

PART-B**(30 Marks)****Predict which machine would break down next.**

Dataset consists of anonymized features of machines environment. You must predict whether machine will breakdown. The data is collected from a fleet of machines of the same type. Each machine starts with different degrees of initial wear and manufacturing variation which is unknown to the user.

Data

Name of the dataset is “Machine_Data”.

“NA” in the data set are populated as “-1”

You will be evaluated based on the following:

- Data Understanding
 - Understand the data statistics
- Data Preprocessing
 - Perform necessary data preprocessing, data transformations and feature engineering
- Exploratory analysis
 - Explore any underlying patterns of the Machines
 - Come up with comprehensive Visualizations that show the patterns
 - Discuss the insights from the Visualizations
- Model Building
 - Apply ML algorithms to predict whether a machine will breakdown or not
 - Apply Regularization, and Dimensionality Reduction Techniques if necessary
 - Use all the relevant classification algorithms that you have learnt so far
- Analyzing the results
 - Choose appropriate metric for the given problem.
 - Report a table depicting the consolidated results of models built along with their train and test metrics – Report the details of the experimentation performed
- Challenges/Further scope
 - Report challenges faced if any, and scope for further improvement
- Deliverables (Group Submission)
 - Entire R code (Including various data preprocessing and modeling experiments) with appropriate comments
 - Power point presentation
- Individual contribution to problem solving & individual performance during presentation