Solar Power Forecasting

Diana C. Lopera - 101495151
Zarina Dossayeva - 101385370
Juan Henao - 101488061

Full Stack Data Science Systems

Agenda

01

About the problem

Solar Power Forecasting

02

Significance

Solar Power Plants

03

Past Projects

State of the art

04

ML Canvas

Our approach

Agenda

05
Business Case

06
Model Benchmark

07
Model Deployment

08
Demo and Q&A

01

About the problem

Solar Power Forecasting

Persona

- **Name:** Jorge Ramirez
- **Age:** 32
- **Location:** Armenia, Colombia
- **Occupation:** Small Farm Owner
- **Family Status:** Married with two children

Jorge relies on solar power to operate his irrigation system and is keen to predict the power output from his solar generator more accurately. This would allow him to plan his irrigation schedules more efficiently, ultimately reducing his dependence on costly diesel generators.

Jorge: “I need to know when the sun will power my farm, not just when it shines.”

Solar Power Forecasting

In Colombia, a country blessed with vast solar potential yet cursed with energy rationing, small solar power generators face a dilemma. Solar intermittency is one of the major barriers for solar energy implementation.

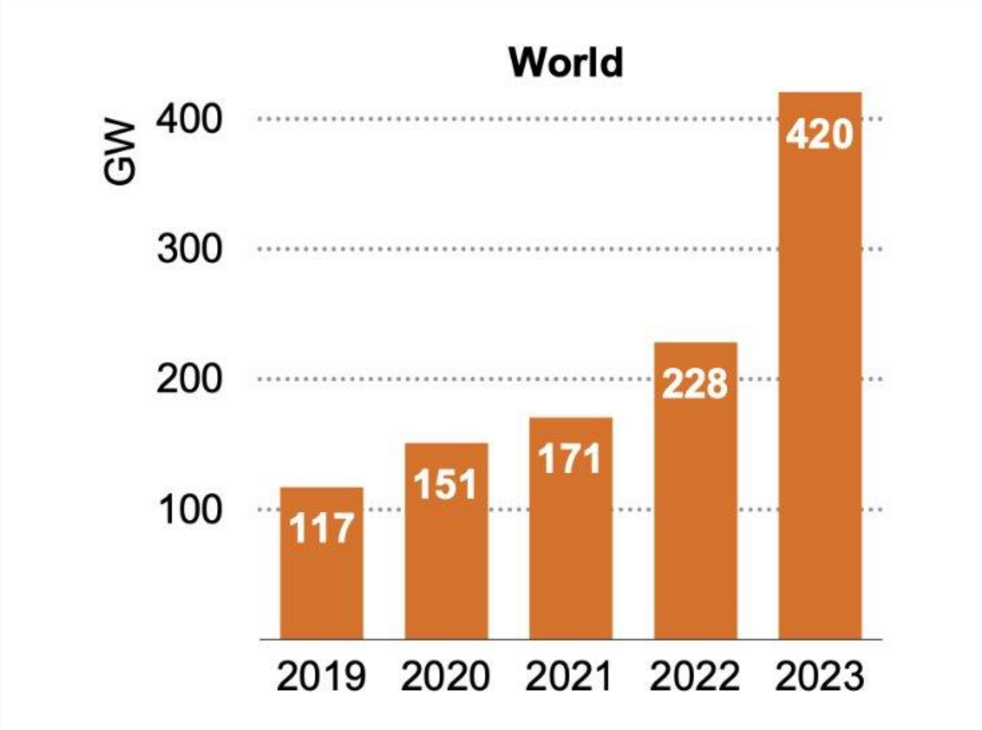
- **Nowcasting**
- **Short-term**
- Long-term

Information used for the solar power forecast usually includes the Sun´s path, the atmospheric conditions, the scattering of light and the characteristics of the solar energy plant.

Small solar power generators are key players in the energy transition but lack the tools to predict their power output, making it difficult to plan, reduce reliance on fossil fuels, and achieve true energy independence.

