Predictive Aircraft Maintenance:

Detecting Imminent Part Failure with Cox Regression and Advanced Ensemble Learning Methods


2Lt Abraham Eaton
2Lt Joshua Couse

Advised By: Dr. Allison Chang
Dr. Chelsea Curran
Ms. Cindy Engholm
Mr. Yan Glina

Sponsored By: MIT Lincoln Laboratory,
Group 43,
Air Traffic Control Systems

Problem Statement

Problem Description: The United States Air Force spends $50B+ yearly in aircraft operations and maintenance. Maintaining complex and aging aircraft requires large maintenance crews and around the clock operations. Over the past several years, aircraft mission capable rates have been declining while maintenance costs have increased. How can the Air Force use machine learning to predict when components will require maintenance in order to keep planes flying and our nation safe?

Project Goal: Develop an end to end machine learning pipeline that utilizes historical maintenance records to predict near term component failure.

Project Timeline:

Jan – April	May – June	July – Aug
Problem Framing Data Exploration	Preliminary Modeling Feature Engineering Feature Selection	Model Building Parameter Tuning Presenting Results

Data

Air Force Aircraft Maintenance Log

- Details a specific part removal, replacement, installation or inspection, and date completed
- Indicates type of malfunction, and type of maintenance (scheduled v. unscheduled)
- 3.5 Million records spanning 12 years
- Anonymous Tail Numbers
- 800k Blank Values
- No Dependent Variable

Data Cleaning Pipeline

- 167k Observations
- Filtered data to include accurate tail numbers
- Condensed records into daily maintenance summary by aircraft tail
- Completed time-series for missing dates
- Generated additional features
  - Total breakages, total man hours, total unscheduled maintenance jobs, etc
- Created indicator for part breakage

