A Project Report on

SOLAR RADIATION FORECASTING

in partial fulfillment for the award of thedegree

Of

BACHELOR OF TECHNOLOGY
in
COMPUTER SCIENCE AND ENGINEERING

Submitted by

P. CHANDU SRUJAN	**160102167**
P. POOJA SOWMYA	**160102179**
N. V N L NIKHITHA	**160102153**
N. BHASKAR VAMSI	**160102151**

Under the esteemed Guidance of

Sri T V K P Prasad,
Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SRKR ENGINEERING COLLEGE (A)

ChinnaAmiram, Bhimavaram, West Godavari Dist., A.P.

[2019 – 2020]

SRKR ENGINEERING COLLEGE (A)

BONAFIDE CERTIFICATE

This is to certify that the project work entitled **"SOLAR RADIATION FORECASTING"** is the bonafide work of **"P. CHANDU SRUJAN- 160102167, P. POOJA SOWMYA - 160102179, N. V N L NIKHITHA- 160102153, N. BHASKAR-160102151"** who carried out the project work under my supervision in partial fulfilment of the requirements for the award of the degree of Bachelor of Technology in Computer Science and Engineering.

SUPERVISOR
Sri T V K P Prasad
Assistant Professor

SELF DECLARATION

We hereby declare that the project work entitled “SOLAR RADIATION FORECASTING” is a genuine work carried out by us in B.Tech., (Computer Science and Engineering) at SRKR Engineering College(A), Bhimavaram and has not been submitted either in part or full for the award of any other degree or diploma in another institute or university.

1. P. CHANDU SRUJAN	**160102167**
2. P. POOJA SOWMYA	**160102179**
3. N. V N L NIKHITHA	**160102153**
4. N. BHASKAR VAMSI	**160102151**

ABSTRACT

As world using fossil fuels for generating electricity. But in the next 100 years, these coal deposits will get depleted and should aware of it. These fuels are led to global warming and depletion of ozone layer which will causes a great threat to mankind. We are having a possibility of replacing these fossil fuels with the non-conventional resources. Solar energy is a renewable energy and will generate a large amount of electricity. The main goal is to switch to Solar energy and non-conventional resources instead of these fossil fuels and led to many practices for addressing to people. As world is looking for friendly environmental resources and the quantity of solar light strikes the earth per hour will meet the requirements of an humans for a year. As installing solar panels costs high, we are trying to find where the solar radiation is greatly produced. For this we considered NASA dataset to predict solar energy radiation. Along with the solar energy we are having other energy resources like Hydro energy, Nuclear energy, wind energy to generate electricity.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Solar energy can be available in anywhere in the planet and this energy is received from the sun by the Earth. The energy received from the sun will cause less harm and can able to convert to electricity. This usage of solar energy can gain more potential energy for natural resources and climate protection. So, this will route the future energy supply. There having many experiments performed on generating electricity from solar energy. It’s proved that small amount of electricity is produces during light exposure to electrolytic cell.

India is having large amount of solar energy for producing green electricity due to its primary geographical location. Due to some technical issues in measurements, the radiation data is not available in all locations. On the other side of coin, having the extensive networks of weather stations provide reliable long-term climate data which closely correlated with the solar radiation. As a result of which various solar radiation models have been developed with respect to climate data.When sunlight strikes the solar panels, it converts sun’s energy to Direct current electricity which can be sent to inventor and it converts to Alternating current electricity.

The generated solar energy greatly affected by unfavorable weather conditions. To predict the solar energy the data should be consistent. For prediction, we are having many algorithms. But the solar radiation prediction should require continuous data. For these continuous data, to get high accuracy, we are having the Time series prediction. Many machine learning algorithms do not have this capability, as they tend to be restricted to a domain that is defined by training data. So those kinds of algorithms will not suite for time series. Solar radiation prediction is a problem within time series prediction that have been received considerable attention, as such predictions can inform the expected yield from crop in a given year or the energy that can be generated from a solar panel.

As time series is a value of variable at equal interval of time. This time series is the understanding on the observed data to predict for the next. We are using these Time series applications in our daily life like sales forecasting, Yield and Economic forecasting etc.

2. PROBLEM STATEMENT

Today energy is something that is taken for consideration. Nonrenewable energy has a great impact on environment. As world looks for environment friendly resources, solar energy comes out as a clean energy resource. The main goal is to build a model that predict future values of solar radiation using previous data more accurately. This makes information regarding one's potential for switch to solar radiation that makes it available for wider audience and decrease of other fuels which will cause damage to the air pollution. The goal is to forecast the solar radiation and install solar panels to generate electricity and used for various purposes.

3. LITERATURE SURVEY

Manal marzouq et al., [1] had developed an ANN based model as it is difficult to obtain all the necessary data. In that, authors felt of considering the parameters like temperature, relative humidity and solar irradiance and so on will be the most available. And then tried to combine the above mentioned parameters for finding the relevance between them and also individually. Then he calculated the feasibility of incremental combination method and finds out the best model using the number of neurons in the hidden layer. A multilayer perceptron back propagation training algorithm with one hidden layer is used for planning daily values of solar irradiance. Authors used the measured data between 2009 and also 2015.

