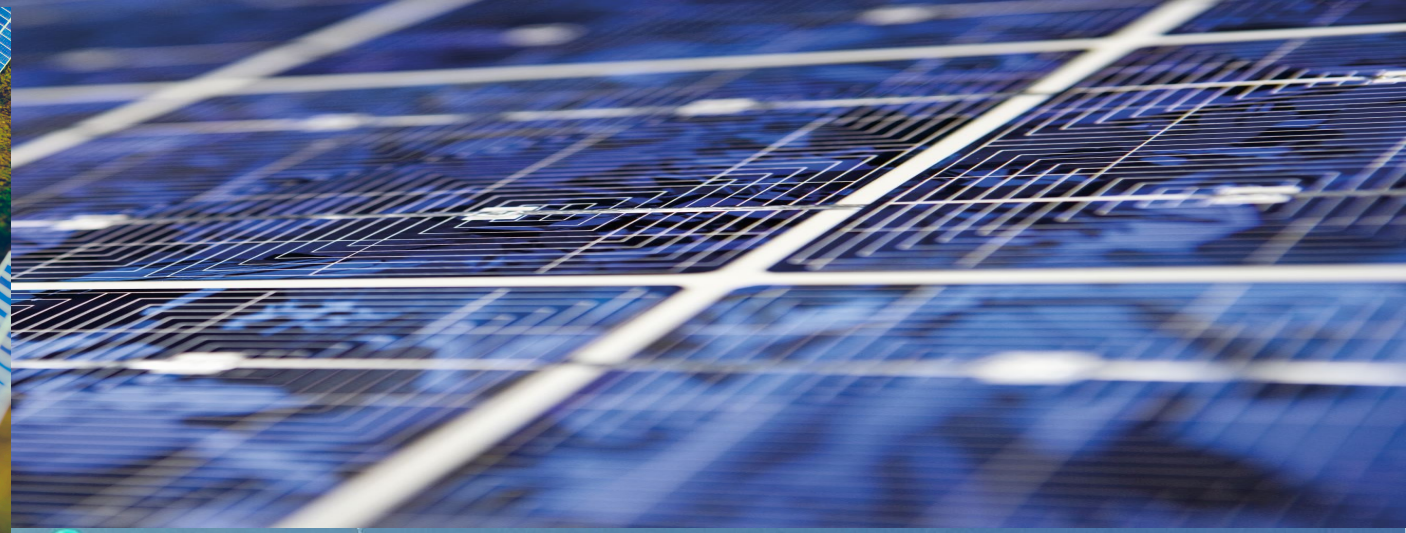


ENERGY PREDICTION FROM SOLAR PHOTO VOLTAIC CELLS

ABSTRACT

- The justification for home and business to invest in Solar Photo-Voltaic Cells require data to shows that the average daily or monthly production of energy from the Photo- Voltaic cells will offset the cost of purchasing power from the electric company. For this reason, there has been an increase to predict how much energy can be generated from Solar Photo-Voltaic cells. My plan is to design a neural network model to achieve this goal. Data for use on this project include 14 categorical variables and I obtain from Kaggle online community for data science but originally collect by HI-SEAS research station in Hawaii. Previously one group I know of, has carrier out similar research with this dataset using decision tree algorithms.

Introduction

- There is an increasing demand in alternative energy sources due to escalating prices of fossil fuels and pollution concerns.
- However, a bigger part of the population in the world has lacked the information they need to solve this problem.
- How can data scientist can help to pass on information to people so that we can come up with an optimal solution to this crisis.
- Data scientist can design models that can predict how much energy can be generated by alternative sources such as Solar Photo-Voltaic Cells. Then this information can be passed on to consumers to make an informed decision.