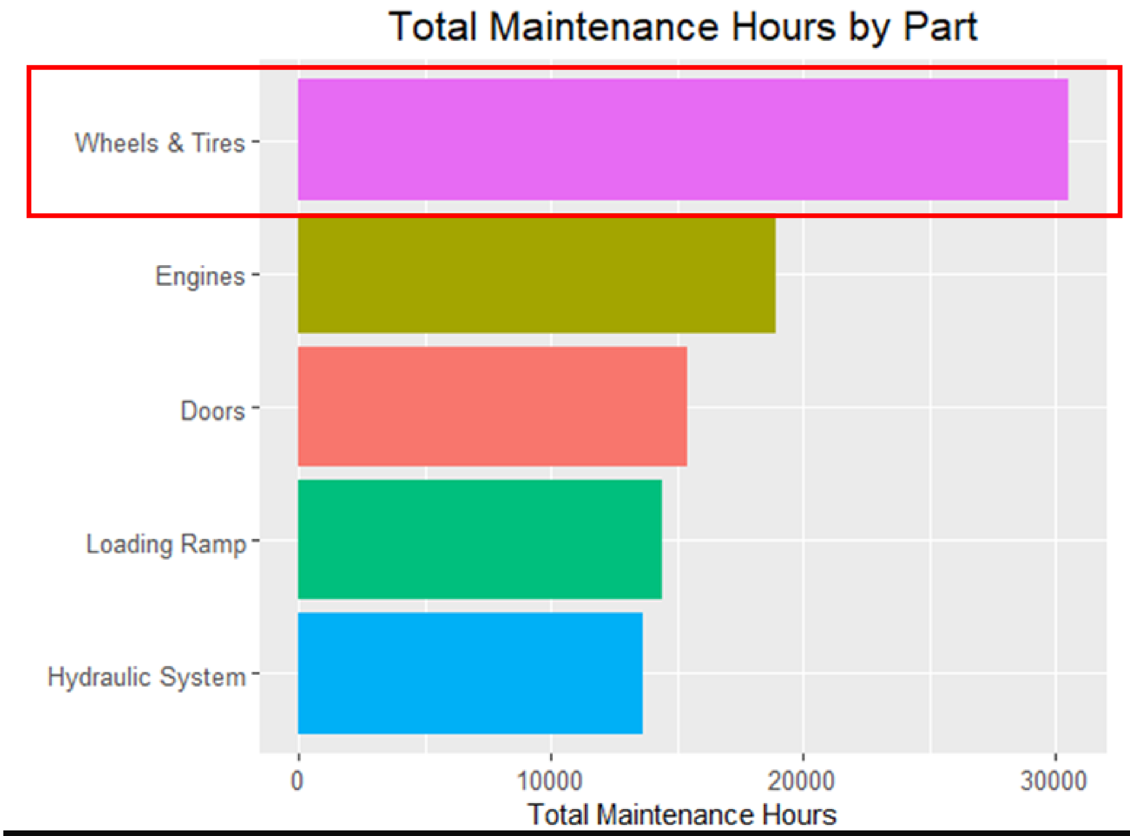
Final Dataset

Modeling Objectives

- **Predict rarely occurring failures of specific aircraft components**
- **Prediction must be early enough to allow for maintenance crew and supply adjustments**
- **Focus on components with largest unscheduled maintenance requirements**
- **Maximize failure detection rate while limiting false positive rate to tolerable measure**

Modeling

1. Decide Which Component Failure to Predict



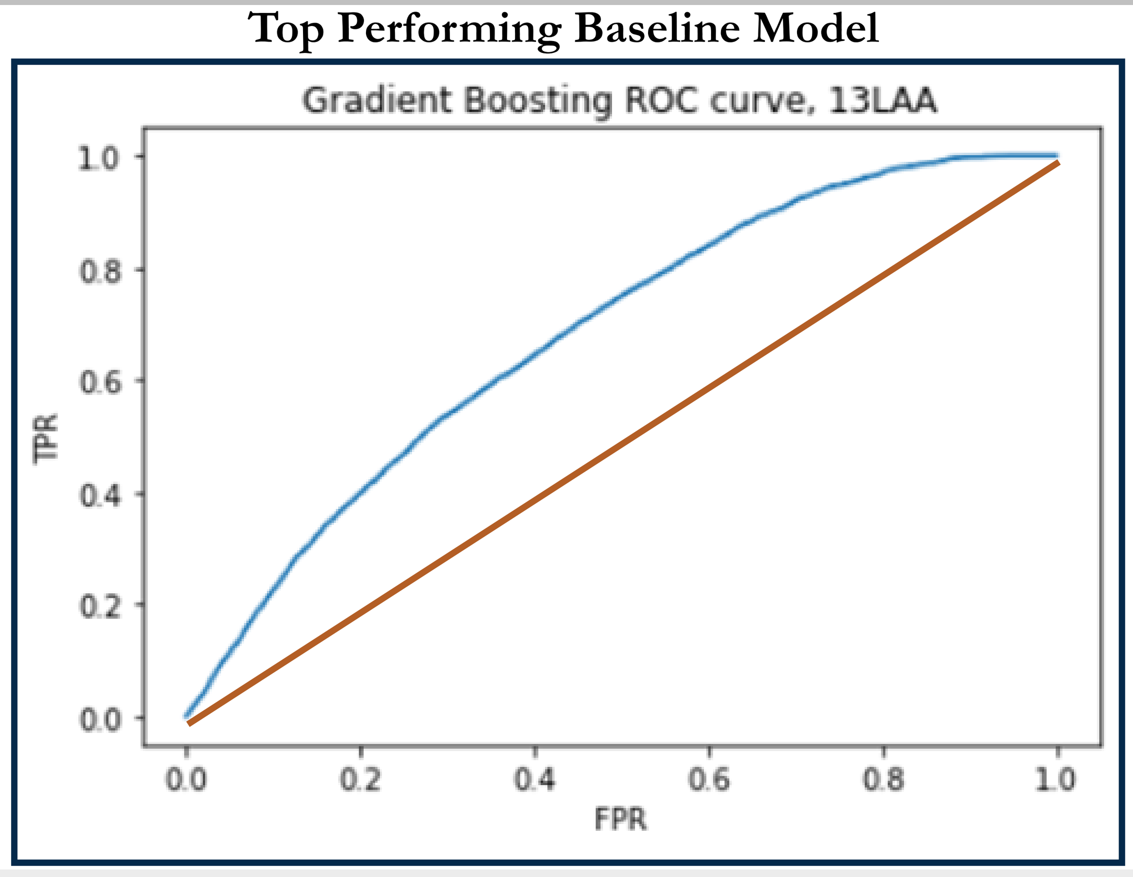
Wheel and Tire failures occur frequently and require the most total man-hours out of all components

