**PREDICTION OF SOLAR ENERGY USING DEEP LEARNING**

**A TECHNICAL REPORT**

***Submitted by***

**A.AADHITYA(16C001)**

**M.AKASH(16C007)**

**V.BHUJITH MADAV(16C015)**

**14CS7C0 – CAPSTONE COURSE - II**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**



**THIAGARAJAR COLLEGE OF ENGINEERING, MADURAI – 625 015**

**(A Government Aided ISO 9001: 2008 Certified Autonomous Institution affiliated to Anna University)**

**OCTOBER 2019**

**THIAGARAJAR COLLEGE OF ENGINEERING**

**(A Government Aided ISO 9001: 2008 Certified Autonomous Institution affiliated to Anna University)**

**MADURAI – 625 015**



**BONAFIDE CERTIFICATE**

Certified that this technical report **“PREDICTION OF SOLAR ENERGY USING DEEP LEARNING”** is the bonafide work of **“A.AADHITYA(16C001),M.AKASH(16C007),V.BHUJITH MADAV(16C015)** who carried out the project work under my supervision.

**SIGNATURE**

## Dr. G.S.R EMIL SELVAN,

Associate Professor,

Computer science and Engineering,

Thiagarajar college of Engineering,

Madurai-15

Submitted for Viva - Voce conducted at Thiagarajar College of Engineering, Madurai – 15 on ……………..

**Examiner - I Examiner - II**

**ABSTRACT:**

Solar energy is the most abundant energy resource on the planet. The solar energy that strikes the Earth’s surface in one hour is approximately the same as the amount consumed by all human activities in one year. Today, photovoltaic (PV) energy provides 0.1% of the total worldwide electricity production. In the IEA Solar PV Roadmap vision, PV is expected to deliver 5% of the global electricity consumption in 2030, rising to 16% in 2050. Solar power is widely acknowledged to be the fastest-growing energy industry in the world. A major concern surrounding solar power is the variability and unpredictability of sunlight. If it is overcast or cloud cover present during the day, then the photovoltaic cells are unable to produce electricity. This inherent variability poses issues with grid reliability and the expenses associated with operating the solar units. All of these factors make it difficult to predict the photovoltaic output of solar panels, and solar forecasting is a method used to address this issue.

|  |  |  |
| --- | --- | --- |
| **TABLE OF CONTENTS** | | |
| **CHAPTER NO** | **TITLE** | **PAGE NO** |
|  | **ABSTARCT** | **iii** |
|  | **LIST OF FIGURES** | **vi** |
|  | **LIST OF TABLES** | **vii** |
|  |  |  |
| **1** | **INTRODUCTION** |  |
|  | 1.1 Packet | 1 |
|  | 1.2 Packet Classification | 1 |
|  | 1.3 Need for Packet Classification | 2 |
|  | 1.4 Machine Learning Algorithms | 2 |
|  | 1.5 Problem Statement | 2 |
|  |  |  |
|  |  |  |
| **2** | **LITERATURE REVIEW** |  |
|  | 2.1 Traditional Approach | 3 |
|  | 2.2 Approach using Machine |  |
|  | Learning | 5 |
|  |
| **3** | **REQUIREMENTS** |  |

3.1 Hardware Support 4

3.2 Software Support 5

3.3 Dataset 5

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | |  | **TITLE** | **PAGE NO** | | **CHAPTER NO** | |  |  |  | | **4** | **OVERVIEW OF PROPOSED DESIGN** |  | |  | 4.1 Work Overview | 6 | |  | 4.2 Data Collection | 6 | |  | 4.3 Dataset Preprocessing | 7 | |  |  |
| |  |  |  | | --- | --- | --- | | **5** | **IMPLEMENTATION** |  | |  | 4.1 Work Overview | 6 | |  | 4.2 Data Collection | 6 | |  | 4.3 Dataset Preprocessing | 7 | |  |  |  | | **6** | **REFERENCES** |  | |  |  |  | |  |  |  | |  |  |  | |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |

1. **INTRODUCTION:**
   1. **SOLAR FORECASTING**

Analyzing the movement of the sun, cloud cover, various atmospheric factors such as irradiance, temperature and other factors relating to the solar power plant and predicting the amount of solar energy that can be produced on a particular day is called solar forecasting. Solar energy is a sustainable resource amongst other renewable energy resources. The integration of solar PV plants into power grids has received much attention. The implementation of large scale grid connected solar PV plants has shown significant issues to the power networks such as system stability, reliability, electric power balance, reactive power compensation and frequency response .Solar PV power forecasting has emerged as a brilliant way to address these issues.

