**A**

**Project Report**

**on**

**Big Data Analytics**

**Solar radiation prediction & its implementation on a webpage**

**Submitted by-**

***Shashank Sai Palani Anirudh Jilla***

***1800183C202 1800137C202***

***B.Tech CSC B.Tech CSC***

**Under the guidance of**

**Dr. Yogesh Gupta**

Associate Professor

## 

**Department of Computer Science and Engineering**

**SCHOOL OF ENGINEERING AND TECHNOLOGY**

**BML MUNJAL UNIVERSITY GURGAON-122413, INDIA**

***May, 2021***

# **Acknowledgement**

This project was pursued as part of Big Data Analysis coursework. This acknowledgement stands in support to mention the names of all the people that have been an active and passive role in this project. This project has been pursued by Shashank Sai Palani and Anirudh Jilla together as a group and all the related coursework has been developed and documented by them, hence the primary acknowledgement is theirs. The next acknowledgement is to our course instructor Dr.Yogesh Gupta who has guided us all the way through this project at various aspects and has given us suggestions throughout the work being done and we like to acknowledge his contribution in helping us develop this project.

Thanking You

Shashank Sai Palani

Anirudh Jilla

Index

[**Acknowledgement**](#_2xbfa76xrzok) **1**

[**Abstract**](#_f2gaifs9iyxx) **3**

[**Motivation**](#_arck14hsrv2p) **5**

[1. INTRODUCTION](#_5upwnxfhv9qf) **6**

[2. PROBLEM STATEMENT](#_h9v03sct8xxt) **6**

[4. METHODOLOGY](#_5w0t0gybog8c) **7**

[5. RESULTS AND DISCUSSION](#_9x29gfqd3gke) **8**

[6. CONCLUSIONS AND FUTURE WORK](#_yr4zhsozrnxj) **11**

# **Abstract**

The objective of this project is to conduct an Exploratory Data Analysis on data of our problem statement and accordingly discuss the various observations that can be concluded upon from these Analyses. Then we have to research the possible solutions that could be developed using various tools/resources that are available to be able to develop an efficient model/system that can predict the future outcomes based on the learnings from the past history data and feature parameters as the inputs. Our project is based on prediction of Solar Radiation using a given set of parameters such as Temperature,Pressure, Humidity, Wind Direction, Wind Speed to achieve this functionality we use the power of Machine Learning models that take in the input to be as parameters of the before mentioned features and predict the amount of solar radiation that could be observed based on the given parameters. This functional Machine Learning model is to be deployed on a webpage that is hosted using any of the free frameworks such as Flask, Django, Streamlite and be made publicly accessible.

# **Motivation**

The main motivation behind this project is the problem statement, the capability to be able to predict the amount of Solar Radiation that would fall for a given set of defined parameters, this data would prove invaluable to the person that would depend on Solar Radiation data for their ideas. When a company that installs solar panel farms or an agro products company wants to know a better place to set up their unit where they can have access to harness the energy of solar radiation in abundance that would help them in many ways. This project would help them in deciding the favourability of a location based on its solar radiation index that has been predicted by the model based on the parameters of that location such as Temperature, Pressure, Humidity, Wind Speed, accordingly they would be able to decide whether the location would be able to give them a viable solar irradiance factor after installation. Similarly an agro based company can use this prediction data to decide on the location to grow their organic product farms that require sunlight and this prediction helps them to calculate the yield that would grow with the predicted solar radiation.

# **1. INTRODUCTION**

Climate change and the energy crisis have motivated the use and development of sustainable energy sources, with solar energy being identified as one of the most abundant and promising candidates for bulk power generation. Yet, an inherent characteristic common to all renewable energy sources is that power generation is fully dependent on weather and meteorological parameters, and therefore the solar power output cannot be fully controlled or planned for in advance. To ensure secure integration of photovoltaic (PV) systems into the smart grid, accurate PV forecasting is a critical element of energy management systems. If accurate PV forecasting is lacking, any unexpected fluctuations in solar energy capacity may have significant impacts on the daily operations and physical health of the entire grid and may negatively affect the quality of life of the energy consumers.

