Solar Energy Prophecy

N V Ganapathi Raju, Gidde Akhil Reddy

Department of Information Technology

Gokaraju RangaRaju Institute of Engineering and Technology, Hyderabad

Abstract: In this modern era the data generated is huge and diverse. An organized date is required to ensure efficient usage of the solar energy. Progressive analytics is achieved by analyzing and predicting future events. Solar energy is the most profitable source of renewable energy and accurate prophecy of solar energy will improve the quality of service. The production of solar energy depends mainly on weather conditions which generally change over time. This work presents forecasting the solar energy. Prophecy of Solar energy using the statistical modelling method named ARIMA time series model. This model is used to understand past data and predict future data. With the wealth of historical solar energy data, solar radiation prophecy can be calculated. After detailed study on the Arima model and output results have proved that this Arima model is best suited to forecast the solar energy. These predicted results will help to manage to use quality services of the solar energy.

Index Terms: Prediction, Forecasting, Arima, time series.

1.**Introduction**

The demand for the electricity is huge and its demand is increasing constantly day by day. The usage of fossil fuels is limited with many disadvantages and the

usage of this is decreased almost. The demand for the energy is going to be massively huge in near future. All these circumstances have created a demand for usage of renewable energy Providing an accurate amount of required energy is one important aspect to be considered. Among all the renewable energy sources the solar energy is most promising and sustainable energy source.
It depends mostly on many factors such as temperature,
humidity, atmospheric status and location. The desired amount of solar energy is forecasted and this will help the power supply management to supply required amount of energy. Different methods have been adopted in the literature for predicting solar energy.

The time series forecasting model which uses historical measured data (ARIMA). Consequently, the time series data had to be in stationary. If it is not in stationary it has to be converted to stationary so that we can work with ARIMA model the most well-researched models in this arena is ARIMA, an acronym for Autoregressive Integrated Moving Average. Arima Model will work with time series to understand past data which usually helps to better predict future data. This model usually assumes the future values are based on the past data. Instead of examining actual values, it examines the difference between the values in the time series.AR stands for Autoregression which displays a changing variable that regresses when it gets lagged on its own values. I mean Integrated which tells that the data values will be changed when the difference between the previous values and the current values.MA Stands for Moving average which shows the dependency between an observation and an error in lagged values when a model like moving average is applied. ARIMA model has parameters and each parameter has its own notation. The notation is described with p, d and q. p describes number of lagged observations. d indicates the number of times the observations are differenced. q indicates order of moving average. In this model the stationary is achieved by differencing the data

2.LITERATURE SURVEY

Related work done by Mariam AlKandari, Imtiaz Ahmad
From my findings they have worked on a model that adds both Machine Learning (ML) methods with theta statistical methods for high accuracy in predicting solar power production from the valuable renewable energy plants. It includes LSTM, GRU, Auto-LSTM and Auto GRU. They combined four methods such as simple average, weighted average with linear and non-linear, finally the inverse approach. They have included structural and data diversity to enhance the accuracy of MLSHM.Their results shows that accuracy is high with proposed MLSHM when compared with regular models.

Related work done by C. Vennila, Anita Titus.
They has proposed a hybrid model that adds machine learning and other statistical Hybrid Model (MLSHM)approach which is used for predicting generation of solar energy. This method has proved that it showed reduced placement cost. The ensemble model has outperformed over the traditional models. Their future scope was to improve all aspects by implementing several deep learning models.

Related work done by Jose Manuel Barrera Alejandro Reina Reina Their approach is on ANN structures that tested in various topologies provide solar energy predictions Uses data obtained from an open data source. Their solution provides that open data can be trained for different locations so that this will provide abstract layer on energy production when compared rather than radiation data. Theirs results shows that they have achieved mean squared error of 0.040 when it is compared to mean squared error 0.555 in other literature.

Related work done by Emil Isakson, Mikal Karpe Conde
they have experimented comparing the time series models with different ML methods for the prediction of solar energy in five different locations across Sweden. They figured out that the working with the time series data is tough when working with non-stationary time series and mentioned machine learning models works better in this criterion and concluded Artificial Neural Networks and Gradient Boosting Regression Trees are the best working models.

3.Methodology

In this work, Solar data set is considered for predicting the solar energy. The dataset for forecasting is obtained with the help of industry management from Kaggle community. The dataset consists of 3153 rows and 8 attributes.

