A DATA-DRIVEN FRAMEWORK FOR COMPLEX SYSTEM PROGNOSTICS DEVELOPMENT

KATARINA VUCKOVIC(COLLINS AEROSPACE)
SHASHVAT PRAKASH (COLLINS AEROSPACE)
BEN BURKE(INSIGHT)

NOV 2023

phmsociety conferences

Collins Aerospace

OUTLINE

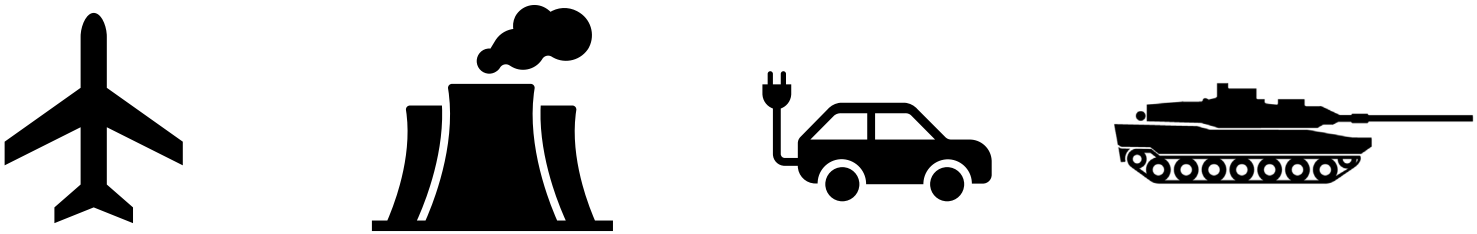
- Introduction
  - Problem Statement
- Framework
  - Opportunity Identification
  - Dataset Preparation
  - Model Development
  - Performance Evaluation
- Numerical Example
- Conclusion and Future Work

INTRODUCTION

PROBLEM STATEMENT

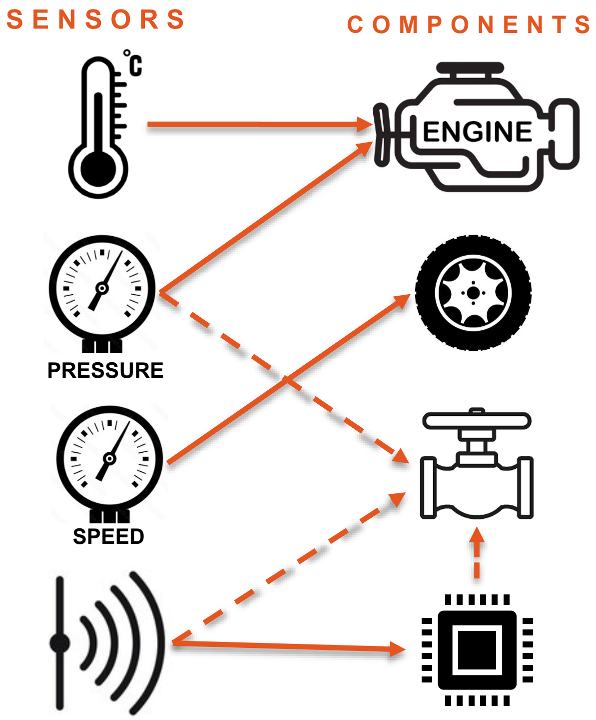
MISSION: DEVELOPING SYSTEM LEVEL ANALYTICS

- Single component are easier to model when they are treated as an isolated system.
- Component level prognostics do not capture the intricate correlations and interdependencies of the component within the complex system.
- Digital Twins have gained interested and over-time they may be able to model these systems at a detailed level.



DATA DRIVEN PROGNOSTICS

SYSTEM IS A COLLECTION OF COMPONENTS AND SENSORS



- How to identify correlations between sensors and components?
  a) Physics/system architecture-based models (obvious)
  b) Data-driven analysis (not so obvious)
- Propose Framework:
  - Data driven approach to identify relationship between sensors and component removals.

