Predictive Analysis & Data-Driven Decision-Making for Operational Efficiency in Marine Propeller Manufacturing

Caleb Cox, Yoonyoung Choi, Jason M. Dong, Sreekar Gudipati, Prathik Iyengar, Siddhanth Kumar, Monish Muralicharan, Tuan Minh Pham, Agrim Sharma, Yuze Tang

PURDUE UNIVERSITY | The Data Mine

YAMAHA

Setting the Context

Yamaha stands as a leader in marine propeller manufacturing, setting the industry standard for excellence and innovation. Despite a meticulous manufacturing process, occasional defects or manufacturing errors may arise, leading to flaws in propellors – referred to as scrap - which result in unnecessary loss of time and financial resources.

Central Questions:

- How can we leverage the available data to understand why scrap occurs?
- Can we predict the possibility of scrap in real-time?
- Is there a seamless way to explore the available data?

Key Deliverables

- Machine Learning model capable of identifying reasons for scrap production and providing real-time predictions.
- Data Analysis platform to seamlessly explore and identify KPIs.

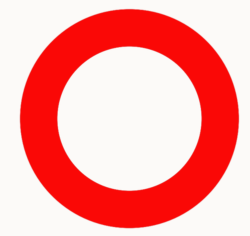
References & Acknowledgements

- "Feature-Selected Tree-Based Classification," in IEEE Transactions on Cybernetics, vol. 43, no. 6
- Kaggle's "Comprehensive Guide on Feature Selection"
- The Data Scientist's "Understanding Tree-Based Models"
- Thank you for all the support and assistance:
  - Our TA: Jebran Syed
  - Our Mentors: William Irwin & Batuhan Ak
  - The Data Mine Staff

Data Preparation

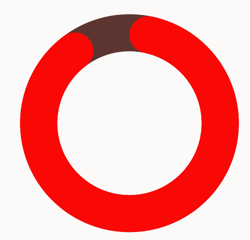
In order to build the model, we must first ensure data integrity and conformity through exploratory data analysis and statistical techniques.

- Removal of null, duplicate and outlier values.
- Scaling of the numerical data by normalization.
- Addressing dataset imbalance through statistical synthesis of records.



180,000 datapoints

25 distinct features



97:3 Ratio Non-Scrap to Scrap

Exploratory Data Analysis

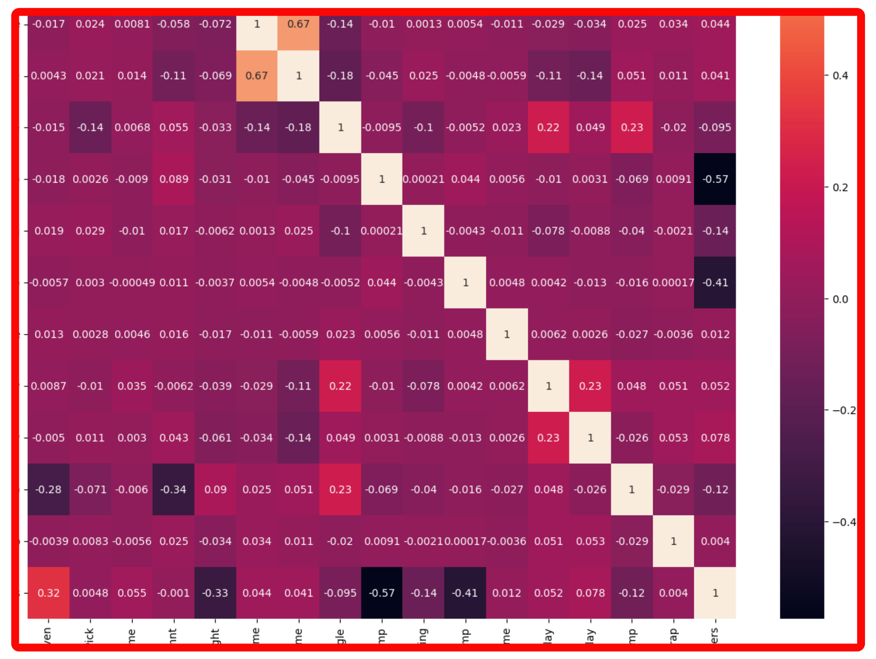
Exploring the processed data is an important step towards identifying important anomalies and patterns.

Identify correlations to support feature selection

Plotting feature interactions to identify predictors

Notable Results: Through these steps, we were able to identify unnecessary variables (Fig. 1) and key variable interactions that could contribute to scrap prediction. For example, increased time spent between certain processes showed a positive correlation with scrap.

FIG 1. HEATMAP TO DETERMINE LINEAR CORRELATIONS BETWEEN FEATURES



Data Analysis Platform

We developed an interactive, light-weight platform for exploratory data analysis of the manufacturing data.

- Empowers Yamaha users to view important manufacturing KPIs.
- Includes dataset customization and feature interactions
- Provides visualizations of key dataset features.

Model Selection & Development

We selected tree-based models to classify whether a propeller was scrap or non-scrap

- Effective for imbalanced datasets (scrap vs. non-scrap data)
- Efficiently handles complex relationships between various features and classification outcomes

Model Performance

We observed an F1 score of 94% with the ExtraTrees model. This means the model correctly identified scrap propellers while minimizing the misclassification of good propellers as scrap 94% of the time. (Fig 3.)