3.1 Data Cleaning and Data Preparation:

Initially data cleaning and data preparation is performed. So that, it could be in a usable format by fixing structures, removing outliners, handling missing data, removing all the duplicates, irrelevant observations. The dataset has 15-minute time interval, it is been converted into daily format so that forecasting can be performed easily. DC power is been considered in this case. Unwanted attributed are removed from the dataset and grouping of data is performed with respect to date and max yield is considered from dataset.

Date	DAILY_YIELD
2020-05-15	5754.000000
2020-05-16	6292.000000
2020-05-17	7045.000000
2020-05-18	4998.000000
2020-05-19	6449.000000
2020-05-20	8249.000000
2020-05-21	7243.000000
2020-05-22	6848.000000
2020-05-23	7966.000000
2020-05-24	7537.000000
2020-05-25	8268.000000
2020-05-26	7456.428571
2020-05-27	6164.000000
2020-05-28	7977.000000
2020-05-29	7564.000000
2020-05-30	6754.000000

Fig 1: Snapshot of the solar energy dataset

3.2 Exploratory Data Analysis

This Analysis is performed to identify the pattern in the dataset. This summarizes the important characteristics using data visualization method. With the help of this analysis, we could figure out some valuable information from the chart before the operations is performed.

Fig 2: Exploratory data analysis graph

3.3 Stationary Test- Augmented Dicky Fuller Test

The data need to be stationary to perform this operation. Here we are testing the dataset by mean and standard variation as the data is generating in timeseries. Initially Rolling Statistics is determined. This is performed to understand and asses the model stability over time. Augmented Dicky Fuller Test helps to know whether the given time series is stationary or not stationary. In this hypothesis test implied with a null and alternative hypothesis and therefore a test statistic is calculated and the p values are reported. If the ADF statistic is lesser than the 5% of the critical value then it is said that the timeseries is stationary.

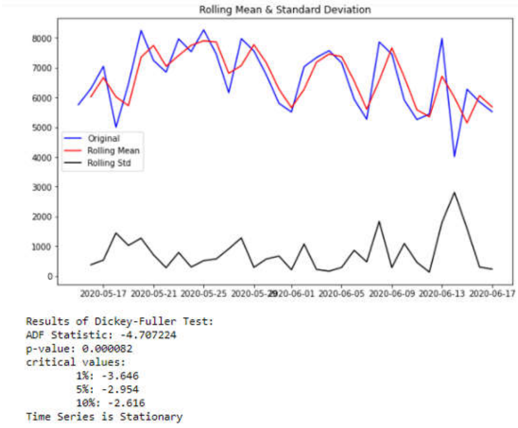

Fig 3: Rolling mean and Standard Deviation chart and results of dicky fuller test.

3.4 Autocorrelation and Partial correlation

Now the data is been split so that the training and testing can be performed. The ACF and PACF plots are used to get p and q values to send in to the ARIMA Model. The ACF gives the lagged values in the time series. PACF will find the correlation of the residuals with the next lag value.Hyper parameter tuning is performed.

