperation and encouragement.

Finally, we also thank all the staff members, faculty of Dept. of IT, CBIT, and our friends, who with their valuable suggestions and support, directly or indirectly helped us in completing this project work.

ABSTRACT

Data Analytics is defined as analysis or process of inspecting cleansing data in order to get useful output we have decided to use data analytics in field of energy to predict the energy consumption of a house our mini project is going to use a set of machine learning algorithms and we are going to draw the inference which one is better the usage being that if we can determine which machine learning algorithm is better which in turn helps in order to predict correct values so that we can save money on appliances energy consumption

BY:

S.YADAVENDRA REDDY

R.NO:160116737120

B.SANJAY

R.NO:160116737321

CONTENTS

	Page No
Abstract	v
Contents	vi
List of Figures	vii
1.Introduction	1
1.1. Overview	1
1.2. Aim of Project	1
1.3. Organization of the Report	2
2. Technologies Used	3
2.1. About Python	3
2.1.1. Introduction	3
2.1.2. Functions	3
2.1.3. Modules	3
2.1.4. Exception Handling	4
2.2 About MachineLearning	5
2.2.1 Regression	5
2.3. About Scikit-learn	6
2.3.1.Introduction	6
2.3.2.Syntax	6
2.4.About Pandas	7
2.4.1.Introduction	7
2.4.2.Syntax	7
2.5 matplotlib and accuracy checking	8
2.5.1 Matplotlib	8
2.5.2 Accuracy score	8
3. Software Requirement Specification	
3.1. Introduction	9
3.1.1 Purpose of the document	9
3.2. Users and their Characteristics	9

3.3. Software and Hardware Requirements	9
4. Software Design	
4.1 Usecase diagram	10
4.2 Class diagram	11
4.3 Activity diagram	12
4.4 Flow chart	13
5. Implementation	
5.1. Introduction	14
5.2. Python	14
5.3. Scikit-learn	14
5.4. Multioutput regressor	15
6. Results	16
6.1. Introduction	16
6.2. Testing Objectives	16
6.3. Output Screens	17
7. Conclusion and Future Scope	23
Bibliography	24

LIST OF FIGURES

Figure Name	Page No
Fig.6.1 Output1	17
Fig.6.2 Output2	18
Fig.6.3 Output3	19
Fig.6.4 Output4	20
Fig 6.5 Output5	21
Fig 6.6 Output6	22

1. INTRODUCTION

1.1.OVERVIEW

The necessity of the energy is important as the technology is advancing and as per Some reports there might be a point where we wont have energy to use .This mini Project focuses on the using machine learning algorithms to determine which model To use for energy conservation for better results.

1.2.AIM OF THE PROJECT

The aim of the project is to suggest which machine learning algorithm is best to obtain results so that we can use model to accomplish the task of energy conservation by large means

Conditions:

- A dataset is given which has time stamp data(A data which is recorder at time interval)
- there is a house whose energy consumption depends on relative humidity temperature and other factors.

ORGANISATION OF REPORT

The organization of the report is as follows:

Chapter 1 deals with the Introduction of the project and gives the details about the project in an abstract view.

Chapter 2 deals with the information about Python, Scikitlearn,matplotlib, pandas and its utilization details are discussed in brief.

Chapter 3 deals with the Software Requirements Specifications which is a specification of the project software and hardware requirements.

Chapter 4 deals with the Implementation part which includes the tools and software that are used.

Chapter 5 deals with the Testing of the project and screenshots of the project

Chapter 6 explains the Conclusion and further scope of the project.

2. TECHNOLOGIES USED

2.1 ABOUT Python

2.1.1 INTRODUCTION

Python is a widely used general-purpose, high level programming language. It was initially designed by Guido van Rossum in 1991 and developed by Python Software Foundation. It was mainly developed for emphasis on code readability, and its syntax allows programmers to express concepts in fewer lines of code.

Python is a programming language that lets you work quickly and integrate systems more efficiently.

There are two major Python versions- Python 2 and Python 3. Both are quite different.

It uses interpreter instead of compiler. # in the front of the line indicates that it is comment.

It is strongly indented language.

2.1.2 FUNCTIONS

A function is a block of code which only runs when it is called.

You can pass data, known as parameters, into a function.

A function can return data as a result.

Creating a function:

In python we use def keyword while defining a function.

Example:

```
def example():  
    print("this is an example")
```

Calling a function:

To call a function, use the function name followed by parenthesis:

```
example()
```

2.1.3 MODULES

A module is a file containing Python definitions and statements. A module can define functions, classes and variables. A module can also include runnable code. Grouping related code into a module makes the code easier to understand and use.

We can use any Python source file as a module by executing an import statement in some other Python source file. When interpreter encounters an import statement, it imports the module if the module is present in the search path.

A search path is a list of directories that the interpreter searches for importing a module.

For example, to import the module `calc.py`, we need to put the following command at the top of the script, example:

```
import math

print (math.sqrt(4))
```

2.1.4 EXCEPTION HANDLING

The try block lets you test a block of code for errors. The except block lets you handle the error. The finally block lets you execute code, regardless of the result of the try and except blocks. When an error occurs, or exception as we call it, Python will normally stop and generate an error message. These exceptions can be handled using the try statement:

try:

```
    print(x)
```

except:

```
    print("An exception occurred")
```

Since the try block raises an error, the except block will be executed. Without the try block, the program will crash and raise an error.

We can define as many exception blocks as we want.

2.2 ABOUT Machine learning

2.2.1 Introduction

Machine learning is a field of artificial intelligence that uses statistical techniques to give computer systems the ability to "learn" (e.g., progressively improve performance on a specific task) from data, without being explicitly programmed.