KEY ISSUES

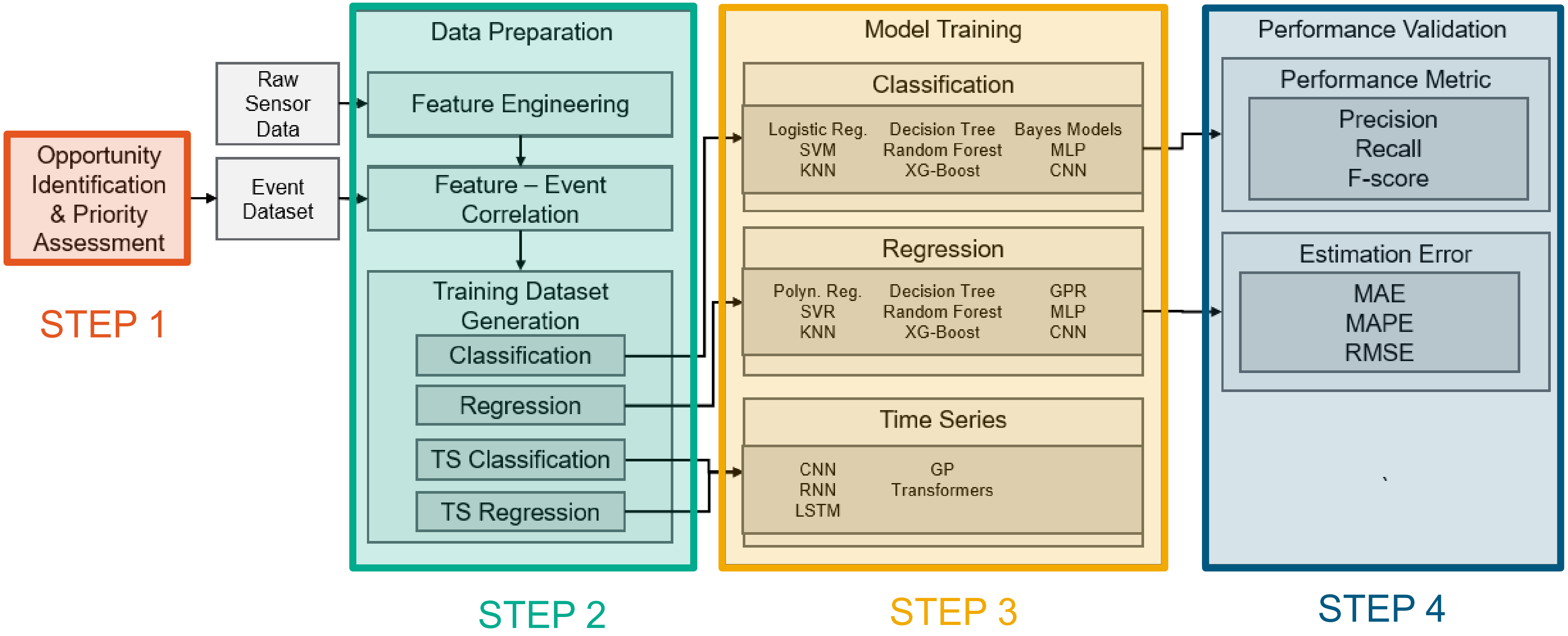
SUMMARY

- The proposed framework is intended to address three key issues:
    1. Ensure system coverage
    2. Identify relevant features for a given component/event
    3. Reduce prognostic development time

