

AnomaData Project Report

1. Project Overview

This project aimed to predict anomalies in machinery using a binary classification approach. The provided dataset contained 18,398 rows with 62 columns. The primary target variable was 'y'.

2. Steps and Workflow

- 1. Data Preparation: Missing values and outliers were handled, and the 'time' column was engineered.
- 2. Class Imbalance: Addressed using oversampling.
- 3. Feature Engineering: Extracted features from the 'time' column.
- 4. Model Training: A Random Forest model was trained on balanced data.
- 5. Evaluation: The model achieved a perfect classification on the test set, with ROC-AUC = 1.0.

3. Results

Classification Report:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	3655
1	1.00	1.00	1.00	3655
accuracy			1.00	7310
macro avg	1.00	1.00	1.00	7310
weighted avg	1.00	1.00	1.00	7310

Confusion Matrix:

AnomaData Project Report

[[3654 1]

[0 3655]]

ROC-AUC Score: 1.0

4. Deployment Plan

The model pipeline has been saved, including the scaler and trained model. It can be deployed in a production environment to predict anomalies in real-time data.