Motivation

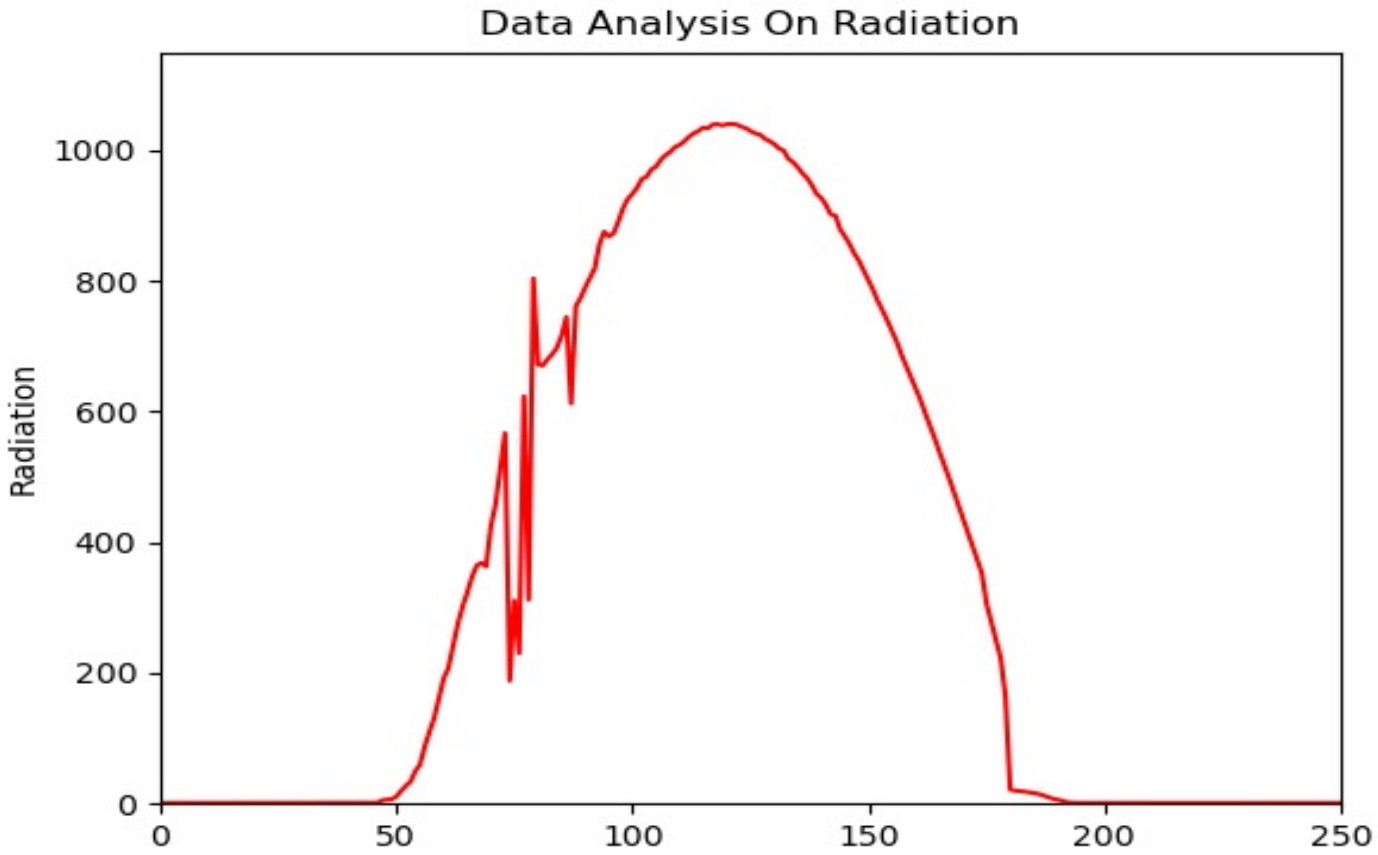
- Clean energy production: This will help reduce pollution in the atmosphere. By 2035 the contribution of power from renewable energy is expected to account for almost one third of total electricity out, with Solar Photo-Voltaic and wind as the main contributor.
- Reduction in global temperatures: Will help to reduce global warming.
- Reduce cost to home and business owners: Most government have awarded rebates to consumers who invest in alternative clean energy sources. This has led consumer in generating energy from their own installed systems.
- Stability of the electrical grid: Stability of power production in Photo-Voltaic Power Plant is essential for large scale grid connected systems since it affects the control and operation of the electrical grid
- Prediction of Solar energy production is very significant in order to attain the previously mention benefits
- The main motivation and objective of this research is to predict the hourly production and, consequently the monthly and yearly