1. **SOLAR FORECASTING TECHNIQUES**

* There are three major methods :

Statistical time series method ,Physical methods, Ensemble methods.

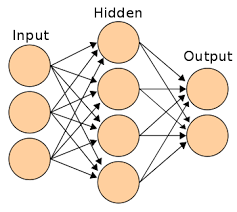
* The statistical approach is comprised of five sub-models i.e., (i) Artificial Neural Network (ANN), (ii) Support Vector Machine (SVM), (iii) Regression models. These statistical methods are highly dependent on the historical data, and the ability to extract valuable information from the past data.
* The physical methods consist of three sub-models i.e., (i) Numerical Weather Prediction (NWP), (ii) Sky Imagery, and (iii) Satellite-Imaging models.
* The ensemble approach ,as the name suggests combines the unique and positive features of various models to mitigate the limitations of the individual model and overall improve the performance.

**3.STATISTICAL TIME SERIES METHODS:**

* Time series is a set of continuous or sequential observations made on a specific parameter over subsequent periods of time. The statistical methods can establish the relation between the meteorological parameters and irradiance.

**3.1 Artificial Neural Networks(ANN):**

* Artificial Neural Network (ANN) is known as Artificial Intelligence (AI) system is used to mimic the human brain. The accuracy of a ANN models relies on parameters, training algorithm, and structure of the model. The complex nature of most practical problems can be effectively solved via ANN. The modeling of ANN can be classified into three stages:
* (1) Design phase consists of ANN type, number of neurons in each layer, number of layers, input and output parameters, training, and validation of selected datasets
* (2) Training phase takes place when the data were presented to ANN and the weight was modified until a predetermined condition was satisfied and finally
* (3) Validation phase, where the ANN model was built using unseen data, excluding those used in the training phase.
* Upon successful validation test, the model is ready to perform its designed function or further modification was required in the previous stage



**3.2 Support Vector Machines:**

* Support Vector Machines (SVMs) were categorized as a machine learning technique. This approach was widely used in prediction, classification, pattern recognition, and regression analysis. These models can prevent over-fitting of training data, dismiss iterative tuning of model parameters, require a few kernels, make faster computations, and have good generalization and convergence.
  1. **Regression Models:**
* Regression analysis refers to the method that determines a functional relationship of the model between the dependent and independent parameters. Regression analysis is called as ‘univariate regression’ when only one response parameter is involved and called ‘multivariate regression’ where two or more parameters are considered. The regression analysis represents a repetitive process so that the outputs are utilized to analyze, verify, criticize, and modify the input. Two regression analysis methods i.e., simple linear regression and multiple linear regression arewidely used with the complexity of correlation between parameters.

**4. PHYSICAL MODELS:**

* The interactions between solar radiations and atmospheric components such as aerosols and gases act as the physical model for solar irradiance prediction.
  1. **NUMERICAL WEATHER PREDICTION:**
* NWP tools were helpful in the enhancement of prediction models for renewable sources. The objectives of these NWP tools are to provide information about atmosphere conditions for a given time-scale. These NWP models can be categorized into two categories. Global models simulate the features of the atmosphere to a worldwide scale ,while mesoscale models simulate the features of the atmosphere for a limited area .NWP models were used widely to predict atmospheric state up to 15-days ahead. The benefit of this forecast model is that they are based on deterministic physical models.

**4.2 SKY IMAGERY:**

* Sky imager is a digital camera that provides high quality images of the sky from horizon-to-horizon and it is applied for determining clouds and various features and aspects regarding the cloud. Their built-in cameras are either oriented upwards or downwards in obtaining a direct image or in capturing reflections from a spherical mirror. The efficiency of cloud tracking functions depends on sky imagers quality. The solar predictions based on sky imagery analysis has four elements i.e
* Obtaining the sky image near the forecast site by using Sky Imager device
* Sky image data are analyzed to recognize clouds,
* Cloud motion vector is estimated by utilizing consecutive images
* Cloud location and motion vector data are used for short-term probabilistic and deterministic cloud cover, irradiance, and power predictions.
* Sky imagery techniques were used for very short-term forecast of future cloud patterns in solar plants.

