**Early warning and predictive maintenance system**

**Objective -**  To predict and reduce the number of false reject since it leads to unplanned downtime and unnecessary maintenance check.

**Approach** –

1. Basic EDA to understand dimensionality and variable types of data.
2. Created dependent variable “False Reject” based on definition provided in the assignment.
3. Checked for missing values, Outlier and Skewness of data.
4. Converted variable “PROD\_TIMESTAMP” to date time.
5. Data is highly imbalanced as only 374 records (0.08%) are false rejects out of 4,56,124.
6. Classification metric chosen for Validation – F1 score
7. Machine learning model implemented – a. Random Forest

b. Xgboost

c. Cat-boost

1. Tried above algorithms with resampling approach - ADASYN (Oversampling) and Undersampling.
2. Performed label encoding on high cardinality variable “TOPLEVELSERIALNUMBER”.
3. Created “N-1” dummy variables for “PARTNUMBER”,” CHARDESC”.
4. Created multiple features on time virile like – Month, Week, Day, Hour, Minute, Second,

Week etc.

1. Performance table for below algorithms –

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **S.No** | **Model** | **Sampling technique** | **Train F1 score** | **Test F1 score** | **Remarks** |
| 1 | Random Forest | None | 0.8313 | 0.8220 | Well balanced score with train and test |
| 2 | Xgboost | None | 0.8694 | 0.8223 | Performed decent across |
| 3 | Catbost | None | 0.9255 | 0.8681 | Performing best on test set |
| 4 | Random Forest | Downsampling | 0.9336 | 0.8465 | Sign of over fitting |
| 5 | Xgboost | Downsampling | 0.9336 | 0.8465 | Sign of over fitting |
| 6 | Catbost | Downsampling | 0.9552 | 0.8538 | Sign of over fitting |
| 7 | Random Forest | Adasyn (Oversampling) | 0.999931 | 0.99981 | Generalizing too well |
| 8 | Xgboost | Adasyn (Oversampling) | 0.999956 | 0.99982 | Generalizing too well |

1. Feature importance on the basis of Catboost algorithm –
2. Confusion , Precision and Recall Matrix (Catboost without Resampling) –