AaftaabMoosa et al., [2] proposed that predicting of clear sky Global Horizontal Irradiation with the collected data. The authors had used three machine learning algorithms to choose the best among them. The algorithms which they worked on Artificial Neural Network, Random forest and XGB (stands for Extreme Gradient Boosting). In Artificial Neural Network, Each parameter is taken as an input in the input layer and having many hidden layers with their each combination to produce Hourly GHI. And the next is Random forest, the main aim of using this algorithm is group of weak learners will give a robust learner if they combined. In this algorithm, large number of decision trees can be created. By additive principle can combine the base models by the choices. For obtaining the optimized values, they ran into several cross validation models and choose the best values which gives the accurate result. The final algorithm, XGB which works on Ensemble principle. Here main functionality is optimizing loss function at each and every stage.The parameters for overall controlling are categorized into Tree specific, Boosting and Miscellaneous parameters which influence every tree in the model, influencing boosting operation, influencing overall functioning in the model. After comparing the results of each model, they came with a conclusion that XGB is the best algorithm with minimum RMSE.

Mutaz et al., [3] proposed that accurate prediction of solar radiation is depends on the parameters selected. Authors completely used artificial Neural network models for forecasting solar radiation. They tested on twelve models with different models varying the input variables with Artificial Neural networks, NARX and MLP and Radial Basis Function back propagation learning algorithm. As per the calculations of Error rate, authors felt that

Radial basis Function is the best algorithm it can be used for 24 hours tool for predicting solar radiation when meteorological stations are not available. They have used New zealand database.

S. Mohanty et al. [4] proposed that Non Linear Auto regressive Network with Exogeneous Input for forecasting solar radiation. This is mainly used for predicting daily solar radiation with the parameters such as temperature, humidity, duration of sunshine. Marquardt Leven berg algorithm of NARX producing a maximum of regression coefficient. For predicting the time series issues, three models were used by the authors. The process of predicting output is done in the training process of series parallel architecture. For calculating the performances of the models, they used statistical analytics like Mean squared Error. Author collected the data of Bhuvaneswar between 2002- 2005 for predicting the solar radiation.

Chuang Liu et al. [5] proposed that knowledge of solar radiation is crucial for solar energy utilization. The authors applied the algorithm of Least squares (LS) – Support Vector Machines(SVM) for predicting monthly and annually average daily solar radiation in and around china. To calculate the performance of the model, the indices i.e, Root Mean Squared Error , Coefficient of determination, Mean Bias and Mean Absolute Bias Error are applied. China have taken data of 101 meteorological stations for predicting solar radiation. Out of them, 57 stations data is used for training and rest is for testing. They concluded that solar radiation is higher at the west of china compared to the east.

4. SOFTWARE REQUIREMENTS SPECIFICATION

4.1 PURPOSE

The purpose of this project is to know the wider audience regarding solar energy usage as the nonrenewable energy resources getting depleted and the to predict the solar radiation to produce electricity from natural sources.

4.2 SCOPE

The scope of this project is to forecast the solar prediction with minimum loss function and able to predict the solar radiation when there is only accurate data.

4.3 OBJECTIVES

The following are the objectives of the project.

1. Collecting data which affects the radiation prediction
2. Determine the best methodology to design to implement the proposed system.
3. Based on the techniques that are chosen from the literature survey, formulate requirements
4. Process the implemented code to check whether desired requirements are meet or not.

4.4 EXISTING SYSTEM

The following are some existing systems related to our domain.

- Prediction of solar radiation With Artificial Neural Networks
- Statistical methods for predicting solar radiation
- Solar Radiation forecasting using Random Forest

4.5 PROPOSED SYSTEM

- The system focuses on predicting the solar radiation using Holt winter, Seasonal Auto Regression and Moving average and Auto Regressive Integrated and Moving Average
- So that we are able to train the machine from past to make future prediction.
- The proposed system will be able to calculate the relation between the past values to predict the future data.

4.6 REQUIREMENTS

4.6.1 Functional Requirements

- First, we pre-process the data by filling the missing values, Conversions and Preprocessing.
- Validating and testing the data
- Predicting the solar radiation.

4.6.2 Non-functional Requirements

4.6.2.1 Robustness

Robustness is the ability of a computer system to cope with errors during execution and cope with erroneous input. The system should be able to handle the errors during the execution process.

4.6.2.2 Performance

Performance is measured in terms of the output provided by the application. The performance is increased than the models used in the existing system.The system should respond within 15seconds from the user's request. The accuracy of the prediction should be greater than 70 per cent.

4.6.2.3 Costeffectiveness

Thetotalcostrequiredfordevelopingthesystemislesswhencomparedtotheexistingsystems as we used open sourceenvironment.

4.6.2.4 Error handling

In case of any error, the system should display a meaningful error message to the user, such that user can correct his error.

4.6.2.5 Quality issues:

Data provided by the system is based on the dataset given by the user.

4.6.2.6 HardwareRequirements

- Processor : Intel dual coreprocessor
- RAM : 8GB
- DiskSpace : 500GB

4.6.2.7 SoftwareRequirements

- ProgrammingLanguage : Python
- IDE :Anaconda
- UML : Rationalrose
- OperatingSystem : Windows 8 orHigher

5. SYSTEM DESIGN

System design is that the method of characterizing the design, interfaces, components, modules and information to satisfy specified needs for a system. Someone might watch it because the utilization of systems theory to enhance development. There's some overlay and synergism of system analysis with the disciplines, system's design and system's engineering.