**4.3 SATELLITE IMAGING:**

* The concepts of satellite imaging are quite similar to sky imagery models. The cloud pattern is determined from both visible and infrared images taken from satellite-based sensors flying overhead. Satellite imaging provides cloud motion information and its properties. Firstly, a physical driven model forecasts the clear-sky conditions. Next, these clear-sky irradiance models are modulated by predicting the irradiance obtained from the satellite images. Cloud motion vector fields can be generated by mixing with satellite images which can be utilized to forecast the future cloud locations. These methods were used effectively in irradiance predictions in the range of 1-min to 5-h ahead. It is less efficient when the clouds are quickly forming and dissipating.

**5. ENSEMBLE METHODS:**

Ensemble methods were presented to solve the weakness of individual methods and to enhance their strengths and accuracy . These methods combine different models to enhance the forecasting accuracy . The ensemble model was used widely in both statistical and learning machine techniques to obtain aggregated decisions by using multiple predictors.

**5.1 COOPERATIVE ENSEMBLE FORECASTING:**

To improve the performance of a forecasting technique, cooperative ensemble method can divide a forecasting task into a few sub-tasks and hence each sub-task is solved individually. The overall forecasting results were acquired by aggregating the forecast values of all predictors.

**5.2 COMPETETIVE ENSEMBLE FORECASTING:**

The competitive ensemble forecasting models use the multiple predictors with different parameters to build individual forecast models and to form an ensemble forecast model. Results of the forecasting model from the selected models were aggregated by averaging.

**6. FACTORS AFFECTING SOLAR ENERGY PRODUCTION:**

There are various factors affecting solar energy production. Let’s discuss about some of the crucial factors.

1.Irradiance

2.Cloud cover

3.Temperature

**6.1 IRRADIANCE:**

Irradiance is defined as the energy per unit time that strikes a unit horizontal area per unit wavelength interval, where the typical unit is W m−2

In context to solar energy production it is of 3 types

DHI

DNI

GHI

Direct Normal Irradiance (DNI)is the amount of solar radiation received per unit area by a surface that is always held perpendicular (or normal) to the rays that come in a straight line from the direction of the sun at its current position in the sky.

Diffuse Horizontal Irradiance (DHI)is the amount of radiation received per unit area by a surface (not subject to any shade or shadow) that does not arrive on a direct path from the sun, but has been scattered by molecules and particles in the atmosphere and comes equally from all directions

Global Horizontal Irradiance (GHI)is the total amount of shortwave radiation received from above by a surface horizontal to the ground. This value is of particular interest to photovoltaic installations and includes both Direct Normal Irradiance (DNI) and Diffuse Horizontal Irradiance (DHI).

**6.2 CLOUD COVER:**

Obviously light equals power, so the more direct light the panels receive, the more power will be produced. Bright, sunny days will contribute to your system working at peak capacity. But on a day with thick cloud cover, power production will be much lower than average.An unusual phenomenon also exists when there are patches of cumulus clouds drifting through the sun’s beams. Called the **edge of cloud effect**, as the sun peeks out of the spaces in between the clouds, the direct light combined with the reflective light will briefly boost your panels’ power production.

**6.3 TEMPERATURE**

Solar cells perform better in cold rather than in hot climate and as things stand, panels are rated at 25˚C. For each degree rise in temperature above 25˚C the panel output decays by about 0.25%. Thus, in hot summer days panel temperature can easily reach 70˚C or more, which means is that the panels will put out up to 25% less power compared to what they are rated for at 25˚C.

**7. PROBLEM STATEMENT:**

India’s peak demand for power stands at 1.81 lakh MW. Load Dispatch Centers(LDC) is responsible for distribution of power to our households and other commercial purposes Load dispatchers collect the information from various generating stations about how much they will be providing the next day. Then a schedule is prepared with 96 time blocks of 15minutes each by.

A problem is caused in two ways:

1) When a generating station doesn’t provide enough load as mentioned a day before to the LDC. The reason for this demand-supply mismatch may be due to sudden overcast conditions or a gloomy sky or whatever reasons disrupting production. This causes the LDC to request for load from other power plants which is a strenuous process(for eg, a thermal power plant requires time in term of hours to start production increasing the cost of fuel and other variable costs).