# **2. PROBLEM STATEMENT**

2.1 Develop an efficient and deployable solution that can predict the Solar Irradiation pattern using various features referenced in the dataset.

2.2 This project allows the user to predict the amount of Solar Radiation that a particular location might get depending upon various positively correlated features identified during the preliminary and exploratory data analysis phase of the project. It gives an average number to the value of Solar radiation that is predicted by the Machine Learning model using the parameters given.

2.3 The implementation of the project starts with conducting a preliminary data analysis of the various features and types of values present in the dataset, this helps in understanding the types of values present in the and the ones that could be used for prediction computations.

The next step is the Exploratory data analysis, this is where we try to identify the patterns formed by the features of the dataset and also identify the limits of the dataset. We achieve this by various visualizations such as bar graphs, scatter plots, frequency distributions, box plots, etc.,

After a careful study of these visualizations we figure out the features that are positively correlated with Solar Radiation then we stick to these features to be used in the learning.

Once we have identified the positively correlated and directly impacting features we now filter the dataset to contain only the data of these features and fit them into the model using a transform function, later we take this transformed data and input into the model for training and testing. Then according to the precision score and accuracy values we decide on the model that functions best and use it for our real data predictions. This way we come up with a number that depicts an average amount of radiation that could be expected.

# **4. METHODOLOGY**

**4.1** The implementation of the project starts with conducting a preliminary data analysis of the various features and types of values present in the dataset, this helps in understanding the types of values present in the and the ones that could be used for prediction computations. The next step is the Exploratory data analysis, this is where we try to identify the patterns formed by the features of the dataset and also identify the limits of the dataset. We achieve this by various visualizations such as bar graphs, scatter plots, frequency distributions, box plots, etc. After a careful study of these visualizations we figure out the features that are positively correlated with Solar Radiation then we stick to these features to be used in the learning. Once we have identified the positively correlated and directly impacting features we now filter the dataset to contain only the data of these features and fit them into the model using a transform function, later we take this transformed data and input into the model for training and testing. Then according to the precision score and accuracy values we decide on the model that functions best and use it for our real data predictions. This way we come up with a number that depicts an average amount of radiation that could be expected.

**4.2** The Technical features of this project are the Machine Learning models that have been used to power the prediction model such as:

* Linear regression
* Gradient Regression
* Random Forest
* Decision Tree

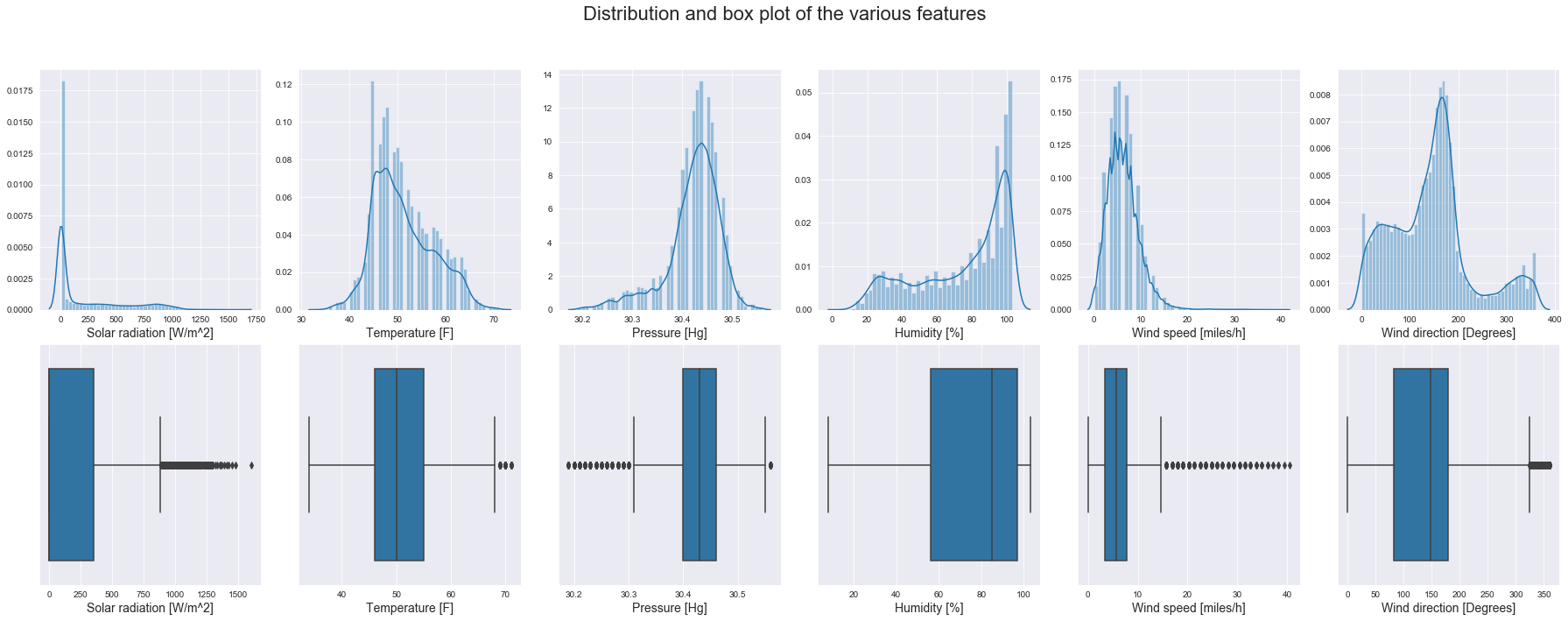
**4.4** The various packages that were used for coding the models and data pre-processing and cleaning part of the project are as follows:

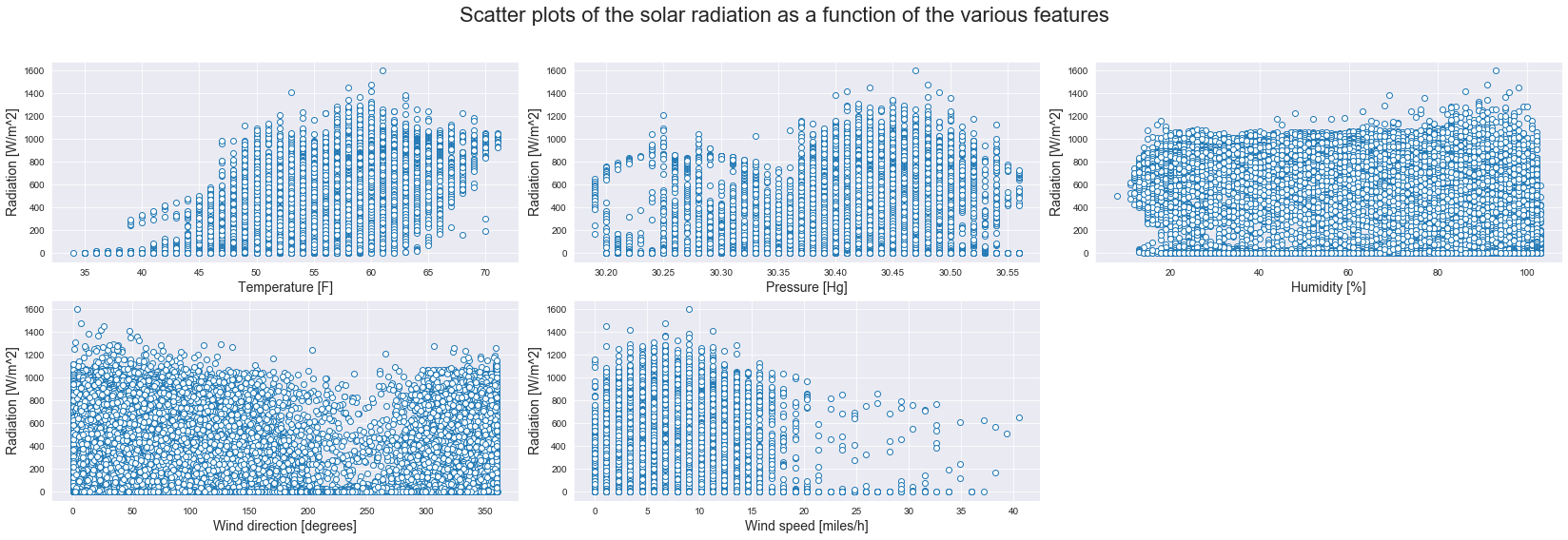
* Numpy : This package supports mathematical and array based computations that
* Matplotlib
* Seaborn
* Pandas
* sklearn