Solar power capacity grows

Annual additions of solar PV (photovoltaic) growing 85% with a capacity of 420 GW globally (IEA, 2024)



Solar PV capacity additions and avoided emissions (Source: Clean Energy Market Monitor – March 2024)

02

Significance

Why is Solar Forecasting important ?

Significance

Efficiency

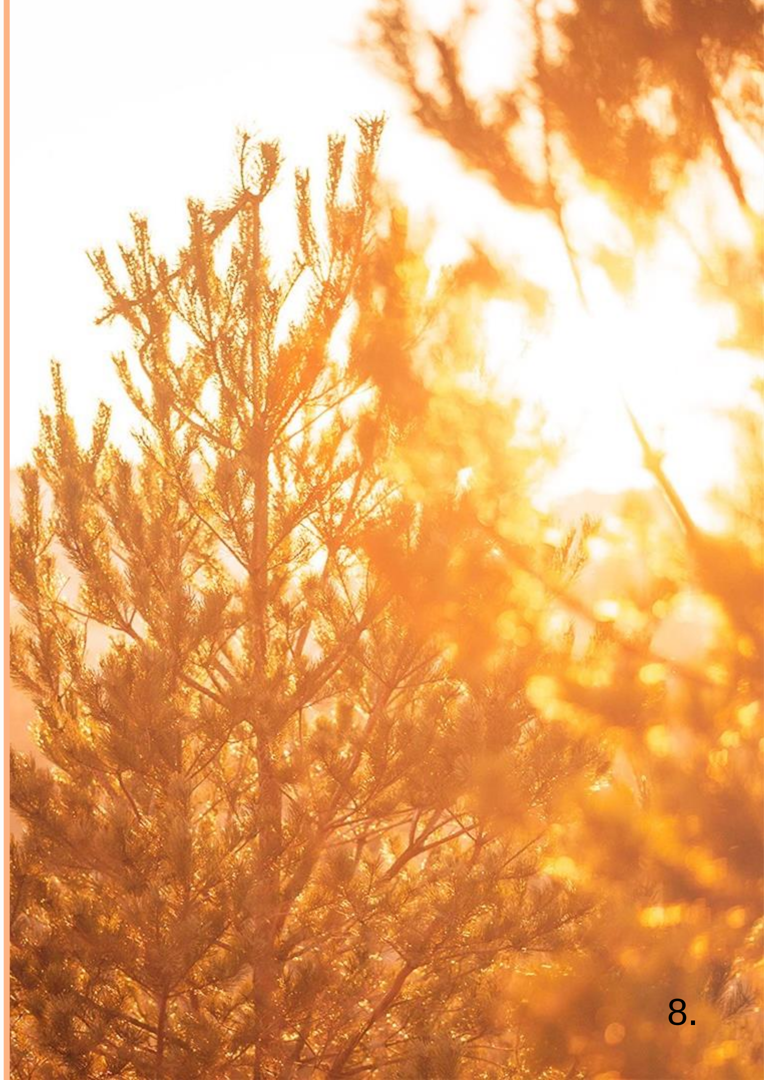
Efficient management of the electric grid.