2) Under normal condition, the [Grid frequency](https://etrical.blogspot.in/2016/12/why-is-it-important-to-maintain-constant-frequency.html) is expected to be constant at 50 Hz. But during peak load period the frequency goes down to 48-48.5 Hz and [off-peak hours](https://etrical.blogspot.com/2016/07/load-curve-and-load-duration-curve.html), the frequency goes up to 50.5-51 Hz. So when a generating station doesn’t provide enough supply when there is a demand, frequency decreases and the station is penalized with Availability Based Tariffs(ABT). Similarly when frequency decrease and a generating station provides with power more than it’s expected to deliver it is provided with an incentive.

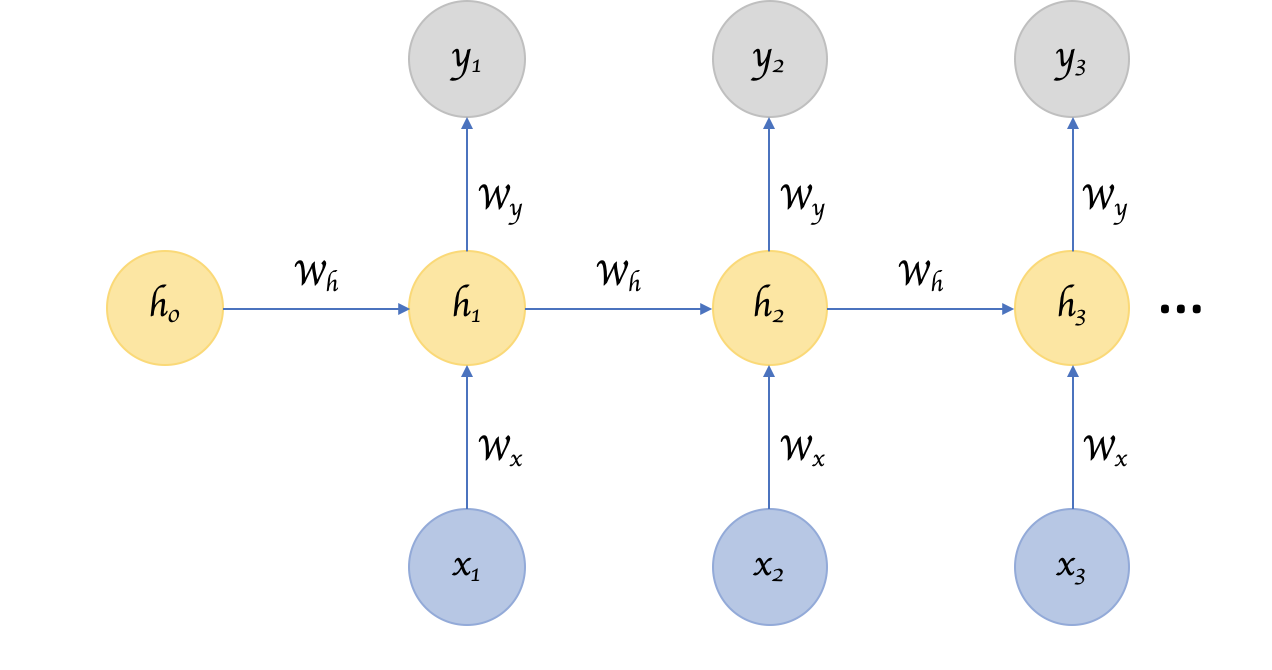
**8. LITERATURE REVIEW**

**8.1 EXISTING SYSTEM:**

Now currently numerical weather prediction models and machine learning techniques such as regression, ANN, support vector machines are in use. Similarly other physical models like sky imagery and satellite imagery are also used. But they don’t provide enough accuracy which we are trying to improve in our proposed system.

**8.2 PROPOSED SYSTEM:**

The method we are suggesting is called Recurrent Neural Network(RNN). Recurrent Neural Networks (RNNs) add an interesting twist to basic neural networks. RNNs are designed to take a series of input with no predetermined limit on size. While RNNs learn similarly while training, in addition, they remember things learnt from prior input(s) while generating output(s). So, the same input could produce a different output depending on previous inputs in the series.



**8.3 BENEFITS OF PROPOSED SYSTEM:**

RNNs, can remember their former inputs, which gives them a big edge over other artificial neural networks.

RNNs have showed better accuracy and clarity in time series prediction.

**9. REQUIREMENTS**

**9.1 Hardware Support**

This project has the following minimum hardware requirement

**Number of Processors** : 2

**Processors** : Intel Core(TM)

**Architecture** : i5/i7

**CPU Speed** :2.00 GHz

**RAM** :4 GB

**9.2 Software and Package Requirement**

The list of software required for the project are:

* Windows / Linux
* Keras
* Python
* Scikit
* Matplotlib

**9.2.1 Windows**

Windows can be used for this project as it supports machine learning and deep learning tools which can be used for classification and prediction purpose.