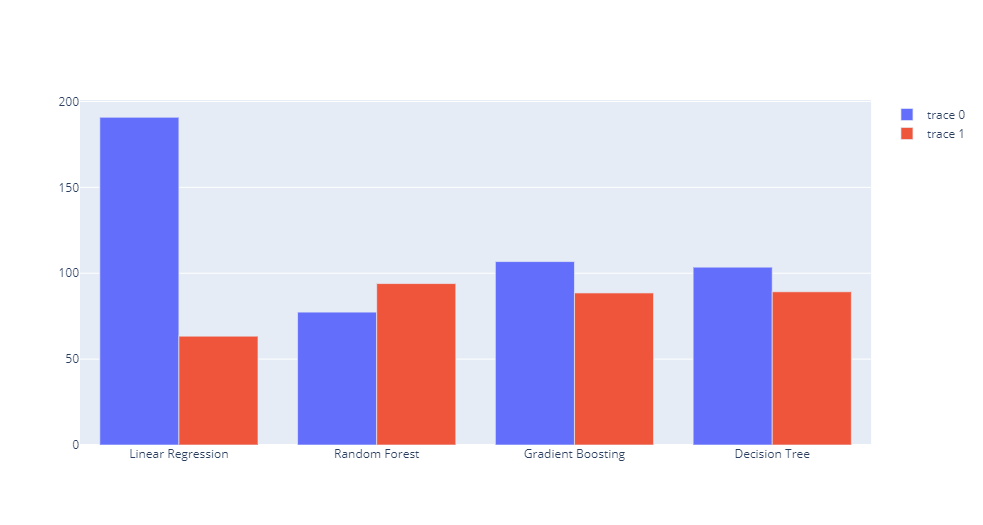
**4.5** Some of the features of our project are the feature engineering and data analysis aspects of it. The features that have been selected to predict the Solar radiation value are some of the aspects that our unique in our project apart from that the models used for the prediction also yield results with very high accuracy, precision and recall scores.