5.1 SYSTEM ARCHITECTURE

System Architecture describes the general framework of the system and therefore the paths in which the structure gives the abstract unity. The below figure 5.1 represents design of the system of the give project. It includes principally four components as shown within the diagram as Input, Pre-processing, Regression, Evaluation and therefore the corresponding sections underneath every part.

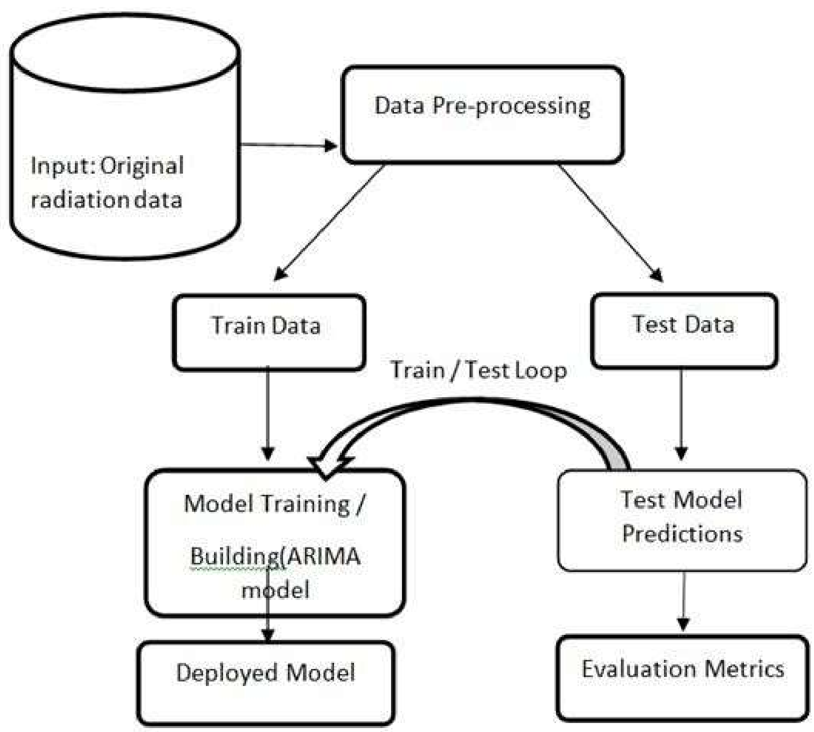

Figure 5.1: system architecture

5.2 UML DESIGN

The Unified Modelling Language could be a basic modelling language. The key of UML is to define a customary way to visualize the means a system has been designed. It's quiet the same as blueprintsemployed in alternative fields of engineering. It's not a programming part, it's rather defines a visual part.

5.2.1 USECASE DIAGRAM

Use case diagrams are typically defined to be thebehavior diagrams used to describe a group of actions (use cases) that some system or systems (subject) will perform unitedly with one or more additional external users of the system (actors).

