2021 International Conference on Artificial Intelligence and Machine Vision (AIMV) | 978-1-6654-4211-4/21/$31.00 ©2021 IEEE | DOI: 10.1109/AIMV53313.2021.9670909

Recommendation and Prediction of Solar energy consumption for smart homes using machine learning algorithms

Anish Dhage
Department of Computer Engineering
Pune Institute of Computer Technology
Pune, India
anish.dhage@gmail.com

Apoorv Kakade
Department of Computer Engineering
Pune Institute of Computer Technology
Pune, India
apoorv.kakadeg@gmail.com

Gautam Nahar
Department of Computer Engineering
Pune Institute of Computer Technology
Pune, India
gtnahar2@gmail.com

Mayuresh Pingale
Department of Computer Engineering
Pune Institute of Computer Technology
Pune, India
mayureshpingale1156@gmail.com

Sheetal Sonawane
Department of Computer Engineering
Pune Institute of Computer Technology
Pune, India
sssonawane@pict.edu

Archana Ghotkar
Department of Computer Engineering
Pune Institute of Computer Technology
Pune, India
aaghotkar@pict.edu

***Abstract*—Solar Energy is a principal source of renewable energy generation. Solar intensity is directly proportionate to solar power generation and it is highly reliant on the weather. A model is proposed that predicts the amounts of solar radiation produced using weather information implemented using various machine learning techniques such as Gradient boosting, SVM, etc. The results allow us to make effective energy consumption plans for smart homes with efficient utilization of solar energy which may provide several economic benefits. Additionally, accurate forecasts would make users more prepared to switch between conventional and renewable sources as required. A comparison study is performed with various machine learning models to determine the best method for building a prediction model. The groundwork for constructing models that could be dispatched to various regions is laid out that will incorporate that geographic location's weather data, and output accurate solar intensity predictions for that area. Furthermore, a recommendation system is proposed for the consumption of thus predicted energy.**

***Index Terms*—renewable energy, smart grid, solar power generation prediction, photo voltaic power, neural networks, machine learning; Gradient boosting, recommendation system**

I. Introduction

The shortage of energy has driven the development of alternative, sustainable, and reliable power sources. Solar energy has become popular due to ease of access. It is inexhaustible and considered a major source of power generation. Solar energy can be used with the help of technologies such as solar heating, photovoltaics, solar thermal energy, etc. Almost, all governments are providing subsidiaries to increase the use of solar energy. The advantage of this technique is that it will eliminate the dependency on non-renewable resources along with a reduction in the rate of pollution.

Solar generation capability is undergoing an exciting increase in India, with its capacity having grown 8 folds from 2,650 MW on 26 May 2014 to over 20 GW on 31 January 2018. The country's solar installed volume touched 31.696 GW on 31 October 2019[1].

In this paper, we present that predicting solar power supports a different set of difficulties for small-to-medium-sized solar deployments than predicting it for huge solar farms. Specifically, the position of massive solar farms is meticulously chosen to be in open places that reduce occlusions, which enables installers to maximize solar production by accurately tuning the orientation of the panels or using "trackers" that continuously adjust the inclination of the panels to follow the sun. Additionally, industrial solar farm operatives regularly wash the panels to keep them dust-free to preserve maximum solar output. At the same time, they also have the professional expertise and resources to correctly design and attune custom models to predict solar energy.

Considering solar energy now only presents a small fraction of grid energy, the lack of planning does not currently pose a problem. But, as solar penetration grows, efficient planning by apartments and utilities will be more crucial to maintain grid stability. In this paper we aim to predict solar energy generation for the next 48 hours or any given time.

hours or any given time. Rooftop solar is getting increasingly popular. However, a precise forecast of solar energy is significant for completely exploiting its advantages. We introduce machine learning methods to predict solar intensity from openly available weather estimates. The predicted solar intensity values can be used to determine solar energy generation under the assumption that intensity is directly proportional to the energy generated.

A system to support the allocation of solar energy based on the predictions of the model will assist in making effective energy consumption plans for smart homes. The energy predicted can be used for a smart energy allocation recommendation system for solar devices in a smart home to recommend the operable one or more electronic devices.Several economic benefits such as subsidy in electricity bills by sending additional energy back to the grid can be gained. Assistance in finding the right time to switch between conventional and non conventional forms of energy is another viable use case of predicting solar energy generation and the allocation system.

The paper contributes to research in the following ways. Two IoT devices are used where first IoT device is coupled with the solar panel to obtain data about the solar panel and to fetch weather data to use for prediction and second IoT device is associated with the electronic devices present in the smart home to obtain energy consumption data and provide smart recommendation for allocation of solar energy. A study of various machine learning techniques is done to select the best approach for prediction of solar intensity along with a formula to obtain energy generated based on predicted intensity. Accuracy of prediction is assessed using a combination of metrics, RMSE, MAE and R2 score. Data obtained by first IoT device on an hourly basis when used in combination with a deep neural network gives the best result.

II. Literature Survey

José Gustavo Hernández-Travieso et al [3] predicted solar power by training an Artificial Neural Network (ANN) multi-layer perceptron. For prediction, they considered the information gathered from the Gran Canaria and Tenerife (Spain) meteorological station. They considered parameters like radiation, humidity, temperature, precipitation, cloudiness, meteor, and wind speed. Their research of energy prediction delivers a mean average error of 0.04 kilowatts per square measure. Their system is based on an Artificial Neural network. They have used a back-propagation training algorithm with twenty-four neurons on the hidden layer and one neuron on the output layer. An ANN may be a machine that works in a very similar way the brain works. To achieve excellent performance, it applies complex interconnections of neurons or processing units. The back-propagation algorithm works by calculating the gradient of the loss function for every weight, iterating backward from the last layer to scale back errors.

Kurt Spokas et al [4] conducted a study taking into account parameters like latitude, longitude, and elevation for multiple different locations, and predictions were made season-wise by an empirical model. This was conducted to maximize agricultural yield. There was high difficulty in predicting the occurrence of irregular events such as precipitation. RMSE values varied widely depending on each season hence not giving a generalized prediction model.

Faizan Jawaid et al [5] provides a comparative study of forecasting solar power between an Artificial Neural Network between machine learning algorithms. The dataset considered in the paper is obtained from the National Oceanic and Atmospheric Administration (NOAA). The collected data includes the periodic values of solar zenith angle, visibility, solar azimuth angle, and dew-point, etc. An Artificial Neural Network model is used with a simple feed-forward network with a backpropagation algorithm. They also implemented Cross-Validation for better training of data.

Navin Sharma et al [6] made use of seven weather prediction parameters and compared three machine learning models namely, SVM-RBF kernel with 4 dimensions, cloudy computing model using sky condition forecast, past predicts future prediction model. This paper also made use of feature selection techniques. They concluded that the SVM model outperforms most pre-existing models for the given purpose but no comparison was made with ANNs in this study.

Seul-Gi Kim et al [7] tries to decrease the reliance on external weather predictions by connecting unannounced weather variables i.e. weather observations with announced weather forecasts. That is done in the first step with the help of an auxiliary model. The model used in the second step is a random forest regression as it provides the best results and provides a day ahead prediction of energy generation. However, the final score is comparatively lower despite giving a more robust result.

Despite the excellent results, the paper [3] and [4] limits its extent to predict the solar radiation reaching earth's surface at a particular location, furthermore relies completely on the input from the meteorological station. In the paper [7] the system relied on inputs provided by PV plants for determining the solar power generated, this could not provide the solution for a location that lacks a PV plant in the neighborhood. A system to predict the solar energy output at any location when provided with weather data and coordinates along with a recommendation system for consumption of the solar energy by a smart home is proposed.

III. Methodology

The process of the system is depicted in Figure 1. This section is divided into five parts. Section A gives dataset specifications. section B provides procedures used for the feature selection method. Section C gives an analysis concerning different models. Section D presents the approach for the prediction of Solar energy and Section E provides the recommendation of predicted energy.

A. Dataset

NSRDB (National Solar Radiation Database) dataset was considered. [2].They collect hourly amounts of meteorological information and that comprises the three common values of solar radiation namely global horizontal, direct normal, and diffuse horizontal irradiance. Data from the years 2014-2017 were combined to give a dataset consisting of a total of 70000 entries. The dataset consisted of 18 features which include, Year, Month, Day, Hour, Minute, DHI, DNI, GHI, Cloud Type, Dew Point, Solar Zenith Angle, Surface Albedo, Wind

Speed, Precipitable Water, Wind Direction, Relative Humidity, Temperature, and Pressure. Four years of data were combined to predict solar intensity (GHI). Around 4/5th of the data was for training and the remaining for testing purposes. Features that do not contribute any important information were removed and the best features that are selected using the heatmap are chosen to predict solar intensity

B. Feature Selection

A variety of features were used for prediction. The interrelation between features was studied and the most important ones were selected. NSRDB datasets were for training and testing purposes. The terms solar intensity, GHI, solar power, and solar energy are used interchangeably as solar energy generation is assumed to be directly proportional to solar intensity.

The NSRDB dataset was used for building the machine learning model.

The dataset had to be cleaned and pre-processed before giving it to the model for training. The features that are present are shown in Table I. The interrelation between the features was studied and the most important ones were selected.

Heatmap depicted in Fig 2 was used for feature selection. It is a graphical representation of data in which each data value represents in a matrix and it has a special color. The value to be predicted is present as the 'GHI' column in the dataset. The most accurate model was selected as the main model.

Normally, low-value shows in low-intensity color and high-value shows in high-intensity color format. It provides the correlation between each feature. The Pearson correlation coefficient measures the correlation between the input features. It is shown in equation 1. [13]

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt[2]{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \qquad (1)$$

Correlation ranges from -1 to +1. Values closer to zero mean there is no direct relationship between the two features. Towards 1, the correlation is positive and -1 is negative.

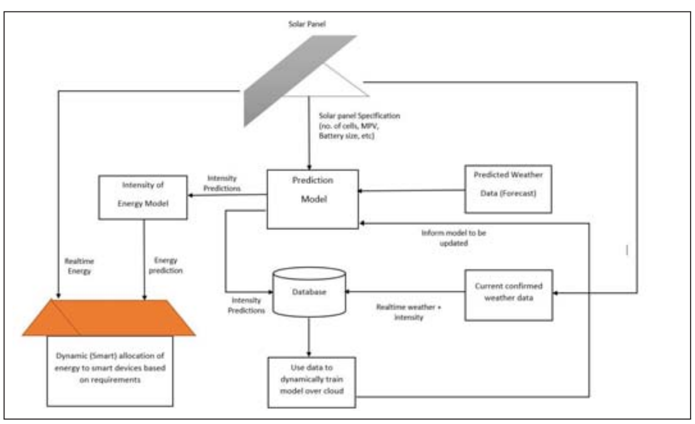

Fig. 1. Overall Process

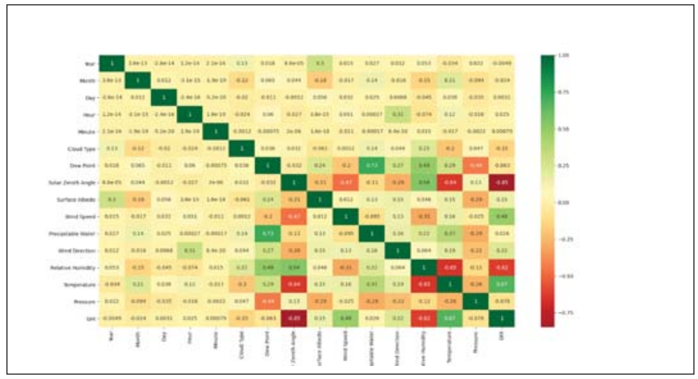

Fig. 2. Heatmap

Hence, the final features selected are shown in Table I. Table I shows the description and range of every feature used in the model.

TABLE I
SELECTED FEATURES

Feature Name	Description	Range
Hour	Hour of the day when feature values and ghi was recorded	0-23
Cloud Type	Categorical variable describing type of cloud recorded	0-9
Dew Point	The atmospheric temperature below which water droplets begin to condense and dew can form	(-14) - 22
Solar Zenith Angle	The zenith angle is the angle between the sun and the vertical.	10.67-169.2
Surface Albedo	Surface albedo is defined as the ratio of radiosity to the irradiance (flux per unit area) received by a surface	0.121-0.139
Wind Speed	The rate at which air is moving in an area	0-10.6
Wind Direction	Direction(angle) in which the wind is blowing	0-360
Relative Humidity	The amount of water vapour present in air expressed as a percentage of the amount needed for saturation at the same temperature.	6.19-100
Temperature	The degree of heat present	2-40

C. Model Selection

A comprehensive study of various machine algorithms was done. Models selected were supervised learning models for regression analysis as the target value, i.e. Solar Intensity is a continuous value. Factors in the selection of the models were the presence of a lesser number of categorical features in the dataset and the presence of dependent features. Another factor considered was that the dataset features have a non-linear relationship with the target value. The models selected considering these factors are given below.

1) Quadratic Regression: he Quadratic regression aims to capture the non-linear relationship between the variables. It is useful when the data does not have a linear relationship with each other. The accuracy of this algorithm is generally high when an appropriate degree is used for training the model. Quadratic Regression gave an R2 score of 0.935.

2) Support Vector Machine (SVM): Support Vector Machine (SVM) is a supervised machine learning algorithm. This model outputs an optimal hyperplane which classifies new examples. Support Vector Regression is different from the normal regression model. It tries to fit the hyperplane within boundary lines.It gave an R2 Score of 0.777.

3) Gradient Boosting Regression: It is a machine learning method that creates a prediction model in an ensemble of weak models, normally decision trees. It incrementally builds the model and generalizes them by allowing optimization of the error function. It teaches many models in a gradual, additive, and serial manner. Gradient Boosting algorithm gave an R2 score of 0.94.

4) K Nearest Neighbors: It is a simple supervised learning method. K-NN assumes the similarity between the new and available data and returns the value based on the highest similarity with existing data i.e. the nearest data points. It is non-parametric and a lazy learner as it does not learn from training data immediately but stores data and performs the action on it at the time of prediction. KNN algorithm gave an R2 score of 0.96.

5) Neural Networks: Neural Networks are widely used for the task of regression because of low error and high R2 score. They can understand the non-linear nature of the dataset. It is viewed as a group of interconnected nodes performing the computations. A fully connected dense neural network model of 4 layers is created having 48,129 trainable parameters and learning rate regularizers to avoid overfitting. Neural Network gave the best R2 score of 0.965.

TABLE II
RECOMMENDATION EXAMPLE

Time (24 hour format)	**Predicted Energy Generation (in Watt)**	**Stored Energy /Residue Energy in battery /cell i.e. Storage Medium (in Watt)**	**Energy Required for: Electronic Device 1 (High Priority) (in Watt)**	**Energy Required for: Electronic Device 2 (Medium Priority) (in Watt)**	**Energy Required for: Electronic Device 3 (Low Priority) (in Watt)**	**Remaining Energy (in Watt)**
0500	0	0	0	0	0	0
0600	0	0	0	0	0	0
0700	50	50	0	0	0	50
0800	150	200	0	20	30	150
0900	250	400	200	200	100	0 (system will recommend allocation for device 3 at a different time)
1000	400	400	100	50	50	200
1100	500	700	200	0	200	300
1200	600	900	500	500	0	0 (system will recommend allocation for device 2 at a different time)
1300	500	500	100	100	0	300
1400	400	700	0	100	150	450
1500	250	700	0	0	50	650
1600	150	800	100	100	100	500
1700	100	600	0	50	0	550
1800	50	600	0	50	50	500
1900	0	500	50	0	0	450
2000	0	450	0	50	100	300

D. Prediction

To predict the solar intensity, the parameter GHI(Global Horizontal Irradiance) was to be predicted. To do so 9 selected feature values, from table I are first taken from the weather forecast and then fed to the system at any desired time. The system outputs the GHI value. An IoT device was used to overcome the challenge of continuously fetching real-time data.
Equation 2 displays the formula used to calculate solar power from the predicted solar intensity value. [14]

$$MaximumPVPanelSolarPower = PanelEfficiency * PanelArea * RadiationValue * (1 - 0.005(AmbientTemperature - 25°C)) \quad (2)$$

This power calculated can then be allotted to various electrical appliances according to user preferences.

E. Recommendation

The energy predicted by the system is then used for recommending appropriate allocation. The allocation is done by taking into account the electronic devices, their priority as provided by the user, and energy requirements. The priority is classified as a high, medium, and low priority. Energy requirements for the electronic devices are provided by IoT devices As seen in Table II, the method in the paper is used for forecasting solar energy production of solar panels in a region, wherein expected solar energy value for a location at a particular interval of time in the future is depicted as predicted energy generation. Recommending the one or more electronic devices may be done by display on a user device associated with a user, using a recommender system to allocate distribution of solar energy among the one or more electronic devices according to a predefined priority input by the user. The recommender system may be an external thing. Table II also provides the hourly energy requirements of 3 electronic devices, for which recommendation may be needed, wherein the devices may be described as a high, medium, and low priority, based on the requirements of the user. The hourly recommendation for each electronic device may be obtained from the second IoT device or an application installed on the user device that may display the energy allocation.

IV. Results

A. Prediction

The results of the various models are tabulated in Table III. It shows the comparison between algorithms used.

TABLE III
RESULT

Model	RMSE	MAE	R2
Linear	143.18	116.87	0.79
Quadratic (degree = 2)	78.97	54.02	0.935
K-Nearest Neighbour	47.765	22.88	0.96
Gradient Boosting Regression	73.72	45.58	0.94
Support Vector Machine	148.16	119.35	0.777
Neural Network (4 – Layer model)	**47.73**	**22.84**	**0.965**

It is clear from table III that Neural Network and K-nearest Neighbour performed the best. KNN considers the nearest neighbors for estimating the prediction and is also better for classification rather than regression problems. Neural Networks can easily identify the non – linear relationship between the data. Hence, they perform better and are chosen for the prediction model.

B. Recommendation

As seen in Table II, the method in the paper is used for forecasting solar energy production of solar panels in a region, wherein expected solar energy value for a location at a particular interval of time in the future is depicted as predicted energy generation. For example, the predicted energy or the expected solar energy value as given in Table II maybe for a particular future date, and the plurality of time interval is depicted as hourly time interval starting from 5 am (0500 hours) in the morning till 8 pm (2000 hours) in the device associated with the electronic devices. As observed in Table II, at around noon (1200 hours), the energy prediction is 600 Watts, whereas the energy needed for electronic devices 1 and 2 at that same time is 1000 Watts. Hence, the system may recommend the user to run the electronic device 2 at some other time. Further, the energy remaining after the entire day can be sent to the grid to maximize savings.

V. Conclusion and Future Work

The machine learning algorithms are implemented to predict the value of solar intensity (GHI) given the input feature values in this paper. The model that was the most accurate was found to be the Neural Network (4–layer) model giving an RMSE value of 47.73.

The problems that we'll be able to solve after getting the predicted value of the solar power include, less dependence on the conventional energy resources and efficient Energy management using an allocation system to dynamically allocate the energy on the go, to maximize efficiency. We are also able to lessen down the skeptical views people have about solar generation by giving near-accurate predictions. Future aspect for this work is to refine neural network architecture to obtain higher accuracy based on deployment location of model, applying a local dataset to ensure best results for intensity prediction model and exploring dynamic allocation system to adjust and support various kinds of situations for smart homes.

References

[1] Ministry of New and Renewable Energy - Government of India. (2020, November). Solar energy current status. https://mnre.gov.in/solar/current-status/

[2] National Renewable Energy Laboratory (NREL)'s National Solar Radiation Database (2020, November) NSRDB Data Viewer. https://maps.nrel.gov/nsrdb-viewer/

[3] José G. Hernández-Travieso, Carlos M. Travieso, Jesús B. Alonso and Malay Kishore Datta,"Solar Radiation Modelling for the Estimation of the Solar Energy Generation.",2014 IEEE

[4] Spokas, K., Forcella, F, "Estimating hourly incoming solar radiation from limited meteorological data.", Weed Science, 54(1). 2006, 182-189.doi:10.1614/WS-05-098R.1

[5] Faizan Jawaid, Khurum NazirJunejo, "Predicting Daily Mean Solar Power Using Machine Learning Regression Techniques", IET Renew. Power Gener., 2017, Vol. 11 Iss. 10, pp. 1268-1273

[6] N. Sharma, P. Sharma, D. Irwin and P. Shenoy, "Predicting solar generation from weather forecasts using machine learning," 2011 IEEE International Conference on Smart Grid Communications (SmartGridComm), Brussels, 2011, pp. 528-533.

[7] Kim, S.-G., Jung, J.-Y., Sim, M.K.," A Two-Step Approach to Solar Power Generation Prediction Based on Weather Data Using Machine Learning.", Sustainability 2019, 11, 15

[8] Can Wan, Jian Zhao, Yonghua Song, Zhao Xu, Jin Lin and Zechun Hu, "Photovoltaic and Solar Power Forecasting for Smart Grid Energy Management." CSEE Journal of Power and Energy Systems, Vol. 1, No. 4, December 2015

[9] Yuanying Chi,Jinxia Zhang and Linpeng Xiao," Solar Power Generation Forecast Based on LSTM",2018 IEEE

[10] Swaroop Ramaswamy and Pradip Kumar Sadhu, "Forecasting PV Power from Solar Irradiance and Temperature using Neural Networks",2017 International Conference on Infocom Technologies and Unmanned Systems (ICTUS 2017),Dec. 18-20, 2017, ADET, Amity University Dubai, UAE

[11] Srinivasan Iyengar, Navin Sharma, David Irwin, Prashant Shenoy, and Krithi Ramamritham,"A Cloud-Based Black-Box Solar Predictor for Smart Homes", ACM Trans. Cyber-Phys. Syst. 2017, 1, 4, Article 21

[12] Voyant, Cyril and Notton, Gilles and Kalogirou, Soteris and Nivet, MarieLaure and Paoli, Christophe and Motte, Fabrice and Fouilloy, Alexis. "Machine learning methods for solar radiation forecasting: A review," 2017, Renewable Energy, Elsevier, vol. 105(C), pages 569-582.

[13] Pearson correlation coefficient. https://en.wikipedia.org/wiki/Pearson_correlation_coefficient

[14] Dewangan, Chaman. (2018). Re: How to calculate PV power output from solar radiation?. Retrieved from:https://www.researchgate.net/post/How-to-calculate-PV-power-output-from-solar-radiation/5b7ebf528b950029a42415c1/citation/download.

[15] Solargis (2020, November) Solar radiation modeling https://solargis.com/docs/methodology/solar-radiation-modeling