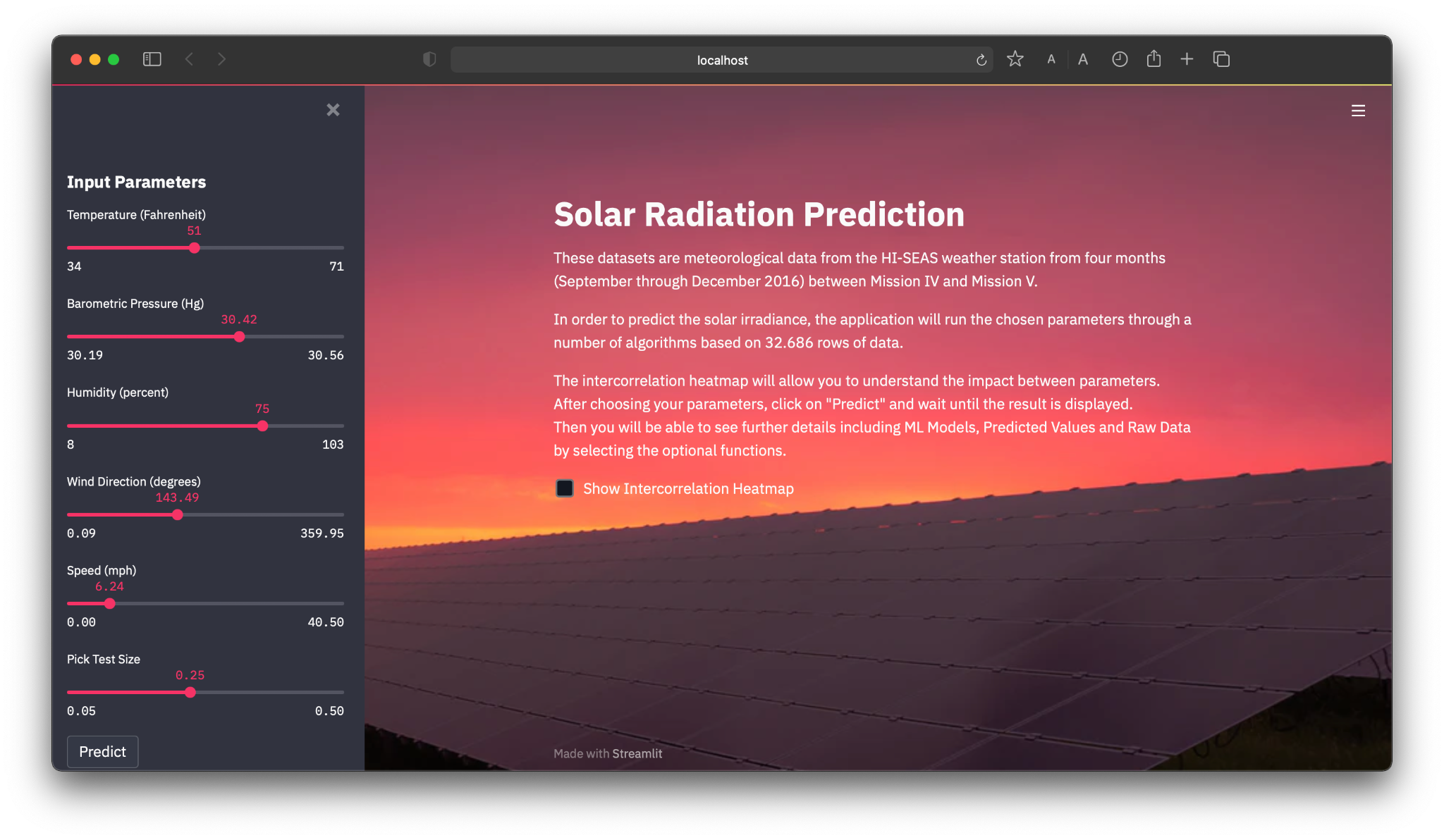
# **5. RESULTS AND DISCUSSION**

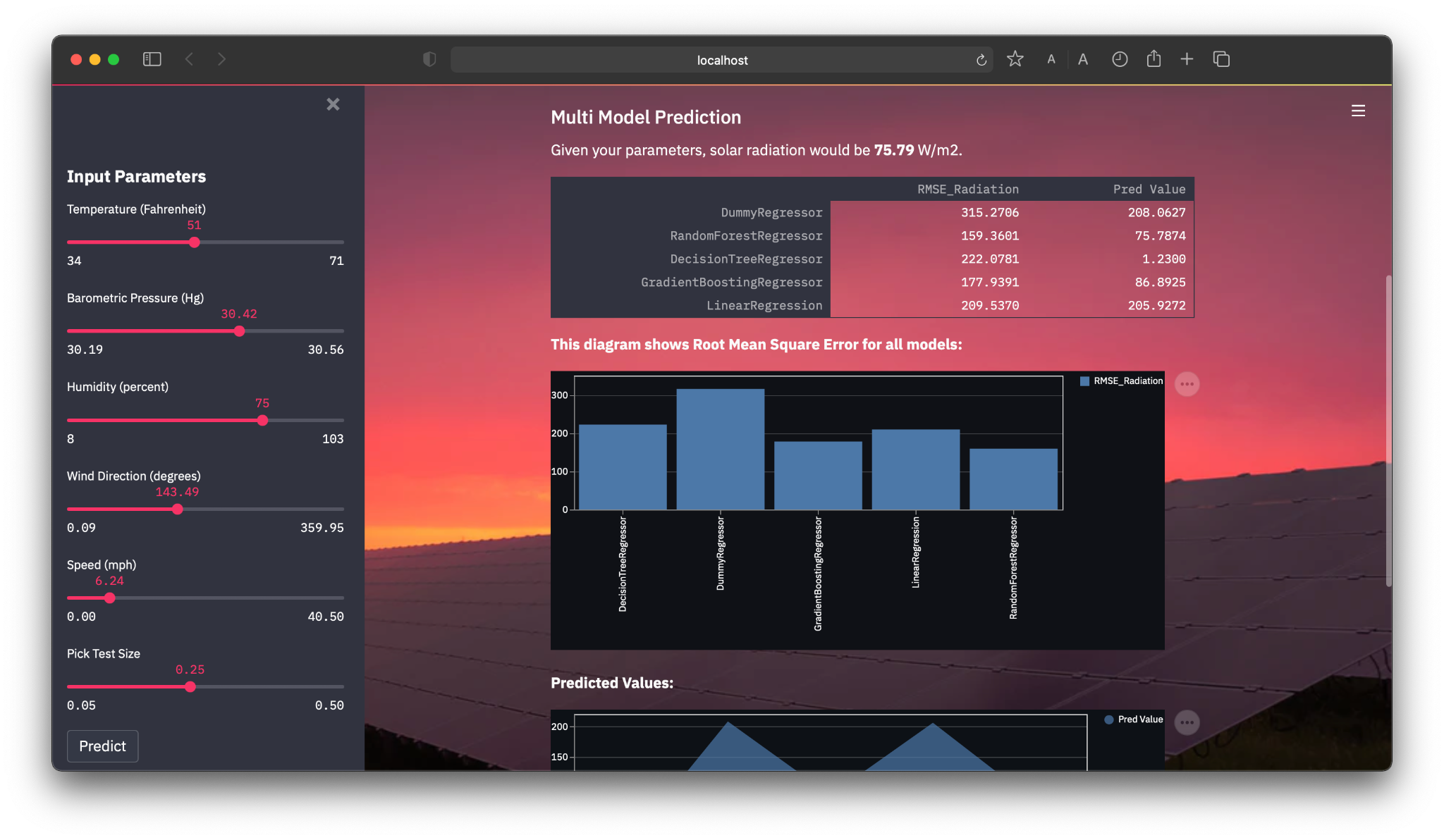
The results yielded by the Machine Learning models in this project are quite accurate and convincing along with the precision scores that follow them. The final prediction program that has been hosted on a webpage.

