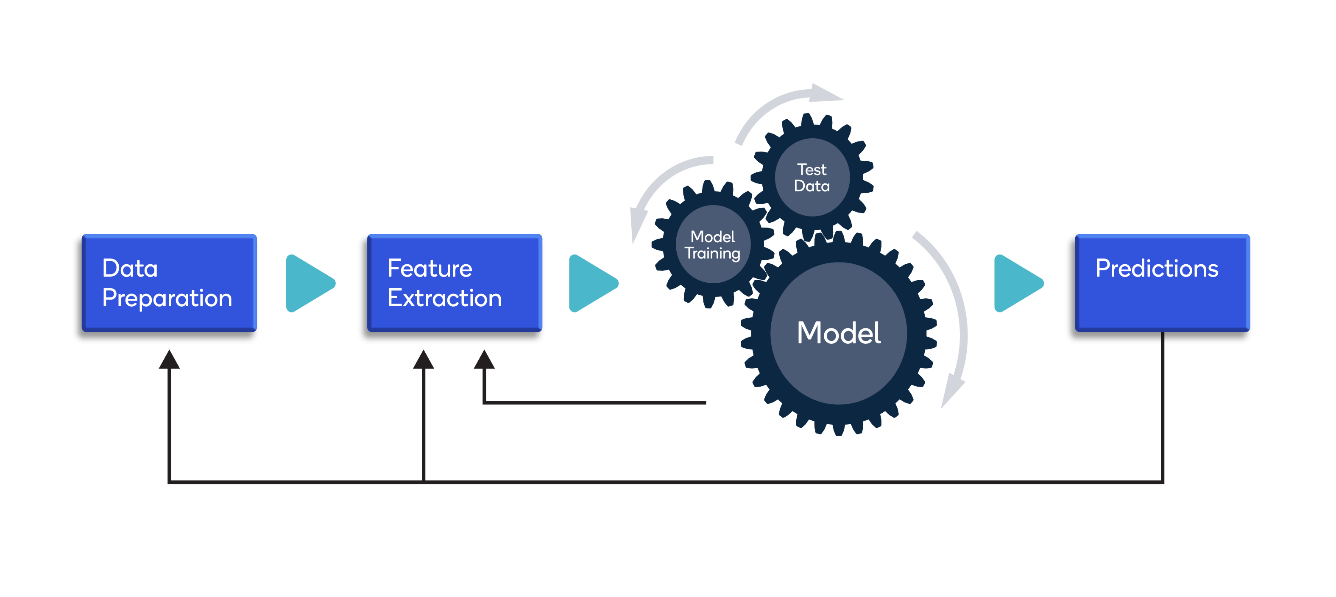
**SOLAR ENERGY PREDICTION**

**PARTICIPANTS AND RESPECTIVE ROLES:**

|  |  |  |  |
| --- | --- | --- | --- |
| **NAME** | **STUDENT\_ID** | **EMAIL\_ID** | **ROLE** |
| Anusha Yaramala | 11594436 | anushayaramala@my.unt.edu | Data Collection and Training data |
| Dilip Goud Vatnala | 11589632 | DilipGoudVatnala@my.unt.edu | Feature Selection and visualization |
| Sai Teja Munja | 11591803 | SaiTejaMunja@my.unt.edu | ML Algorithms |
| Subha Satya Rishitha Kodamalla | 11598564 | subhasatyarishithakodamalla@my.u  nt.edu | Evaluation of model |

**WORKFLOW:**



**ABSTRACT:**

Solar power is a free and clean alternative to traditional fossil fuels.However, nowadays, solar cells' efficiency is not as high as we would like, so selecting the ideal conditions for its installation is critical in obtaining the maximum amount of energy out of it. Generation of energy by a solar panel or cell depends upon the doping level and design of solar PV array but the main factors are the amount of solar radiation falling on the panel, environmental factors like atmospheric temperature and humidity and dust present on the panels. These factors are naturally variable and hence the output of solar cells directly depends on it. Also, the solar irradiance as well as all the above-mentioned factors are variable throughout the day.

Here we are predicting the daily power production of a photovoltaic power station at the University ofIllinois campus from the measurements of a set of weather features like relative humidity, cloudcoverage, visibility, dew point, wind speed, Station pressure, Altimeter by employing various machine learning models.

