AER850: Intro to Machine Learning

Project #1 Due Date: October 20, 2024 23:59 PM

---

1 Background & Introduction

Augmented Reality (AR) based instruction modules through predictive machine learning (ML) algorithms has been a growing field of study, as there is scope of making the AR modules more efficient and accurate primarily for the manufacturing and maintenance sectors within the aerospace industry. A micro-level coordinate system approach is an effective way to utilize ML models to optimize the current AR-based animation techniques, incorporating such models with the existing technology allows technicians to gain better insights while performing their tasks which include: Aircraft maintenance, computer-aided design and modelling, part assembly and routine maintenance for space related applications.

2 Project Outline

This project allows for exploring various classification-based ML algorithms to successfully predict the maintenance step/stage given a specific part and its coordinates. The part utilized for this project is an inverter of the *FlightMax Fill Motion Simulator*. Figure 1 showcases the test subject.

Figure 1: Inverter of the FlightMax Fill Motion Simulator

Based on the test subject identified, 13 unique steps are defined within the process of disassembling the inverter. Each step has precise X, Y and Z axis points. The features for the ML model are the coordinates and the target variable is the step the coordinates belong to. The project is split into seven stages to design and develop ML models to predict the maintenance step based on the coordinates provided.

2.1 Step 1: Data Processing - 2 Marks

Read data from a csv file and convert that into a dataframe, which will allow for all further analysis and data manipulation.
Package(s) required: **Pandas**

2.2 Step 2: Data Visualization - 8 Marks

Perform statistical analysis on the dataset and visualize the dataset behaviour within each class. This will provide an initial understanding of the raw data behaviour. You are required to include the plots and explain the findings.
Package(s) required: **Matplotlib, Pandas, Numpy**

2.3 Step 3: Correlation Analysis - 15 Marks

Assess the correlation of the features with the target variable. A correlation study provides an understanding of how the features impact the target variable. A common correlation method used is *Pearson Correlation*. You are required to include the correlation plot, and explain the correlation between the features and the target variables, along with the impact it could have on your predictions.
Package(s) required: **Seaborn, Pandas**

2.4 Step 4: Classification Model Development/Engineering - 20 Marks

Prepare the data to create three classification models (based on ML algorithms). The dataset needs to be split into training and testing categories to develop the models. For each ML model, utilize grid search cross-validation to assess the hyperparameters that give you the best results. You are required to explain your selected choice of classification algorithms. In addition to the three classification models with grid search cross-validation, you must make one model based on using *RandomizedSearchCV*. this will provide another method of

determining the best hyperparameters to optimize your results.
Package(s) required: **scikit-learn** - All ML models can be accessed from the sklearn package

2.5 Step 5: Model Performance Analysis - 20 Marks

Compare the overall performance of each model based on f1 score, precision and accuracy. You are required to provide an explanation to what these metrics mean and which metric to prioritize for this use-case. Based on the selected model, create a confusion matrix to visualize the performance of your model. Include the confusion matrix in the report as well.
Package(s) required: **scikit-learn (Metrics, Confusion Matrix)**

2.6 Step 6: Stacked Model Performance Analysis - 5 Marks

Using scikit-learn's *StackingClassifier*, combine two of the previously trained models to analyze the impact of model stacking on overall performance. Evaluate the performance of the stacked model based on their f1 score, precision and accuracy along with a confusion matrix to provide a clear visual representation. Include this confusion matrix in the report for detailed performance analysis. If a significant increase in accuracy is observed, discuss how combining complementary strengths of the models contributed to the improvement. Conversely, if the change is minimal, explain why you think the stacking models had limited effectiveness.
Package(s) required: **scikit-learn**

2.7 Step 7: Model Evaluation - 10 Marks

Package the selected model and save it in a *joblib* format, this allows you to call the model to predict the class based on random set of coordinates given.
Package(s) required: **Joblib**
Based on the data set provide, you are required to predict the corresponding maintenance step:

[9.375,3.0625,1.51], [6.995,5.125,0.3875], [0,3.0625,1.93], [9.4,3,1.8], [9.4,3,1.3]

3 Submission - 20 Marks

The major goal of the project is to understand the process of developing ML models and evaluating them.

3.1 GitHub Submission

You are required to create a new repository on GitHub using the template provided and regularly push changes to the repository as you work on the project. Please read the description in the template repository for instructions on how to clone and push changes.

GitHub Template Link

3.2 Report Submission

Please submit a report discussing your observations on the results you found in Steps 2-6.

By the end of this project, you will have a comprehensive understanding of the ML model pipeline that is required for any data-driven challenge.