DASSA HACKATHON

Team Name: MUSTATicans

Team Members:

- Santhosh Pattamudu Manoharan
- Saaijeesh Sottalu Naresh
- Niranjan Cholendiran

Data Exploration

- Period ID is just the continuous occurrence, and it changes after a faulty machine is fixed
- Good/Bad - If it becomes bad, then it remains bad until the period ID changes
- Few sensor data were same entirely

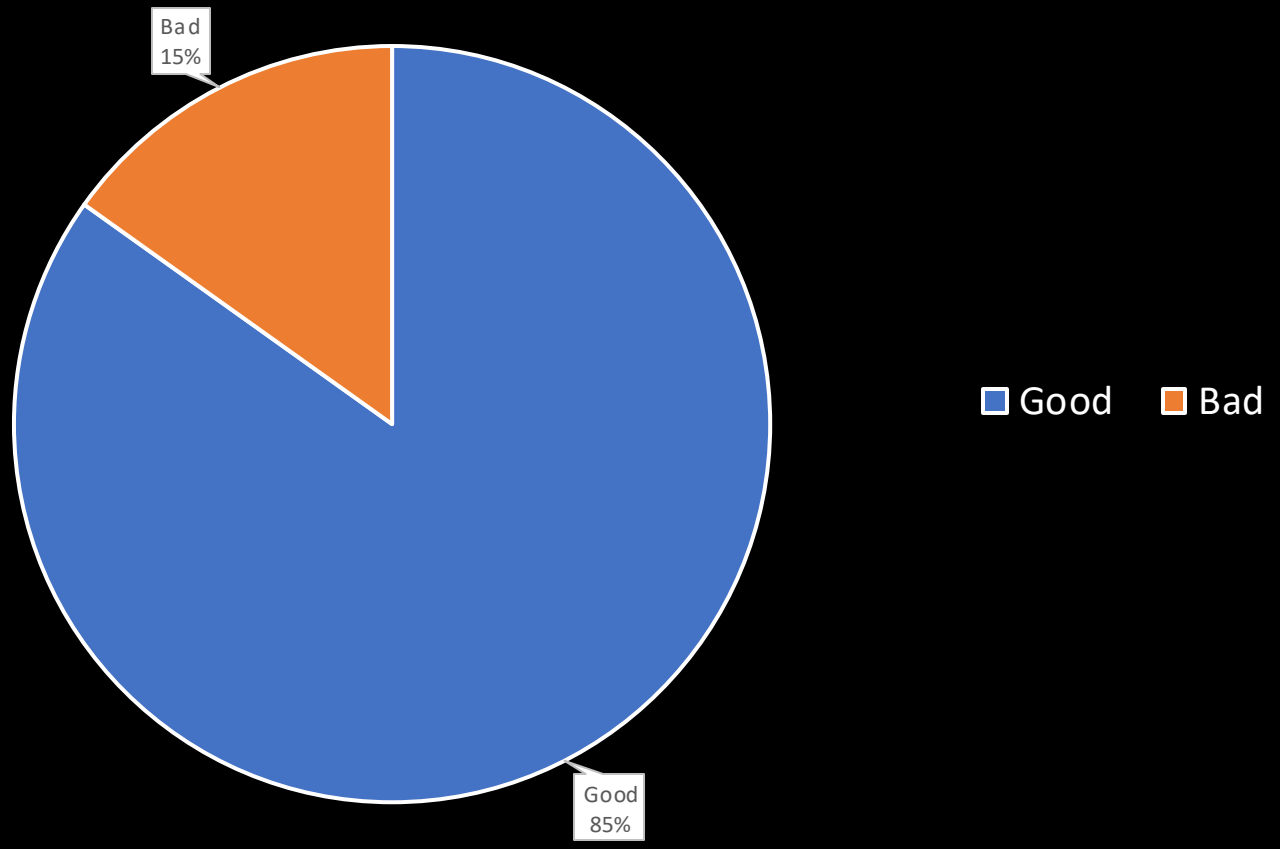
Period Co	Cycle ID	B_2	B_3	B_4	B_5	B_9	B_10	B_14	B_15	B_16	B_17	B_18	B_19	B_20	B_21	B_22	B_23	B_24	B_25	Good/Bad	timestamp	Time_Until_Last_Repair_in_mins
1	1	-0.0007	-0.0004	100	518.67	14.62	21.61	1.3	47.47	521.66	2388.02	8138.62	8.4195	0.03	392	2388	100	39.06	23.419	0	3/1/2020 0:00	0
1	2	0.0019	-0.0003	100	518.67	14.62	21.61	1.3	47.49	522.28	2388.07	8131.49	8.4318	0.03	392	2388	100	39	23.4236	0	3/1/2020 0:05	5
1	3	-0.0043	0.0003	100	518.67	14.62	21.61	1.3	47.27	522.42	2388.03	8133.23	8.4178	0.03	390	2388	100	38.95	23.3442	0	3/1/2020 0:10	10
1	4	0.0007	0	100	518.67	14.62	21.61	1.3	47.13	522.86	2388.08	8133.83	8.3682	0.03	392	2388	100	38.88	23.3739	0	3/1/2020 0:15	15
1	5	-0.0019	-0.0002	100	518.67	14.62	21.61	1.3	47.28	522.19	2388.04	8133.8	8.4294	0.03	393	2388	100	38.9	23.4044	0	3/1/2020 0:20	20
1	6	-0.0043	-0.0001	100	518.67	14.62	21.61	1.3	47.16	521.68	2388.03	8132.85	8.4108	0.03	391	2388	100	38.98	23.3669	0	3/1/2020 0:25	25