Sustainability

Schedule maintainance in the most convenient way

Smart consumption

Plan ahead, saving more money



03

Past projects

State of the art

Already existing

Current business

AESO: https://www.aeso.ca/aeso/
UNERGY: https://unergy.io/en

ARIMA

Statistic and probabilistic models

Sky observations

Estimating cloud motion and optical depth

Machine Learning

New era of probabilistic models

Satellite based

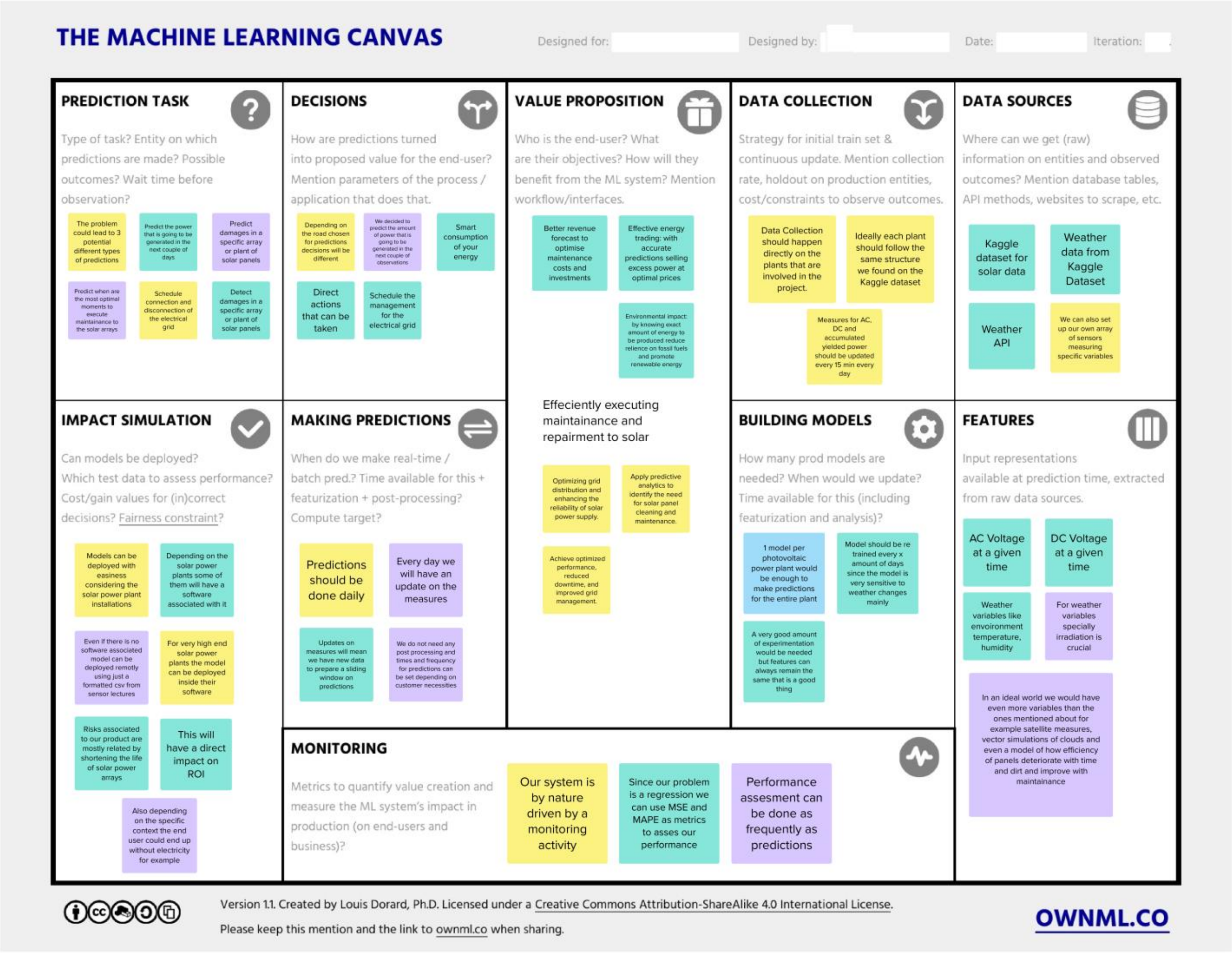
Leverage information using geostationary satellites

Mixed Models

Combining 2 or more of the mentioned

04

ML Canvas



THE MACHINE LEARNING CANVAS

Designed for: Designed by: Date: Iteration:

PREDICTION TASK

Type of task? Entity on which predictions are made? Possible outcomes? Wait time before observation?

- The problem could lead to 3 potential different types of predictions
- Predict the power that is going to be generated in the next couple of days
- Predict damages in a specific array or plant of solar panels
- Predict when are the most optimal moments to execute maintainance to the solar arrays
- Schedule connection and disconnection of the electrical grid
- Detect damages in a specific array or plant of solar panels

DECISIONS

How are predictions turned into proposed value for the end-user? Mention parameters of the process / application that does that.