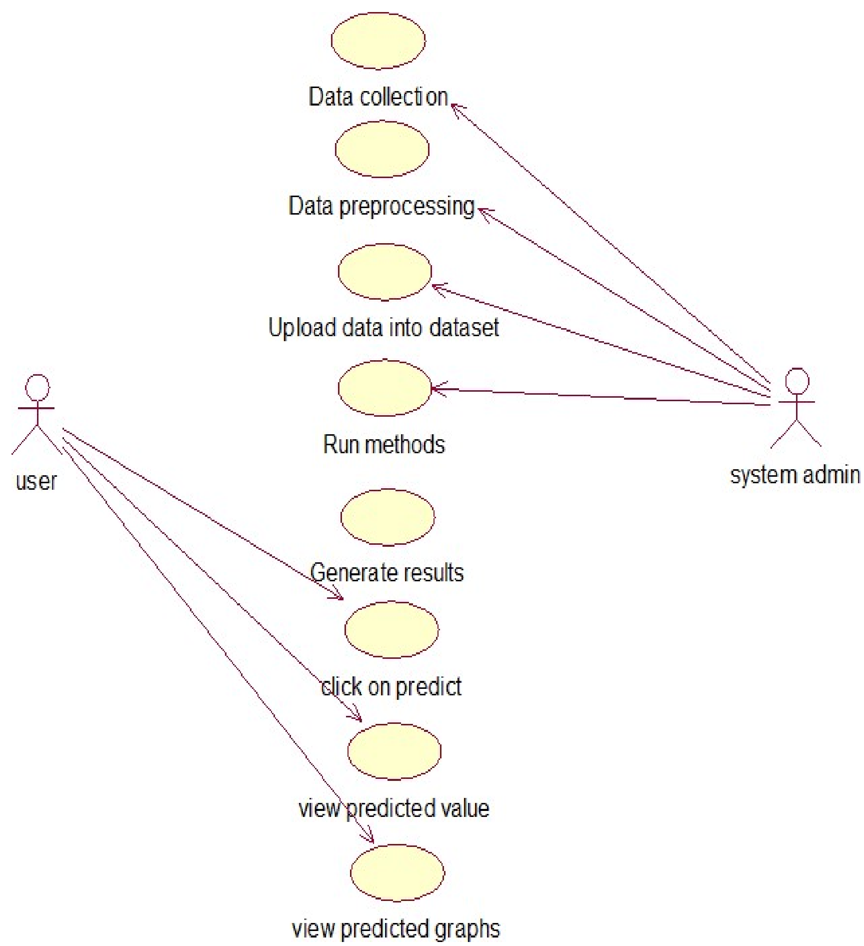

Figure 5.2: Use case diagram

5.2.2 CLASS DIAGRAM

The below figure 5.2 is the class diagram of the project. Here algorithm class is generalized into 3 classes namely Holt winter, SARIMA, ARIMA as these are the algorithms used in the project. These classes will also have various multiplicities between them.

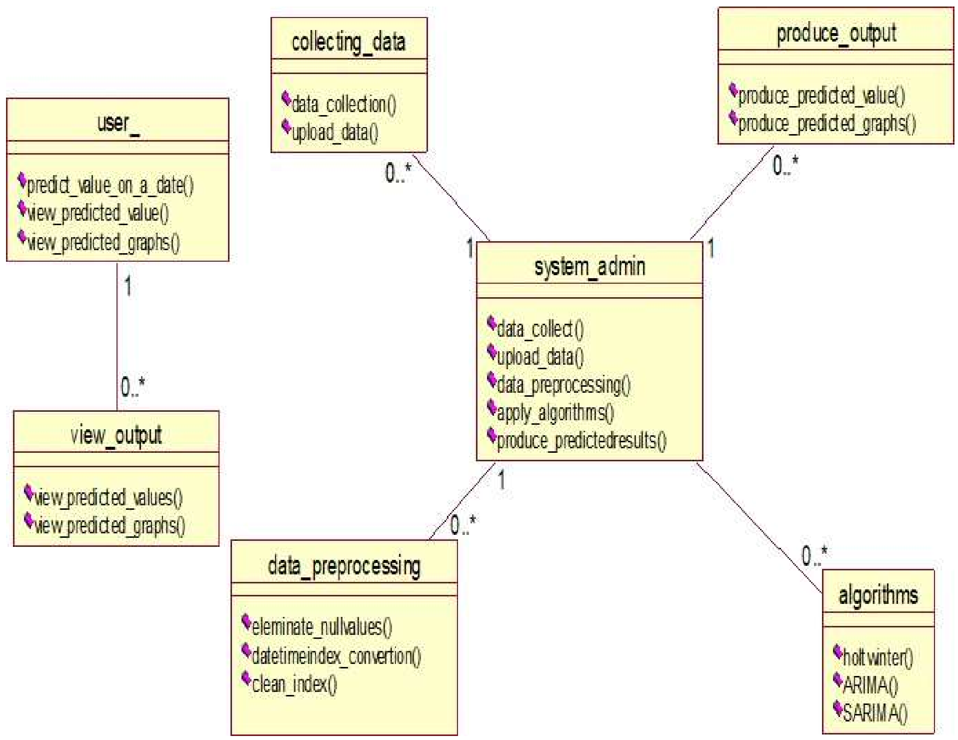

Figure 5.3: Class diagram

5.2.3 SEQUENCE DIAGRAM

The figure 5.3 is the sequence diagram of the project. Here it includes a step by step process of our project implementation steps. Here there are two actors. They are user and the admin. The objects depicted in the figure are user interface, system and admin interface.