The name machine learning was coined in 1959 by Arthur Samuel. Machine learning explores the study and construction of algorithms that can learn from and make predictions on data – such algorithms overcome following strictly static program instructions by making data-driven predictions or decisions, through building a model from sample inputs. Machine learning is employed in a range of computing tasks where designing and programming explicit algorithms with good performance is difficult or infeasible; example applications include email filtering, detection of network intruders, and computer vision.

Machine learning is closely related to (and often overlaps with) computational statistics which also focuses on prediction-making through the use of computers. It has strong ties to Mathematical optimization, which delivers methods, theory and application domains to the field. Machine learning is sometimes conflated with [data mining](#), where the latter subfield focuses more on exploratory data analysis and is known as unsupervised learning.

Within the field of data analytics, machine learning is a method used to devise complex models and algorithms that lend themselves to prediction; in commercial use, this is known as predictive analytics. These analytical models allow researchers, [datascientists](#), engineers, and analysts to "produce reliable, repeatable decisions and results" and uncover "hidden insights" through learning from historical relationships and trends in the data.

Machine learning is of 3 types:

1) Supervised learning

2) Unsupervised Learning

1) Supervised learning: The type of machine learning in which the dataset is given and we train the algorithm so that it can identify/predict the values

2) Unsupervised learning: In the Unsupervised Learning the data doesn't have any labels and The data is group into clusters so that it can identify the outliers and so on.

Regression analysis:

In statistical modeling, **regression analysis** is a set of statistical processes for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables (or 'predictors'). More specifically, regression analysis helps one understand how the typical value of the dependent variable (or 'criterion variable') changes when any one of the independent variables is varied, while the other independent variables are held fixed.

2.3 ABOUT Scikit-learn

2.3.1 INTRODUCTION

Scikit-learn (formerly **scikits.learn**) is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, *k*-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy. **Scikit-learn** (formerly **scikits.learn**) is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, *k*-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

2.3.2 SYNTAX

To use Scikit-learn we download all packages from Anaconda which contains scikit-learn
Using Anaconda we can use the IDE (integrated development environment) for our project
To import scikit-learn we must use

Import sklearn

calling of object depends on the data if we want to split data as train/test we can use for models which we want to use.

2.4 ABOUT Pandas

2.4.1 INTRODUCTION

Python is a great language for doing data analysis, primarily because of the fantastic ecosystem of data-centric Python packages. Pandas is one of those packages, and makes importing and analyzing data much easier. In this article, I have used Pandas to analyse data on Country Data.csv file from UN public Data Sets of a popular 'statweb.stanford.edu' website. As I have analyzed the Indian Country Data, I have introduced Pandas key concepts as below. Before going through this article, have a rough idea of basics from matplotlib and csv.

Importing Data with Pandas:

The first step is to read the data. The data is stored as a comma-separated values, or csv, file, where each row is separated by a new line, and each column by a comma (.). In order to be able to work with the data in Python, it is needed to read the csv file into a Pandas DataFrame. A DataFrame is a way to represent and work with tabular data. Tabular data has rows and columns, just like this csv file Indexing DataFrames with Pandas Indexing can be possible using the pandas.DataFrame.iloc method. The iloc method allows to retrieve as many as rows and columns by position.

Indexing Using Labels in Pandas

Indexing can be worked with labels using the pandas.DataFrame.loc method, which allows to index using labels instead of positions.

Data Frame Math with Pandas

Computation of data frames can be done by using Statistical Functions of pandas Tool

Pandas Plotting

Plots in these examples are made using standard convention for referencing the matplotlib API which provides the basics in pandas to easily create decent looking plots.

2.4.2 SYNTAX

To need import pandas module in order to use it,

```
import pandas as pd
```

There are many functions available in pandas while are mainly useful for manipulating arrays, series, dataframes.

pd.read_csv("filename") is used to read a csv file.

Let data be a dataframe,

`data.sort_values (by=['attribute'])` is used to sort dataframe according to the given attribute.

`data.to_csv('filename.csv',index=True)` dataframe will be stored in the specified file.

Let data be a series,

`data. Head(n)` is used to get first n values present in the data.

`data.append (data2)` will append data2 series in data.

`Data.str.replace('initial', 'final')` will replace 'initial' with final.

2.5 Matplotlib and accuracy checking

2.5.1 Introduction to matplotlib

Matplotlib is a python library using which we can visualize the results which are applied using machine learning models Using this we have generated two graphs in which we give test values of output and predicted values of y by overlapping 2 graphs we get the difference in values and probably the results.

2.5.2 Accuracy score

While implementing machine learning models in scikit-learn based on the problem we have different models which give different values so we use a function called as r2 score in `sklearn.metrics` which gives us the accuracy of a regression model as the project solves regression models.If you want to find accuracy without `r2_score` we use `rmse`(root mean square error) this determines how much the prediction is wrong if `rmse` is low the model is more accurate

3. SOFTWARE REQUIREMENT SPECIFICATION

3.1 INTRODUCTION

The requirements specification is a technical specification of requirements for the software products. It is the first step in the requirements analysis process it lists the requirements of a particular software system including functional, performance and security requirements. The requirements also provide usage scenarios from a user, an operational and an administrative perspective. The purpose of software requirements specification is to provide a detailed overview of the software project, its parameters and goals. This describes the project target audience and its user interface, hardware and software requirements. It defines how the client, team and audience see the project and its functionality.

3.1.1 PURPOSE OF DOCUMENT

This software requirement specification describes all the requirements elicited for Using scikit-learn to build models and train them then we test the models and is intended to be used by the members examining the project and implementing and verifying the application. Unless otherwise noted all requirements are of high priority and are committed.

