ReDI School of Digital Integration

DATA SCIENCE FINAL PROJECT

ReDI School of Digital Integration

COURSE: DATA SCIENCE

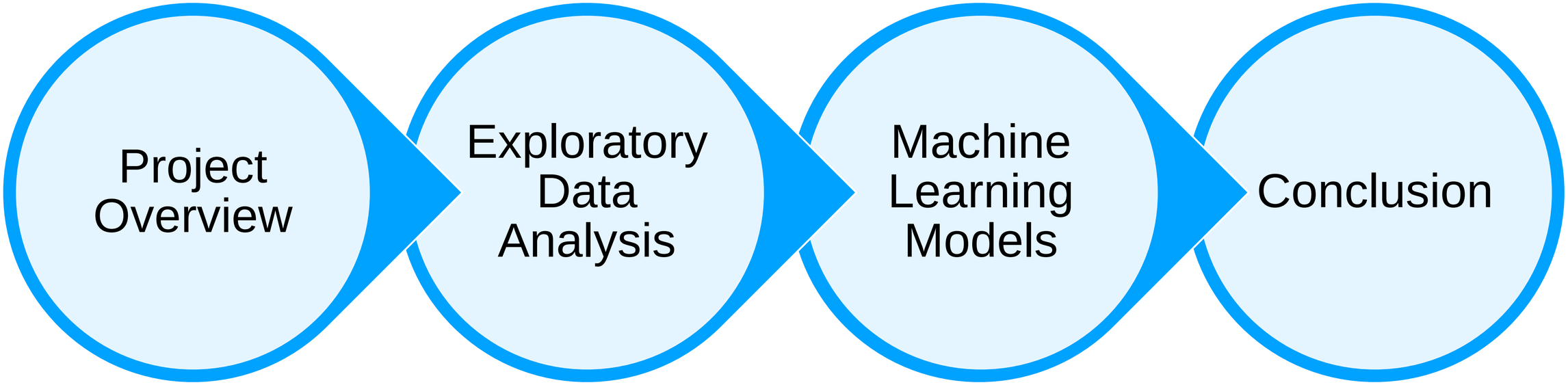
ReDI School of Digital Integration

TITLE : SOLAR PHOTOVOLTAIC ENERGY PREDICTION ÚSING MACHINE LEARNING MODELS

Presenter: Felicia Ojinji

June, 2021

Contents


Project Overview
Exploratory Data Analysis
Machine Learning Models
Conclusion

Project Summary

- The utilization of solar photo-voltaic (PV) systems for power generation has been on the rise owing to the environmentally friendly nature, durability, cost effectiveness and reliability such systems offer.

- Solar PV systems are highly dependent on meteorological conditions like solar irradiation, ambient temperature, humidity, windspeed etc. The changing nature of these meteorological conditions influence the generated output power with variations and uncertainties, creating an unbalanced demand (load) and supply (generation) link.

- Predicting the solar energy output is one of the ways to mitigate this imbalance, save operations cost and improve the planning and smooth operation of the system.

- Therefore this projects aims at developing Artificial Neural networks (ANN) and Linear regression(LR) models for predicting solar photovoltaic energy output, using weather variables as input and the solar energy as target.

- The Mbita WeTu Solar hub, a roof top solar PV system in Kisumu, Kenya is used as the case study off-grid solar PV system in this project.

- Weather and solar energy data of the PV system from 01/01/2020 to 31/ 01/2020, has been collected from Solcast.com and the MATLAB Simulink model of the 30kWp system respectively.

- The training data set, made up of 8670 data points of weather and power output in hourly frequency have been divided into two parts in which 70% is allocated for training, 30% for testing. The mean squared error (MSE), r-squared (R^2), have been used to evaluate the performance of the models.