**PROBLEM STATEMENT:**

Now-a-days climate change is visible and affecting us in many ways hence we are going for eco-friendly generation methods like Solar where weather conditions play a major role in the generation. So, it’s important to analyze these features to predict the accurate energy produced to meet the power needs.

**DATA SPECIFICATION**:

The dataset consists of **10 Columns** and **637 rows**and type is **float.**

• Cloud Coverage: The fraction of the sky covered by clouds on average when observedfrom a particular location. (%)

• Visibility: A measure of the horizontal opacity of the atmosphere (Miles)

• Dew Point: The atmospheric temperature below which water droplets begin tocondense, and dew can form. (0C)

• Relative humidity: The amount of water vapor present in air expressed as apercentage of the amount needed for saturation at the same temperature. (%)

• Station Pressure: the pressure that is observed at a specific elevation and is the truebarometric pressure of a location. (Inch Hg)

• Altimeter: Altitude measurement (inch Hg)

• Wind speed (Mph)

• Temperature (0C)

Here **solar energy** is predicting column.

**Dataset** :

Table

Description automatically generated

**DESIGN:**

1.Data process

2.feature selection based on correlation matrix and multivariate analysis

3.visualized how the features depending on each other

4.trained the data

5.performed diff ml algorithms

6.Evaluated the models using r2square, RMSE method

**DESIGN FLOW CHART:**

**Diagram

Description automatically generated**

**IMPLEMENTATION OF DATA PREPROCESSING TECHNIQUES:**

1.Missing values: We checked for missing values in the data and there are no null values.

2.Box Plots: We found out that there are no outliers in the data.

3.Atribute Selection: We found the dependencies between the features individually using correlation matrix and multivariate analysis and found that all features are affecting the model. We tried removing features that are highly scattered, but it did not help us in improving our accuracy since we have a smaller number of features. So, all features are included.

**Feature Selection:**

A picture containing background pattern

Description automatically generated

**Multi-variate analysis :**

A picture containing table

Description automatically generated

**MODELING:**

We have used ML models such as XGB , LGBM, Decision Tree Regression and Linear Regression. Out of this four we found the better results with XGB and LGBM.

**XGBooster**- XGB works with structured and tabular data and also works well with complicated datasets. As dataset Consists of 637 rows so it works better with model.

**LGBM Regressor** - Light GBM is a gradient boosting framework that uses tree-based learning algorithm. Moreover, LGBM is light and speed. Another reason why we chose this algorithm is because of its focuses on accuracy of results.

**Decision Tree Regression**- We used Decision Tree because it is one of the most commonly used, practical approaches for supervised learning. As our data consists of Dependent variables Decision tree gives the better results for Dependent Variables.

**Linear Regression-** Linear regression is a supervised machine learning algorithm that can best fit with the linear data. As few of the features in the dataset we have chosen varies linearly we have chosen this model.

**EVALUATION**:

Metrics that we used for evaluation are RMSE, R2 square method. Here are the results for our models, XG booster has less RMSE and higher R2 square hence its better model.

|  |  |  |  |
| --- | --- | --- | --- |
| **MODEL** | **RMSE(Train data)** | **RMSE(Test data)** | **R2 SQUARE** |
| Decision Tree | 0.000000 | 9508.842278045735 | 0.29912512351447773 |
| LGBM regressor | 1974.7142555057587 | 5972.637544853962 | 0.6879621126982283 |
| XGBooster | 52.166325735015334 | 5698.865193004202 | 0.7159127279383681 |
| Linear regression | 8516.746462239387 | 8277.34478421072 | 0.40068230552382045 |

**Best of line fit:** From below plots, we can say that XG booster has better model fit.

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

**Linear regression Decision tree XG Booster**

**REFRENCES**:

<https://github.com/Srijha09/Prediction-of-Energy-Consumption>

<https://github.com/anantgupta129/Solar-Power-Generation-Forecasting/blob/main/solarpower_generation_forecasting_ann.ipynb>

<https://github.com/mz256/ashrae_energy_forecast/tree/main/notebooks>

We used above references for data cleaning and visualization techniques. The above references are based on deep learning models like LSTM here, we predicted using machine learning algorithms like Decision tree, linear regression, XG Booster, LGBM regressor to predict the model.

**Scope of the Project:**

Incorporating more features could give us a better accuracy. The factors such as number of solar panels, type of panels and their angle of installation can be included in determining the solar energy generated by the power station.

**SOURCE CODE:**