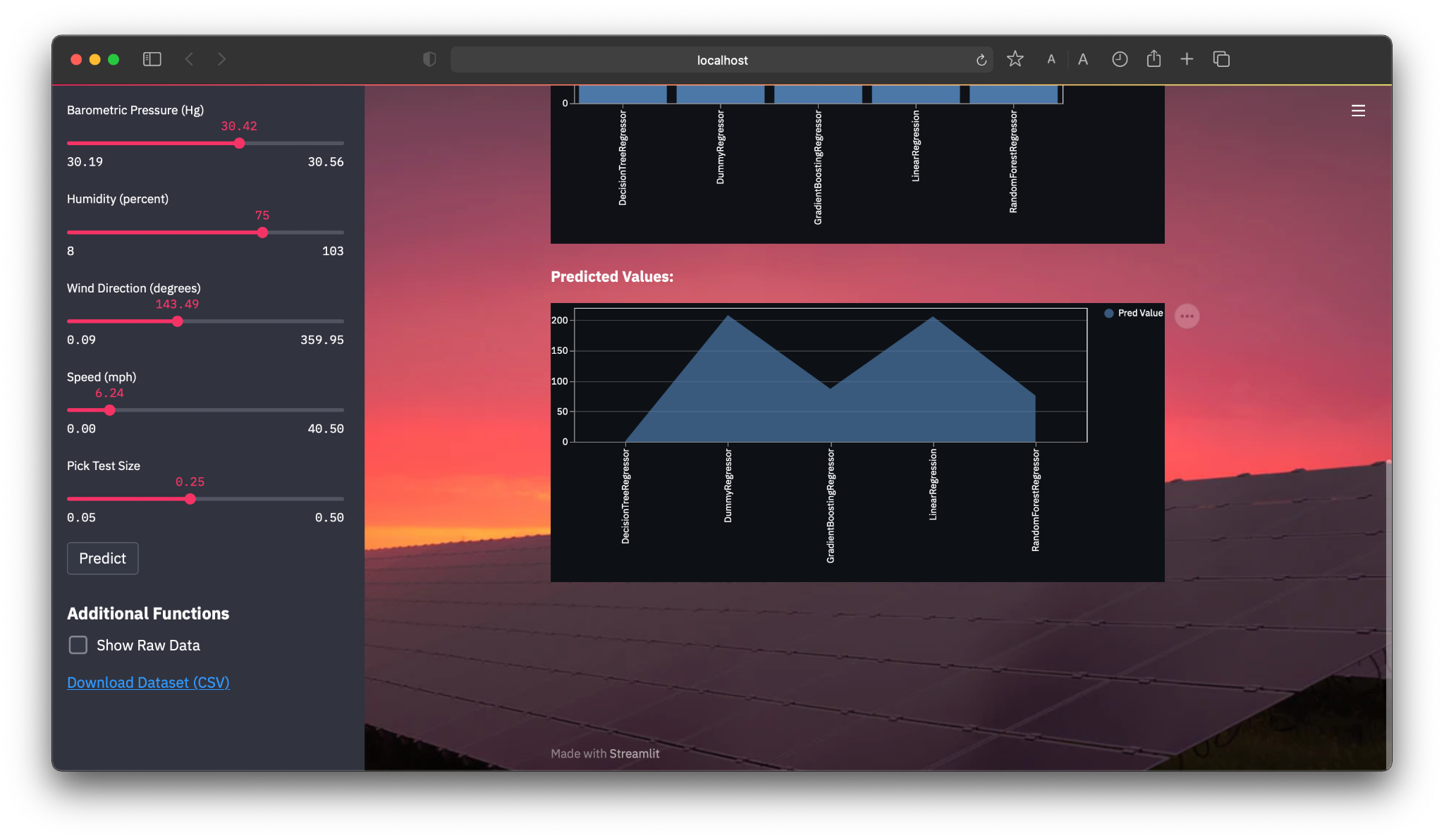












# **6. CONCLUSIONS AND FUTURE WORK**

The objective of this project has had a great impact in widening our knowledge and technical skills regarding the concepts of Machine Learning and Model Prediction. We have understood the nuances that exist in trying to analyze the data using various methods such as preliminary and exploratory data analysis, this is most important in order to identify the perfect features that allow for the prediction of the required value. Also we have understood the metrics required to judge the performance of various models and finalize on the decision of the model to be implemented for prediction.

The results may appear very insignificant in terms of the scale at which the model has been deployed yet it does resemble the fundamental characteristics in defining a prediction model, such as feature engineering and appropriate feature detection that allows to predict the data when some static parameters are passed into the model, also the exploratory data analysis part that has allowed to understand the various trends and patterns in the dataset and the degree of correlation that the features share with each other. However, there is a lot of scope for development and future work, the project can be replaced with a much better and diversified dataset that allows to include much better dynamic and technically sound feature that follow a continuous pattern such as the weather data around the surrounding the region and also dynamic data synchronization that can render prediction as fast as the data keeps updating itself. We could also broaden the sphere of work to include faulty predictions detection, calculation of the lifetime of a given prediction result that would allow the user to know how long the prediction lasts for a given location. Therefore , the future is highly uncertain when compared to broader aspects that could be consolidated , but the fundamental work is similar to our analysis in the project.