CS235 Fall'22 Project Proposal: Solar Power Generation Prediction

MANISH CHUGANI #1, NetID: mchug002
SUMEDHA GIRISH ATREYSA#2, NetID: satre002
ALISHA KULKARNI #3, NetID: akulk018
NAVEEN DEVADI #4, NetID: ndeva005
TEJAS MILIND DESHPANDE #5, NetID: tdesh006

Additional Key Words and Phrases: Deep Learning, Neural Networks, Wildfire Prediction, Data Mining, Data Analysis

ACM Reference Format:
Manish Chugani #1, Sumedha Girish Atreysa#2, Alisha Kulkarni #3, Naveen Devadi #4, and Tejas Milind Deshpande #5. 2022. CS235 Fall'22 Project Proposal: Solar Power Generation Prediction. 1, 1 (October 2022), 2 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The benefits of solar energy as we know are innumerable, especially in the field of generating electricity. Solar power is a pollution-free, renewable source of electricity which makes it highly significant in today's day and age. Even with its vast benefits, solar power accounts for just 3.6% of the power generated across the globe. Predicting the generation of solar power plays a crucial role in increasing efficiency, thereby increasing its production. Estimating the amount of power generated will lead to improved grid management, efficient solar panel maintenance and quicker identification of faulty equipment in the system. The proposed approach compares various data mining techniques used to predict the generation of solar power.

1.1 Project Type

The proposed approach falls under the category of a Software project as it implements existing methodologies and compares the results obtained.

2 PROBLEM DEFINITION

Harnessing solar energy is not very easy because of various factors like conversion rate of solar radiation to electricity, weather conditions, landscapes, cost and maintenance of equipment, etc. In order to optimize the process of harnessing solar energy while ensuring systemic efficiency and reduced maintenance costs, monitoring and deriving patterns from past data is highly essential. Using strong data mining techniques and statistical analysis, the prediction of power generated on a daily basis by a solar power plant will benefit the consumers and equipment manufacturers alike since any kind of mishaps can be mitigated and maintenance and solutions can be planned ahead of time. The proposed

Authors' addresses: Manish Chugani #1NetID: mchug002; Sumedha Girish Atreysa#2NetID: satre002; Alisha Kulkarni #3NetID: akulk018; Naveen Devadi #4NetID: ndeva005; Tejas Milind Deshpande #5NetID: tdesh006.

Manuscript submitted to ACM

approach plans to compare various data-driven estimators and arrive at a conclusive methodology that can be followed with the least possible erroneous technique so that fewer resources are wasted and coherence and productivity are maintained.

3 DATASET DESCRIPTION

The dataset titled Solar Power Generation[1] contains data from two solar power plants for a period of 34 days. It has two pairs of files, each pair having data about the amount of power generated by the plant and a dataset having sensor readings for the day. The sensor data which is gathered at a plant level consists of the ambient temperature, module temperature, and irradiation. The dataset, which is gathered at the inverter level where each inverter has multiple lines of solar panels attached to it, consists of the amount of DC Power and AC Power generated in 15 minute intervals and the daily yield and total yield of the plant. The daily yield is a cumulative sum of the power generated on that day until that point in time. The total yield is a cumulative sum of the power generated by the inverter until that day at that point in time.

4 PROPOSED APPROACH

Since the dataset contains sensor data from 2 plants and each plant contains one sensor for the entire plant, the proposed approach is to use the 'DATE_TIME' field to map the corresponding sensor values as features to the 'DC_POWER', 'AC_POWER' and 'DAILY_YIELD' as labels. The idea then is to use the features in estimators for regression such as Multi-Variate Linear Regression, Support Vector Regression and RANSAC Regression and compare it with the proposed model which is a non-linearity based Shallow Multi-Layer Perceptron/Feed-Forward Artificial Neural Network.

5 EVALUATION PLAN

The proposed approach aims to solve a regression analysis problem and predict the power generated by the plant in the next few days when given the sensor readings on a particular date and time. Given the magnitude of the values of 'DAILY_YIELD', the proposed evaluation metrics for 'DAILY_YIELD' are Mean Absolute Error(MAE) and Root Mean Squared Error(RMSE) since a small variation in predicted power would not be amplified by these metrics. For the estimation of 'AC_POWER', Mean Squared Error(MSE) is to be used to amplify any inaccuracy in the estimator since the generated values are relatively smaller in magnitude. The labeled dataset will be split into 'train', 'test' and 'validation' sets with the split being 80%, 10% and 10% respectively in order to verify the bias and variance of the estimator.

6 PROJECT TEAM & PROJECTED LABOR DIVISION

Throughout the quarter, each team member intends to implement one data mining methodology. The results of each method will be evaluated and the results will be presented in the concluding stages of this project.

[1] https://www.kaggle.com/datasets/anikannal/solar-power-generation-data