FRAMEWORK

FRAMEWORK OVERVIEW

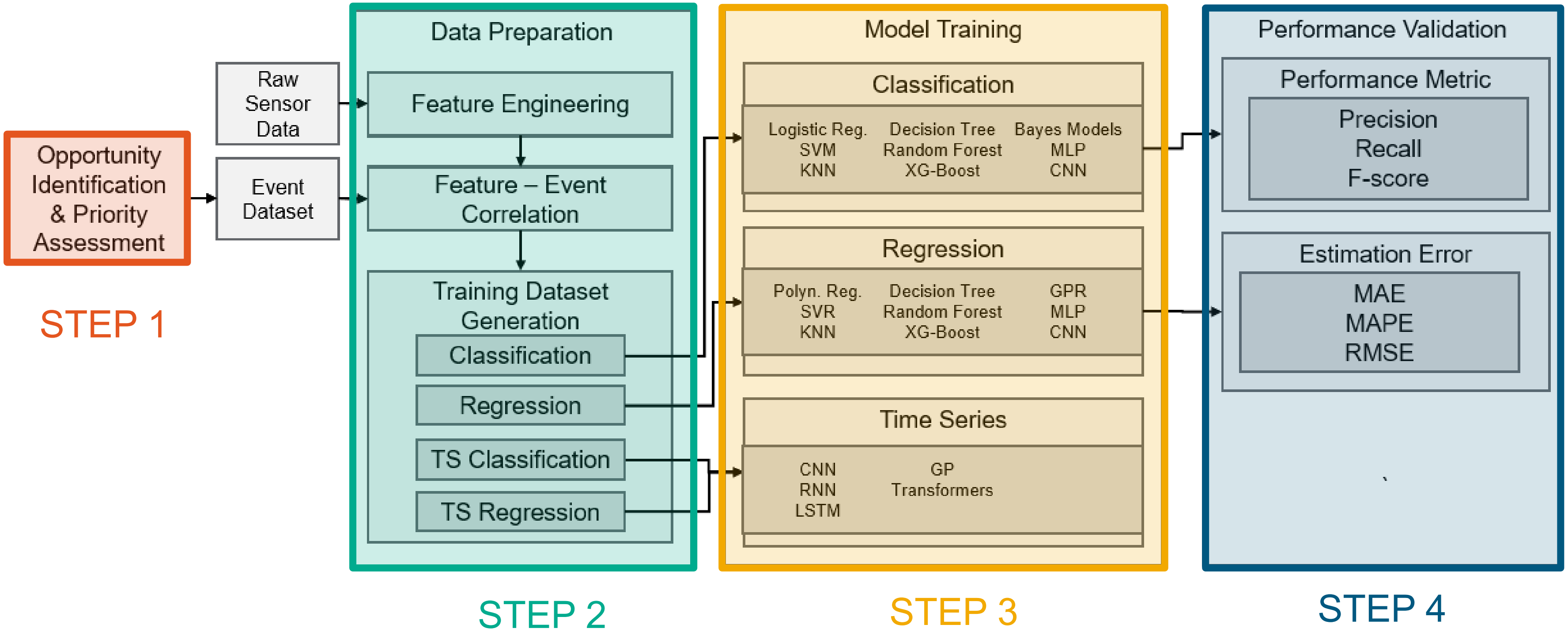
ANALYTIC DEVELOPER WORKFLOW PROCESS



Collins Aerospace

FRAMEWORK OVERVIEW

ANALYTIC DEVELOPER WORKFLOW PROCESS



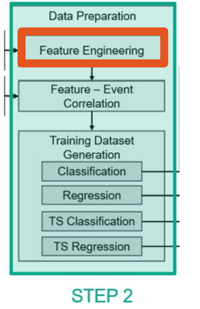
STEP 1: OPPORTUNITY IDENTIFICATION

Opportunity Identification & Priority Assessment

STEP 1

BUSINESS ASPECT OF PROGNOSTIC DEVELOPMENT

- How should we prioritize analytics in an incremental development process?
- Priority Criteria Guidelines:
    - Customer Request
    - Cost-Benefit Analysis
    - Safety Critical Components
    - High Correlation (Quick Wins)



STEP 2: DATASET PREPARATION

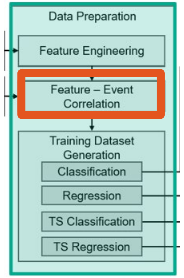
FEATURE CORRELATION

- **Feature Engineering**
  - Apply statistical transformations to raw flight signals to generate features.
  - Transformations reduce raw signals to single values per flight.
  - Features are tracked over time and saved in feature-specific datasets.