Problem Formulation

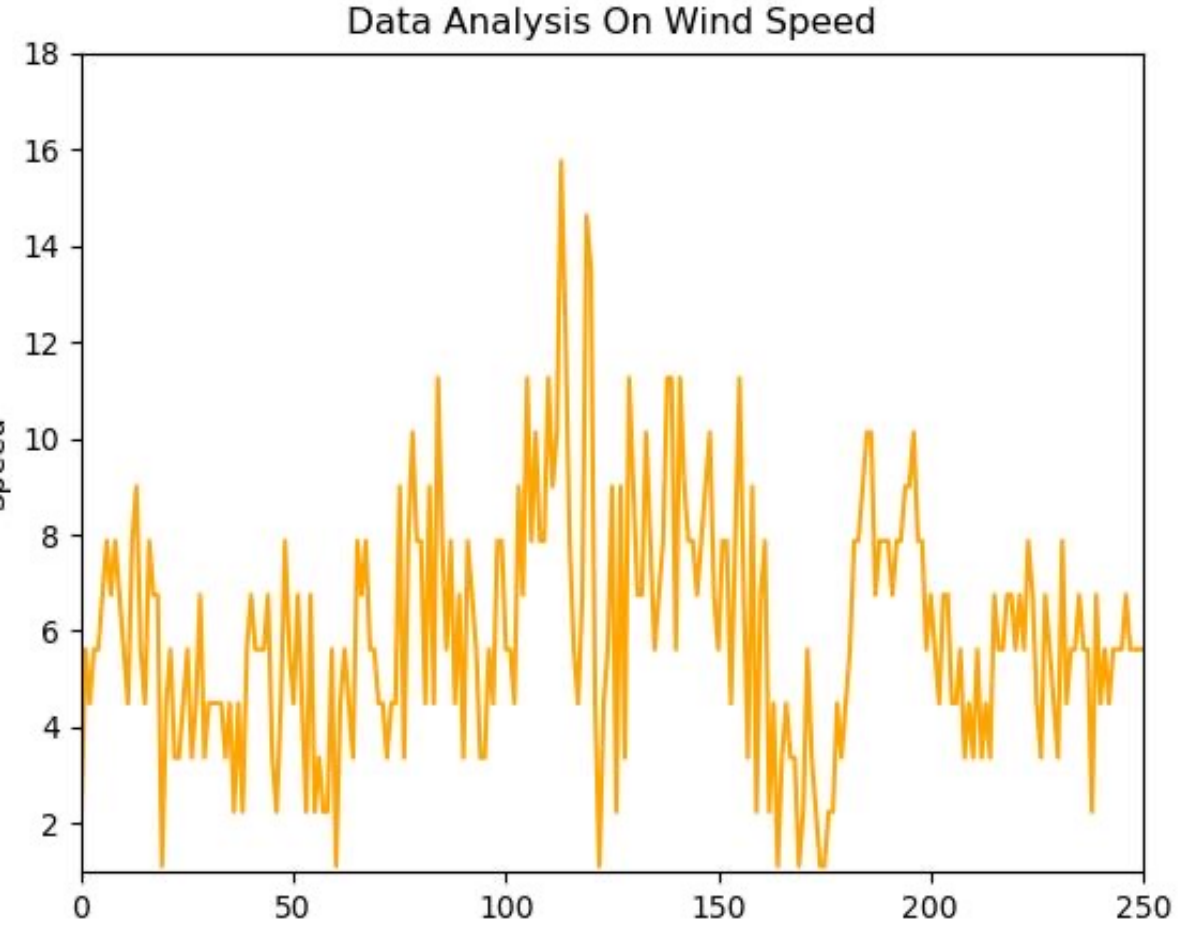
- There is previous work on this problem. And it found on github page below. https://github.com/Ghadamiyan/Solar-Radiation-Prediction/blob/46e835e717ee4aa9f161ae21b02d0f774c62debd/solar-radiation-prediction.ipynb
- In their work the used decision tree RandomForestRegressor models from sklearn. RFR uses a number of classifying decision tree.
- Data: It was collected by HI-SEAS research station located in Hawaii. But made available by Kaggle community datasite.
- Data Collection:
- The meteorological data consists of air pressure, time, humidity, wind speed, daily temperature, wind direction and global solar radiation. This data was recorded by a meteorological station name Space Exploration Analog and Simulation Weather Station,