Data Cleaning & Preprocessing

1. Replaced nulls and strings in the sensor data with its median.
    1. Why Median?

2. Data Type handled
3. Normalization

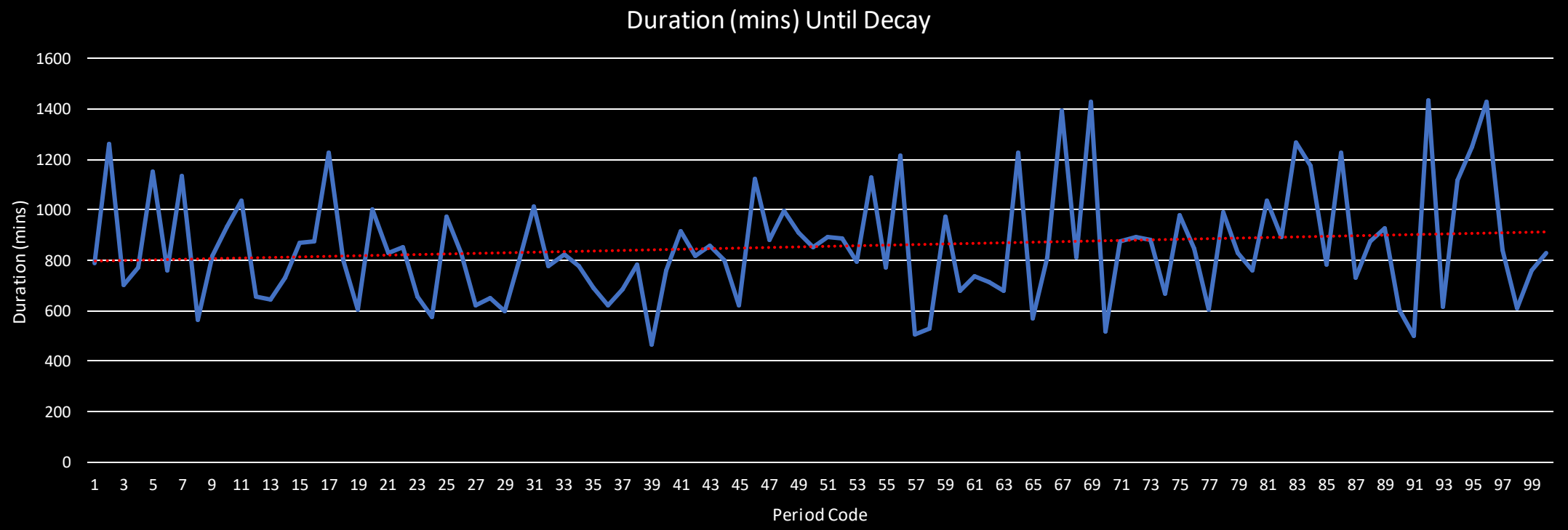
