

Reports

- Link

<https://github.com/cccccircke/Final>

- Any code from other resources

<https://www.kaggle.com/code/rebecca610654/tpsaug22-eda-which-makes-sense/edit>

- Methodology (Data pre-process, Model architecture, Hyperparameters, ...)

- ✓ 在缺少測量值的情況下計算有條件的產品故障率，並將其與無條件產品故障率 0.212608 進行比較。
- ✓ 當 $\text{abs}(z) > 2.5$ 且 $\text{p-value} < 2\%$ 時，缺失 measurement_3 和缺失 measurement_5 的條件故障率明顯偏離平均故障率，可以在模型中使用特徵 missing_3 和 missing_5
- ✓ 特徵工程：我們可以通過添加按產品代碼分組的測量聚合統計數據作為新特徵，將產品代碼用於特徵工程
- ✓ 主要 NA 用 HuberRegressor 處理
- ✓ 其他 NA 用 KNNImputer 處理
- ✓ Model architecture 用 linear_model.LogisticRegression 和 GroupKFold(n_splits=5)預測

● Summary

■ 109550106_Final_train

用 HuberRegressor 填滿主要 NA 後，再用 KNNImputer 田次要 NA，主次由參考 LINK 選定特徵，將填滿的 TRAIN 經由 WoEEncoder 轉換最後用 linear_model.LogisticRegression 和 GroupKFold(n_splits=5)訓練 MODEL

■ 109550106_Final_inference

用 109550106_Final_train 訓練好的 model 預測

● Result



109550106.csv

Complete (after deadline) · 1h ago

0.59114

0.59077