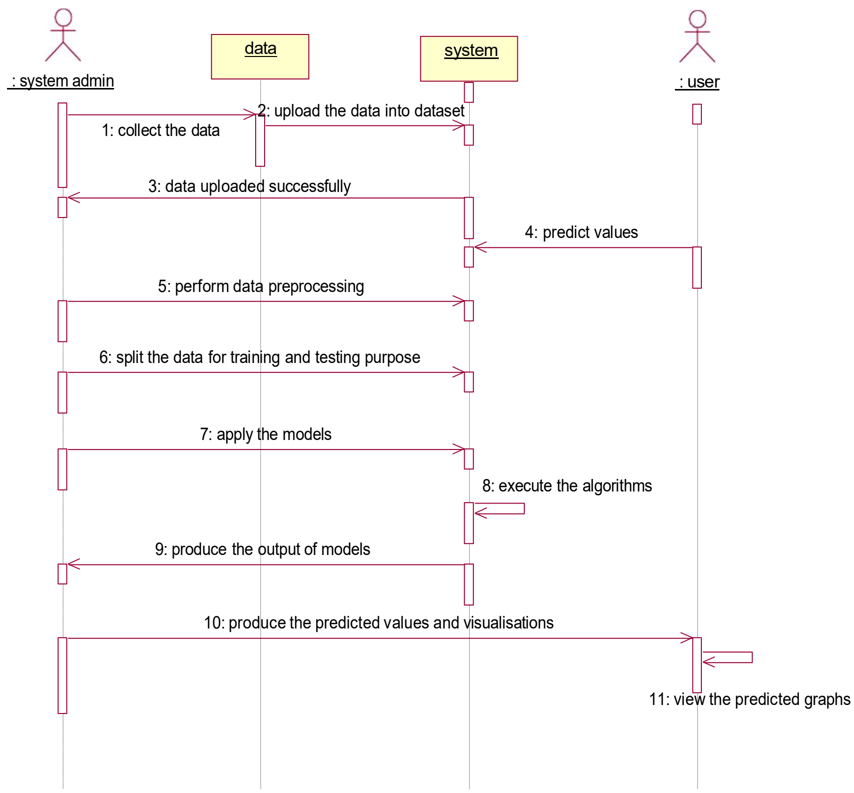

Figure 5.4: Sequence diagram

5.2.4 ACTIVITY DIAGRAM

The figure 5.4 is the activity diagram of the project. It consists of one start stage and one stop stage and it includes a step by step procedure of predicting the solar radiation. Here the user first wants to predict the price. Then the admin will apply the algorithms and produce the result to the user.

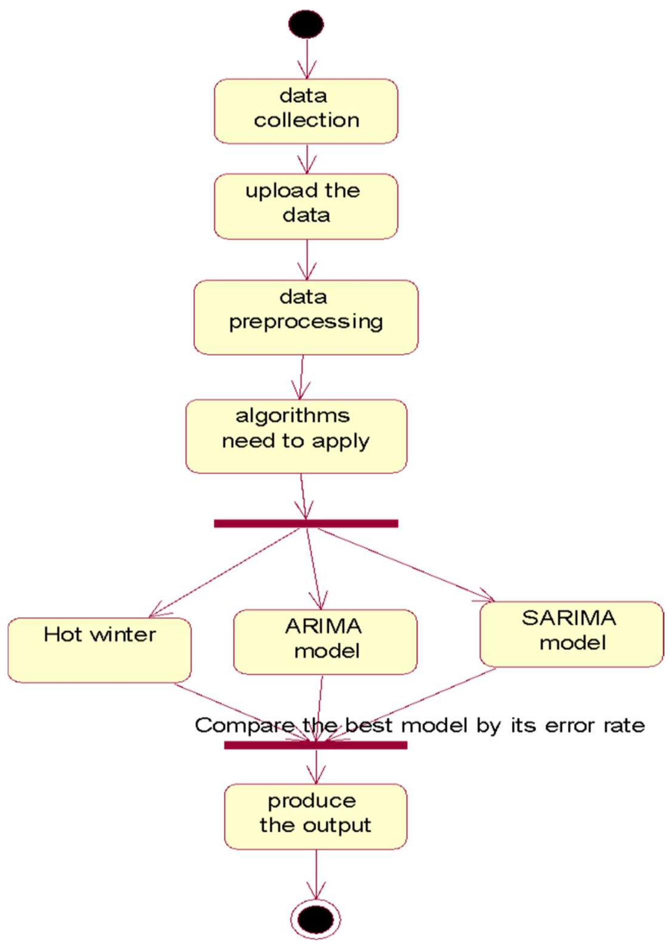

Figure 5.5: Activity diagram

5.2.5 STATE CHART DIAGRAM

A state diagram is in a different way of expressing dynamic data about a system. It's basically used to describe the externally visible behavior of a system or of a single object.

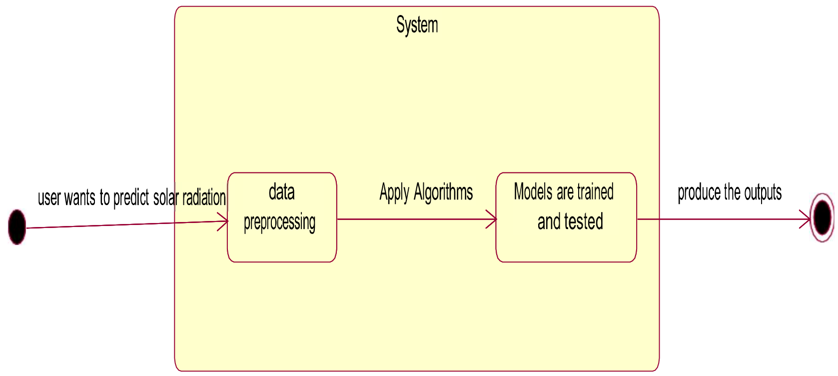

Figure 5.6: State chart diagram

6. METHODOLOGY

6.1 DATASET

The data is collected from a website which contains day to day solar radiation details. The columns include UNIX time, Data, Time, Radiation, Temperature, Pressure, Humidity, Wind direction, Speed, Sunrise time and Sunset time.

6.2 ALGORITHMS

The following are the various algorithms that are used to predict the price of Bitcoin.

6.2.1 Auto Regressive Integrated Moving Average (ARIMA)

ARIMA models is used to describe the autocorrelations in the data. It is the generalization of simpler Auto Regressive Moving Average and adds the motion of integration. Differencing is the subtraction of the current value from the previous and this is used to transform a time series into stationary. The AR in ARIMA indicates that the evolving variable is regressed on its own lagged (i.e., prior) values. The MA part indicates that the regression error which is a linear combination of error terms whose values occurs at various times in the past. The I (integrated) indicates that the data values have been replaced with the difference of their values and the previous values and this differencing process may have been performed more than once.

