(Introduction)

Data science is a broad area that employs scientific methods, processes, algorithms, and systems to extract information and insights from structured and unstructured data, as well as to extend that knowledge and actionable predictive analytics through a wide range of application domains. Data mining, machine learning, and big data are all linked to data science. Data science unifies statistics, data processing, informatics, and their associated processes in order to use data to understand and interpret real phenomena. Three in the context of mathematics, statistics, computer technology, information science, and domain knowledge, it employs techniques and hypotheses from a variety of fields.

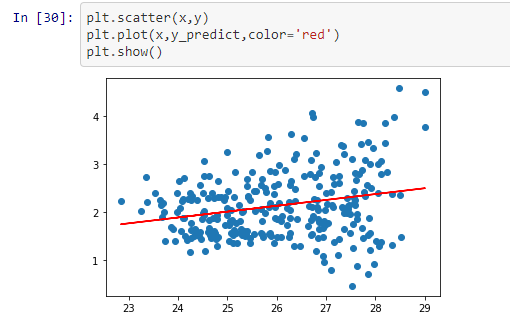
(about our dataset)

A dataset is collection of related sets of information that is composed of separate elements but can be manipulated as a unit by a computer. (about our dataset/how cleaned the dataset)

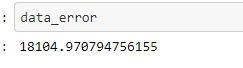
(models we used)

According to the dataset we will be calculating and predicting the Solar Irradiance (G score) for a given date and time. So to predict that we tried several models until we get the correct one.

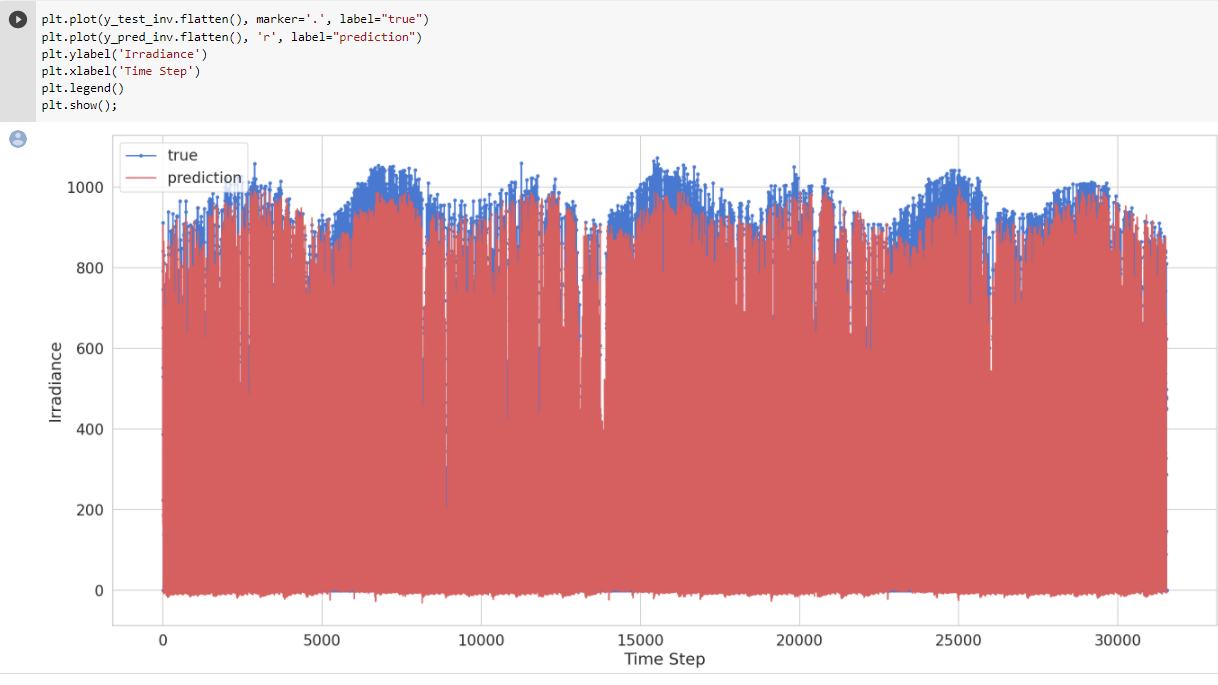
At first we used Linear regression model. Linear regression model is a simple model which has a linear approach to modelling the relationship between a scalar response and one or more explanatory variables. In our model seasonal forecasting has been used to model the relationship between the variables. So our results were more spread. Because of that we had to reject the Linear regression.

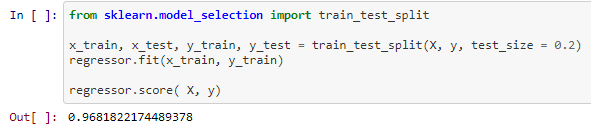


After that we moved on to another model named ARIMA model. Auto Regressive Integrated Moving Average model a model that uses time series analysis. It takes past values from a dataset and calculates them with a help of a formula and predicts future values. We had to reject that model because the predicted data was not accurate enough and also the data error was bit high.



After that we tested LSTM model which is special kind of recurrent neural network that is capable of learning long term dependencies in data. So we implemented it to our model and tested with our dataset, but the results were not accurate. There was a considerable gap between the test data and the predicted data. So we had no choice only to reject that model also.



Finally, we moved onto Decision Tree regression model. This model is also considered as a classification model, and formed as a tree. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. After implementing and training this model with our dataset the results were better. The accuracy between the test data and the predicted data were above 90%.