Collins Aerospace

STEP 2: DATASET PREPARATION

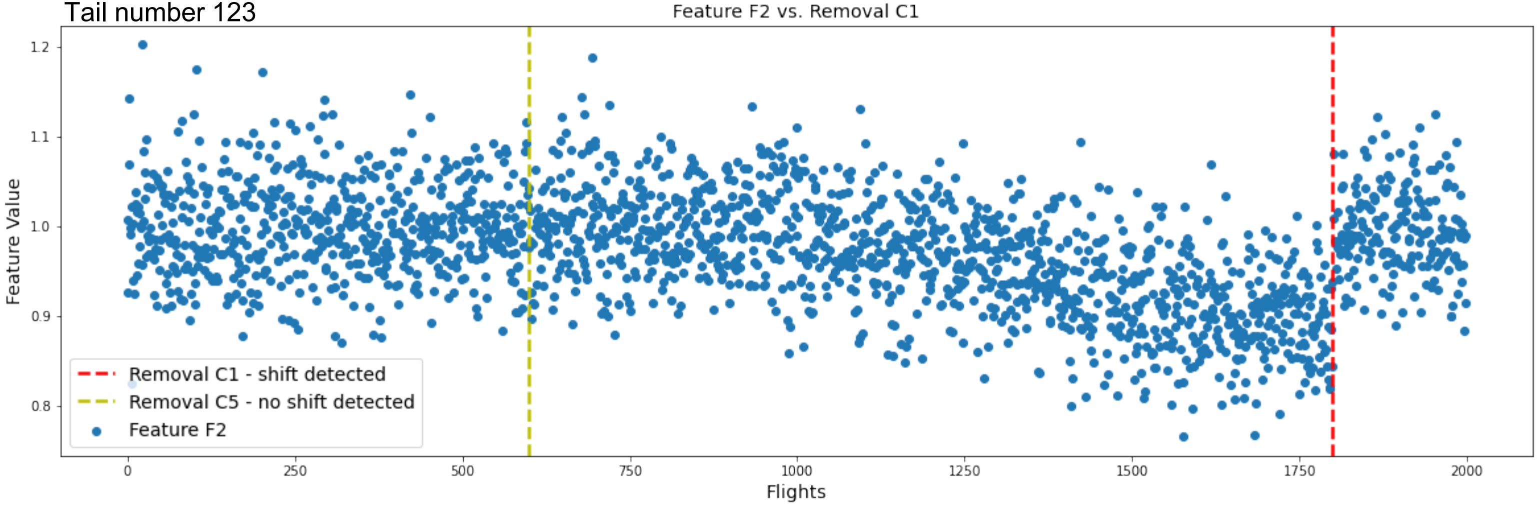
CORRELATION



STEP 2

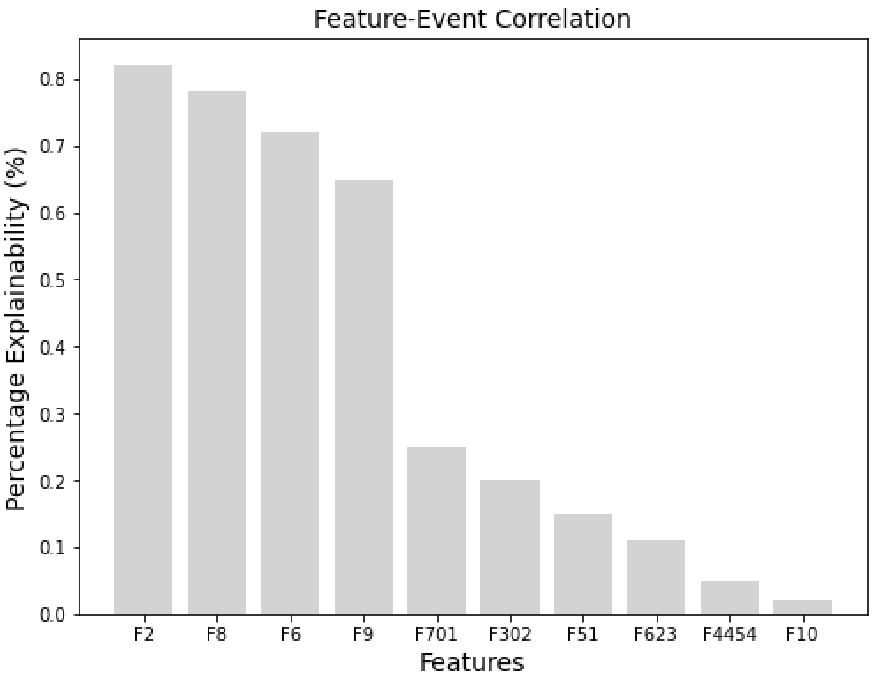
- **Feature-Component Correlations**
  - Perform correlations between unique features and component.
  - Statistically evaluate differences before and after events to identify correlated events.
  - Correlated events are saved for future reference and investigation.