**9.2.2 Keras**

It is the deep learning package which is used to build the prediction model.

**9.2.3 Python**

Python is open source high level programming language which includes lots of packages that can be used to run machine learning algorithms easily. Here we python version 3.6.

**9.2.4 Scikit**

Scikit is a tool used for data mining and data analytics purpose. It can be used with python by import sklearn package. It contains many classification and regression algorithms inbuilt in it.

**9.2.5 Matplotlib**

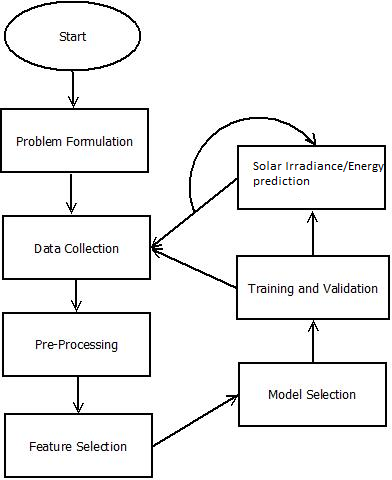
Matplotlib is a package in python programming language which is used to plot graphs based on results inferred from our experiments. It can be used in python by importing the package matplotlib.

**9.3 Data Set**

The data used was solar irradiance measurements from  [Loyola Marymount University](https://midcdmz.nrel.gov/apps/go2url.pl?site=LMU) from April 06, 2010 to May 05, 2016, collected from a RSR. This dataset was found via the National Renewable Energy Laboratory's (NREL) list of Measurement and Instrumentation Data Centers (MIDC).

**10. OVERVIEW OF PROPOSED SYSTEM**

**10.1 Work Overview**

****

## 10.2 Data cleaning

The raw data had the following issues:

* No date time index
* Some negative values (such as -99999) for features that should only have positive values
* Some outliers
* Missing data
* Unneeded columns

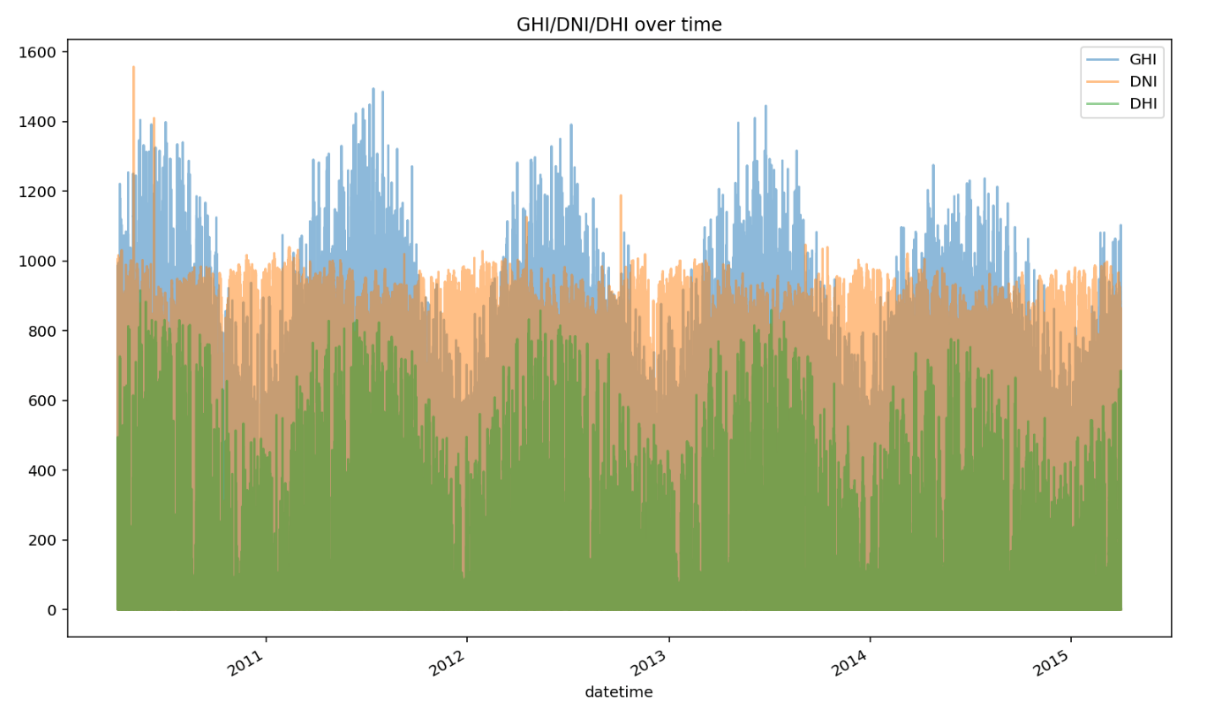
To get the data into usable form, we did the following steps:

* Written a custom function to convert existing time features to a date time object
* Set negative values to 0
* Removed and imputed outliers
* Dropped dates with missing data
* Dropped unneeded columns

## 10.3 Exploratory Data Analysis

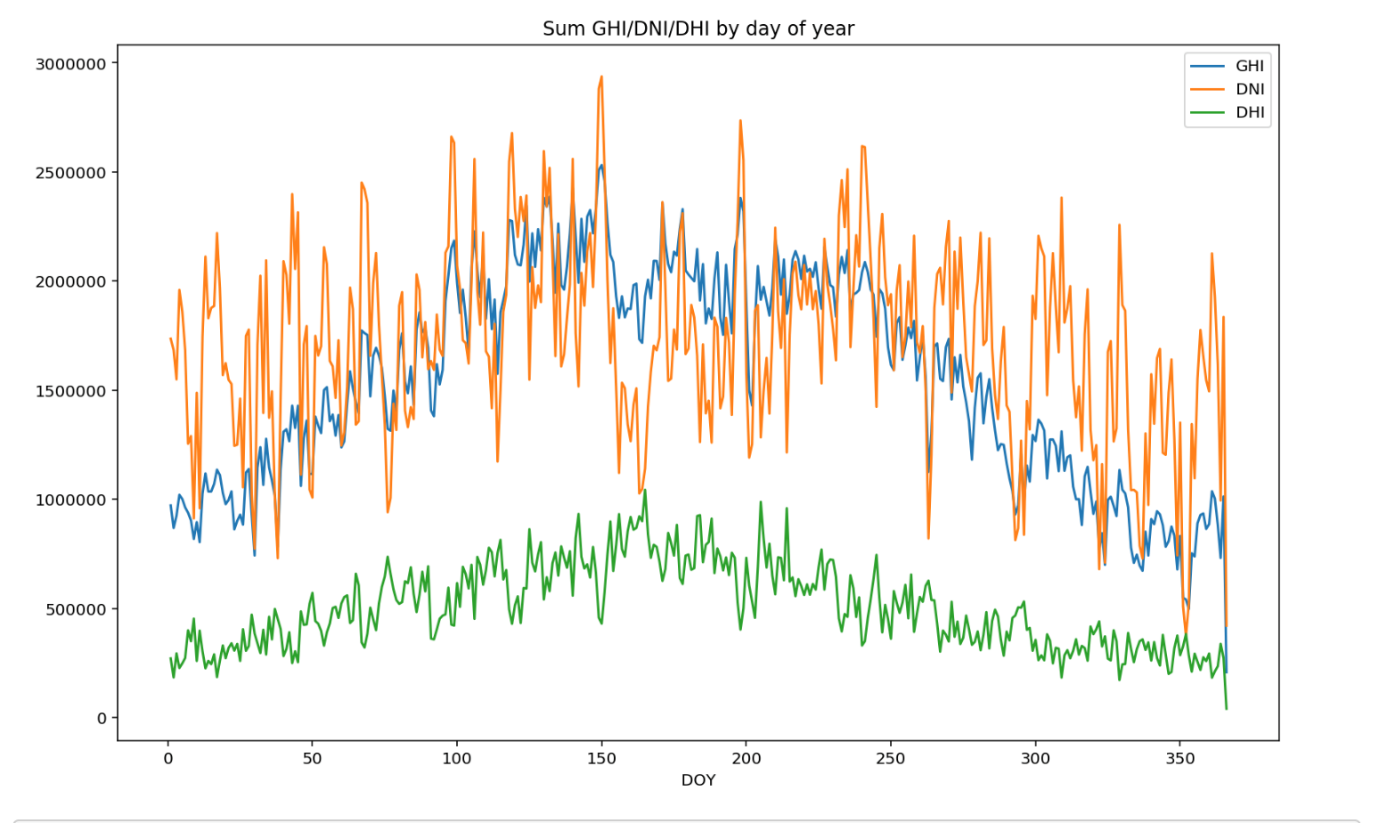
### **10.3.1 Predictors over Time**

This plot shows the irradiance metrics (DNI, DHI, and GHI) over time. Clearly there is a seasonal effect as values peak in the middle of the year (summer), and decline in the winter.