3.2 USERS AND THEIR CHARACTERISTICS

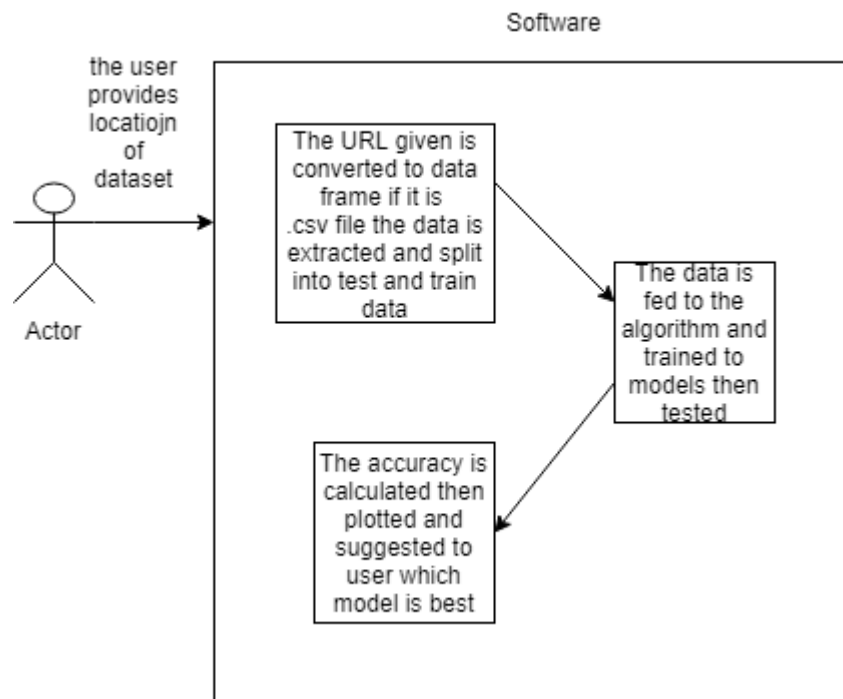
Data Analysis for energy conservation.

3.3 SOFTWARE AND HARDWARE REQUIREMENTS

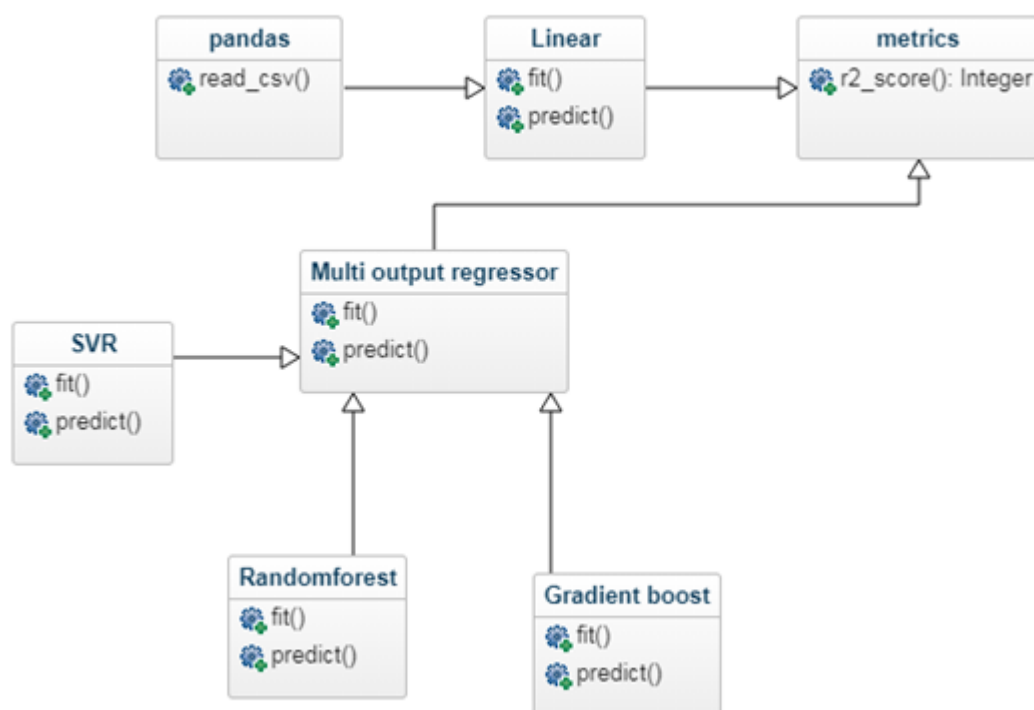
Operating System	Windows 10,Linux
Programming Language	PYTHON3(Anaconda is used as it has IDE),R
Processor	Intel(R) Core(TM) i3 CPU M 350 @2.27GHz
RAM	1 GB or more
Disk Space	1GB or more