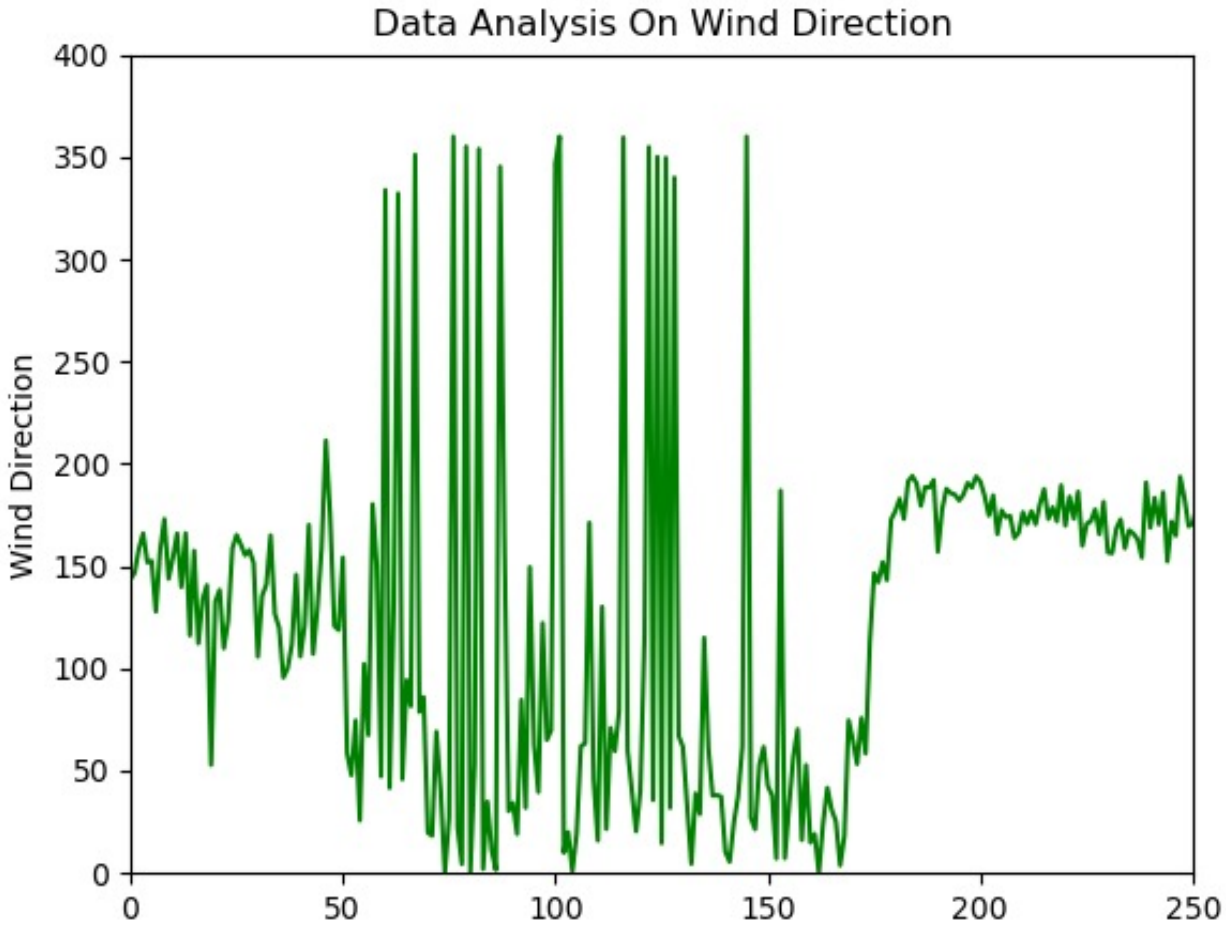
Problem Formulation

- The figure below show the graph of daily global solar levels in watts per square meter W/m²

