Data Mining

Project Report

Title:

Solar Power Generation Forecasting

Submitted to:

Ms. Beenish Akram

Submitted By:

Hira Firdous [2020-CE-47]

Department of Computer Engineering

University of Engineering and Technology, Lahore

Table of Contents

Introduction:

The escalating use of non-conventional energy sources for electricity generation plays a pivotal role in diminishing reliance on fossil fuels, reducing production costs, mitigating environmental pollution, and curbing greenhouse gas emissions. Among renewable energy sources, solar photovoltaic (PV) systems stand out due to their abundant accessibility and sustainability. However, the intermittent nature of solar PV, coupled with its impact on grid stability, presents challenges. This work delves into forecasting short-term power generation from solar plants using numerical weather prediction (NWP) and explores the optimization of prediction approaches. Various models, including neural networks and recurrent neural networks (RNN) with long short-term memory (LSTM), are discussed, highlighting the significance of accurate predictions for optimizing power dispatch plans and achieving business profits. The report uses a dataset taken from github repository to train a ANN, using heatmaps, correlations understanding the relation between each parameter and emphasizing evaluation metrics such Root Mean Square Error and coefficient of determination (R2) to assess model accuracy.

Problem Statement:

The model aims to enhance accuracy in solar power generation predictions, optimizing energy production and distribution. This will contribute to improved resource utilization, grid management, and overall efficiency in renewable energy systems.

Dataset:

For this project, I will utilize a dataset sourced from GitHub, encompassing historical solar power generation data, meteorological information (including sunlight intensity, temperature, and wind speed), as well as details about the time of day and geographical factors. The dataset has been curated to include a comprehensive array of attributes crucial for accurate solar power generation forecasting. The information is drawn from various sources, including solar farms, weather stations, and pertinent environmental databases. This diverse and well-structured dataset, available on GitHub, ensures a robust foundation for developing a forecasting model that is both accurate and representative of the multifaceted factors influencing solar power generation.