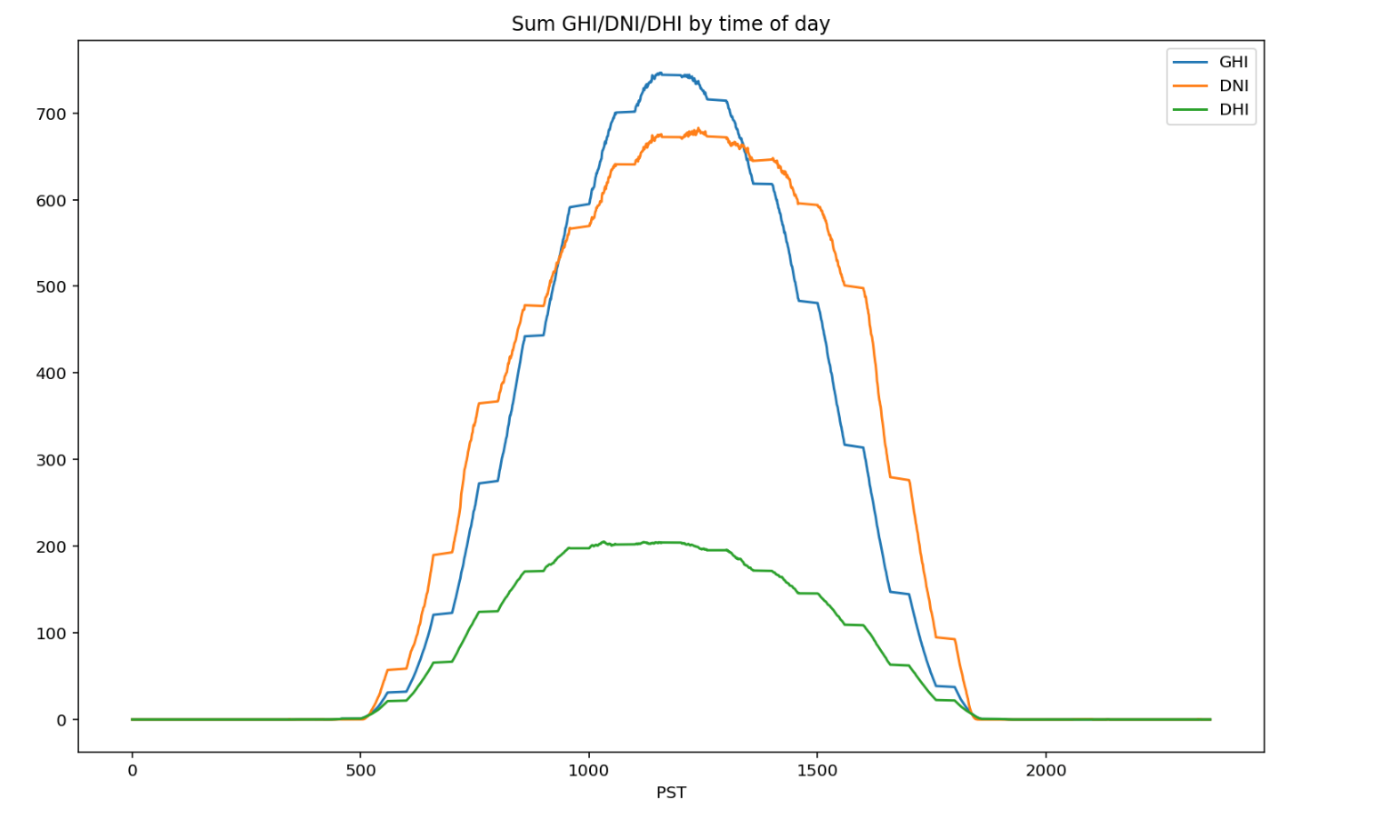


### **10.3.2 Predictors over day of year (average)**[¶](https://render.githubusercontent.com/view/ipynb?commit=d2f48b994abee13a87835880888c943437735b2d&enc_url=68747470733a2f2f7261772e67697468756275736572636f6e74656e742e636f6d2f73616d63686161612f63617073746f6e655f7265706f2f643266343862393934616265653133613837383335383830383838633934333433373733356232642f335f546563686e6963616c5f5265706f72742e6970796e62&nwo=samchaaa%2Fcapstone_repo&path=3_Technical_Report.ipynb&repository_id=153551684&repository_type=Repository#Predictors-over-day-of-year-(average))

Predictions over day of year support seasonability.



