AASD 4016 Full Stack Data Science Systems.

Students: Diana C. Lopera ID 101495151, Zarina Dossayeva ID 101385370, Juan M. Henao ID 101488061,

Project: Solar Power Forecasting

Introduction

For the project of this subject, we approached the solar power forecasting problem. The scope of this problem involves being able to predict the amount of power that is going to be generated by a solar photovoltaic plant in the next couple of days. The solar power forecasting can be divided into "nowcasting" for prediction of some minutes and a couple hours, short-term predictions for the following hours and days and Long-term for predictions greater than a week. This problem is very significant because it allows both great and small photovoltaic power plants to schedule ahead and program the connections and disconnections to the electric grid, and also manage efficiently the maintenance of the power plant, this is crucial since these kinds of installations have very slow ROI (usually 15 – 30 years).

Data

The dataset used for this project was found on Kaggle https://www.kaggle.com/datasets/anikannal/solar-power-generation-data. This dataset contains observations about AC, DC, Irradiation, and temperature taken for 2 different power plants every 15 minutes for 34 days. This data set was used for the initial development and testing of our product.

Past Projects

Regarding the state of the art in this topic we found several points that are worth discussing. First, there are already some companies that offer some kind of service like the one developed on this project, for example AESO https://www.aeso.ca/aeso/ which is a Canadian company operating in North America. Also there are several paths to come up for a solution for the power forecasting, classic statistical models like ARIMA regression, monitoring by satellite and of course machine learning algorithm, so why do this if there are already persons in the market? simple because the biggest market for this is yet to be explored, countries that have a high solar photovoltaic potential like Colombia, Ecuador, and Brazil are still lacking solutions like the one we are proposing that can be easily implemented.

ML – Canvas:

Our creative and design process was highly influenced by the tool of ML-Canvas. The complete image of our ML-Canvas can be found at the end of this document.

Model Benchmark:

To develop our model, we trained 3 different models. 2 of them with the main purpose of benchmarking and a final one to be deployed on our MVP. The first model was a simple Artificial Neural Network, afterwards we moved on to the initial development of an LSTM model, and we further fine tuned it into a second version of LSTM model. In the end we tested 3 different models that improved performance for each iteration. It's worth to mention that ARIMA regression models are also very famous regarding this topic and are widely used, they could also be a very good point of comparison for benchmarking, a table with a summary of the errors related to each of the models is presented below. Minimum value and Maximum value in the ranges of the predicted value are displayed for comparison.

	ANN	LSTM	LSTM_2	Min_Value	Max_Value
Average Error [W]	83,026.00	31,513.00	19,381.00	0.00	200,000.00

Table 1. Error for different models

Model Deployment:

Model deployment was executed using Flask, Docker and GCP. The initial notebook was turned into a python script and then turned into an API using Flask. Finally, the product was deployed on Google Cloud Platform. Although the product works as expected there are a couple of things that we should mention regarding the challenges that can be found for this kind of product. First, the product is prone to give wrong predictions when weather behaves unexpectedly, this problem can be addressed by correctly scheduling a re training action on the API of our product. Also, the product currently does not consider how drastically the efficiency of solar power cells drops when maintenance is needed. It's very well known that cleaning is a main characteristic that affects the efficiency of solar power cells, this aspect can lead to wrong predictions even if all the other variables behave as expected. In an ideal world also, we would be able to include more metrics that would enrich our feature space allowing the model to have an even better performance since during training it was noticeable how the model is hungry for data.

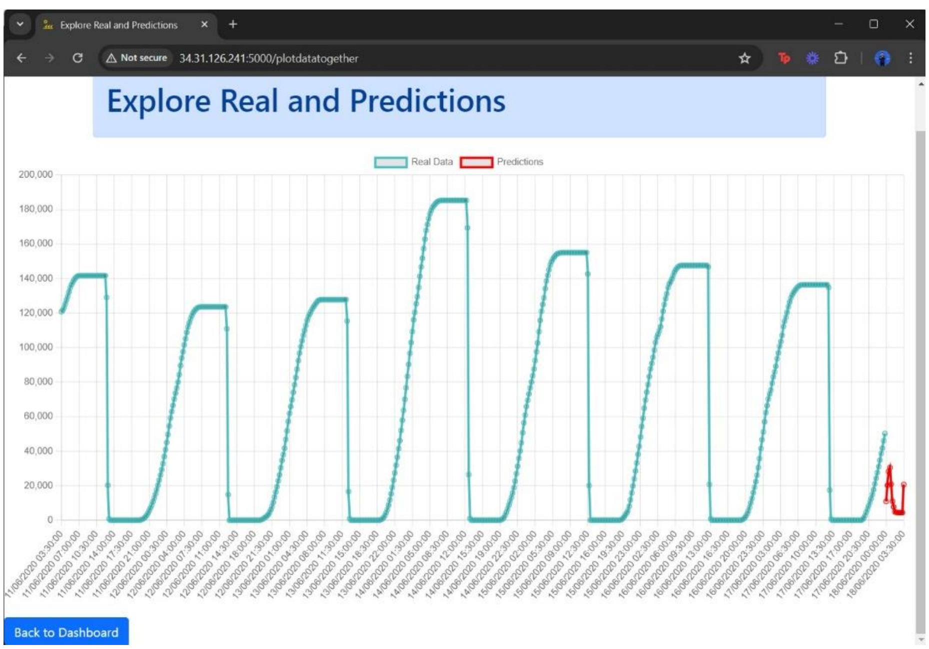

Figure 1. Historic data and predictions on our App

Conclusions:

The main conclusion of our product is how easy it is to implement and the great benefits that it can draw out of the solar power plants, specially helping maintain a good health of the hardware along time which is a crucial characteristic of solar power installations due to their long ROI. In future versions, automatic re training, and recommendations regarding actions that can be taken to prevent damage to the solar power plants can be included to the product.

Annex:

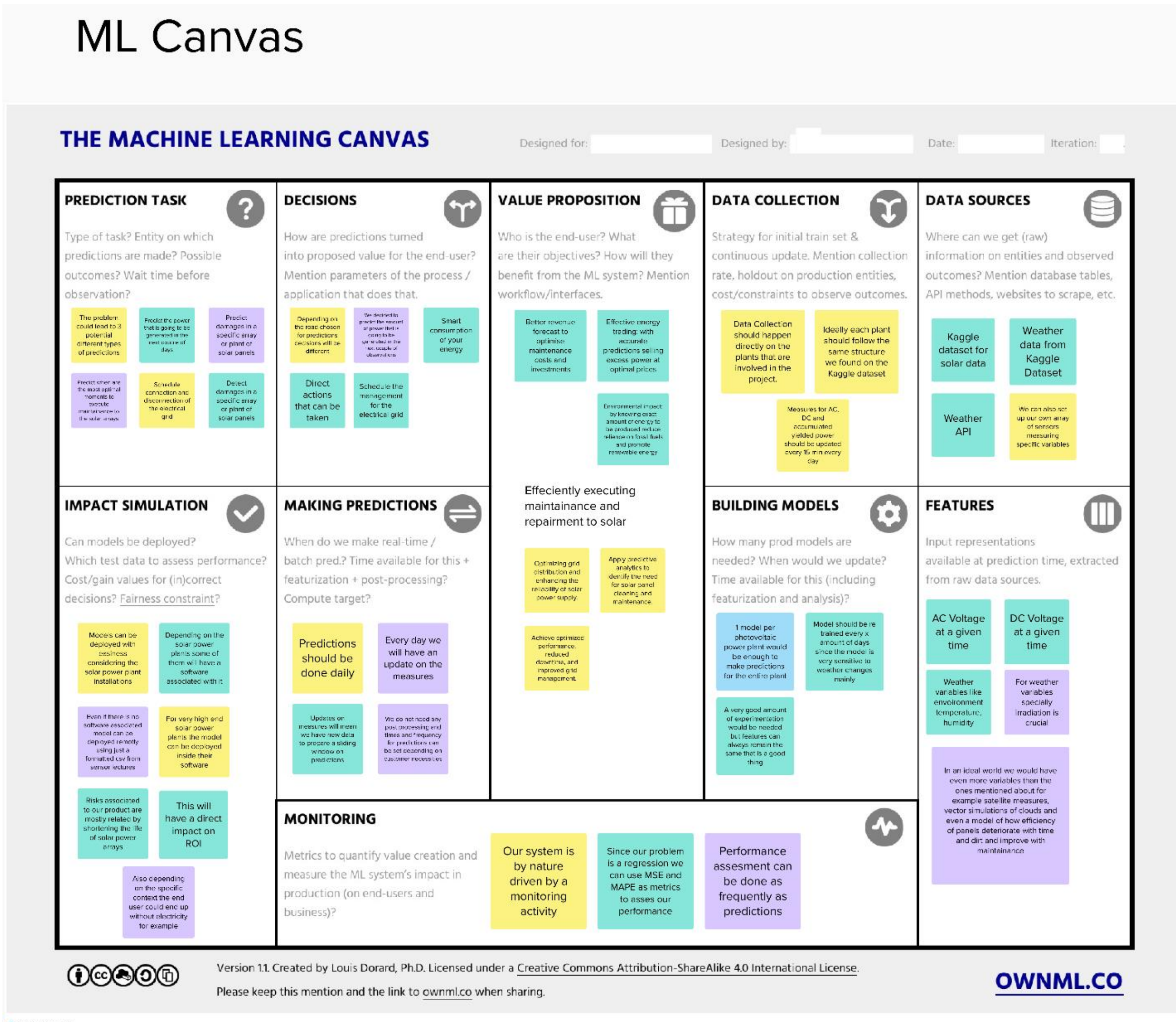

Figure 2. ML-Canvas

Links:

Visit our app: http://34.31.126.241:5000/

Visit our Github repository: https://github.com/catalinalopera/PowerForecasting/tree/main