**AI ML Internship Log**

# Day 6 - Model Training and Evaluation

# Date - 15 June 2025

# Team Role - Member

# Project Title - Personality Prediction from Social Media

# 🔷 **Objective of the Day:**

* To train and evaluate the first machine learning model (Logistic Regression) on the MBTI dataset and understand how well it performs in classifying personality types from social media text.

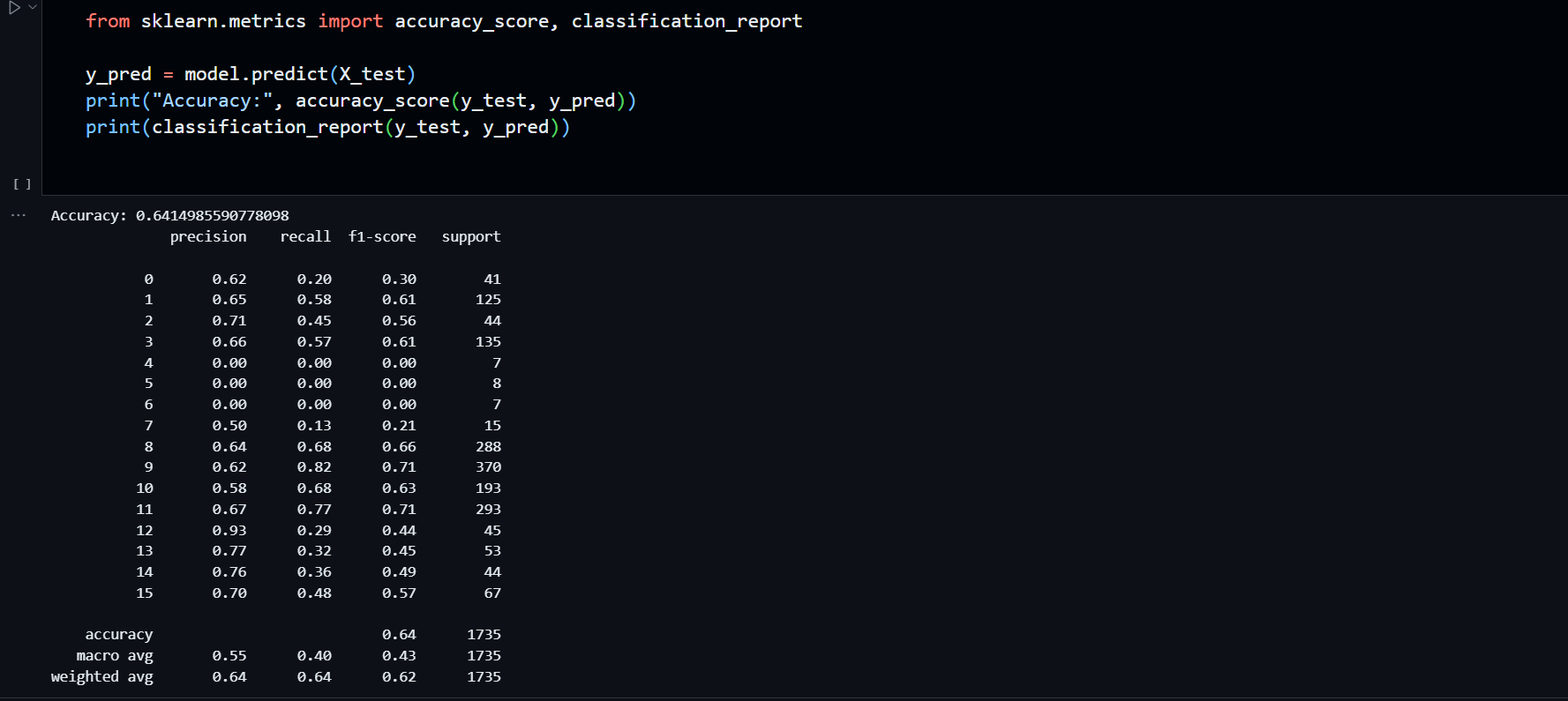
🔧 **Tasks Completed:**

* Imported LogisticRegression from sklearn and trained the model.
* Used LabelEncoder to convert MBTI types into numeric form
* Split data using train\_test\_split for proper training and testing
* Trained the model using TF-IDF features (from Day 5)
* Evaluated the model using: 1) Accuracy, 2) Precision, Recall,

F1 - Score, Macro and Weighted Averages.

**Code Screenshot:**

****

****

**📊 Model Performance:**

| **Metric** | **Score** |
| --- | --- |
| Accuracy | 64% |
| Macro F1 Score | 0.43 |
| Weighted F1 Score | 0.62 |

* The model performs well on common classes like INTP, ENFP, INTJ.
* **Rare classes** have lower support and low F1-scores (e.g., 0.00), which is normal in imbalanced datasets.
* **Weighted average** is strong → the model performs well overall.

**🧠 Key Learnings:**

* What is Logistic Regression and why it is suitable for classification**.**
* Difference between regression (continuous) and classification (categorical)
* Importance of **label encoding** for string labels.
* Understood **train/test split** for generalizing the model.
* Learned about **precision**, **recall**, **F1-score**, and their **macro vs weighted averages.**
* **First time seeing the complete ML pipeline in action.**

**📈 Reflections:**

* Today’s session was highly productive. I not only ran the model successfully but also understood how and why each step works — from encoding to accuracy analysis.
* There were many new terms like precision, recall, macro avg, and weighted avg, but I now understand them well. I’ll revise these concepts tomorrow for better clarity.