However, similar performance could not be achieved when tested with unseen data due to the lack of, and highly variable nature of, the data.

However, this model lays a solid foundation for future advancements with the integration of more data and ML expertise.

DecisionTree	0.840724	0.877436	0.858688
RandomForest	0.930686	0.937044	0.933854
ExtraTrees	0.938981	0.950238	0.944576

FIG 3. TREE-BASED MODEL SCORES ON TEST DATASET

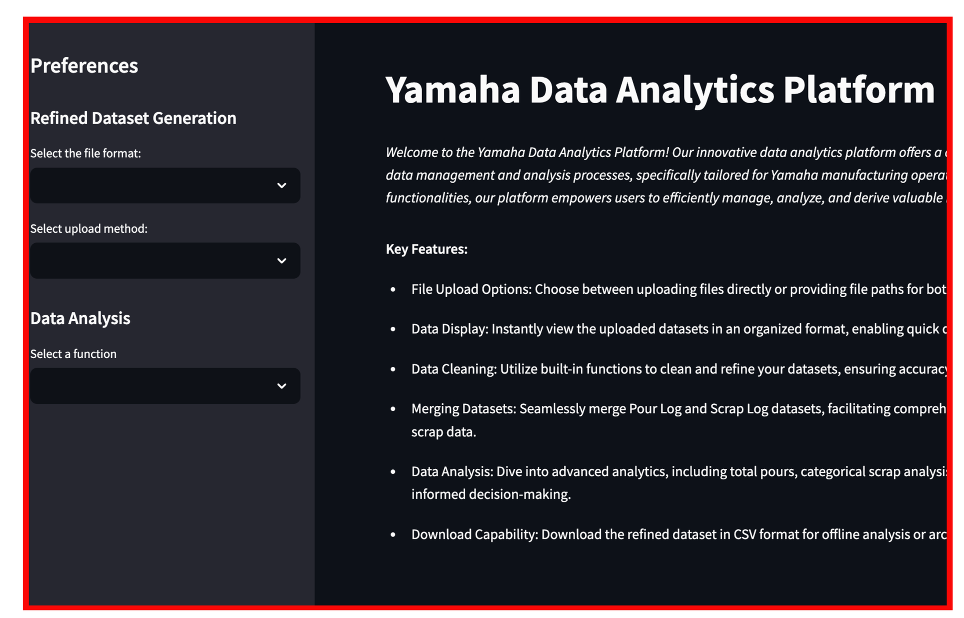


FIG 2. HOME PAGE OF THE DATA ANALYSIS PLATFORM

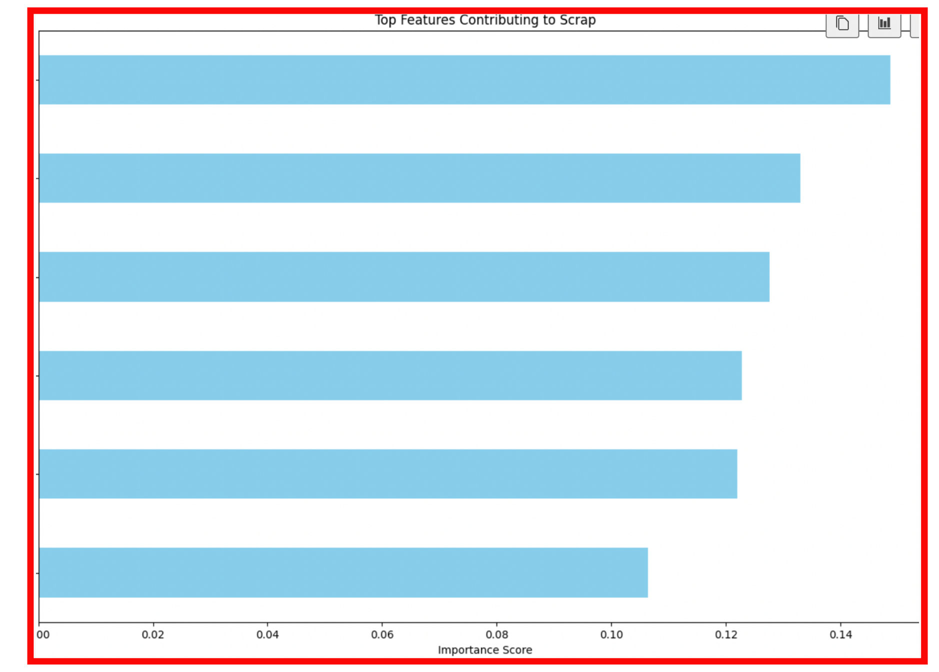


FIG 4. MODEL RANKING FEATURE IMPORTANCE WHICH PINPOINTS THE REASONS FOR SCRAP

Business Value Delivered

- Reduced technical scrap by about 50% compared to scrap rate in January 2023
- Identified interactions between process variables and casting defects

Future Goals

- Improve performance on unseen data through training and domain-expert backed feature transformation.
- Assist Yamaha in model implementation to produce real-time outputs on the manufacturing floor.
- Capture information like feature importance to observe trends and implement Real Time ML (RTML)