- Depending on the road chosen for predictions decisions will be different
- We decided to predict the amount of power that is going to be generated in the next couple of observations
- Smart consumption of your energy
- Direct actions that can be taken
- Schedule the management for the electrical grid

VALUE PROPOSITION

Who is the end-user? What are their objectives? How will they benefit from the ML system? Mention workflow/interfaces.

- Better revenue forecast to optimise maintenance costs and investments
- Effective energy trading: with accurate predictions selling excess power at optimal prices
- Environmental impact: by knowing exact amount of energy to be produced reduce relience on fossil fuels and promote renewable energy

Effeciently executing maintainance and repairment to solar

- Optimizing grid distribution and enhancing the reliability of solar power supply.
- Apply predictive analytics to identify the need for solar panel cleaning and maintenance.
- Achieve optimized performance, reduced downtime, and improved grid management.

DATA COLLECTION

Strategy for initial train set & continuous update. Mention collection rate, holdout on production entities, cost/constraints to observe outcomes.

- Data Collection should happen directly on the plants that are involved in the project.
- Ideally each plant should follow the same structure we found on the Kaggle dataset
- Measures for AC, DC and accumulated yielded power should be updated every 15 min every day

DATA SOURCES

Where can we get (raw) information on entities and observed outcomes? Mention database tables, API methods, websites to scrape, etc.

- Kaggle dataset for solar data
- Weather data from Kaggle Dataset
- Weather API
- We can also set up our own array of sensors measuring specific variables

IMPACT SIMULATION

Can models be deployed? Which test data to assess performance? Cost/gain values for (in)correct decisions? Fairness constraint?

- Models can be deployed with easiness considering the solar power plant installations
- Depending on the solar power plants some of them will have a software associated with it
- Even if there is no software associated model can be deployed remotly using just a formatted csv from sensor lectures
- For very high end solar power plants the model can be deployed inside their software
- Risks associated to our product are mostly related by shortening the life of solar power arrays
- This will have a direct impact on ROI
- Also depending on the specific context the end user could end up without electricity for example

MAKING PREDICTIONS

When do we make real-time / batch pred.? Time available for this + featurization + post-processing? Compute target?

- Predictions should be done daily
- Every day we will have an update on the measures
- Updates on measures will mean we have new data to prepare a sliding window on predictions
- We do not need any post processing and times and frequency for predictions can be set depending on customer necessities

BUILDING MODELS

How many prod models are needed? When would we update? Time available for this (including featurization and analysis)?

- 1 model per photovoltaic power plant would be enough to make predictions for the entire plant
- Model should be re trained every x amount of days since the model is very sensitive to weather changes mainly
- A very good amount of experimentation would be needed but features can always remain the same that is a good thing

FEATURES

Input representations available at prediction time, extracted from raw data sources.

- AC Voltage at a given time
- DC Voltage at a given time
- Weather variables like envoironment temperature, humidity
- For weather variables specially irradiation is crucial
- In an ideal world we would have even more variables than the ones mentioned about for example satellite measures, vector simulations of clouds and even a model of how efficiency of panels deteriorate with time and dirt and improve with maintainance

MONITORING

Metrics to quantify value creation and measure the ML system's impact in production (on end-users and business)?

- Our system is by nature driven by a monitoring activity
- Since our problem is a regression we can use MSE and MAPE as metrics to asses our performance
- Performance assesment can be done as frequently as predictions

