Problem Framed in ML:

1. Are there outliers in the dataset.
2. What are the feature which can be eradicated easily.
3. Are there any categorical variables in the dataset.
4. Which features are important in predicting the Power(kW).
5. What are the data points which can be visualized as an error/outlier.
6. What algorithms should be used for the regression problem.

Setting up the Python Environment:

The Python environment setup typically involves the following steps:

Python Installation: Install Python on your system. You can download the Python installer from the official Python website (https://www.python.org) and follow the installation instructions.

Virtual Environments: Create a virtual environment using tools like virtualenv or conda. A virtual environment provides a self-contained Python installation and allows you to install specific versions of packages without affecting the system-wide Python installation.

Package Manager: Set up a package manager such as pip or conda. These tools allow you to install, upgrade, and manage Python packages and libraries within your virtual environment.

Dependencies and Libraries: Install the necessary packages and libraries required for your project.

Editor or IDE: Choose an integrated development environment (IDE) or text editor for writing and running your Python code. Popular options include PyCharm, Visual Studio Code, Atom, or Jupyter Notebook.

Code Execution: Once your Python environment is set up, you can execute your Python scripts or run Jupyter Notebooks within the configured environment.

Running the Code:

->The ML Pipeline file

Through running the code, we will get a csv file that contains the required features with the predicted feature (Power(kW)) at the end.

->The Jupyter .Ipynb file

By running the python script, the output we get is the metrics, accuracies of the different compared models and the visualization of some of the columns for a better understanding.

Metrics Considered:

For the regression project, mostly 3 metrics are considered:-

Mean Squared Error(MSE)

Mean Absolute Error(MAE)

R2 Score

Model comparison and interpretation:

Mainly 4 algorithms are considered:-

Random forest Regressor

Lasso Regression

Gradient Boosting

|  |  |  |
| --- | --- | --- |
| Random forest Regressor | Lasso Regression | Gradient Boosting |
| 99.6 | 89.1 | 96.4 |
| Between balanced and overfitted  But nearer to balanced. | balanced | Between balanced and overfitted  But nearer to balanced. |

Conclusion:

All the 3 models can be used for the prediction and analysis but Random Forest regressor can be chosen to be the best predictor. To get a more variance result we can use Lasso Regression.