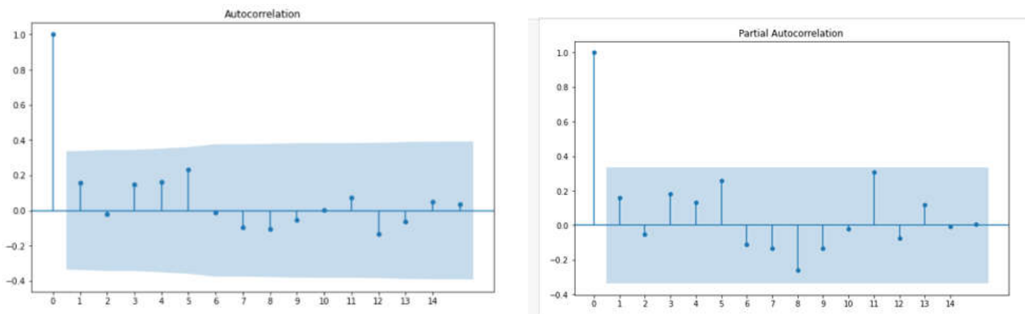

Fig 4. Autocorrelation and partial autocorrelation plots

3.5 Hyper Parameter Tuning

Here, rather than selecting the parameters, we will let the machine decide to select the optimal parameters. Ideally, the machine was asked to perform this scan and upon operation it selects the best model architecture which suits. The parameters are called as hyperparameters Hyperparameter tuning is performed for Arima model and the best model is considered for the parameters 0, 0,2.

```
Performing stepwise search to minimize aic
 ARIMA(0,0,0)(0,0,0)[0]             : AIC=556.170, Time=0.01 sec
 ARIMA(1,0,0)(0,0,0)[0]             : AIC=464.611, Time=0.00 sec
 ARIMA(0,0,1)(0,0,0)[0]             : AIC=inf, Time=0.01 sec
 ARIMA(2,0,0)(0,0,0)[0]             : AIC=465.844, Time=0.01 sec
 ARIMA(1,0,1)(0,0,0)[0]             : AIC=inf, Time=0.03 sec
 ARIMA(2,0,1)(0,0,0)[0]             : AIC=inf, Time=0.03 sec
 ARIMA(1,0,0)(0,0,0)[0] intercept   : AIC=450.188, Time=0.00 sec
 ARIMA(0,0,0)(0,0,0)[0] intercept   : AIC=449.307, Time=0.00 sec
 ARIMA(0,0,1)(0,0,0)[0] intercept   : AIC=447.386, Time=0.02 sec
 ARIMA(1,0,1)(0,0,0)[0] intercept   : AIC=448.466, Time=0.02 sec
 ARIMA(0,0,2)(0,0,0)[0] intercept   : AIC=446.102, Time=0.02 sec
 ARIMA(1,0,2)(0,0,0)[0] intercept   : AIC=448.626, Time=0.02 sec
 ARIMA(0,0,3)(0,0,0)[0] intercept   : AIC=448.009, Time=0.03 sec
 ARIMA(1,0,3)(0,0,0)[0] intercept   : AIC=449.982, Time=0.03 sec
 ARIMA(0,0,2)(0,0,0)[0]             : AIC=inf, Time=0.02 sec

Best model:  ARIMA(0,0,2)(0,0,0)[0] intercept
Total fit time: 0.280 seconds
```

Fig. 5 Hyperparameter tuning best model

3.6 Prediction Accuracy and Test Results

Initially we are training the dataset, then testing the data and prophecy of the dataset. From 30days of available data, predicting the next three days.

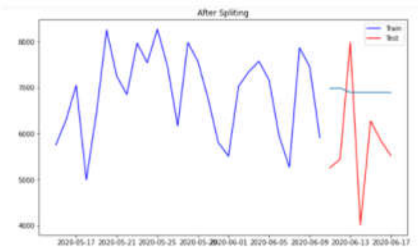

Fig 6 :test and train results

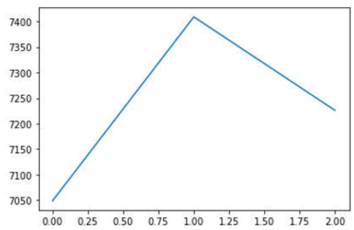

Fig 7: Prediction chart of next three days

4.CONCLUSION

In this project, we have predicted the solar energy based on the previous data using time series model called ARIMA. This model, we can forecast a mape of

28.4 percent which is good in range in time series prediction analysis. The efficiency of the model needs to be still improved. To improve the accuracy, I would like to implement RNN (Recurrent Neural Network) algorithm sequentially, in future, I want to compare the present model with LSTM (Long Short-Term Memory).

REFERENCES

1. Barrera, Jose & Reina Reina, Alejandro & Mate, Alejandro & Trujillo, Juan. (2020). Solar Energy Prediction Model Based on Artificial Neural Networks and Open Data. Sustainability. 12. 6915. 10.3390/su12176915.

2. Chaturvedi, D. & Singh, Isha. (2016). Solar Power Forecasting: A Review.

3. C.Vennila,AnitaTitus,T.sriSudha,U.Sreenivasulu,N.PanduRangaReddy,K .Jamal,P.jagadesh,Assefa Belay. Forecasting Solar Energy Production using Machine Learning.

4. Wang Buwei, Che Jian Feng, Wang Bo. A Solar Power Prediction Using Support Vector Machines Based on Multi-source Data Fusion

5. Alireza Zendehboudi, M.A. Baseer, R. Saidur. Application of support vector machine models for forecasting solar and wind energy resources
6. Mariam Alkandari, Imtiaz Ahmed. Solar Power generation forecasting using ensemble approach based on deep learning and statistical methods
7. Imane Jebli, Fatima-Zahra Belouadha , Mohammed Issam Kabbaj, Amine Tilioua. Deep Learning based Models for Solar Energy Prediction
8. Chih-Feng Yen, He-Yen Hsieh, Kuan-Wu Su1, Min-Chieh Yu Solar Power Prediction via Support Vector Machine and Random Forest
9. EMIL ISAKSSON MIKAEL KARPE CONDE .Solar Power Forecasting with Machine Learning Techniques
10. https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/
11. https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/
12. https://otexts.com/fpp2/arima.html