05

Business Case

Solar Power Generation Forecasting

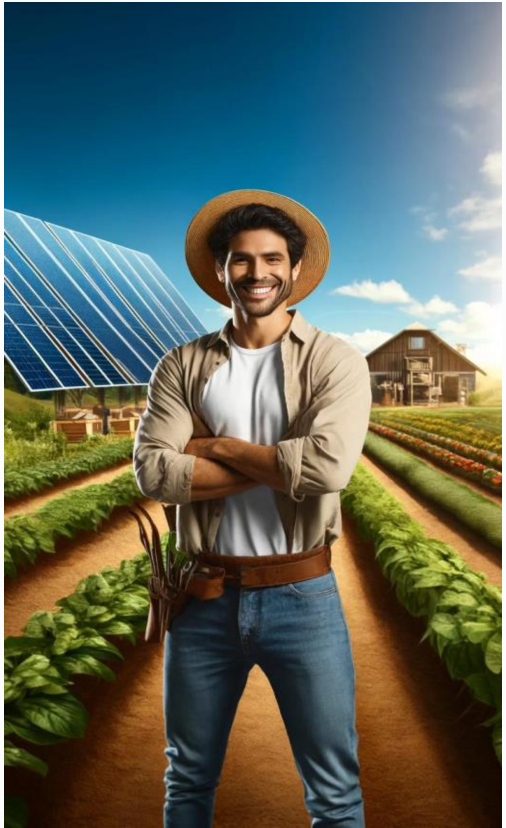
Empowering Energy Independence

Imagine the future

Peak efficiency, minimized waste

Challenge today

Unpredictability hampers small generators



Our solution

Forecasting tool empowers planning

Success story

Jorge optimizes irrigation, cuts costs

Revenue streams

01	02	03
Direct Sales	Licensing	Consulting
Subscription fees from users	Agreements with HW and solar installation companies to built-in our tool	Providing analytics services to study energy patterns
Monthly/Annually fee	Sell licenses	Sell services

Subscription plans

01

Basic

Standard forecasting features for small solar installations with basic support

$50/month

02

Professional

Advanced forecasting features, API access, and priority customer support, for medium-sized solar power generators

$200/month

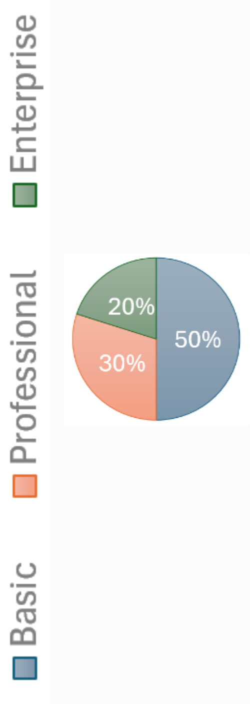
03

Enterprise

Full customization, dedicated support, and enterprise-
level integrations for large-scale installations

