LITERATURE SURVEY

- A. Bajpai and M. Duchon, "A Hybrid Approach of Solar Power Forecasting Using Machine Learning," 2019 3rd International Conference on Smart Grid and Smart Cities (ICSGSC), 2019.

  In this study, multiple machine learning models are used, and different weather parameters are analyzed for solar power forecasting. The accuracy is evaluated using historical weather data. The study proposes a novel approach of PV forecasting where a model is built based on the principle of pipelining clustering, classification and regression algorithms. Clustering and classification algorithms group the data-points with similar weather conditions. Subsequently, segmented regression is applied to these groups.Weather forecast of the next day is utilized to determine the models used for forecasting.

  Dataset : The weather data includes multiple weather metrics for the year from 2013 to 2016 observed in 1 minute intervals. The data includes system output for the period from 8 February 2013 to 10 October 2016. The readings are divided in two parts, generated power and feed-in power in watts. Along with the measured power readings, the cell temperature is also recorded. The frequency of observed readings is random, varying from 6 readings per minute to 1 reading in a 5 minutes interval.

  Results : With the hybrid model, they combined clustering, classification and regression according to specific weather conditions improving the overall quality of the prediction. The hybrid model performed better than the random forest regression model.

  Drawbacks:
  - takes much more time and effort to train and tune that many specialized models as compared to training the universal one.
  - The PV system forecasting accuracy depends on the weather forecasts. The more accurate weather forecasts we use, the better will be the predictions.
  - The distance between the weather station and the PV device can affect the prediction accuracy, especially, during frequent weather variations.

  Future scope:
  In future, the clusters can be inspected and analyzed to show some light on the observed behavior and can be improved further with the refinement of specialized models.

- D. Solanki, U. Upadhyay, S. Patel, R. Chauhan and S. Desai, "Solar Energy Prediction using Meteorological Variables," 2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE), 2018.

  This paper presents a solar energy prediction model consisting of a mathematical model which enables to compute the amount of solar energy generation for the next seven days (including present day) by considering weather data and plant specifications. The factors that affect the solar energy generation are ambient temperature, solar irradiance, efficiency of panels, and efficiency of plant components like: inverter, cables, type of PV module installation etc

  Dataset : The factors taken into consideration are module temperature and its derating factor, solar irradiance, panel efficiency, type of installation and overall plant losses. In order to make mathematical models more reliable, performance factors and radiation reflection factors have been taken into consideration.

  Results : Analysis showed a generalized idea about the impact of temperature, irradiance, plant design and efficiency on the PV system output. Graphical analysis interprets that during normal weather days, the model gives 70% accurate prediction for any healthy PV system.

  Drawbacks : Any uncertainty in weather condition results into vulnerability in mathematical model result

- M. Abuella and B. Chowdhury, "Solar power forecasting using artificial neural networks," 2015 North American Power Symposium (NAPS), 2015.

  The paper presents an artificial neural network model to produce solar power forecasts. Sensitivity analysis of several input variables for best selection, and comparison of the model performance with multiple linear regression and persistence models are also shown.

  Dataset : The objective is to determine the solar forecasts in hourly steps through a month of forecast horizon. The target variable is solar power. There are 12 independent variables available from the European Center for Medium-Range Weather Forecasts (ECMWF) that are used to produce solar forecasts. These are: Total column liquid water of cloud (tclw) - (kg/m2), Total column ice water of cloud (tciw) - (kg/m2), Surface pressure (SP) - (Pa), Relative humidity at 1000 mbar (r) - (%), Total cloud cover (TCC) - (0-1), 10 meter U wind component (U) - (m/s), 10 meter V wind component (V) - (m/s), 2 meter temperature (2T◦) - (K), Surface solar radiation down (SSRD) - (J/m2)- accumulated field, Surface thermal radiation down (STRD) - (J/m2)- accumulated field, Top net solar radiation (TSR) - (J/m2) -accumulated field, Total precipitation (TP) - (m) - accumulated field.

  Results : The artificial neural networks model outperforms the multiple linear regression analysis MLR model and the persistence model. The performance of the ANN depends on how well it is trained and on the quality of the data that is used. The feed-forward ANN with 14 weather variables and with hourly step size for forecasts performed better than the recursive neural networks.

Drawbacks : In the clear sky hours, the model produces more accurate forecasts than cloudy hours. The more accurate weather forecasts we use, the more accurate solar power forecasts will be produced. Using the classification variables and the interactions between the variables enhances the performance of the MLR model significantly but this is not the case for the ANN model. With additional historical data, the model performance will improve

- Mohammed, Azhar & Aung, Zeyar. (2016). Ensemble Learning Approach for Probabilistic Forecasting of Solar Power Generation. Energies. 9. 1017. 10.3390/en9121017.

.In this study, we propose three different methods for ensemble probabilistic forecasting, derived from seven individual machine learning models, to generate 24-h ahead solar power forecasts. We have shown that while all of the individual machine learning models are more accurate than the traditional benchmark models, like autoregressive integrated moving average (ARIMA), the ensemble models offer even more accurate results than any individual machine learning model alone does.