About DATA

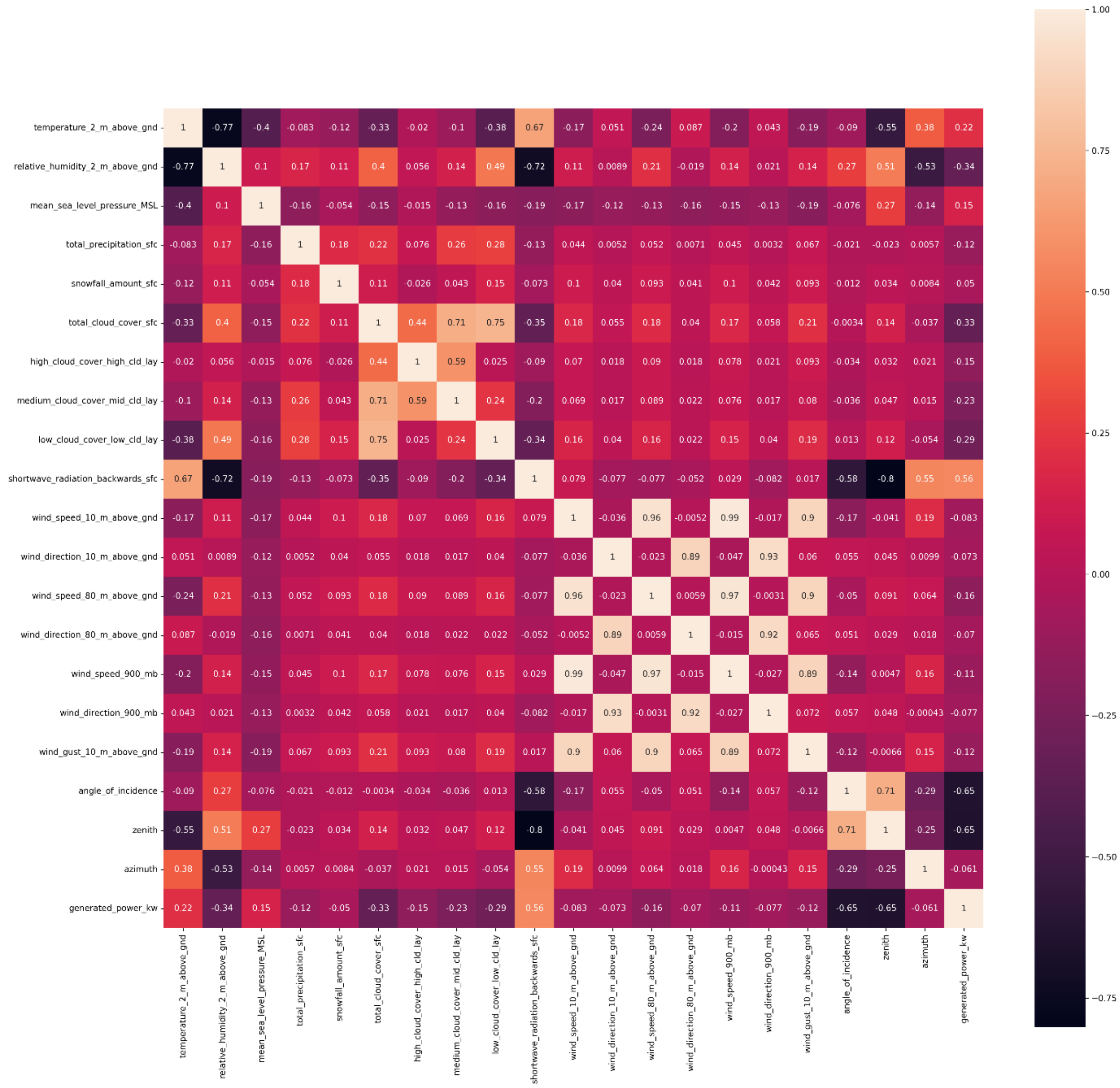
- distance-to-solar-noon, in radians.
- temperature, daily average temperature, in degrees Celsius.
- wind-direction, daily average wind direction, in degrees (0-360).
- wind-speed, daily average wind speed, in meters per second.
- sky-cover, in a five-step scale, from 0 to 4, being 0 totally clear and 4 completely covered.
- visibility, in kilometers.
- humidity, in percentage.
- average-wind-speed-(period), average wind speed during the 3-hour period de measure was taken in, in meters per second.
- average-pressure-(period), average barometric pressure during the 3-hour period de measure was taken in, in mercury inches.
- power-generated, in jules for each 3-hour period.

Project Elaboration

Understanding of data:

For the better understanding of data and correlation between parameters **heatmap** which is generated A heatmap is a visual representation of a correlation matrix, often used in data mining to uncover patterns and relationships within datasets. By assigning colors to cells based on the strength and direction of variable correlations, heatmaps provide an intuitive way to identify clusters and dependencies among variables.

Following is the heatmap which is being generated from the data:



Observations

Following are some observations from the above heatmap:

- High Correlation between Zenith and Agnle of Incidence of 0.71
- Shortwave radiation backwards and Generate Power KW has corr of 0.56
- Relative Humidity and Zenith are +ve corr (0.51)
- Relative Humidity and Low Cloud Cover are + ve correlated (0.49)
- Angle of Incidence and Zenith are -vely correlated with Genarted Power (-0.65)
- -ve corr between Zenith and temperature of -0.55
- High negative corr exists btw Shortwave radiation backwards and Zenith (-.8)
- Shortwave radiation backwards and Relative humidity are -vely correlated (-.72)
- Relative humidity and Temperature are -vely correlated (-.77)

Overview of Model Implementation:

Model which is used, is a simple ANN model sequential model. Summary of the model is as follows:

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense (Dense)                (None, 32)                672
_________________________________________________________________
dense_1 (Dense)              (None, 64)                2112
_________________________________________________________________
dense_2 (Dense)              (None, 1)                 65
=================================================================
Total params: 2,849
Trainable params: 2,849
Non-trainable params: 0
_________________________________________________________________
```

Two hidden layers having 32 neurons and 64 neurons respectively are integrated on the ANN. Keeping the model simple.

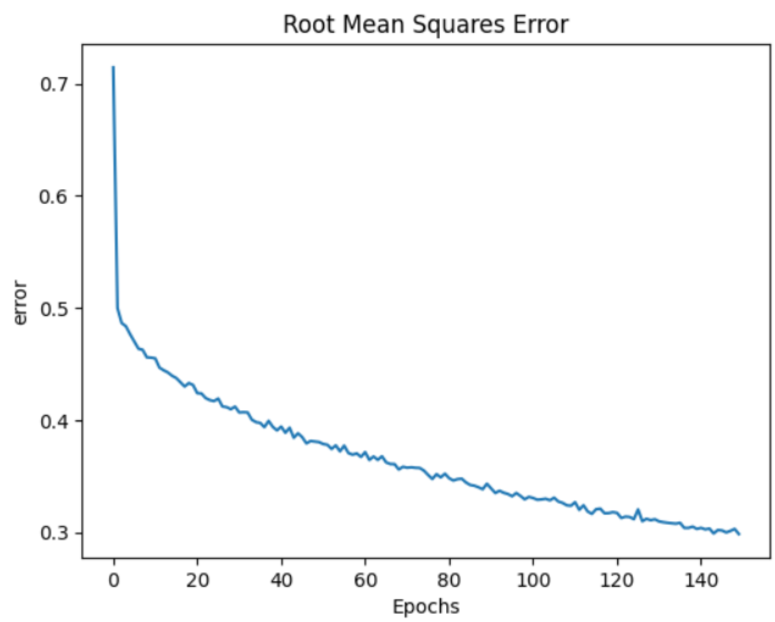
Input and Output:

Input will be 19 parameters and the output will be single predicted value which will be the overall power which can be or will be generated keeping the track of certain parameters.

Evaluation matrices:

Root Mean Square Error (RMSPE):

The graph generated by while training on 150 epochs and batch size of 32 is as follows:



It clearly can be observed from the graph that our error is reduced to very much at ending epochs.

Coefficient of determination (R2)

R-squared measures the goodness of fit of a regression model. Hence, a higher R-squared indicates the model is a good fit, while a lower R-squared indicates the model is not a good fit. And the value of r2 score of testing and training is as follows:

```
[23] from sklearn.metrics import r2_score
     r2_score(y_pred_orig, y_test_orig)

     0.7446386275581713
```

```
[24] r2_score(train_pred_orig, y_train_orig)

     0.9071195637047451
```

Further work proposed:

For now values of the parameters is inserted manually but this whole process can be automated by IOT(internet of things). Main idea is to get the value of those parameters from the sensors and different iot devices embedded there and just pass those parameters to this model and it will predict the power generated.

References:

[1] "Power output forecasting of solar photovoltaic plant using LSTM," *Green Energy and Intelligent Transportation*, vol. 2, no. 5, p. 100113, Oct. 2023, doi: https://doi.org/10.1016/j.geits.2023.100113.

[2] Saradindusengupta, "Solar-Power-Forecasting/Master Thesis.pdf at master · saradindusengupta/Solar-Power-Forecasting," *GitHub*. https://github.com/saradindusengupta/Solar-Power-Forecasting/blob/master/Master%20Thesis.pdf

[3] Anantgupta, "Solar-Power-Generation-Forecasting/solarpowergeneration.csv at main · anantgupta129/Solar-Power-Generation-Forecasting," GitHub. https://github.com/anantgupta129/Solar-Power-Generation-Forecasting/blob/main/solarpowergeneration.csv

[4] "ANALYSIS OF FACTORS FOR FORECASTING ELECTRIC POWER GENERATION BY SOLAR POWER PLANTS", doi: https://ela.kpi.ua/bitstream/123456789/54368/1/eete2020-4_62_p64-69.pdf

[5]I. Ashraf and A. Chandra, "Artificial neural network based models for forecasting electricity generation of grid connected solar PV power plant," International Journal of Global Energy Issues, vol. 21, no. 1/2, p. 119, Jan. 2004, doi: 10.1504/ijgei.2004.004704.