ARIMA is a statistical regression model. It is used in time series forecasting applications. ARIMA makes predictions by considering the lagged values of a time series into stationary. The model is originated from the autoregressive (AR) and moving average (MA) models and with their combination known as ARMA. For making predictions with the ARIMA model, we had to follow a step by step procedure to be able to feed the data to the ARIMA model. It is essential to analyze to build any kind of time series model. Let $y_1,...,y_t$ define a unit variant time series, with each y_i belongs to R. An AR model of order p, denoted as AR(p), the current(predicted)value of a time series is expressed as follows:

$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + ... + \varphi_p y_{t-p} + \epsilon_t \qquad (1)$$

Where c is a constant, $\varphi_1,...,\varphi_p$ are the parameters of the model and ϵ_t is a variable producing white noise. The order q, denoted as MA(q), expressing the time series as a linear combination of its current and previous values. More concretely, MA(q) is defined as follows:

$$y_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + ... + \theta_q\epsilon_{t-q} \qquad (2)$$

where μ is the mean of y, 1, . . . , q are the model parameters and i corresponds to a random variable producing white noise at time point i. Representing AR and MA processes p and q, an ARMA(p,q) process can be defined. Generally, the process of differencing is the transformation of a time series to a stationary process by eliminating trends and seasonality, so that Mean is stabilized. This is achieved simply by calculating the difference between the consecutive observations. An ARIMA model with parameters p,d,q, denoted as ARIMA(p,d,q), performs d differentiations to the original time series values, i.e., yt is converted to y. By using the above mentioned methodologies, we can able to predict the solar radiation.

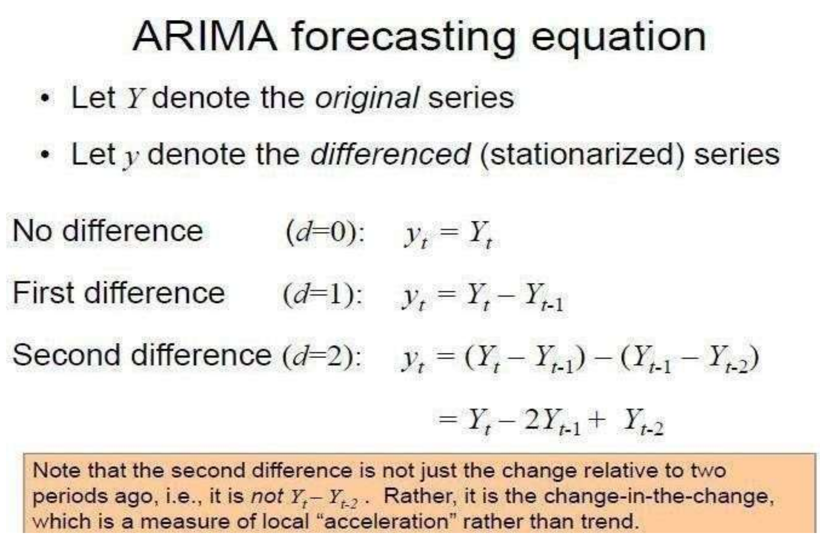

Figure 6.1: ARIMA Model Difference Equations

6.2.2 Seasonal Auto Regressive Integrated Moving Average(SARIMA)

Seasonal Autoregressive Integrated Moving Average (SARIMA) method for time series forecasting which is used for univariate data containing trends and seasonality. Specifically having the parameters to specify the autoregression (AR), differencing (I) and moving average (MA) for the seasonal component of the series and the parameter which is added is for the seasonality period. Seasonal differencing is similar to regular differencing. In regular differencing we subtract the consecutive terms where as in seasonal differencing we subtract the value from previous season.

A seasonal ARIMA model uses differencing at a lag equal to the number of seasons to left out the additional seasonal effects. So, the representation of the model is as SARIMA(p,d,q)x(P,D,Q), where, P, D and Q are SAR, order of seasonal differencing and SMA terms respectively and 'x' is the frequency of the time series.

$$\Phi P(B\ s\ )\varphi(B)\nabla D\ s\ \nabla dZt = \alpha + \Theta Q(B\ s\ )\theta(B)at.$$

6.2.3 Holt Winter Algorithm

The Holt-Winter algorithm is a type of time series forecasting method. We use time series forecasting methods because they predict future values using previous data more accurately. The Holt-Winter method comprises of few smoothing equations in order to smooth a time series, exponential smoothing methods regarding this algorithm are single, double, triple exponential smoothing — the first type of smoothing is mainly used for predicting data without trend or any seasonal pattern, second type of smoothing is used to predict data whenever a trend is present, third type of smoothing is used whenever the trend or the seasonalities (or both)are present. In this algorithm we use m for denoting frequency of seasonality. This method consists of two variations that are different types of seasonal components which are additive and multiplicative methods. The additive method is mainly used whenever seasonal differences are nearly constant throughout series, while the multiplicative method is useful whenever seasonal differences are changed corresponding to level of series. Every year, seasonal components will add up to approximately zero. Using multiplicative method, the seasonal components are viewed in form of percentages, the series is adjusted seasonally by dividing by the seasonal components. Every year, the seasonal components increases to nearly m..