Target Value Distribution



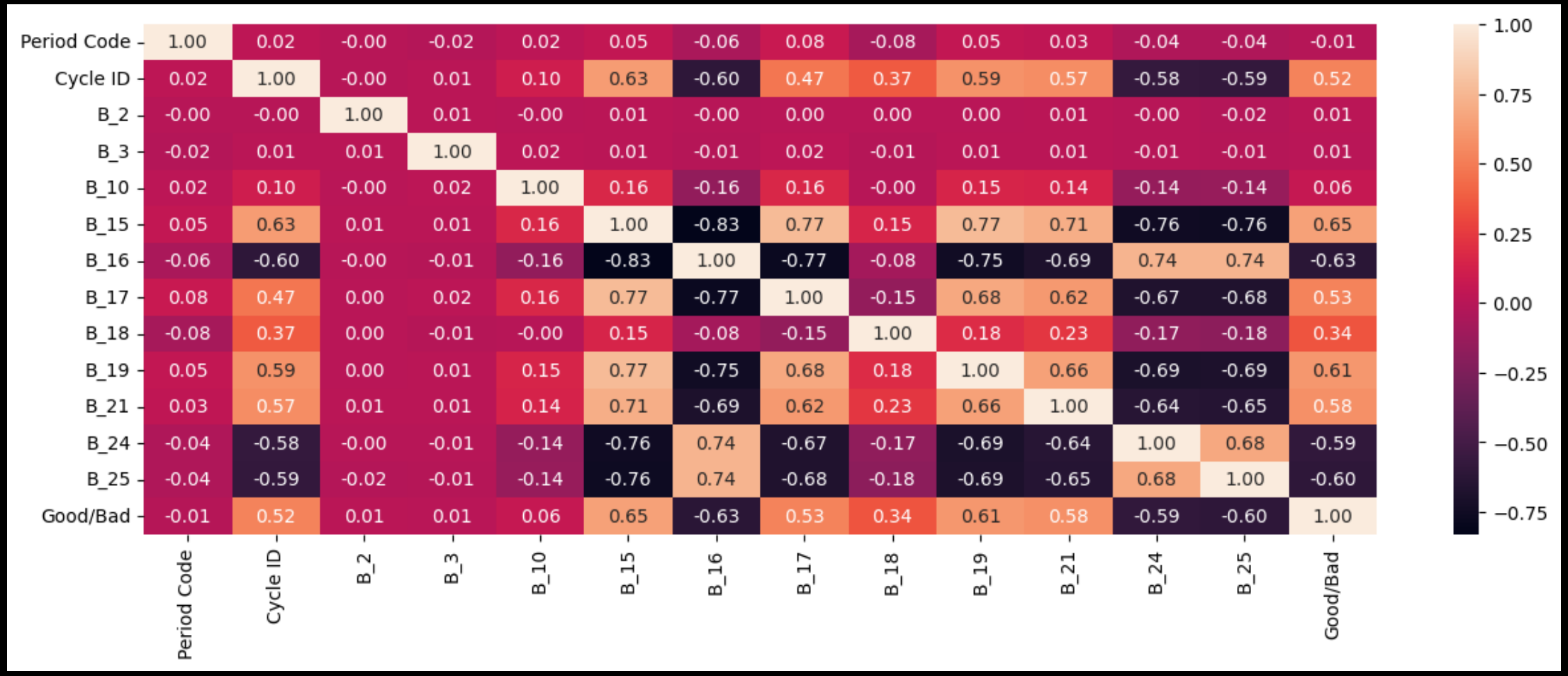
Time Decay

- Added “Time_Until_Last_Maintenance” column
  1. Gives the decay time of the machine
  2. It captures the time information

