# 

# Student Project Proposal

|  |  |
| --- | --- |
| Project Title | Booz Allen – Aerospace Market – DSMS |
| Industry Sponsorship (if Any) | DC Metro / Eastern |

## Project Description

**Problem definition**

*[50-100 word description of the problem which you will solve]*

|  |
| --- |
| Turbofan engine is a pivotal component of the aircraft. Engine components are susceptible to degradation over the life of their operation which affects the reliability and performance of an engine. In order to direct the necessary maintenance behavior, remaining useful life prediction is the key.  The U.S. DoD’s Joint Artificial Intelligence Center ([JAIC](https://www.ai.mil/about.html)) has designated Predictive Maintenance as one of its two founding National Mission Initiatives (NMIs). By leveraging rich data sources and machine learning techniques, predictive maintenance has huge potential to drastically reduce costs, increase mission readiness and even save lives of service members. |

**Key Research Questions/ Technological constraints that the Project will Answer**

|  |
| --- |
| 1. What are the current state-of-the-art methods for predicting engine life that are published in the literature? (Benchmarking State of the Art) 2. Are you able to devise a better method by combining techniques from multiple papers or implementing your own technique? (ML Model building) 3. Once a model is selected, can this model be readily exposed to end users and services across an enterprise data platform? Can the model be easily updated if improvements are made? (ML Model deployment) |

**Final deliverables at the end of the project**

*[Please list the desired technical deliverables from the project team in as much detail as possible]*

|  |
| --- |
| 1. Scripts to ingest and clean data, train a model, evaluate the predictions, and facilitate hyperparameter tuning 2. A script or instructions to deploy the best performing model as an inference service that returns predictions for a selected set of test data |

**Key activities/ technologies the project team may be expected to undertake/ work with**

*[E.g. What kind of technology stack will you work with, the datasets you may need to work on, what kind of analysis you may be expected to undertake, etc.]*

|  |
| --- |
| Dataset:   * Primary: Turbofan Engine Degradation Dataset 1 – less data, more examples * Secondary: [Turbofan Engine Degradation Dataset 2](https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#turbofan) – more data, less examples   Models and Methods: Downloadable with the datasets above.  Pre-Processing: pandas, NumPy, SciPy  Model Training: scikit-learn, Keras/Tensorflow or PyTorch/fast.ai  ML App building and deployment on AWS or GCP |

**Expected learning outcomes**

*[What do you expect to learn from the project? Please mention the technical skills you will imbibe over the project.]*

|  |
| --- |
| 1. Applying variety of machine learning techniques to a real-world, domain specific, and high-impact problem 2. Developing repeatable and reproducible machine learning solutions using containerization technologies such as Docker 3. Employing best practices for tracking training and validation datasets, model architectures, parameters, and results |

|  |  |
| --- | --- |
| Team Size: | 1 |
| Member names: | Safia Babikir Bashir |

### 

## Tentative Time plan

Submit a tentative time plan (table/chart or text) regarding breakdown of the work that will be conducted between in the second half of your cohort, from week 6 onward.

## System Design

From the System design perspective, outline the following:

* Data
* Process (Models, iterations)
* Outcome (output and recommendations)

What are the system design considerations for your deployable ML model? Describe the iterations, delivery formats and limitations you may face and some solutions to overcome the limitations

* Should the model be deployed to run in batch, or to be hit from an api or some sort of streaming process as events are generated?
* What sort of infrastructure will be required for training? If it is a model that requires a lot of resources, where is the best place to train?

## Ethical Considerations

Are there any ethical considerations of your project? Consider the data source, the intended outcome, and/or the eventual use cases.

* Did you modify anything about your plan based on these considerations?
* Can you anticipate any issues that might arise during the process?

**Turbofan Engine Degradation Dataset**

Turbofan engine is a modern gas  turbine engine used by the NASA space exploration agency. NASA has created the following data set to predict the failures of Turbofan engines over time. The data set is available at [PCoE Datasets](https://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/).

The data set includes time series for each engine. All engines are of the same type, but each engine starts with different degrees of initial wear and variations in the manufacturing process, which is unknown to the user. There are three optional settings that can be used to change the performance of each machine. Each engine has 21 sensors collecting different measurements related to the engine state at runtime. Collected data is contaminated with sensor noise.