AutoAI Model Building in IBM Watson Studio

Assignment Description

In this assignment, I used IBM Watson Studio's AutoAI tool to build a machine learning model for predicting diabetes based on health-related features from the PIMA Indian Diabetes dataset. The AutoAI pipeline automatically handled data preprocessing, feature engineering, model selection, and hyperparameter tuning. The goal of the project was to evaluate different models and identify the best-performing one for binary classification (diabetic or not). IBM AutoAI selected the Random Forest Classifier as the best model based on accuracy and F1-score.

Challenges & Solutions

- Challenge: Initially, I faced difficulty in uploading the dataset due to format issues.
Solution: I preprocessed the dataset using Excel and ensured it was in `.csv` format with clean headers.

- Challenge: Understanding the AutoAI pipeline steps (feature transformation, model generation, etc.) was a bit overwhelming at first.
Solution: I referred to IBM documentation and tutorials to clearly understand each step in the AutoAI workflow.

- Challenge: Model accuracy was fluctuating with minor changes in data.
Solution: I ensured data balancing and removed outliers before feeding into AutoAI, which improved consistency.