**10.3.3 Predictors over time of Day (Average)**



## 10.3.4 Correlation coefficients of all features

## The correlation matrix shows some multicollinearity between variables, as many as weak correlations.

## pic

## 10.3.5 Correlation coefficients of predictors

## The closer look at correlation of irradiance metrices.

## pic

## 11.Feature Engineering[¶](https://render.githubusercontent.com/view/ipynb?commit=d2f48b994abee13a87835880888c943437735b2d&enc_url=68747470733a2f2f7261772e67697468756275736572636f6e74656e742e636f6d2f73616d63686161612f63617073746f6e655f7265706f2f643266343862393934616265653133613837383335383830383838633934333433373733356232642f335f546563686e6963616c5f5265706f72742e6970796e62&nwo=samchaaa%2Fcapstone_repo&path=3_Technical_Report.ipynb&repository_id=153551684&repository_type=Repository#Feature-Engineering)

No features were added, but time elements (hour of day, datetime index) were taken from original features.

## 12. Modeling

### **12.1 Resampling**

Before modeling, the data was resampled to 15 minute increments (mean) to save run time. Lag amounts were in intervals of 15 minutes (96 = 24 hours, 672 = 1 week).

### LSTM (long short-term memory) RNN (recurrent neural network) in Keras

Predictions were made using an LSTM (long short-term memory) model. Data was lagged by 1 day and 1 week periods. Specific predictor features used were day of year, time of day, hour, lagged DNI, air temperature, and humidity.

### **12.2 Train test split**

Data was split at the year 2012. This resulted in about a 2:1 train-test-split.

### Hyperparameters

Hyperparameters used for the LSTM were:

* LSTM cells = number of hours predicting
* epochs = 10
* batch\_size = 12
* dropout = .3

### **12.3 Additional steps**

After fitting each model and model results were saved for later use.

## 13 . Model evaluation[¶](https://render.githubusercontent.com/view/ipynb?commit=d2f48b994abee13a87835880888c943437735b2d&enc_url=68747470733a2f2f7261772e67697468756275736572636f6e74656e742e636f6d2f73616d63686161612f63617073746f6e655f7265706f2f643266343862393934616265653133613837383335383830383838633934333433373733356232642f335f546563686e6963616c5f5265706f72742e6970796e62&nwo=samchaaa%2Fcapstone_repo&path=3_Technical_Report.ipynb&repository_id=153551684&repository_type=Repository#Model-evaluation)

* Models were scored on RMSE and r2 score.
* Predictions were inverse scaled to return predictions to original scale.

## 14. Predictions and results

Here are example results for modeling using the hyperparameters above. The plot area is just the last 300 hours of data (about 8 days).

## 24 hours predictions vs actual values

## pic

## 168 Hours predictions vs actual values

## pic

## 15. Implementation

## 15.1 Data Cleaning for Solar Forecasting

## 

## 

## 

## 

## 

## 

## 

## 

## 15.2 Building the Prediction Model

## 

## 

## 

## 

## 

## 

**16. References**

1. Sobri, Sobrina, Sam Koohi-Kamali, and Nasrudin Abd Rahim. "Solar photovoltaic generation forecasting methods: A review." *Energy Conversion and Management* 156 (2018): 459-497.
2. Lawin, Agnidé Emmanuel, Marc Niyongendako, and Célestin Manirakiza. "Solar Irradiance and Temperature Variability and Projected Trends Analysis in Burundi." *Climate* 7, no. 6 (2019): 83.
3. Wan, Can, 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* 1, no. 4 (2015): 38-46.
4. Inman, Rich H., Hugo TC Pedro, and Carlos FM Coimbra. "Solar forecasting methods for renewable energy integration." *Progress in energy and combustion science* 39, no. 6 (2013): 535-576.
5. Sun, Xinpei, and Tao Zhang. "Solar power prediction in smart grid based on nwp data and an improved boosting method." In *2017 IEEE International Conference on Energy Internet (ICEI)*, pp. 89-94. IEEE, 2017.