7. TESTING AND RESULT ANALYSIS

7.1 TESTING

In machine learning, testing can be done in two ways. One way is to divide the total dataset into certain proportions of training data and testing data. Then with the help of training data, various machine learning models are trained. Then after training the models, the trained models can be able to behave like an experienced model prior to the case of training and correct dataset. Then with the help of this trained models and the initial testing data, we can test the system.

The other way is giving a new data tuple to the trained models and identifying the result. But actually, this is a bit complex task. Because after loading the dataset into the machine learning kernel, the data is transformed into various forms. After various pre-processing methods like scaling, normalization, log transformations etc. the data is transformed abruptly. Then the transformed data is used to train the machine learning models. For testing, we need to take care that the same pattern of data tuple is given, which is a monotonous task. Because, explicitly we need to perform those operations that are done on the original data. After this only, we need to provide this tuple to the trained models and get an output.

These are the ways to test a machine learning model. Then based on the tested output and the actual testing data, we can say how the model is performing. There are number of performance metrics to calculate the testing performance of trained models. The following is the testing pattern of our project.

7.1.1 Dataset

The actual dataset used in our project is named as "solarprediction.csv". The following are the varioustesting scenarios that occurred while testing this component.

Test case	**Expected result**	**Obtained result**	**Result**
Load the data using "solarprediction.csv"	Data is stored in a data frame	Data is stored in a data frame	Success
Load the data using "solardetection.csv"	Invalid solardetection.csv file	Invalid solardetection.csv file	Success

Table 7.1: Testcase for loading dataset

7.2 OUTPUT PLOTS

The following are various outputs of the predictions after applying the machine learning algorithms.

7.2.1 ARIMA

The following figure describes the result of ARIMA model

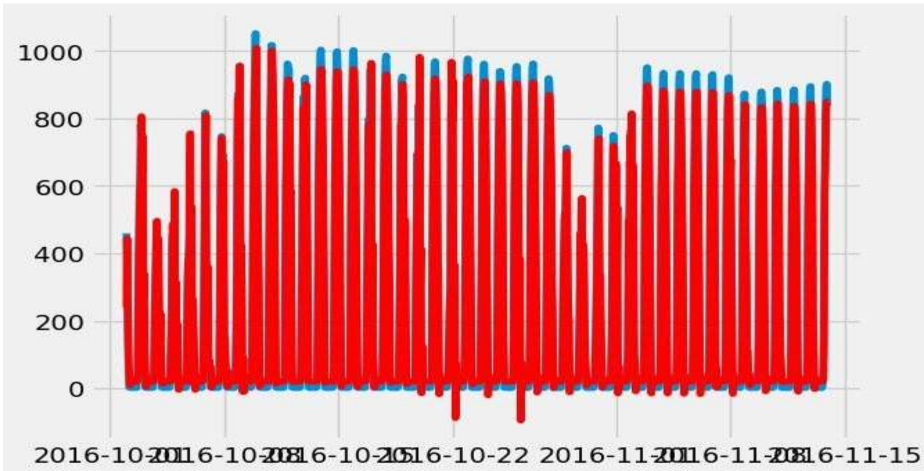

Figure 7.1: Result of ARIMA

The following are the metrics for the ARIMA model

ARMA Model Results

Dep. Variable:	Radiation	No. Observations:	1032
Model:	ARMA(1, 1)	Log Likelihood	-6291.946
Method:	css-mle	S.D. of innovations	107.417
Date:	Sat, 02 Feb 2019	AIC	12591.892
Time:	02:00:25	BIC	12611.649
Sample:	10-02-2016	HQIC	12599.390
	- 11-13-2016		

	coef	std err	z	P>|z|	[0.025	0.975]
const	239.7954	41.564	5.769	0.000	158.331	321.259
ar.L1.Radiation	0.8826	0.015	58.571	0.000	0.853	0.912
ma.L1.Radiation	0.4695	0.023	20.587	0.000	0.425	0.514

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	1.1330	+0.0000j	1.1330	0.0000
MA.1	-2.1299	+0.0000j	2.1299	0.5000

Figure 7.2: AR,MA values of ARIMA

7.2.2 SARIMA

The following figure describes result of SARIMA model.