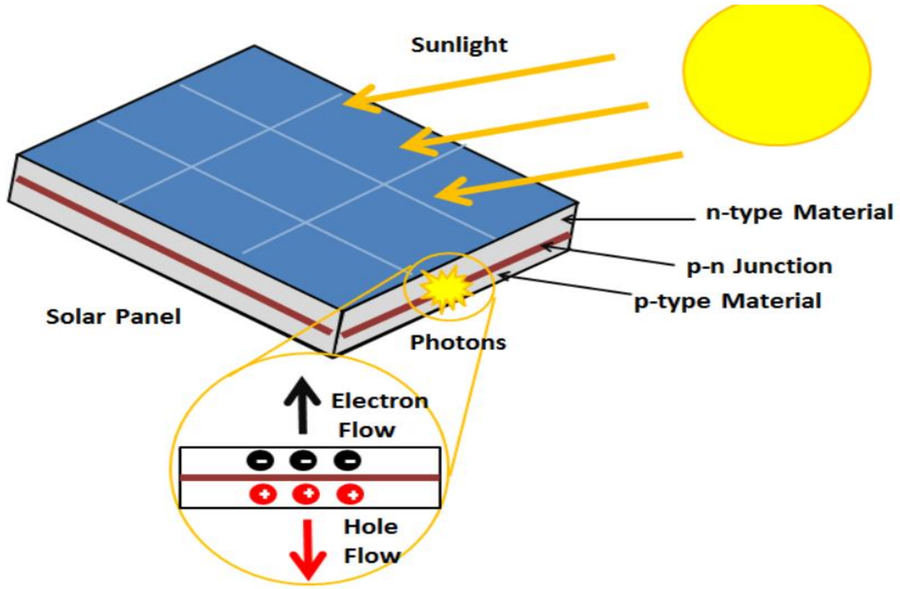
- From the performance evaluation results, the ANN model presents a better prediction with R^2 and MSE values of 0.927 and 3355kWh respectively compared to the LR model which has an R^2 value of 0.824 and MSE value of 8097kWh.

- Hence, the ANN model presents 12.5% increase in R^2 value and 58% decrease in MSE, compared the regression model.

Project Overview

Solar Photovoltaic (PV) Energy

- The light and heat that radiates from the sun is referred to as solar energy.
- Solar photovoltaic systems convert solar irradiation to electrical energy through the photovoltaic effect.
- Photovoltaic effect: The process of a material generating voltage and electric current when exposed to sunlight.



Project Overview

Why is solar photovoltaic energy important?

- There is no greenhouse gas emissions into the atmosphere from solar energy generation (clean energy generation).
- Solar energy is renewable because sunlight is abundant, reliable and will never run out.
- Environmentally friendly and sustainable form of electricity generation. It is good for the climate.

Problem Statement:

- PV power generation is known to be variable and intermittent in nature because of its high dependency on meteorological factors.
- The variable nature of solar photovoltaic systems disturbs planning, stability, and operations such as scheduling and unit commitment, making it difficult to match loads with the generated power, according to their priority.
- Load mismatch leads to surplus or deficit energy supply, hence extra cost.

Project Overview

Objective

- Develop Artificial Neural Network(ANN) and Linear Regression models for predicting solar energy generation, using Python Libraries.

Significance of the project:

- Solar Energy predictions will aid in load prioritizing in order to avoid or reduce the usage of energy storage systems (batteries), elongate the life span and save operations cost.

Project Methodology

- Data preparation: Pandas, NumPy
- Data visualization: Matplotlib, Seaborn
- Machine learning: Scikit Learn, Keras,

Exploratory Data Analysis – Data Preparation

Historical solar irradiation/weather data for Mbita, Kenya (year 2020) – Solcast.com

Solar energy output data – MATLAB Simulink model of the 30kWp Mbita Wetu Solar Photovoltaic Power System

- The rows present weather and solar energy recorded every hour from 01-01-2020 to 31-12-2020.
- 8760 rows (hours) * 18 columns (weather and solar energy data)

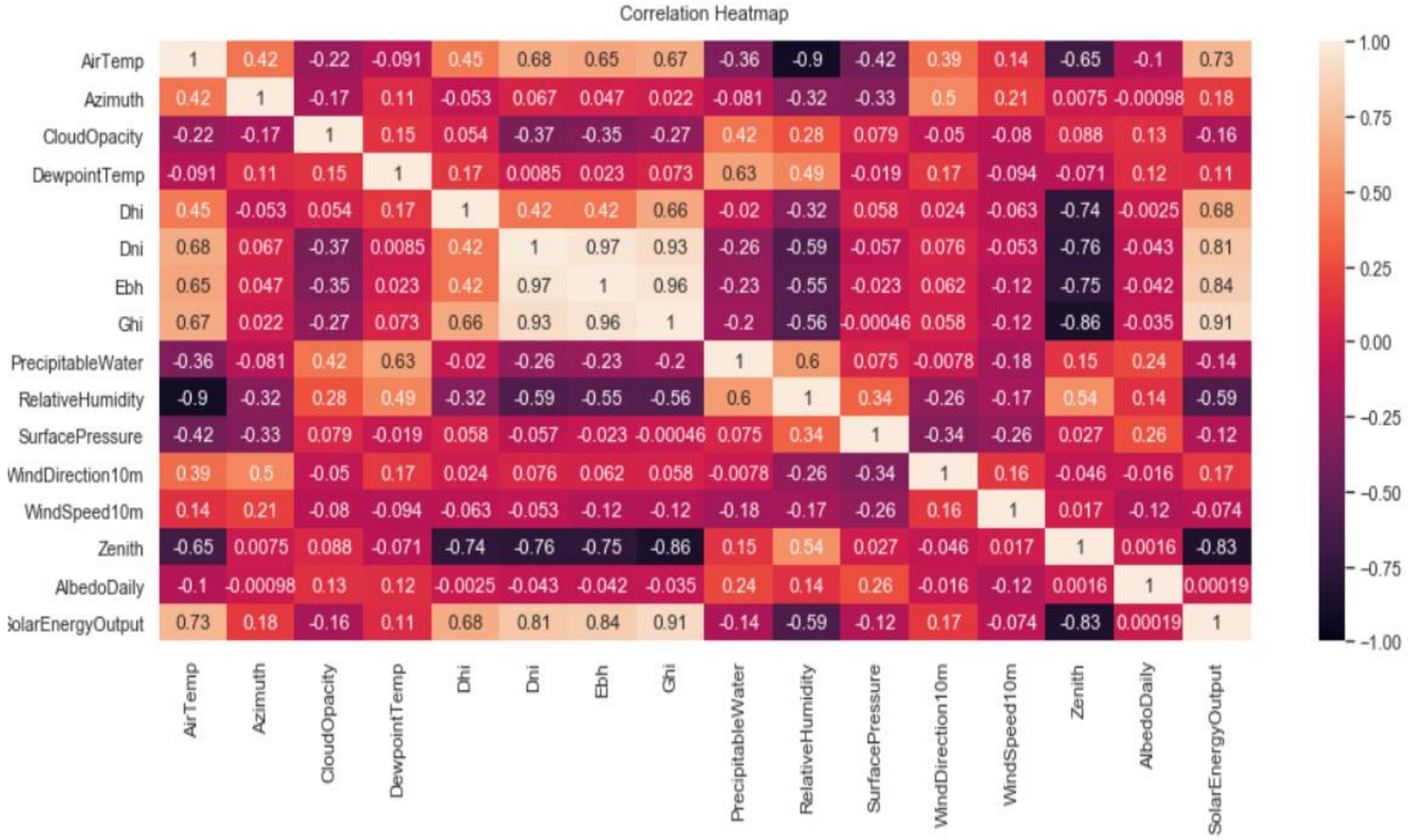
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8760 entries, 0 to 8759
Data columns (total 18 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   AirTemp             8760 non-null   float64
 1   Azimuth             8760 non-null   int64
 2   CloudOpacity        8760 non-null   float64
 3   DewpointTemp        8760 non-null   float64
 4   Dhi                 8760 non-null   int64
 5   Dni                 8760 non-null   int64
 6   Ebh                 8760 non-null   int64
 7   Ghi                 8760 non-null   int64
 8   PrecipitableWater   8760 non-null   float64
 9   RelativeHumidity    8760 non-null   float64
 10  SnowWater           8760 non-null   int64
 11  SurfacePressure     8760 non-null   float64
 12  WindDirection10m    8760 non-null   int64
 13  WindSpeed10m        8760 non-null   float64
 14  Zenith              8760 non-null   int64
 15  AlbedoDaily         8760 non-null   float64
 16  SolarEnergyOutput   8760 non-null   float64
 17  Date                8760 non-null   datetime64[ns]
dtypes: datetime64[ns](1), float64(9), int64(8)
memory usage: 1.2 MB
```

- No missing data

Exploratory Data Analysis – Data Visualization

Heat Map for feature selection

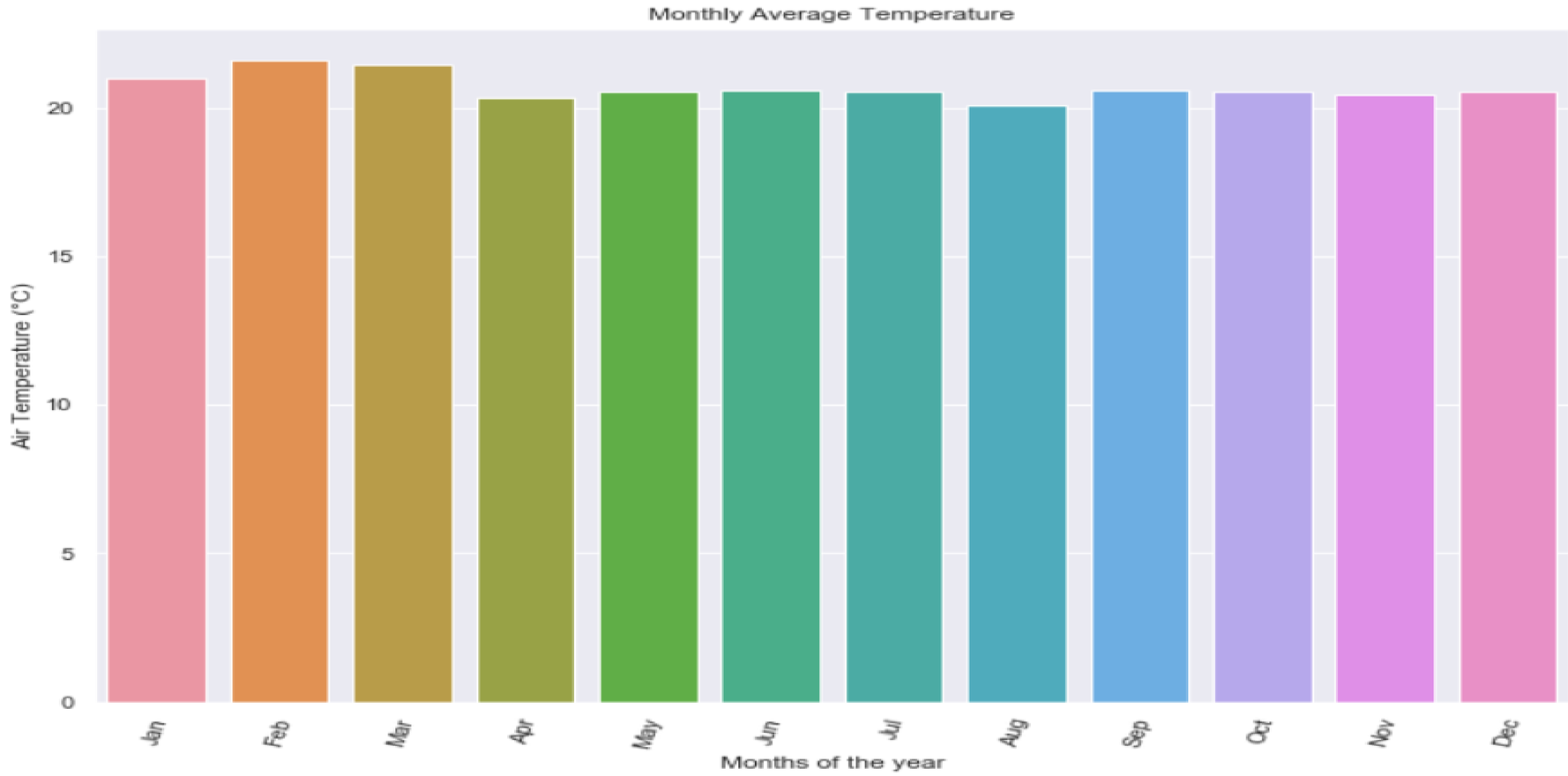


Highly correlated features to Solar Energy Output

- Dhi – Diffused horizontal irradiation
- Dni - Direct normal irradiation
- Ebh - Direct beam irradiation
- Ghi – Global horizontal irradiation
- AirTemp – Air temperature
- RelativeHumidity – Relative humidity
- Zenith – Zenith angle of the sun

Exploratory Data Analysis – Data Visualization

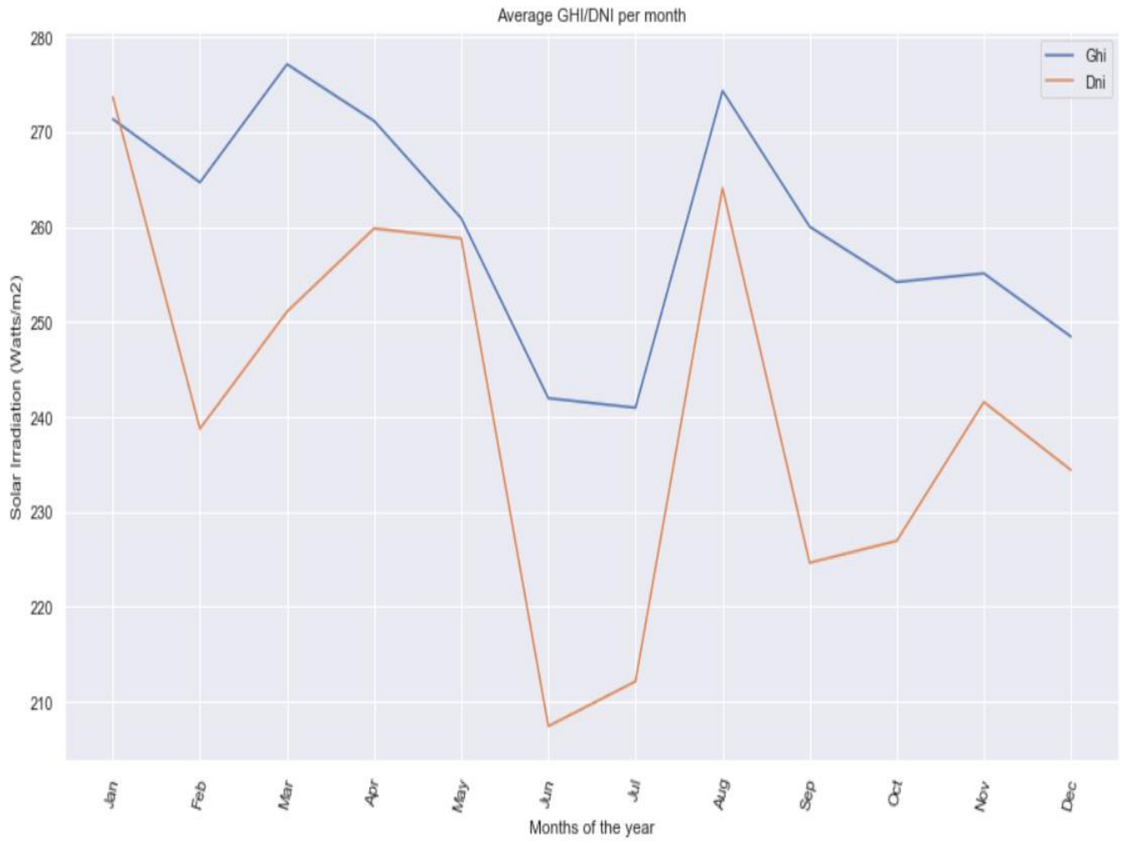
Monthly average temperature through out the year

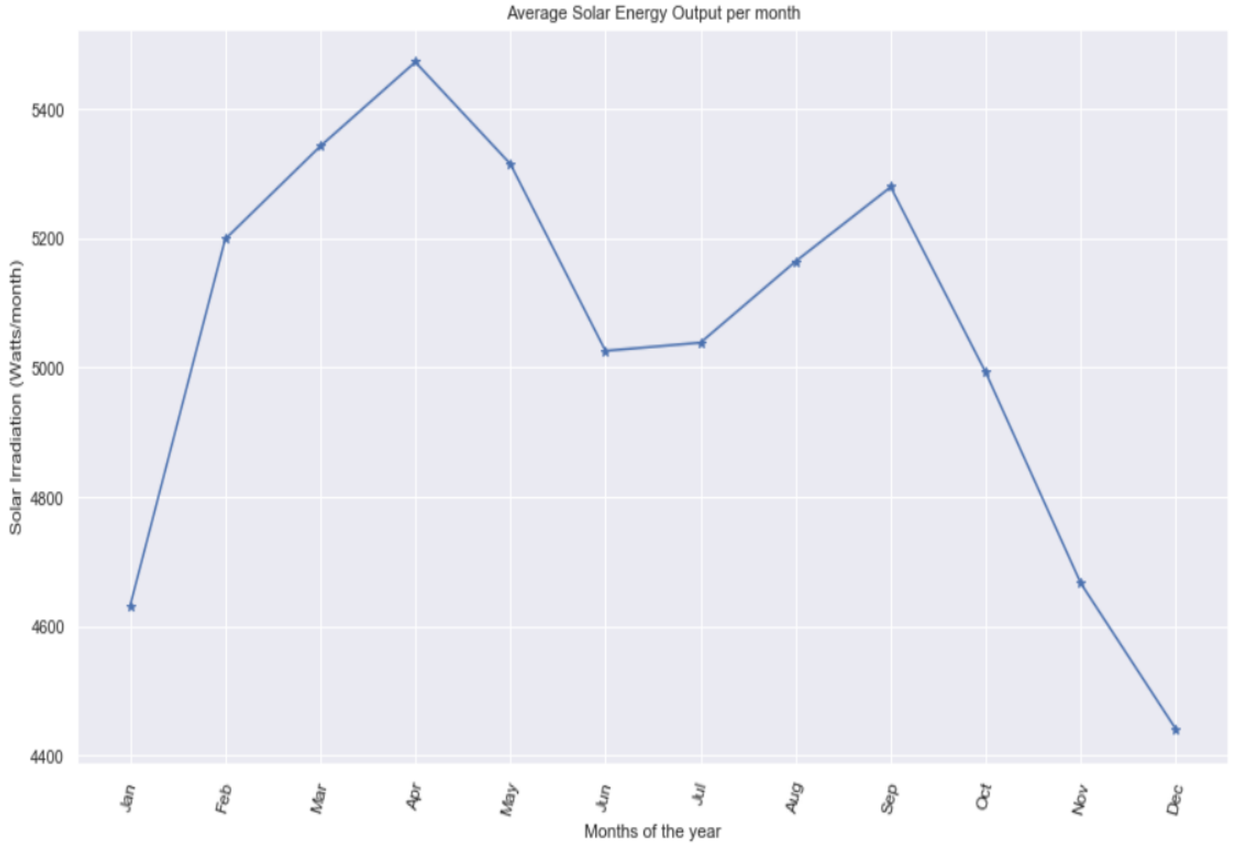


Observation: The temperature in Kenya is within the range of 20-22°C throughout the year.

Exploratory Data Analysis – Data Visualization

Average monthly solar irradiation and energy production trend through out the year





Observation: Solar Irradiation reduces from April – August hence solar energy production also reduces

Machine Learning Models for Solar Energy Prediction

Model Input:

Air Temperature

Diffused Horizontal Irradiation

Direct Normal Irradiation

Relative Humidity

Zenith Angle of the sun

Data Division:

Training Data – 70%

Test Data: 30%

Models:

Artificial Neural Networks

Linear Regression

Model Evaluation Metrics

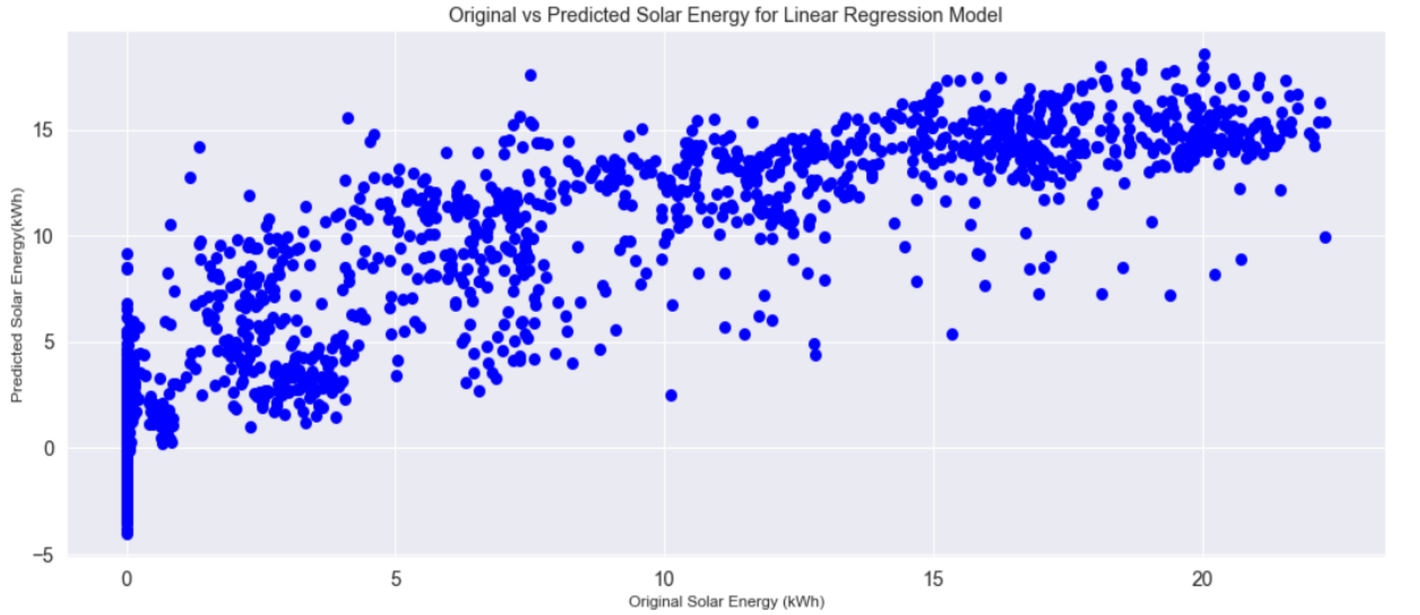
R-squared (R^2)

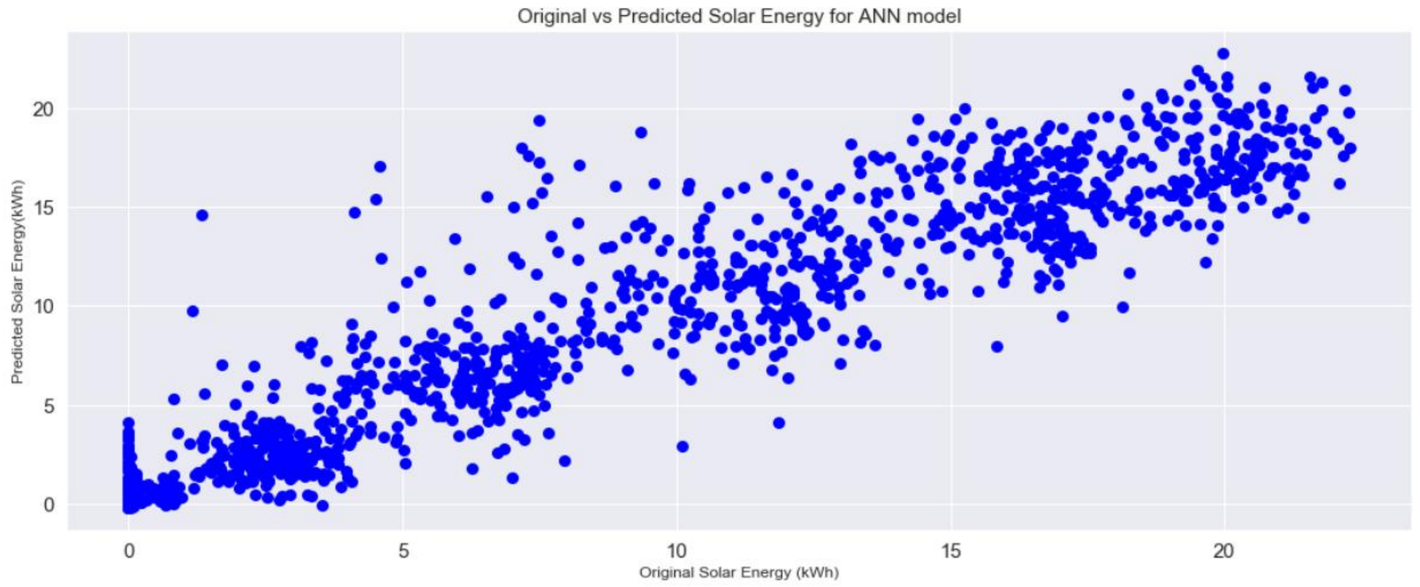
Mean Squared Error (MSE)

Model Target/ Output:

Solar Photovoltaic Energy

Results: Scatter plots showing original versus predicted solar energy





Conclusion – Model Performance Evaluation

Evaluation Metric	Regression Model	Artificial Neural Networks Model
R-squared (R^2)	0.824	0.927
Mean Squared Error	8097466.16 (8097kWh)	3355851.44 (3355kWh)

- The ANN model performs better with 12.5% increase in R^2 value and 58% decrease in MSE, compared the regression model.
- ANN is therefore recommended for more accurate solar energy ouput predictions.

ReDI School of Digital Integration

Thank you / Danke!

Felicia Ojinji

Data Science