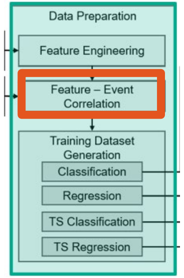
Tail number 123



STEP 2: DATASET PREPARATION

FEATURE SELECTION & EXPLAIN-ABILITY





STEP 2

- Percentage of explain-ability - how many removals of a given component type does a feature explain?
- Example:
  - Component C1 removed 10 times
  - Feature F2 displayed a shift in performance 8 times
  - Explain-ability of C1-F2 is 80%

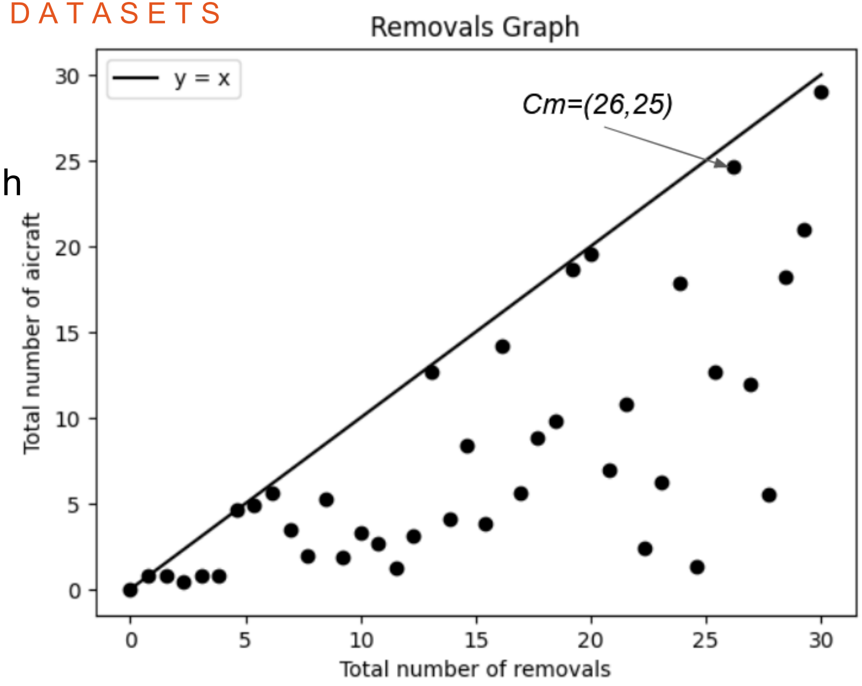
Collins Aerospace

DATA CHALLENGES

EXPLORING CHALLENGES WITH AVAILABLE DATASETS

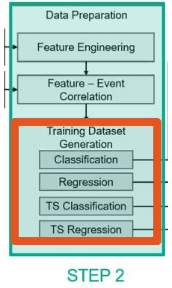
- Requirements:

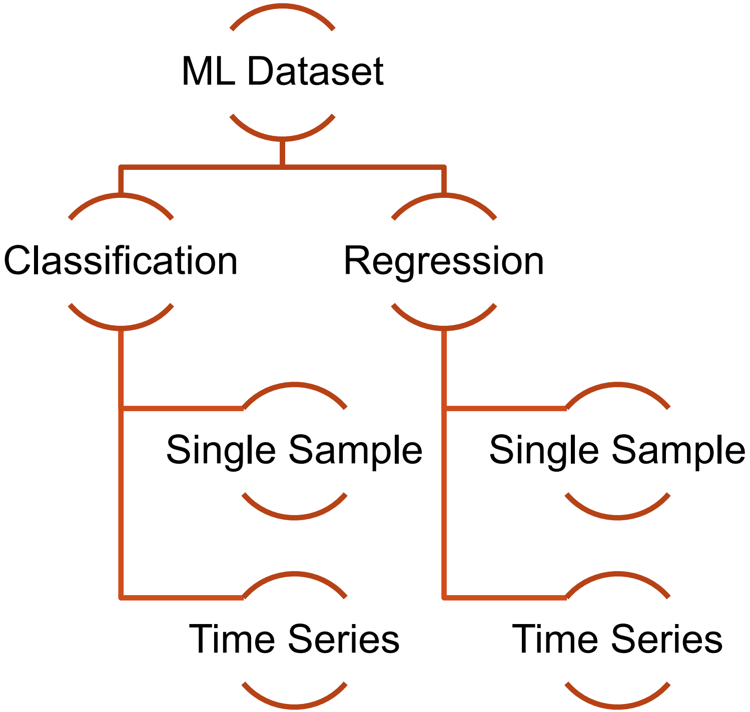
1. The number of component removals is large enough to develop a training dataset
2. The same component is removed on multiple aircraft to avoid aircraft bias.



ML TRAINING DATASET PREPARATION

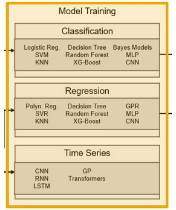
CUSTOMIZED ML LIBRARY





- Classification
  - Healthy vs. Not Healthy
  - Challenge: labeling
- Regression
  - Remaining useful life (RUL) estimate
  - Challenge: harder to estimate
- Single Sample
  - Independent prediction for every new sample collected
- Time Series
  - Predication based on some specified window of data.

Collins Aerospace

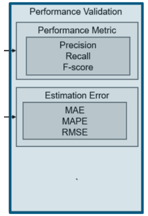


STEP 3

STEP 3: MODEL TRAINING

CUSTOM BUILT AUTO-ML LIBRARIES

Classification	Regression	Time-Series Classification/Regression
Logistic Regression	Polynomial Regression	CNN
Decision Trees	Decision Trees	Gate Recurent Unit (GRU)
Random Forest	Random Forest	Recurrent Neural Network (RNN)
XGBoost	XGBoost	Long-Term Short-Term Memory (LSTM)
Support Vector Machine (SVM)	Support Vector Regression (SVR)	Transformers
k- Nearest Neighbor (KNN)	KNN	GP for time-series
Bayes Models	Gaussian Process Regression (GPR)	
Multi-Layer Perceptron (MLP)	MLP	
Convolutional Neural Network (CNN)	CNN	



STEP 4

STEP 4: PERFORMANCE ANALYSIS

HOW DO WE TEST AND VALIDATE THE MODELS?

CLASSIFICATION MODELS

$$\text{Precision} = P(\text{ImpendingFailureEvent} | \text{PrognosticAlert}) = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalsePositive}}$$

$$\text{Recall} = P(\text{PrognosticAlert} | \text{ImpendingFailureEvent}) = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalseNegative}}$$

		True: Event	True: No Event	
Predicted	Alert	True Positive (FP)	False Positive (TP)	$P = \frac{TP}{TP + FP}$
Predicted	No Alert	True Negative (TN)	False Negative (FN)	
			$R = \frac{TP}{TP + FN}$	

REGRESSION MODELS

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |x_i - \hat{x}_i|$$

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{x_i - \hat{x}_i}{x_i} \right| * 100\%$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \hat{x}_i)^2}{n}}$$

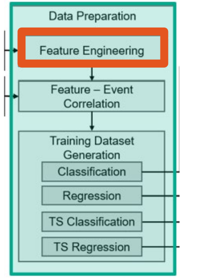
SIMULATION EXAMPLES

DATASET

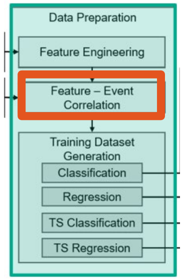
EXAMPLE FROM OUR DATASET

- Dataset cover all sensors and all components on an aircraft fleet.
- Define component as C_i^m
  - i - the component type.
  - m – the number of removals of component C_m
- Define Feature as F_n

Features	20,000
Component Types	2900
Total Removals (All Component Types)	500,000



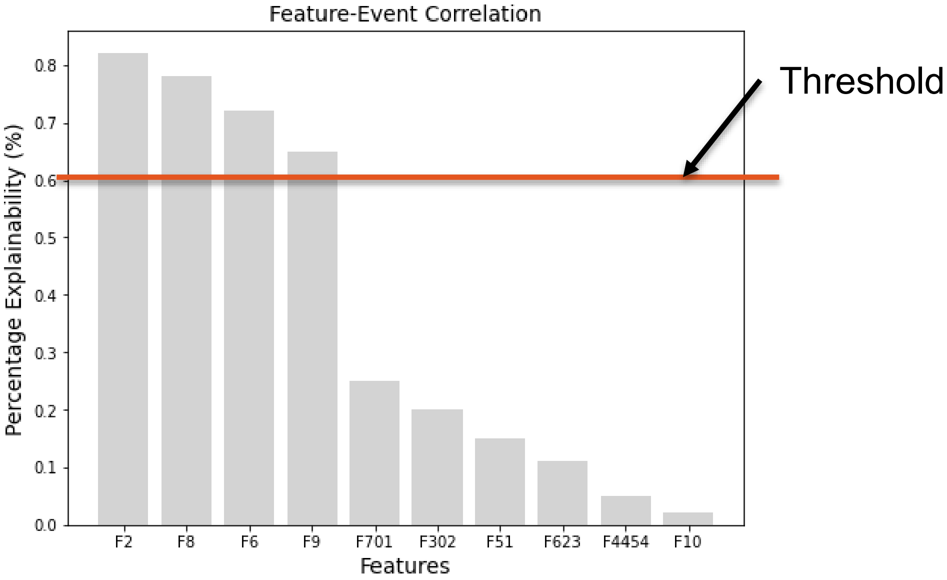
STEP 2



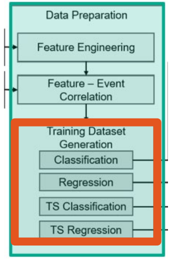
STEP 2

FEATURE SELECTION

1. Assume: Using the “opportunity identification” method, we selected to develop prognostic for component C1
2. Perform Feature-C1 Correlation
3. Pick the most correlated features
   - F2
   - F8
   - F6
   - F9



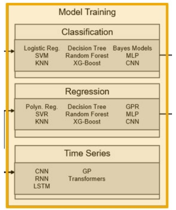
ML TRAINING DATASET PREPARATION



STEP 2

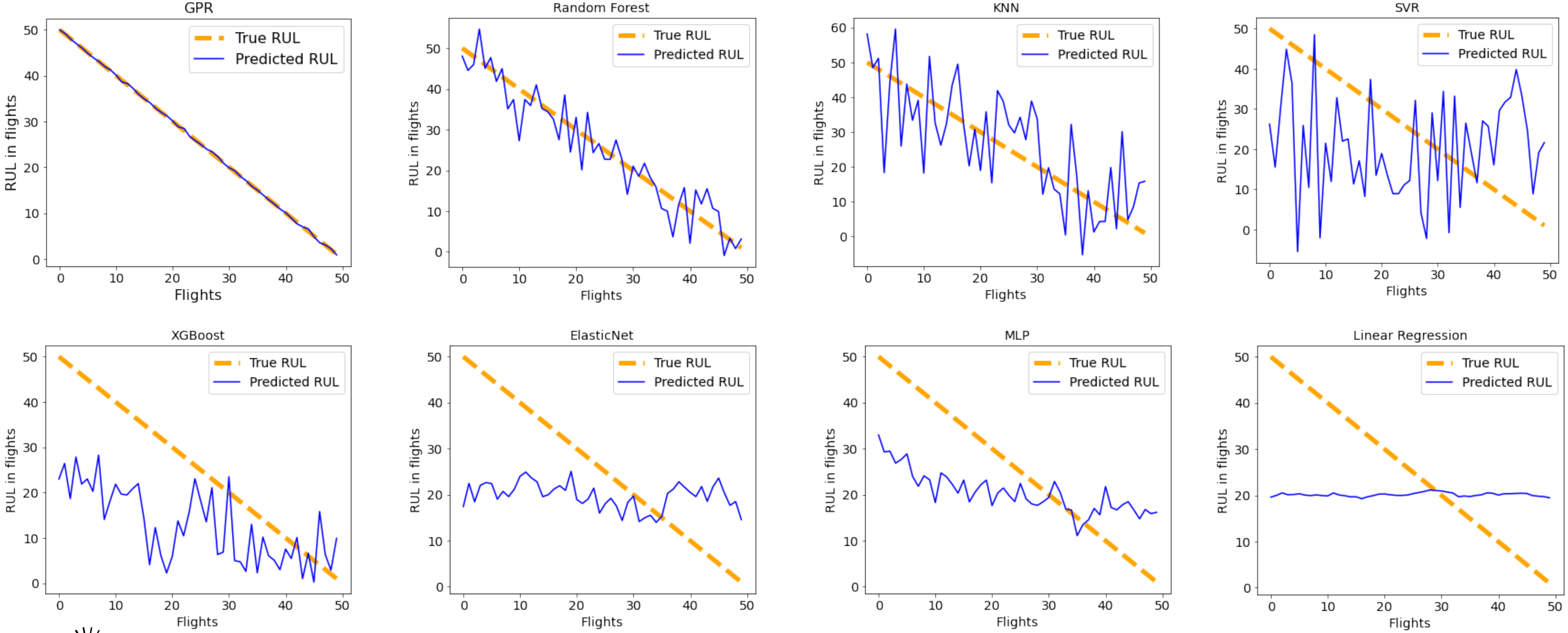
- Single sample RUL regression model.
- Assumption: over the last 50 flight prior to removal. The component transitions from healthy to degraded and finally failed.
- Dataset contain 50 aircraft
- Train/Test Split
  - 45 aircraft are selected for training and 5 for testing.
  - Each tail has one or more component C1 failures.

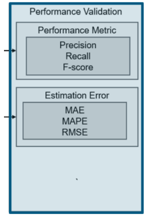
PREDICTION EXAMPLE: TEST SAMPLE 1



STEP 3

PREDICTION ILLUSTRATIONS FOR ONE REMOVAL IN TEST DATASET





STEP 4

PERFORMANCE: NUMERICAL RESULTS

FOR THE ENTIRE TEST DATASET

Models	MAE	RMSE	MAPE
Gaussian Process Regression (GPR)	0.1	0.2	0.5%
Random Forest	4.6	5.61	55.3%
K-Nearest Neighbor (KNN)	10.1	12.2	119.4%
Support Vector Regression (SVR)	11.8	14.1	150.1%
XGBoost	12.1	14.1	154.8%
ElasticNet	12.5	14.5	171.0%
MLP	12.5	14.4	171.8%
Linear Regression	12.5	14.5	170.9%

Collins Aerospace

CONCLUSION & FUTURE WORK

CONCLUSION

SUMMARY

- Discussed challenges of complex systems
- Introduced a data-driven framework for developing prognostic models in complex systems
    - Opportunity identification
    - Feature engineering & feature correlation
    - Auto-ML modeling tool
    - Performance Evaluation
- Demonstrated performance of the framework on a numerical example

FUTURE WORK

WHAT'S NEXT

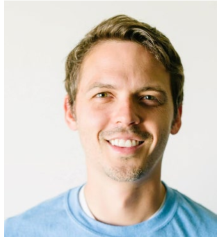
- Expanding the Auto-ML library
    - Adding unsupervised models for unlabeled dataset
- Improving on feature engineering
    - Currently features are combined at the ML level
    - Future Enablement: Combine features at the features engineering level
- Introduce better performance metrics

CONTRIBUTORS

BIOGRAPHY

Katarina Vuckovic received her B.S. in Aerospace Engineering (2017), B.S. in Electrical Engineering (2017), and M.S. in Electrical Engineering (2019) from Florida Institute of Technology. She is currently pursuing her Ph.D. studies in Electrical Engineering at the University of Central Florida. She has been with Collins Aerospace for six years working as a systems engineer on wireless communication systems, aircraft automation applications, and prognostics and health management of aircraft components.

Shashvat Prakash received a B.S. in Mechanical Engineering from University of Illinois, Urbana, an M.S. in Mechanical Engineering from Carnegie Mellon University, and a Ph.D. in Mechanical Engineering from the Georgia Institute of Technology. He has worked on satellite attitude and orbit control at NASA Goddard, combustion and control of aviation turbines at General Electric, and prognostics and health management (PHM) of aircraft components at Raytheon Technologies-Collins Aerospace. Currently a Principal Engineer, he has over 13 years of experience in the commercial aviation industry.



Ben Burke received a B.S. in Communications from Brigham Young University - Idaho and a M.S. in Data Science and Business Analytics from the University of North Carolina at Charlotte. He has worked in the cloud, data and data science space as a consultant for almost a decade, helping clients create unique, native, and distinct solutions harnessing cloud architecture, data engineering and data science skills. Currently he is a Sr. Data Scientist with Insight, who provides consulting services for Collins Aerospace.