2. Build Variety of Models and Evaluate Performance

- Our team tested a variety of machine learning methods to determine which approaches performed the best on our prediction task.
- After finding mediocre results using classification methods we decided to reframe our problem as time-to-event prediction task
- Using a Cox regression model allowed us to take advantage of the time-varying covariates in our survival analysis

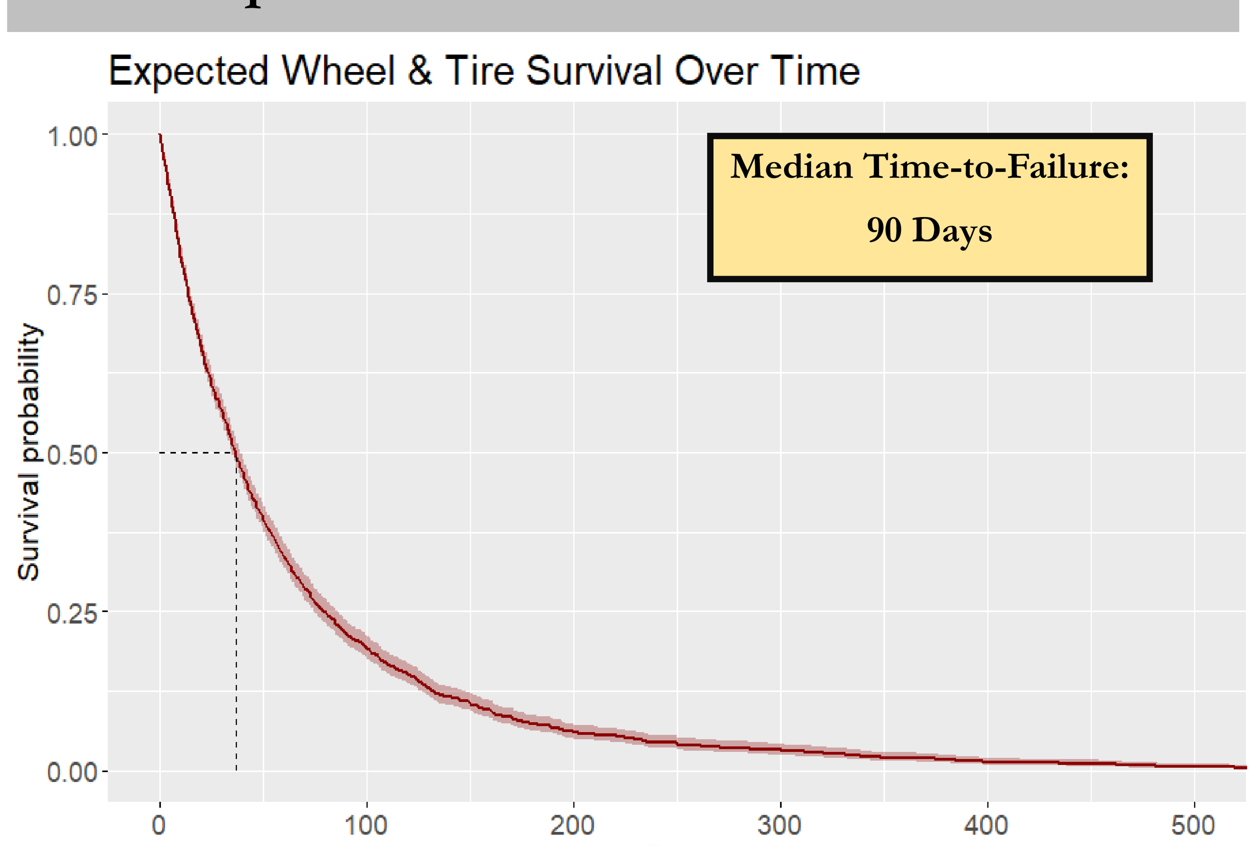
Model	AUC
CART	0.53
Logistic Regression	0.62
Random Forest	0.66
Boosting	0.68