Dataset : The initial training dataset consists of 12 months of hourly solar power data from April 2012 to March 2013. The testing dataset consists of 15 months of data from April 2014 to June 2015. After the forecasting has been done for each month in the testing dataset, that month's data are incrementally added into the training dataset.These 12 input variables are listed below. • tclw: Total column liquid water, vertical integral of cloud liquid water content. Unit of measurement: kg/m2 . • tciw: Total column ice water, vertical integral of cloud ice water content. Unit: kg/m2 . • SP: Surface pressure. Unit: Pa. • r: Relative humidity at 1000 mbar, defined with respect to saturation over ice below −23 ◦C and over water above 0 ◦C. For the period in between, a quadratic interpolation is applied. Unit: %. • TCC: Total cloud cover. Unit: zero to one. • 10u: 10-meter Uwind component. Unit: m/s. • 10v: 10-meter Vwind component. Unit: m/s. • 2T: two-meter temperature. Unit: K. • SSRD: Surface solar radiation down. Unit: J/m2 . • STRD: Surface thermal radiation down. Unit: J/m2 . • TSR: Top net solar radiation, net solar radiation at the top of the atmosphere. Unit: J/m2 . • TP: Sum of convective precipitation and stratiform precipitation. Unit: m. The output variable is the solar power generated in each farm at each hour. This value is normalized to lie between zero and one as the nominal power generated in each of the solar farms is different.

Results : The findings of this study show that combining the results from individual machine learning-based regression models gave exceedingly better performance than the individual models Energies 2016, 9, 1017 16 of 17 themselves. These results are consistent across the forecasting horizon of 15 months. Furthermore, grouping the data based on individual hours of the day results in lower error rates in comparison to the results where the data are not grouped. We use three different strategies to combine the results to generate probabilistic forecasts. It is found that Method III, which assumes the normal probability distribution and incorporates additional features, offers the best results.

Drawbacks : The field of probabilistic forecasts is still new, and various evaluation metrics are still being developed. In this study, we have used RMSE and MAE to evaluate the point forecasts and pinball loss score to evaluate the probabilistic forecasts. To better evaluate the model, there is a need to explore various other metrics in the field of probabilistic forecasts, like the continuous rank probability score (CRPS). Furthermore, the models can be further improved by relaxing the assumption that the probabilistic forecasts follow a normal distribution, as this assumption is too restrictive.

- A. Dhage, A. Kakade, G. Nahar, M. Pingale, S. Sonawane and A. Ghotkar, "Recommendation and Prediction of Solar energy consumption for smart homes using machine learning algorithms," 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV), 2021, pp. 1-5, doi: 10.1109/AIMV53313.2021.9670909.

  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.

  Dataset: NSRDB (National Solar Radiation Database) dataset was considered. 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.

  Results: The results show that the Neural Network and Knearest 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.

  Drawbacks: Requires refining of 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.

- Rahul, A. Gupta, A. Bansal and K. Roy, "Solar Energy Prediction using Decision Tree Regressor," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp. 489-495, doi: 10.1109/ICICCS51141.2021.9432322.

  This paper aims to predict the output production of Solar Power Plants in KWh. Also, this research paper attempts to find the relation of weather attributes with the output power generated. The Decision Tree Regressor model will be used to forecast the power generation of a Solar Power Plant. This model's predictions will help us determine how a solar power plant can be efficiently used to generate a high amount of power. The motivation behind the study was to develop a system that can effectively predict the output of these photovoltaic cells so that orders for supply and demand gaps are placed timely, ensuring the quality of service and cost savings.

  Dataset: The dataset which we have chosen is from a 10MW solar power plant of GPCL (Gujarat Power Corporation Limited) spread across 35 acres. The dataset contains the amount of solar power generated in 5 years. The dataset contains daily values of solar power generated starting from 1st May 2015. Feature scaling was also performed on the dataset. Weather is also an important parameter for the prediction of SE output. In order to obtain higher and better accuracy, we have used weather attributes from NASA Power Project [18]. The location of the GPCL power plant was entered, which is situated in Patan District of Gujarat (23°53'57.9 "N, 71°13'21.3" E).

  Results: The main challenges that one encounters while planning for a solar based power generation system are firstly Storage of Energy as energy generation only takes place when solar radiations are present and secondly, the weather-dependent nature of this arrangement. The first challenge will eventually be solved when battery technology evolves, especially with the Booming Electric Vehicle industry and the second challenge is what we aim to solve using this study. An accurate prediction ahead in time can help us save carbon emissions as one can reduce dependency on other fossil fuel-based sources of energy and also save up economically. This study on 56 months of data collected from a 10MW installation in Patan District, Gujarat, India, has yielded satisfactory results. This study demonstrates that we can develop a stable grid even with renewable sources of energy which currently is not possible.

  Drawbacks: Several other Artificial Intelligence techniques can be applied to predict the production of Photovoltaic cells like Reinforcement Learning, Deep Learning, etc. We can also use these prediction methods to forecast the power generated through other renewable sources of energy since other sources of energy could also be available in abundance due to differences in geographical conditions.

- M. Yesilbudak, M. Çolak and R. Bayindir, "A review of data mining and solar power prediction," 2016 IEEE International Conference on Renewable Energy Research and Applications (ICRERA), 2016, pp. 1117-1121, doi: 10.1109/ICRERA.2016.7884507

  This paper concentrates on reviewing the data mining methods along with their input data, recording intervals and the number of training and test dataset used for solar power prediction. This is done by comparing 52 reference papers. Techniques like inference system, multilayer perceptron, radial basis functions, recurrent neural networks, back-propagation neural networks, wavelet neural network,s support vector machines, relevance vector machine, extreme learning machine, self-organizing maps, fuzzy theory, genetic algorithm, particle swarm optimisation, k-nearest neighbor classifier, learning vector quantization, non linear autoregressive with exogenous inputs, kernel ridge regression trees, absolute shrinkage and selector operator, regression trees, bagging trees, boosted trees and principal component analysis were compared.

  Dataset : Attributes like air temperature, latitude, longitude, solar angle, solar radiation, relative humidity, wind speed and day length were the common input parameters.

  Results : Artificial neural networks are the most preferred method. Researchers use multiple different input parameters. Seasonal performance of the employed models should be considered for better solar power predictions. Clustering and association techniques should be focused in addition to artificial; neural networks for benchmark tests.