**Milestone 2: Model Development & Training**

**Project Name:** PrognosAI: AI-Driven Predictive Maintenance System Using Time-Series Sensor Data

**Dataset:** NASA Turbofan Jet Engine Data Set

**Prepared by:** Durga Veera Prasad V

**Objective:**

Develop and train an LSTM model to predict Remaining Useful Life (RUL) from sensor sequences, using 5-fold cross-validation to ensure generalization and robustness.

---

**Modules Used and Purpose:**

**numpy:** Array handling and numerical operations.

**tensorflow / keras:** Build and train the LSTM model.

- `Sequential`, `LSTM`, `Dense`, `Dropout`: Define model layers.

- `EarlyStopping`, `ReduceLROnPlateau`: Callbacks for training control.

- `Adam`: Optimizer for model training.

**sklearn.preprocessing.MinMaxScaler:** Scale features and target for stable training.

**sklearn.model\_selection.KFold:** Perform 5-fold cross-validation.

**sklearn.metrics:** Evaluate performance (`r2\_score`, `mean\_squared\_error`, `mean\_absolute\_error`).

**matplotlib.pyplot:** Plot training/validation loss curves.

**joblib:** Save scalers for future use with test data or deployment.

---

**Steps Implemented:**

**1. Data Preparation:**

- Generated synthetic CMAPSS-style sequences (`X`) and RUL targets (`y\_raw`) for demonstration.

- Scaled features (`X\_scaled`) and targets (`y\_scaled`) using `MinMaxScaler`.

**2. LSTM Model Definition:**

- Built a sequential LSTM model with two dense layers and dropout for regularization.

- Compiled using `Adam` optimizer and mean squared error loss.

**3. 5-Fold Cross-Validation:**

- Split data into 5 folds to train and validate the model.

- Monitored training using `EarlyStopping` and `ReduceLROnPlateau`.

- Calculated metrics per fold: Train/Test R², RMSE, MAE.

- Plotted training and validation loss curves.

**4. Final Model Training & Saving:**

- Trained the LSTM model on the full dataset.

- Saved the trained model (`.keras`) and scalers (`.pkl`) for inference.

---

**Deliverables:**

- Trained LSTM model and saved weights.

- Scalers for feature and target normalization.

- Training and validation loss curves.

- Performance metrics (R², RMSE, MAE) from cross-validation.

**Evaluation:**

- Achieved high R² (>95% target on training data for demonstration).

- Verified model convergence via loss curves.

- Ensured cross-validation confirms generalization.