Top Performing Baseline Model



Prediction Variable: Binary indicator of Wheel & Tire failure within 5 days

3. Develop Survival Model

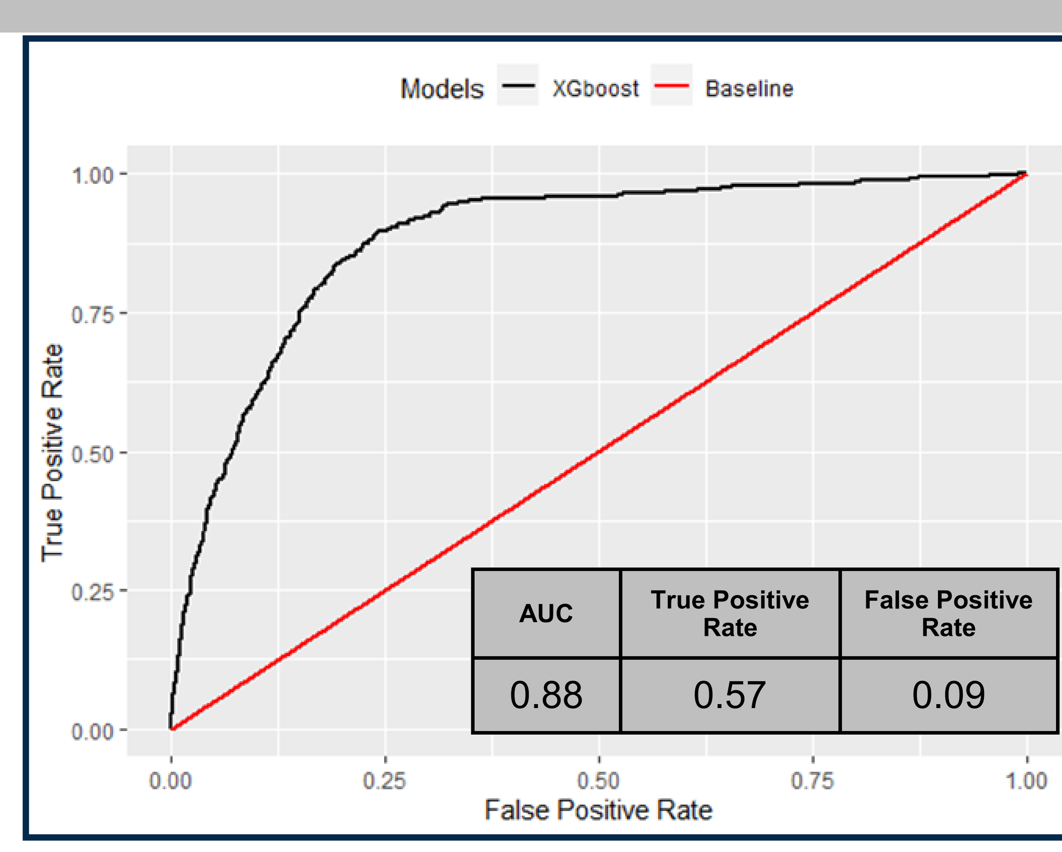


- The survival curve above demonstrates how the wheels & tires of an aircraft deteriorate over time
- A Survival model outputs a survival probability for each day a part is observed providing a statistic of component health
- The median time-to-failure metric represents the point in time when half of the components have failed and is a valuable metric for aircraft maintainers

3. Final Building Model and Performance

Final Dataset + Survival Analysis Predictions

- Utilizing the survival model's predictions as an additional feature to a gradient boosting model improved performance
- Tuning hyperparameter such as depth, learning rate, L2 normalization constant, and training iterations ensures best out of sample model performance



Impact and Conclusion

- Developed effective data management pipeline to transform millions of records into a series of concise and informative daily maintenance summaries
- Our model was able to preemptively identify 2,531 part failures before they occurred, showing potential to save approximately $27M
- Identified an opportunity to add 1,501 available flight hours

Savings:

~$27M

Mission Hours:

1.5K+ Available Flight Hours

LINCOLN LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY