**Predictive Model for Asset Lifecycle Management: A Technical Overview**

**1. Introduction**

This document provides a technical explanation of the predictive model developed to forecast the lifecycle of assets based on data from the Maximo Asset Management system. The primary objective of this model is to estimate the "Life Expectancy" of an asset, thereby enabling proactive maintenance, strategic procurement, and optimized capital planning.

The solution is implemented through a two-stage process:

1. **Data Preparation (prepare\_asset\_lifecycle.py):** Cleaning and feature engineering of raw asset data.
2. **Model Training (train\_lifecycle\_model.py):** Development and evaluation of a machine learning model to predict asset life expectancy.

The model leverages a RandomForestRegressor algorithm, a powerful ensemble learning method known for its high accuracy and robustness. This document details the model's architecture, its performance metrics, and the interpretation of its outputs.

**2. Model Architecture: Random Forest Regressor**

The core of our predictive solution is a **Random Forest Regressor**. This model was selected for its ability to handle complex, non-linear relationships between asset features and their expected lifespan.

A Random Forest operates by constructing a multitude of decision trees during training. For a regression task, each individual tree predicts a value for the target variable (in our case, "Life Expectancy"). The final prediction of the Random Forest is the average (or mean) of the predictions from all the individual trees in the forest. This ensemble approach mitigates the risk of overfitting that a single decision tree might have, leading to more accurate and stable predictions.

The features used to train the model include:

* Asset Age (Years)
* Asset Type
* Manufacturer
* Status
* Location Description

**3. Model Performance Evaluation**

The model's performance was rigorously assessed using a standard set of regression metrics. The data was split into a training set (80%) and a testing set (20%) to ensure that the evaluation reflects the model's ability to generalize to new, unseen data.

Imagen que contiene Texto

El contenido generado por IA puede ser incorrecto.

* **Mean Absolute Error (MAE): 0.80 years**
  + **What it is:** MAE measures the average magnitude of the errors in a set of predictions, without considering their direction. It represents the average absolute difference between the predicted life expectancy and the actual life expectancy.
  + **Mathematical Justification:** It is calculated as the sum of the absolute differences between the predicted values (

) and the actual values (

), divided by the number of samples (

).

* + **Interpretation:** On average, the model's prediction for an asset's life expectancy is off by approximately **0.80 years**. This indicates a high level of accuracy for long-term planning.
* **Root Mean Squared Error (RMSE): 2.38 years**
  + **What it is:** RMSE is the square root of the average of squared differences between prediction and actual observation.
  + **Mathematical Justification:** It is calculated by taking the square root of the mean of the squared prediction errors.
  + **Interpretation:** The RMSE gives a relatively high weight to large errors, meaning it is more sensitive to outliers. The value of **2.38 years** suggests that while most predictions are close to the actual value (as shown by the MAE), there are some instances where the model's error is notably larger.
* **R² Score (Coefficient of Determination): 0.879**
  + **What it is:** The R² score measures the proportion of the variance in the dependent variable (Life Expectancy) that is predictable from the independent variables (the features).
  + **Mathematical Justification:** It represents the ratio of the explained variance to the total variance.

Where

 is the mean of the actual values.

* + **Interpretation:** An R² score of **0.879** is excellent. It signifies that **87.9%** of the variability in asset life expectancy can be explained by our model's features. A value this close to 1 indicates a very strong fit and high predictive power.

A graph with blue and white bars

AI-generated content may be incorrect.

* **What it is:** This chart identifies which features have the most significant impact on the model's predictions. In a Random Forest, feature importance is typically calculated by measuring how much the model's accuracy decreases when a particular feature's values are randomly shuffled.
* **Interpretation:**
  + **Asset Age (Years)** is overwhelmingly the most influential predictor. This aligns with intuition, as the current age of an asset is a primary determinant of its remaining life.
  + **Manufacturer** (e.g., SQUARE\_D, FLUKE, DEKA) plays a critical role, indicating that brand and build quality are strong indicators of an asset's longevity.
  + **Asset Type** (e.g., SECURITY) is also a significant factor, confirming that different categories of assets have inherently different lifecycles.
  + This analysis helps validate the model's logic and highlights the key data points that drive asset lifecycle.

A graph with a blue line

AI-generated content may be incorrect.**What it is:** A residual is the difference between the actual value and the predicted value for a single data point (Residual = Actual Value - Predicted Value). This plot is a histogram of those residuals, showing the frequency of different error sizes.

* **Interpretation:**
  + The distribution is sharply peaked at zero and largely symmetrical. This is the ideal shape for a distribution of residuals, indicating that the model's errors are centered around zero.
  + This means that the model has no systematic bias (i.e., it doesn't consistently over- or under-predict).
  + The vast majority of prediction errors are very small (close to zero), which reinforces the low MAE value.
  + The narrowness of the peak demonstrates the model's high precision. The thin "tails" on either side represent the few outlier predictions that contribute to the RMSE value.

**4. Conclusion**

The developed Random Forest model demonstrates high accuracy and strong predictive power for estimating asset life expectancy. The performance metrics (MAE of 0.80 years and R² of 0.879) confirm its reliability for business applications. The analysis of feature importances and residuals further validates that the model behaves as expected, learning logical patterns from the data. This tool provides a robust, data-driven foundation for optimizing asset management strategies.