**Module 5 Report – Assignment Summary**

**Contributors:**

1. AMANDEEP SINGH -2023AA05162 - 100%
2. SABYASACHI BHUYAN - 2023AB05002 - 100%
3. RUPA CHITTA - 2023AA05103 - 100%
4. SHRIDHAR NARAYAN BAIRI – 2023AA05652 – 100%

**Summary :** This project implements an end-to-end **MLOps pipeline** using the **Fashion MNIST** dataset. The pipeline covers five critical stages:

* Automated EDA using **Pandas Profiling** to generate reports with class distribution, missing values, and feature correlations.
* Preprocessing involved normalizing pixel values and flattening images.
  + Used **SHAP** to generate explainability visualizations and refine the feature engineering pipeline based on feature importance.
* Fine-tuned hyperparameters using **Optuna** to optimize model performance.
* Tracked model performance with **MLflow**, logging parameters, metrics, and artifacts.
* Implemented drift detection using **EvidentlyAI** to monitor changes in data distribution over time.

**Justification for tools:**

• **Pandas Profiling**:

Automated and detailed EDA with comprehensive visual reports.

Identifies data anomalies, outliers, and feature correlations efficiently.

• **SHAP** (SHapley Additive exPlanations):

Provides interpretable feature importance visualizations.

Ensures transparency by explaining how individual features influence predictions.

• **Auto-sklearn**:

Automates model selection, reducing manual effort.

Ensures optimal model selection by comparing multiple ML models.

• **Optuna**:

Fast and efficient hyperparameter tuning with advanced search algorithms.

Helps optimize the model’s performance by identifying the best hyperparameters.

• **MLflow**:

Centralized logging and model versioning.

Tracks performance metrics, ensuring reproducibility.

• **EvidentlyAI**:

Detects data drift to trigger retraining.

**Github link for the project:**

<https://github.com/AmandeepSingh96/bits-mlops-assignment1-group60/tree/master>