Claraestela Torres Rivero

ctorresr

12-760 Fall 2022

Power Output Prediction of Horizontal Photovoltaic Solar Panels

Project Proposal
September 28, 2022

Source: Cornell Chronicle (Jan 6, 2017) [1]

1. Introduction

Sunlight converted into thermal or electrical energy is known as solar energy. This is the most abundant and clean renewable energy source available, and solar technology can capture this energy for a variety of uses, including the production of electricity, interior illumination, and water heating for domestic, industrial, and commercial use. The amount of sunlight that reaches Earth's surface in an hour and a half is enough to supply all the planet's energy needs for a whole year. And two ways of concentrating solar radiation and turn it into electrical energy are Photovoltaic (PV) panels or mirrors [2].

This project predicts, with different machine learning techniques, the solar output of horizontal photovoltaic solar panels with weather and location dataset from Kaggle. This data provides the following columns: location, date, time sampled, latitude, longitude, altitude, year and month, month, hour, season, humidity, ambient temperature, power output from the solar panel, wind speed, visibility, pressure, and cloud ceiling for 12 Northern hemisphere sites where the horizontal photovoltaics are installed over 14 months [3].

2. Problem definition

Currently around 80% of the world's energy comes from nonrenewable energy. Specifically, from fossil fuels (petroleum, natural gas, and coal). This is a big climate change issue since when they are burned, they release large amounts of carbon dioxide, a greenhouse gas, into the air, which traps heat in the atmosphere, causing global warming, biodiversity loss and species extinction, food scarcity, and health issues. Therefore, solar energy is an alternative to tackle this problem, but this depends on sunlight to effectively gather energy. And although this system can still collect energy during cloudy and rainy days, the efficiency drops and since there are marked seasons in most on the United States, and some cities received less unlight that others, this system is still not been widely used. Hence this project will use machine learning techniques including, K Nearest Neighbors KNN, Decision Trees, clustering and Random Forest, to predict the solar output of Horizontal Photovoltaic Solar Panels with information gathered from the Northern hemisphere.

3. Engineering background

This section should explain in more detail the method you are using to solve your problem. For example, if you are using machine learning, image recognition, finite elements, etc, I want to know that you have a basic understanding of the algorithms. Include equations, diagrams, and figures to explain what you are doing and why you are doing it. Cite other people's work within the text and list the references in the reference section.

Include citations to all the web pages, articles, and books that you plan to use to complete your project.

3.1 Detailed background about topic 1

If you need to break up a section into subsections, use Heading Level 2 for the subsections. You can find the style definitions in the **Style** section of the Home tab in Word.

If you are defining mathematical variables, format them as they would be in a textbook.

- v_0 Velocity at time 0
- a_0 Acceleration at time 0

4. Computational methods

Provide a technical description of the computational methods you need to solve your problem. This section should be written for someone who is an engineer, but not a specialist in this area. Yourmust include references and they must be archival references, not web pages.

5. Project plan

6. Implementation

To begin, I will explore the variables that are in the dataset, to understand what they mean, that information I can get from them and their relevance in this project. Following, I will perform pre-processed techniques to clean and get the data ready before implementing the machine learning algorithms. I will also need to find out the collinearity between the predictor variables, highly correlated variables might lead to errors in the prediction process.

moreover, since some of the variables are categorial, I need to encode them to ensure I can account for them and perform quantitative analysis not hem. After this, I will then split the data into train and test, the first one will be used to teach the model and train it and the second one will be used to predict the results. Following this step, I will fit the models. Once this is done, I will make sue of the metrics from the sklearn module to evaluate the models and provide conclusions of the energy output prediction.

7. Conclusions

This project will deliver different models which will give a prediction for solar panel output. Moreover, I will assess these models, validate them, and baes on the results, show which one performs better and gives a trustable prediction. Additionally, this will tell which variables affect the most the target variable, which is the solar output.

8. References

[1] Friedlander, B. (2017, Jan 6). *With three new solar farms, Cornell skims energy from the sun*. Retrieved from Cornel Chronicle: https://news.cornell.edu/stories/2017/01/three-new-solar-farms-cornell-skims-energy-sun

[2] Solar Energy Industries Association. (n.d.). Accessed on sep Sept 27, 2022. Retrieved from Solar Energy: https://www.seia.org/initiatives/about-solar-energy

[3] Kaggle. (n.d.). Accessed on sep Sept 27, 2022. Retrieved from Horizontal Photovoltaic Power Output Data: https://www.kaggle.com/datasets/saurabhshahane/northern-hemisphere-horizontal-photovoltaic?select=Pasion+et+al+dataset.csv%3B+accessed+on+sept+27%2C+2022.