4. SOFTWARE DESIGN

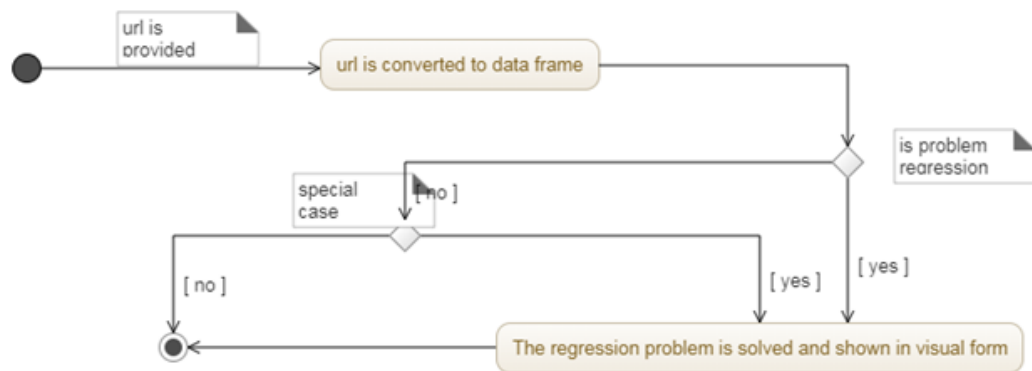
4.1 Usecase diagram



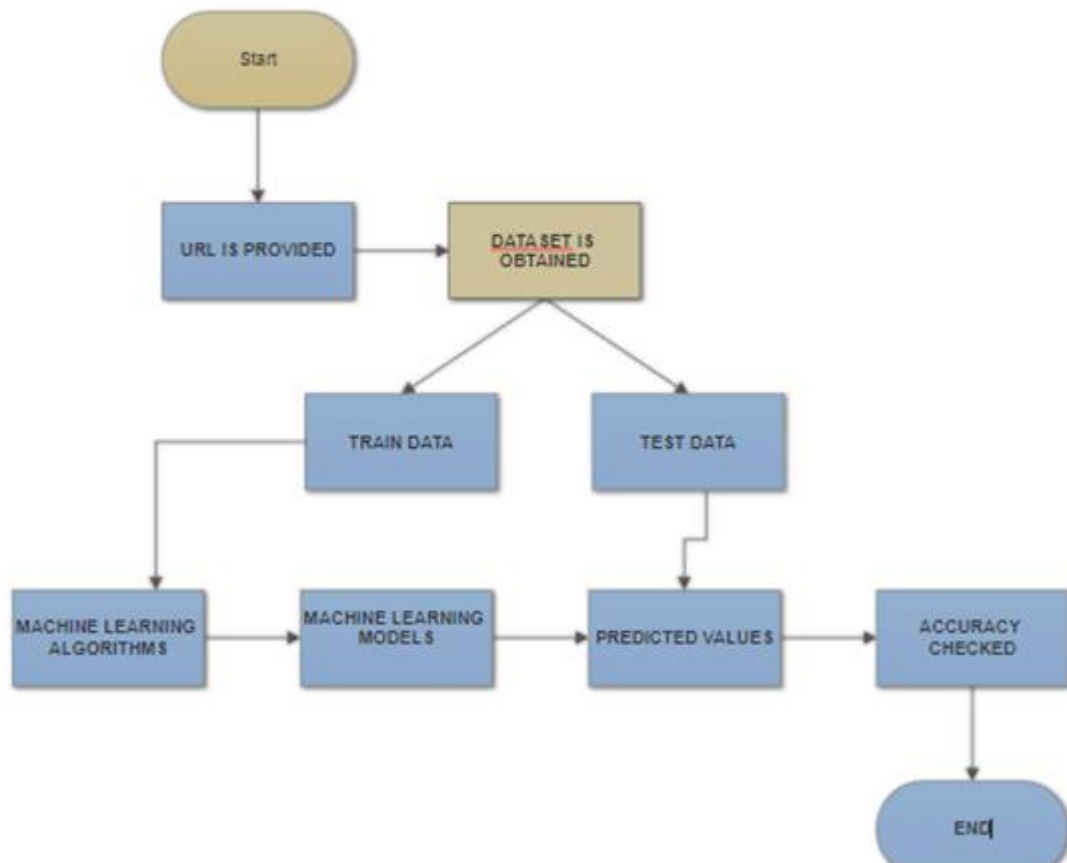
4.2 Class diagram



4.3 Activity Diagram



4.4 Flowchart



5. IMPLEMENTATION

5.1 INTRODUCTION

The success of the software product is determined only when it is successfully implemented according to the requirements. The analysis and the design of the proposed system provide a perfect platform to implement the idea using the specified technology in the desired environment. The implementation of our system is made user friendly.

Any software project is designed in modules and the project is said to be successfully implemented when each of the module is executed individually to obtain the expected result and also, when all the modules are integrated and run together without any errors.

5.2 PYTHON

Python is used to develop this project, so that the project works fast and effectively. Whole project is written in python only, no other languages are used. Different packages and modules are used for easing our work.

```
data=pd.read_csv (location)
```

location is the file url

data has the dataframe of the object

```
X=data[['I1','I2']]
```

```
Y=data[['o1','o2']]
```

X has all input values Y has the output values

Use train_test_split to split test and train values

```
X_train,Y_train,X_test,Y_test=test_train_split(data,test_size=0.2)
```

5.3 SCIKIT LEARN (SKLEARN)

It is used to import in built models so we can train them and test against values so we get some prediction then we check accuracy and analyze which model is better .

From sklearn import linear model

```
tr=lm.LinearRegression()
```

then we train model using data and test it to predict values

```
t=lm.fit (c1, c2)\
```

where c1 and c2 are train data

then we use fit values to predict we will use this for several models and calculate accuracy

5.4 MULTI OUTPUT REGRESSOR

The Multioutput regressor is an object which is used in machine learning algorithms in order to predict the values for multiple outputs. Multi output regressor can store objects like `SVR()`, `RandomForestRegressor ()` etc. and make them accept input like in data frame (since `SVR` takes input in array pattern) .This allows us to predict appliance consumption from lights, appliances etc.

6. RESULTS

6.1 INTRODUCTION

Software testing is a critical element of software quality assurance and represents the ultimate review of specification, design and coding. In fact, testing is the one step in the software engineering process that could be viewed as destructive rather than constructive.

A strategy for software testing integrates software test case design methods into a well-planned series of steps that result in the successful construction of software. Testing is the set of activities that can be planned in advance and conducted systematically. The underlying motivation of program testing is to affirm software quality with methods that can economically and effectively apply to both strategic to both large and small-scale systems.

6.2 TESTING OBJECTIVES

The main objective of performance testing is designed to test whether display is as expected and whether the webpage is functioning properly or not.