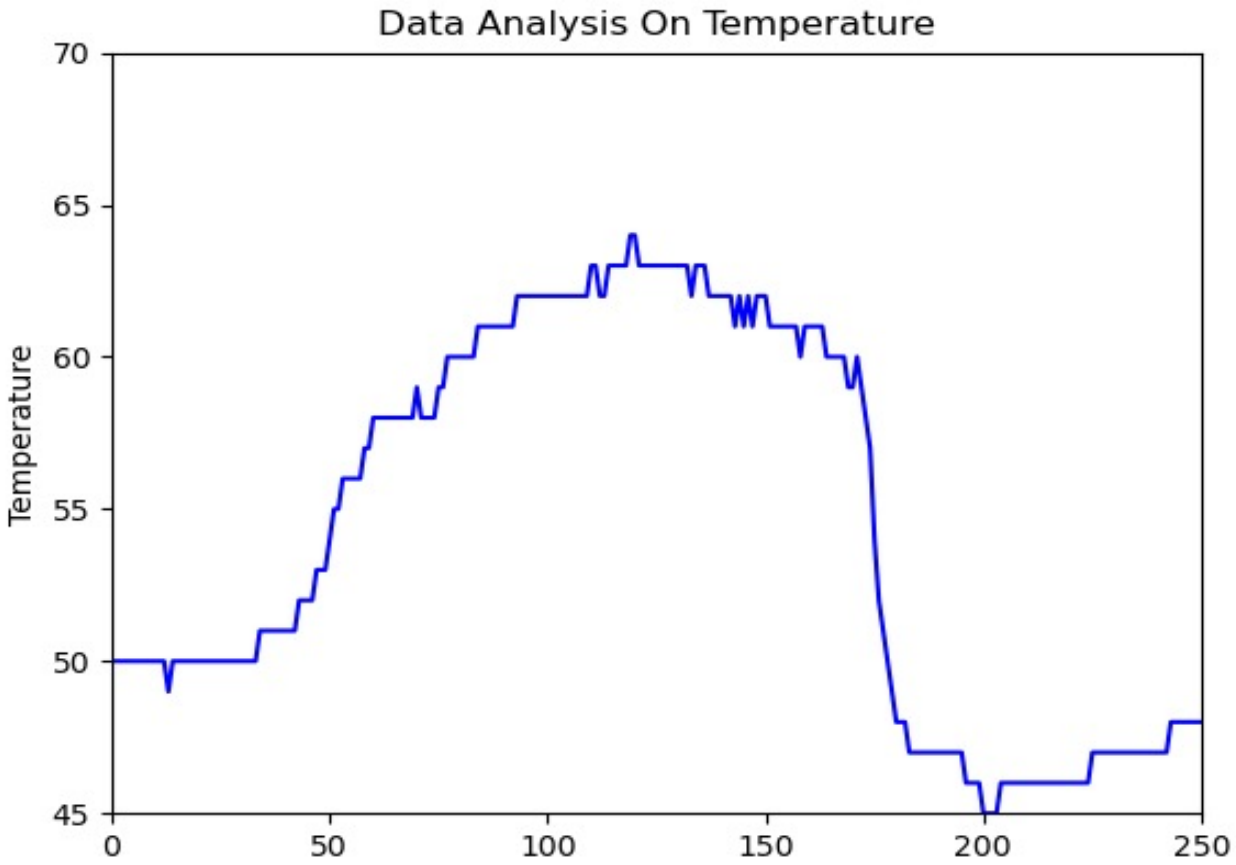


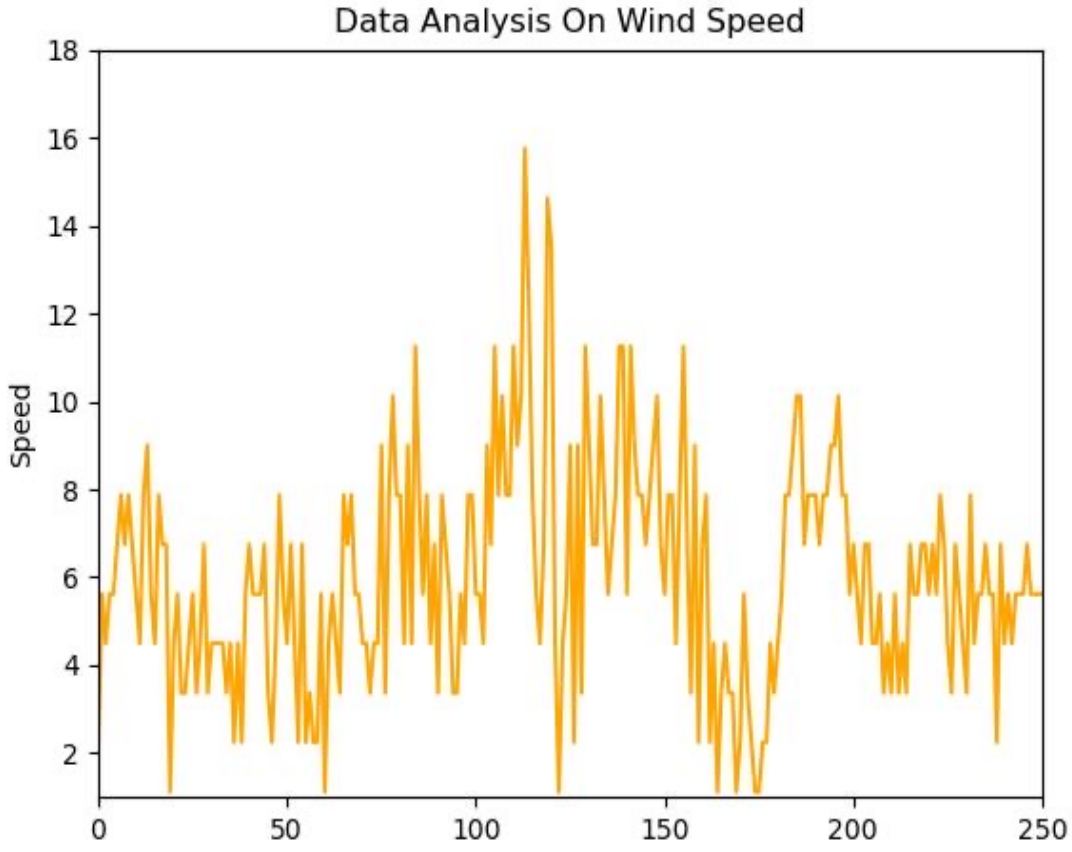
Problem Formulation





Problem Formulation





Problem Formulation

Problem Formulation

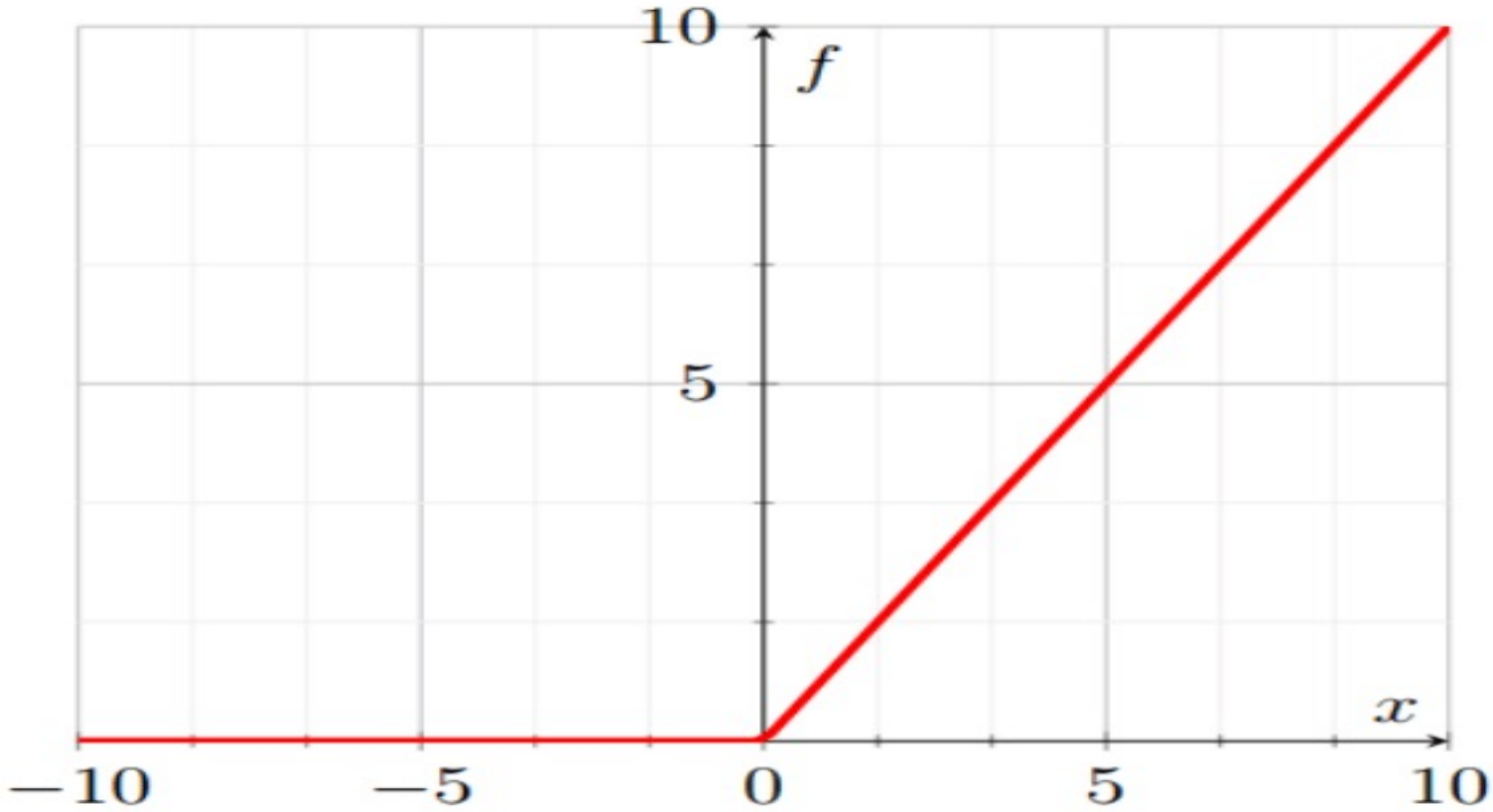
- From the graph and histogram above the data shows a lot of skewness. And after reviewing the histogram it shows that the deference in Solar Radiation value is considerably high. It ranges from 0.15 to 1000.
- Similarly, looking at the temperature, pressure and wind speed, the graphs shows high variation in the data.
- Due to inconsistences in the data, I'm planning to implement a model using the Multi-layer perceptron(MLP).

