**Feature Engineering**

|  |  |
| --- | --- |
| **Features** | **Transformation** |
| **Tracking\_id** | Separated Numeric part from the tracking id . |
| **Datetime** | Separated year, month and time from datetime where time contributes more as compared to other extracted feature. |
| **Cloud level and turbine status** | Both are categorical features , for NAN value a new category is assigned .  Replaced “Extremely Low” with “Low”  In Cloud Level  Label Encoder is used to label the categories. |
| **All numerical Features** | Check the -99 count in each feature, if the count has some prominent values and there are more negative values then replace with the mean of feature.  If there are less negative counts , assign -1 to all negative values.(treat -99 as outlier). |

**Models ,Libraries and Encoders**

|  |  |
| --- | --- |
| **Model** | XgBoost with hyperparameter optimization(Grid Search Cv). |
| **Library & Encoders** | Label Encoder , Pandas , Numpy , Seaborn ,Matplotlib ,Sklearn. |