As the test results are gathered and evaluated they begin to give a qualitative indication of the reliability of the code. If proper output is not obtained, the overall quality of the code is questioned. If, on the other hand, all the results which are not successful, are encountered, and are easily modifiable, then the following conclusion can be made: The tests are inadequate as the requirements mentioned are not compatible. The testing includes:

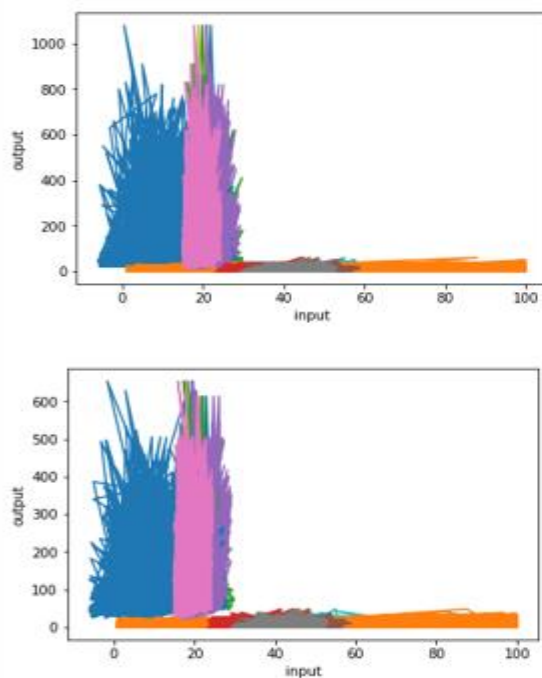
- Checking whether the information is displayed or not.
 - Checking whether all the players data is collected or not.
 - Checking whether all the inputs are correctly taken or not.
1. Verifying if all the pictures are displayed and none of the files are corrupted.

6.3. OUTPUT SCREENS

In the below output we have the accuracy of a Random forest regressor as we can see the accuracy is very less so we can't decide which one to use now itself

OUTPUT:

```
[ 86.  2.]
[402.  3.]
[166.  7.]
the rmse value is
52.665138759301804
the r2 value is
0.5461866694616753
C:\Users\Yadav\Anaconda\lib\site-packages\matplotlib\cbook\deprecation.py:107:
MatplotlibDeprecationWarning: cycling among columns of inputs with non-matching shape
is deprecated.
  warnings.warn(message, mplDeprecation, stacklevel=1)
```



r

Fig 6.1

In the below figure the accuracy is calculate for the Linear model this model gives even less accuracy than Random forest regressor

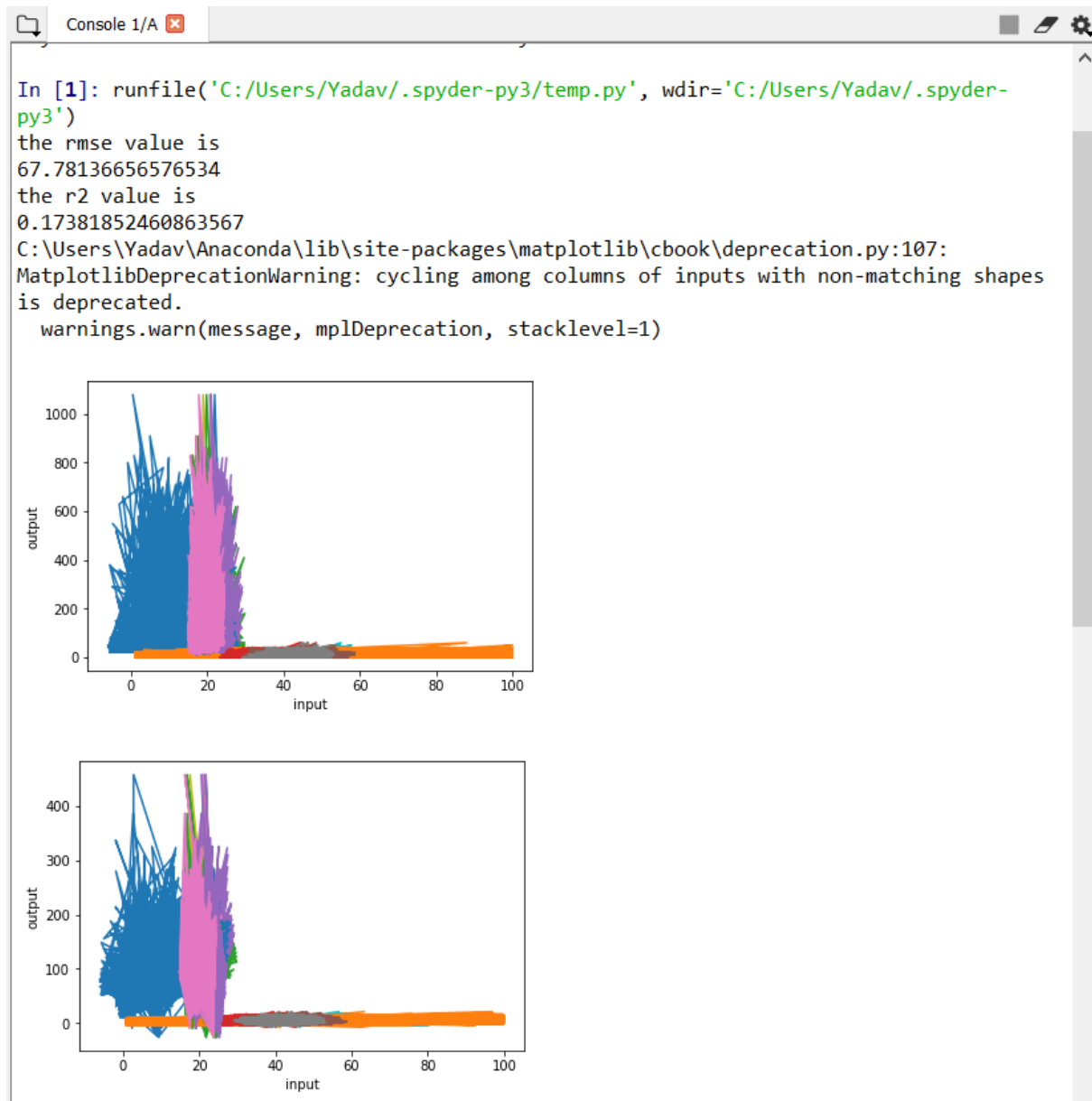


Fig 6.2

In the below figure the SVR model is used the accuracy calculated is very less so we cant tell