Problem Formulation

- Goal/task/ML algorithms:
- The goal is to design an Artificial Neural Net to predict energy production
- I as I note in the previous slide, I will be using the MLP algorithm.
- According to Marvin Minsky and Seymour it is impossible for a single perceptron to learn an XOR pattern
- In addition, I'm planning use other algorithms such as the rectified linear activation function(ReLU), this is a piece wise linear function that will output the input directly otherwise it will out put zero

Problem Formulation

- ReLU function f(x)= X+ = max(0,x)displayed in the graph below. Also
- Called ramp function , similar have wave rectification in electrical
- engineering.



Problem Formulation

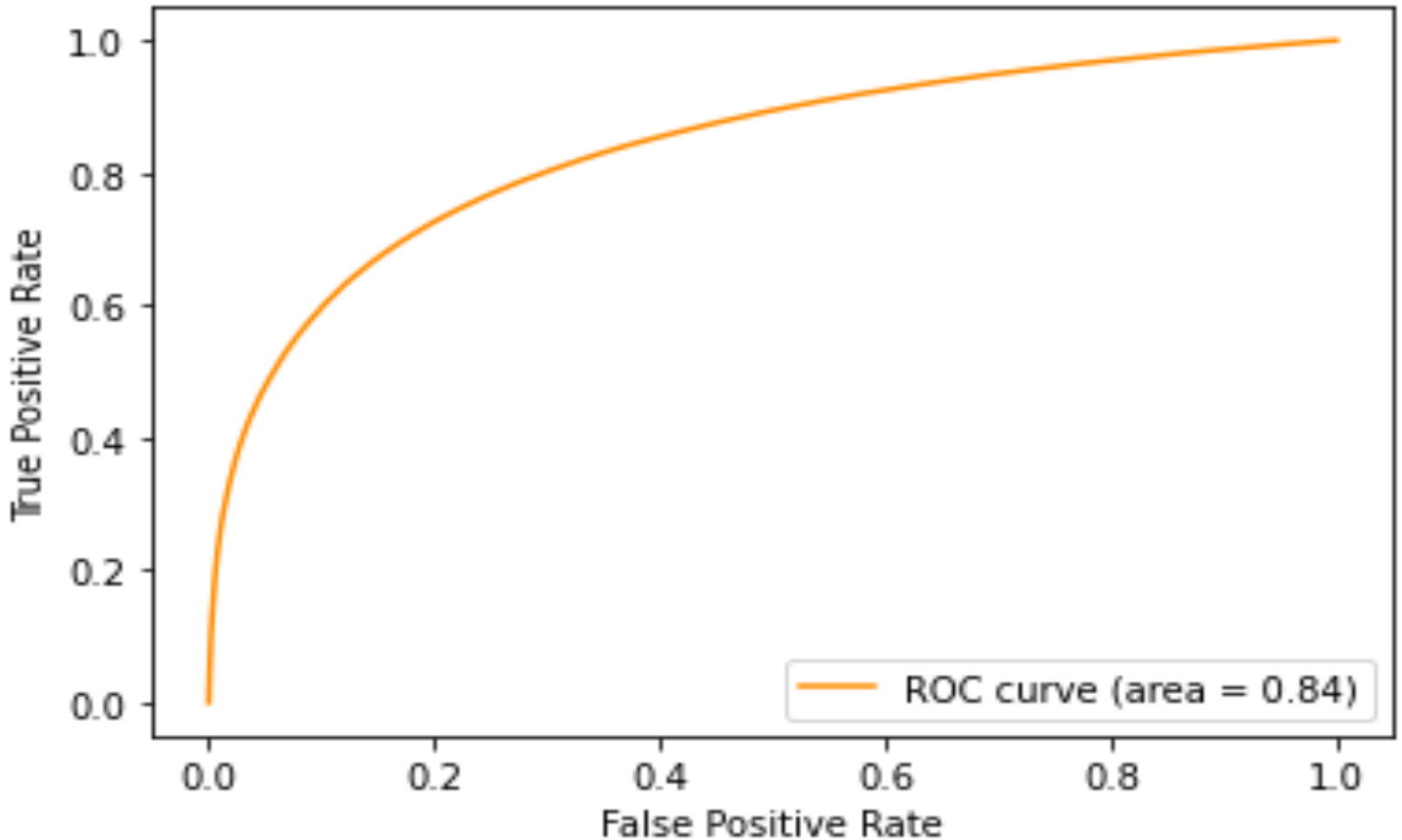
- Also, I will explore the the benefits of the Levenberg-Marquardt and Bayesian Regularization algorithms
- The Levenberg- Marquardt(LMA) also known as the damped least square is used to solve non-least square problems. What make it popular is that it combines Gauss-Newton algorithm and Gradient Descent.
- On the other Bayesian Regularization has the ability to reveal potentially complex relationships. But both techniques can be combined to achieve better results

Problem Formulation

- Libraries to be used include, Pandas , Tensorflow, Sklearn, numpy, Matplotlib and several others, as it may be required.
- Tools to be used:
- Python soft-ware but if time allows, I might try with other soft languages such MatLab or R.
- Hardware:
- I will use Mac Pro 13-inch,2020. Processor 2GHz-Quad core Intel Core i5. Memory 16 GB 3733MHz LPDDR4X
- Graphics: Intel Iris Plus Graphics 1536 MB