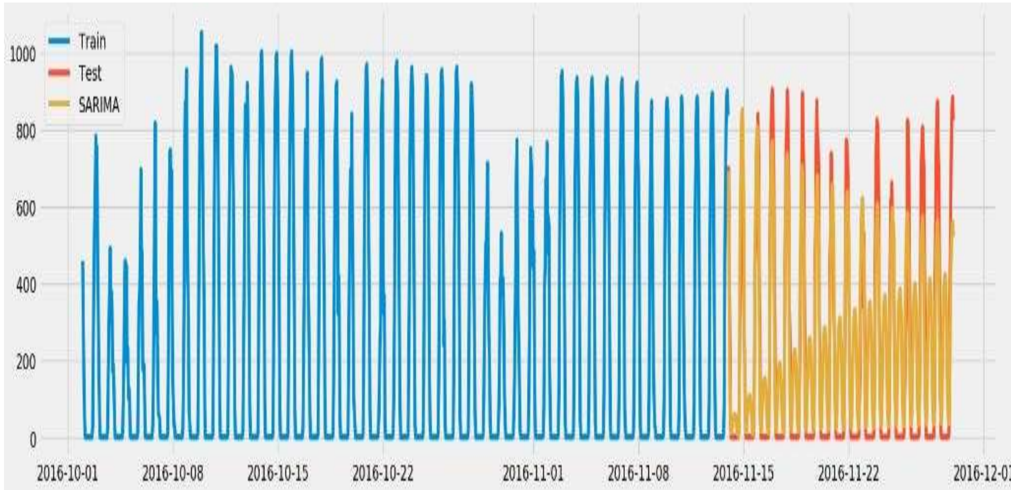

Figure 7.3: Result of SARIMA

The following are the metrics for the SARIMA model

Dep. Variable:	Radiation	No. Observations:	1032
Model:	SARIMAX(1, 0, 1)x(1, 1, 0, 12)	Log Likelihood	-5968.157
Date:	Sat, 02 Feb 2019	AIC	11948.314
Time:	02:00:24	BIC	11977.880
Sample:	10-02-2016	HQIC	11959.540
	- 11-13-2016		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	-0.2406	4.630	-0.052	0.959	-9.315	8.834
drift	0.0011	0.008	0.134	0.894	-0.015	0.017
ar.L1	0.6872	0.022	31.778	0.000	0.645	0.730
ma.L1	-0.1115	0.029	-3.883	0.000	-0.168	-0.055
ar.S.L12	-0.9986	0.003	-393.645	0.000	-1.004	-0.994
sigma2	6770.5516	134.510	50.335	0.000	6506.917	7034.186

Ljung-Box (Q):	289.51	Jarque-Bera (JB):	3340.12
Prob(Q):	0.00	Prob(JB):	0.00
Heteroskedasticity (H):	0.97	Skew:	-0.01
Prob(H) (two-sided):	0.79	Kurtosis:	11.87

Figure 7.4: AR, MA values of SARIMA

7.2.3 Holt Winter

The following figure describes result OF Holt Winter model

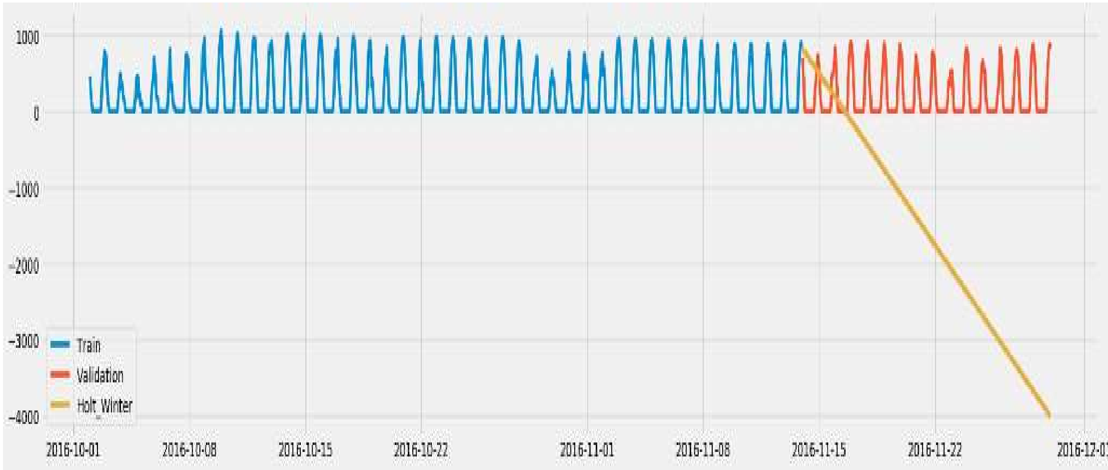

Figure 7.5: result of Holt Winter

8. CONCLUSION

Among the three models, the ARIMA and SARIMA are the best Models indicated the best execution. The working of both the models will execute in a different way based on the data set provided. If the data set containing seasonal values then SARIMA would be the best execution. As the dataset we preferred does not contain those seasonal values ARIMA processed the best execution. As the Sun is one of the source for generating energy which will cause less pollution and if the people would aware of it, then usage of fossil fuels will get reduced.

REFERENCES

[1]M. Marzouq, H. El Fadili, Z. Lakhliai and K. Zenkouar, "A review of solar radiation prediction using artificial neural networks," 2017 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS), Fez, 2017, pp. 1-6.

[2] Moosa, H. Shabir, H. Ali, R. Darwade and B. Gite, "Predicting Solar Radiation Using Machine Learning Techniques," 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2018, pp. 1693-1699.

[3] Mutaz, Turi and Aziz Ahmad. “Solar Radiation Prediction Using Radial Basis Function Models.” 2015 International Conference on Developments of E-Systems Engineering (DeSE) (2015): 77-82.

[4] S. Mohanty, P. K. Patra and S. S. Sahoo, "Prediction of global solar radiation using nonlinear auto regressive network with exogenous inputs (narx)," 2015 39th National Systems Conference (NSC), Noida, 2015, pp. 1-6

[5] Fangping Deng, Gaoli Su, Chuang Liu and Zhengxing Wang, "Prediction of solar radiation resources in China using the LS-SVM algorithms," 2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE), Singapore, 2010, pp. 31-35,