Average: 854Mins

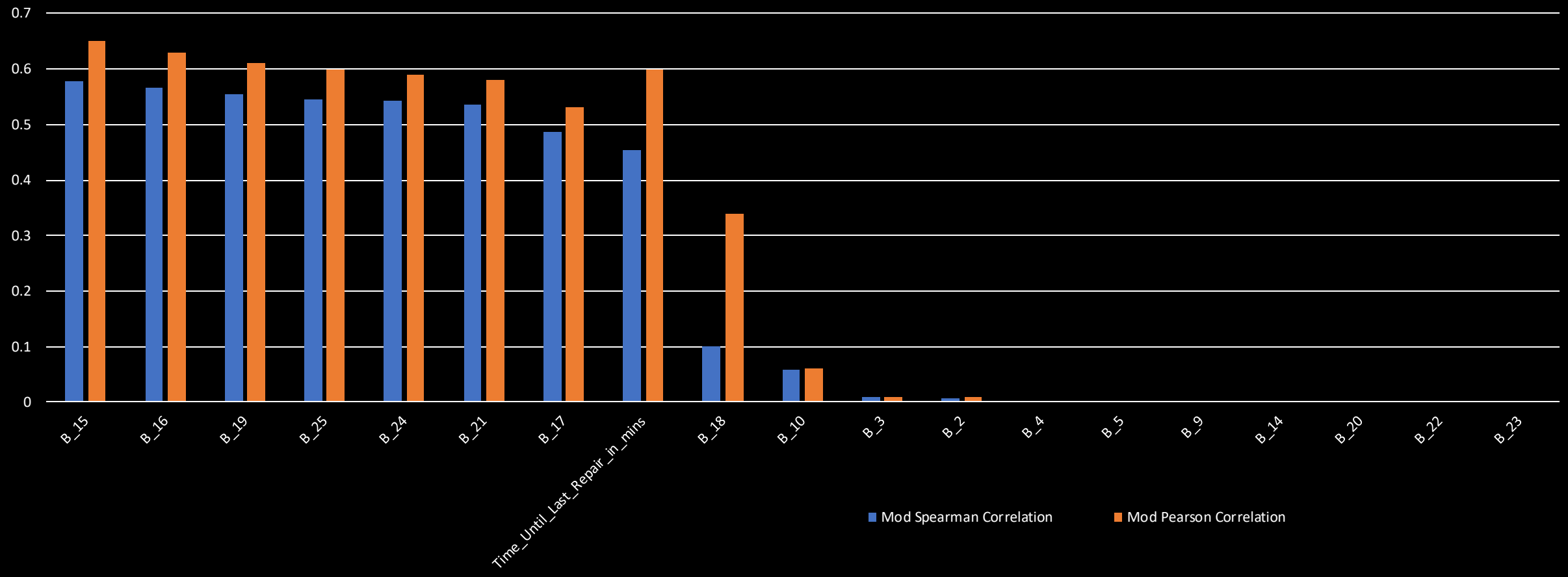


Correlation Between Variables



Non-linear Correlation Between Variables

- Dropped poorly correlated columns



Modelling

- **Model Used**: Random Forest Classifier
- **Why?**
  - Performs better for non-linear relationship data
  - Captures temporal patterns well
  - Robust to outliers
- **About the model:**
  - Ensemble- combining multiple models
  - Bagging- creates different training subsets from different sample training data
  - Output is based on majority voting (parallel process)

Modelling

- **Hyperparameter Tuning:**
  - **Iterative Tuning**
  - **Regularization Techniques – min_sample_splits**

Modelling

- **Evaluation Metrics:**
  - F1 Score (0.86)- Recall(0.84)
    - Balances Precision and Recall. Harmonic mean of precision and recall.
    - Skewed data distribution

Thank You