$500/month

Our growth



Year 3

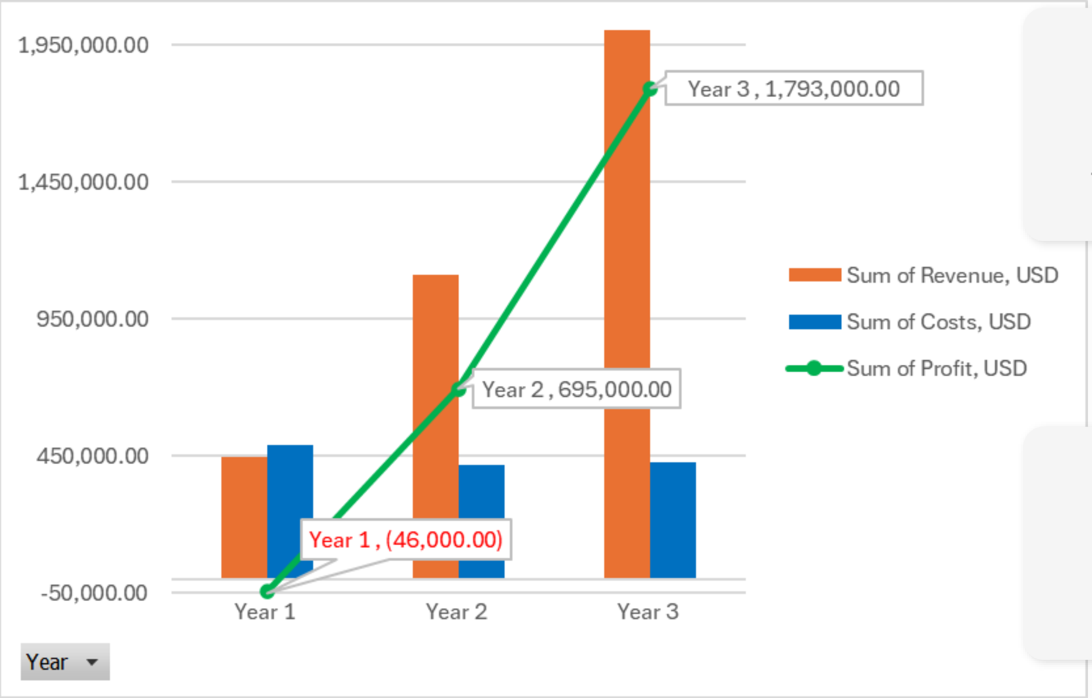
Expansion to 1000 subs

Year 2

Growth to 500 subs

Year 1

Target 200 subscribers



$2.44M

Expected profit for first 3 years

2 years

ROI

The solar power forecasting tool is expected to reach profitability by the second year of operation

Costs

Initial Development Costs Year 1

Fixed Costs (Include staff salaries, office rent, utilities, and software)	$300,000
Data acquisition	$50,000
Software development tools	$30,000
Cloud computing resources	$20,000
Pilot project and testing	$40,000
Marketing and promotional activities	$50,000
Total initial costs	**$490,000**

Ongoing Operational Expenses Year 2 and Year 3

Fixed Costs (Include staff salaries, office rent, utilities, and software)	$300,000
Maintenance and updates	$15,000
Marketing and promotional activities	$50,000 in Year 1, increasing by 20% each year as subscriber base grows.
Customer support team	$30,000
Miscellaneous expenses	$10,000
Total annual operating costs	**$355,000+marketing costs**

06

Model Benchmark

Solar Power Forecasting

Different Approaches

ARIMA Baseline	ANN Baseline	LSTM 1	LSTM 2
Not Implemented	Implemented	Implemented	Implemented
	Mean Error of: 83,026 W	Mean Error of: 31,513 W	Mean Error of: 19,381 W

07

Model deployment

Deployment

Basically, we used all tools learnt in class:

- We created and trained our model within a **Jupyter notebook.**
- We developed the **Python scripts** that form the core of our application.
- We utilized **Flask** to transform our main features into an API.
- We tested all our main features thoroughly using **Postman.**
- We designed the frontend using the **Bootstrap and javascript framework.**
- We built a **Docker image** and deployed it to **Docker Hub**.
- We created an instance on Google Cloud Platform **(GCP) using our Docker image.**
- Our minimum viable product requires a minimum technological resources for now.

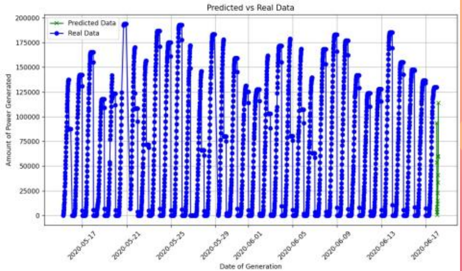
Deployment

Some challenges:

- Configuration of **lag and delay** in our model.
- Prone to have errors when **weather behaves unexpectedly.**
- Have to figure out a good **schedule for re training**
- The model is very hungry for data.
- We faced with some **dependency hells** before to create the docker image.

Demo

SOLAR POWER FORECASTING
In a world inching closer to irreversible climate change, the imperative to pivot towards renewable energy sources has never been more critical.



Predicted vs Real Data
Predicted Data
Real Data
Amount of Power Generated
Date of Generation
200000
175000
150000
125000
100000
75000
50000
25000
0
2020-05-17
2020-05-21
2020-05-25
2020-05-29
2020-06-01
2020-06-05
2020-06-09
2020-06-13
2020-06-17

08

Q&A

Thank you