Case Study 1 – Rare Event Classification

PRESENTED BY:

Konda Sri Bala Bhaskar   B210214CH

Kusumanchi V S S K Lokesh Gupta   B210539CS
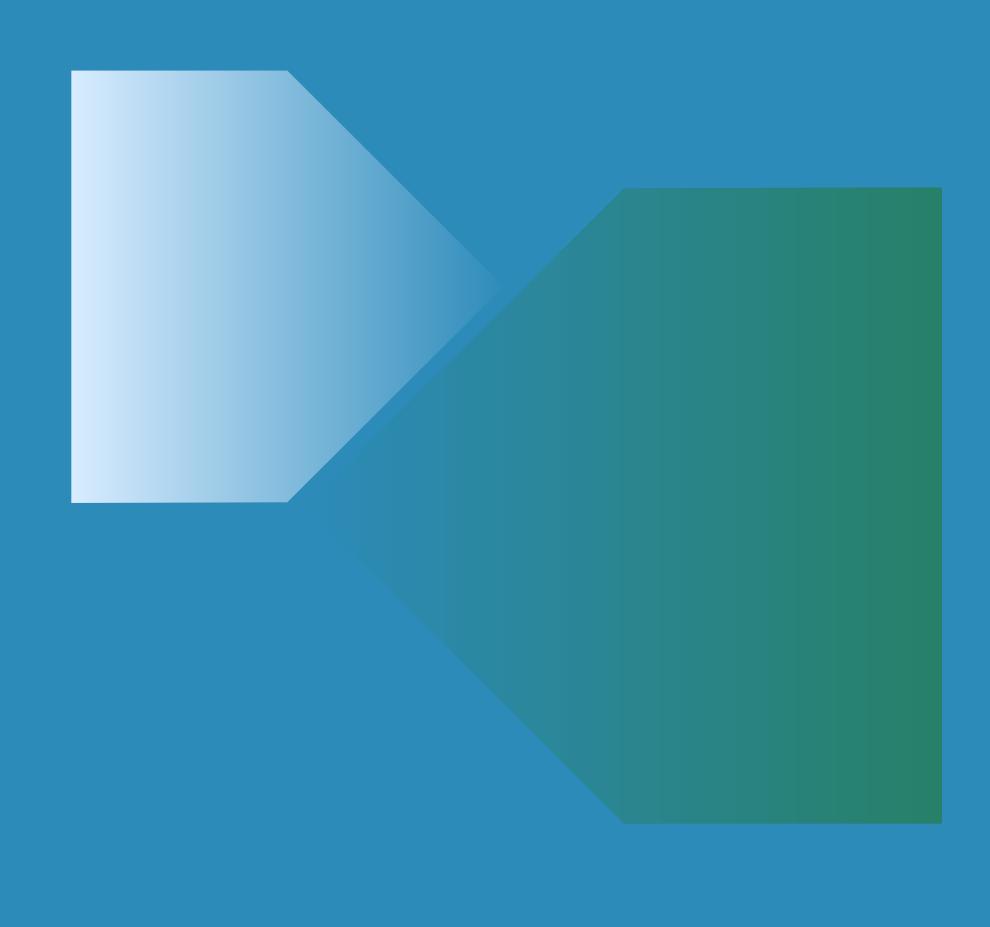
Vedurupaka Venkata Sai   B210437CS

Kavuru Naveen Kumar   B210560CS

Nimmalapudi Leela Krishna   B210569CS

Akash Potturi   B210439CS

Vankadari Krishna Deepak   B210449CS

Agenda

Build a classification model, which can predict failures in the test data set with good precision and recall. The training cutoff date is '21-05-1999'

Milestones

1) **Data overview and Buiseness Impact**
2) **Data cleaning**
3) **Training and testing dataset splitting**
4) **Balancing the data**
5) **Different classification models building**
6) **Accuracy and precision testing**
7) **Selecting the best model**

Data overview

The dataset provided originates from a pulp-and-paper manufacturing industry, consisting of multivariate time series sensor readings.

Buiseness Impact

By deploying an effective classification model, the industry can proactively mitigate the risk of paper breaks, thereby minimizing unplanned downtime, reducing maintenance costs, and optimizing production schedules."

DATA CLEANING

Pandas, a Python library, was employed for data cleaning tasks such as handling missing values, removing duplicates, and outlier detection, ensuring data consistency and integrity

Training and Testing Dataset Splitting

The dataset was split into training and testing sets using techniques like time-based splitting, ensuring temporal integrity in the data.

Balancing the Data

To address class imbalance, techniques such as oversampling, undersampling, or generating synthetic samples were applied, ensuring balanced representation of the rare event of paper breaks.

Different Classification Models Building

Various classification algorithms including Random Forest, Support Vector Machines (SVM), Logistic Regression, and Gradient Boosting were employed to build predictive models, each offering unique advantages and limitations

ACCURACY AND PRECISION TESTING

Model performance was evaluated using metrics such as accuracy, precision, recall, and F1-score, with a focus on precision and recall due to the rare event nature of paper breaks.

SELECTING THE BEST MODEL

Models were compared based on performance metrics, with emphasis on precision and recall, to select the most effective model for predicting paper breaks. Deployment strategies and ongoing monitoring were discussed to ensure real-world effectiveness.

THANK YOU