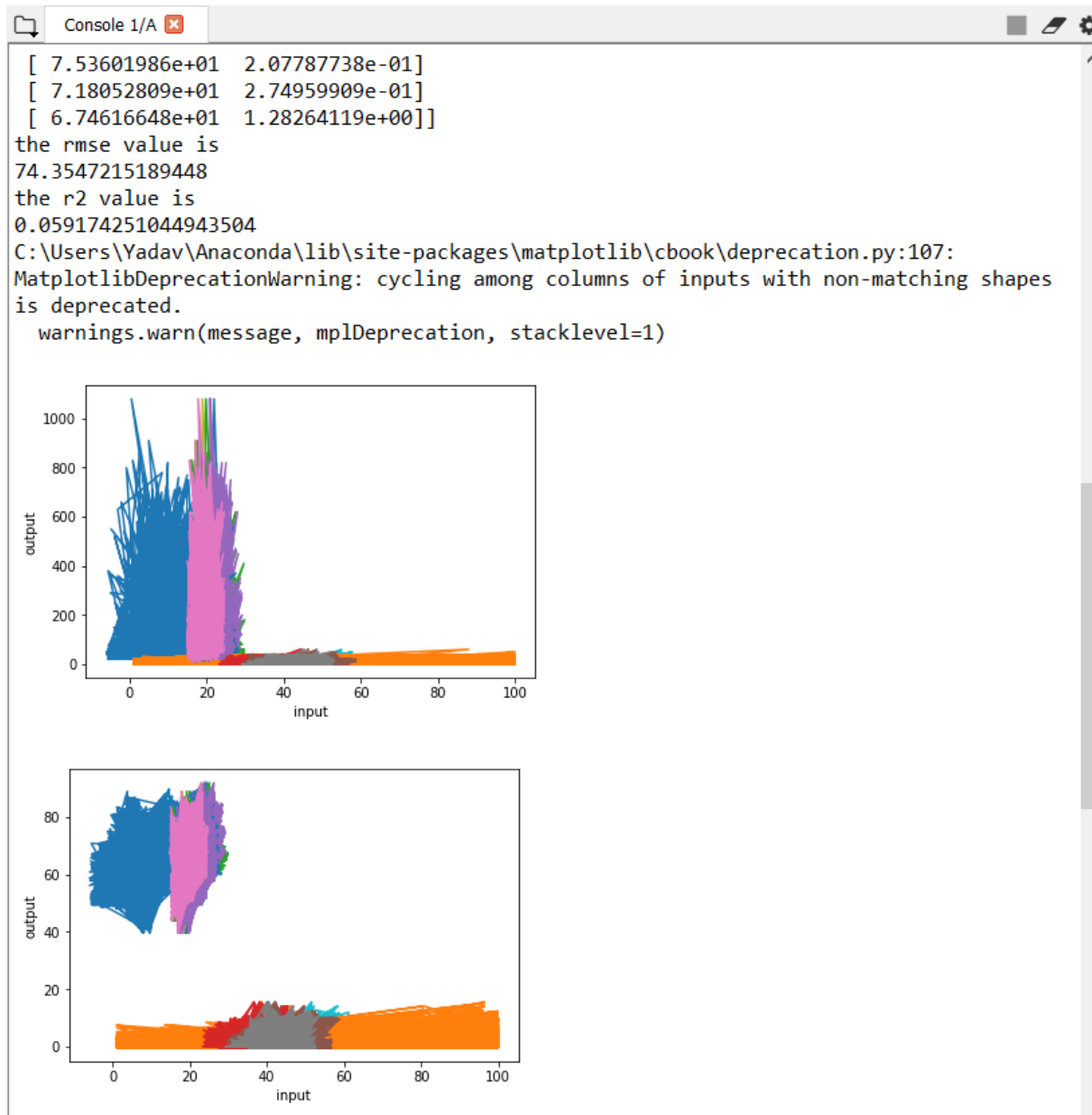


Fig 6.3

This model uses Xgradient regressor the accuracy is almost as Random forest so we can consider it

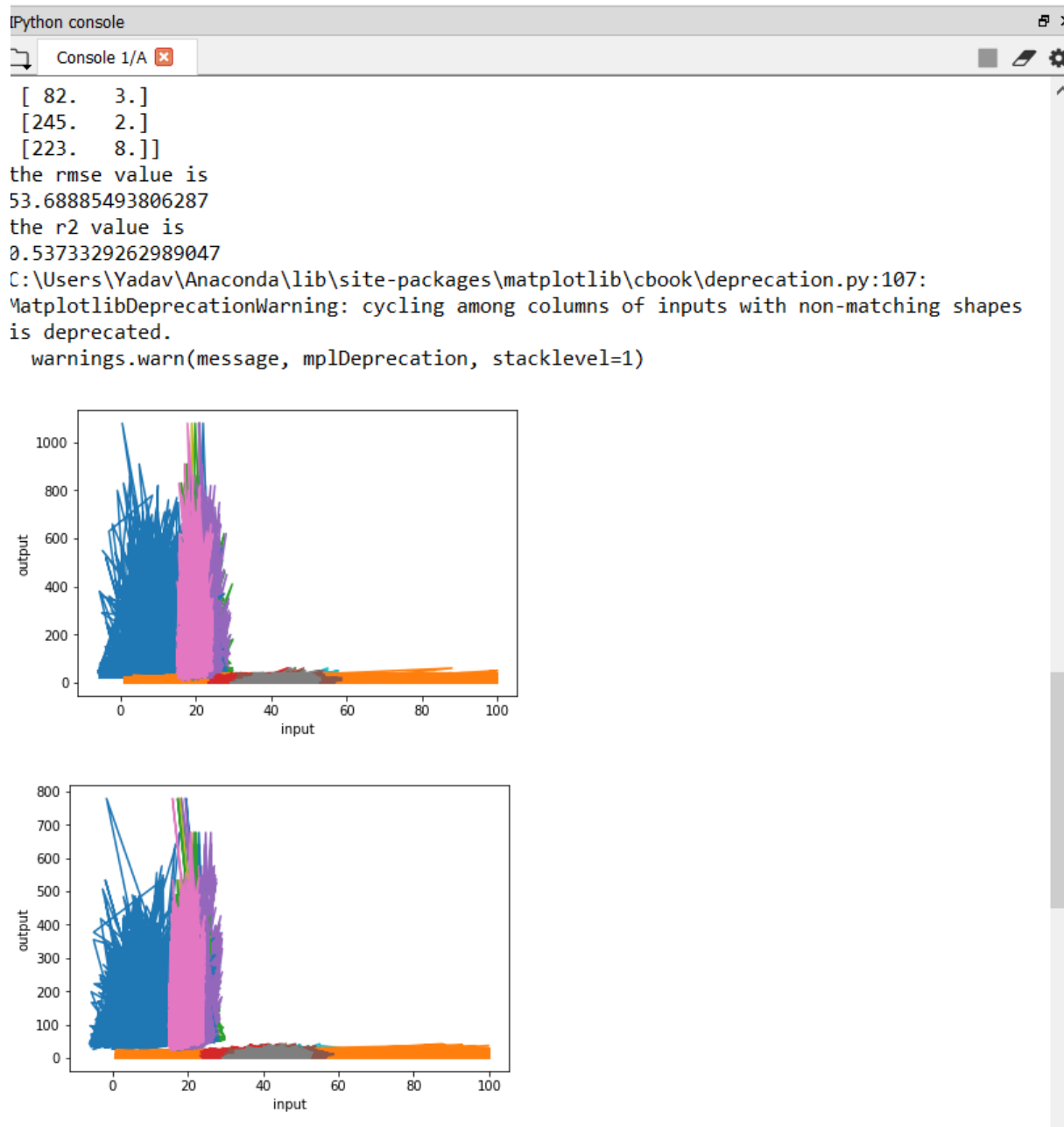


Fig 6.4

The below model is for gradient regressor the accuracy is very less so we cant determine

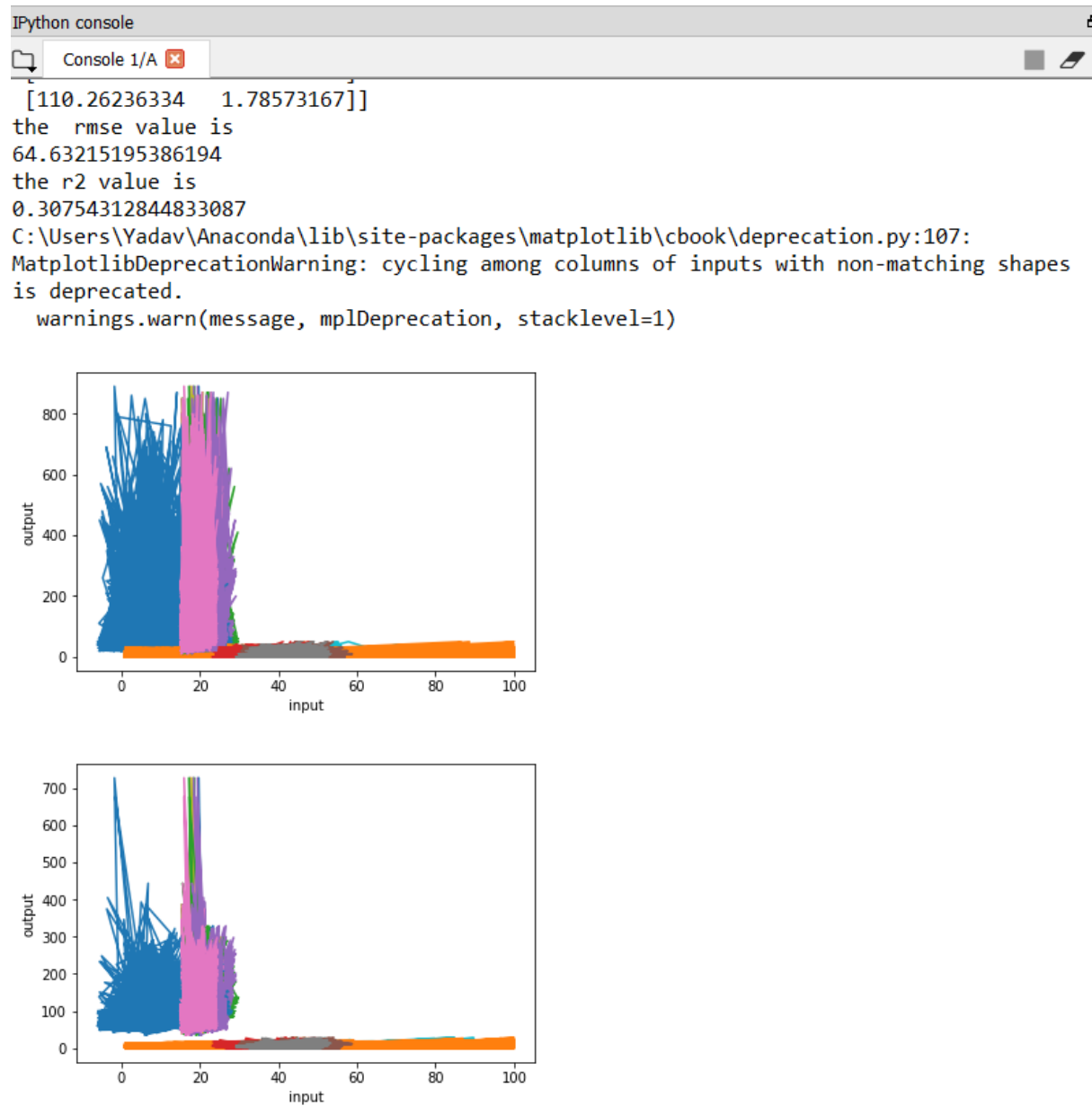


Fig 6.5

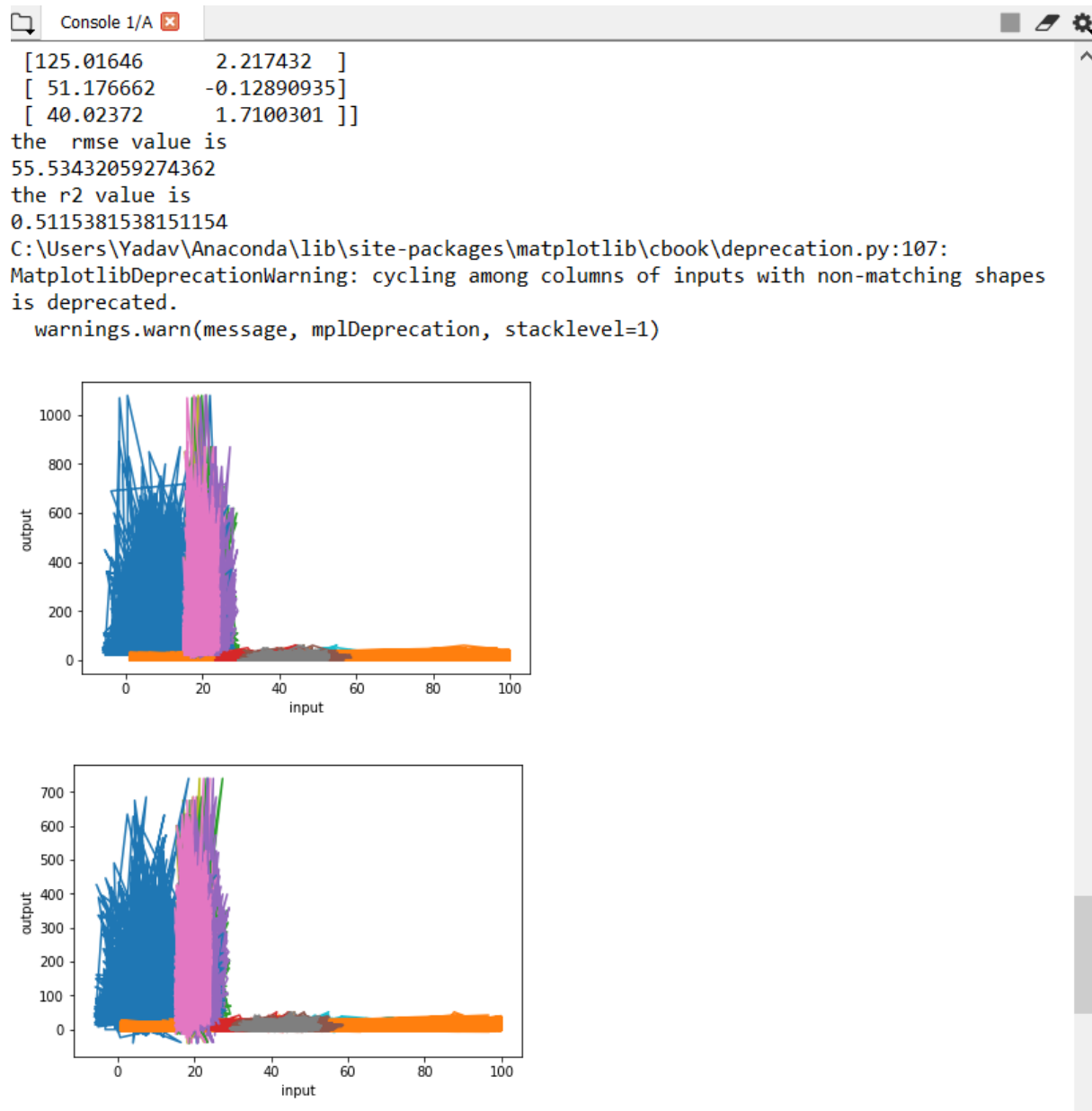


Fig 6.6

From the above graphs and accuracy we can determine that Random forest is the best model for regression of calculating multi values (appliance and lights)

7. CONCLUSION AND FUTURE SCOPE

By using our project one can easily determine which one is necessary algorithm in machine learning to use for energy consumption .The future scope of project is we wanted to have a interactive user interface and use better skills in order to have more accurate predictions as accuracy is about 60 percent.

BIBLIOGRAPHY

Book references

Luis M. Candanedo, Veronique Feldheim, Dominique Deramaix, Data driven prediction models of energy use of appliances in a low-energy house, Energy and Buildings, Volume 140, 1 April 2017, Pages 81-97, ISSN 0378-7788.

Web references:

1. https://www.tutorialspoint.com/python_pandas/
2. <https://www.machinelearningmastery.com>
3. [https://www.anaconda.com\(to download anaconda\)](https://www.anaconda.com(to_download_anaconda))
4. [https://www.uci.edu/Appliances/Energyprediction\(to obtain dataset\)](https://www.uci.edu/Appliances/Energyprediction(to_obtain_dataset))
5. <https://www.stackoverflow.com>