- As a back up, Dell E5470 intel core- i7 vPro with 16GB Memory

Problem Formulation

- Metrics for train and asses the algorithm:
- Multi-layer Perceptron
- Levenberg-Marquardt and Bayesian Regularization
- Classification Accuracy, it is number of correct prediction divided to total number of prediction
- Precision. = True-Positive/(True-positive+False-Positive. It is used in cases where Classification Accurancy is not good indicator.
- Mean Squared Error : $MSE = 1/N + \sum^{n}_{a=1}(y_i - y_j)^2$, N is the number of sample , y_i is the ground truth and y_j is the calculated output value.
- ROC (Receiver operating characteristic curve. It depicts the


1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate
ROC curve (area = 0.84)

Problem Formulation

- Training, Testing and validation.
- 1. Splitting the data
- Train/Test split:
- I intend to use 70% of the data for training.
- Of the remaining data 20%
- Holdout set.
- And 20% for holding out. This is very help in case of sampling bias.
- There are other methods such as K-fold cross validation, nested cross validation and many others that can be used

Work-plan

- Clarification of expectation:
- Need to pinpoint out the goals of the project.
- Estimate how much time I need to accomplish the task.
- What is the date I'm supposed to be done
- What resources do I have, and which ones are missing.
- How to develop the plan:
- Will I need some type of collaboration. If yes, identify it.

Work-plan

- Reverse engineer the outcome:
- Start from the end of the project to beginning braking down steps of how to achieve the goal.
- Identify potential problems that I will be facing.