**AI ML Internship Log**

# Day 11 -Tackling Class Imbalance

# Date - 01 July 2025

# Team Role - Member

# Project Title - Personality Prediction from Social Media

### **✅ What I Did Today:**

* Continued project refinement by analyzing **dataset imbalance** using label frequency counts.
* Found a huge class imbalance (e.g., INFP = 1832 vs. ESTJ = 39).
* Implemented class\_weight='balanced' for all 3 models:
* Re-ran training and evaluation using advanced preprocessing from previous days.

📚 **What I Learned:**

* Imbalanced data can cause ML models to **favor common classes**, ignoring rare ones.
* class\_weight='balanced' helps models give equal attention to all classes by **automatically assigning weights inversely proportional to class frequency**.
* **Macro F1-score** is the best metric to monitor in such scenarios (treats all classes equally).
* Even **0.01–0.02 changes in macro F1 or accuracy are meaningful in NLP** projects.

📊 **Result Summary:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Accuracy** | **Macro F1** | **Weighted F1** |
| Logistic | 0.65 | 0.52 | 0.65 |
| Linear SVM | 0.6028 | 0.47 | 0.61 |
| SVM RBF | 0.65 | 0.52 | 0.64 |

* **Best Model:** Logistic Regression – balanced accuracy + best macro/weighted F1.
* **Improved minority class performance** with better fairness across all 16 MBTI types.

🧠 **Conclusion:**

* Addressing class imbalance class\_weight='balanced' was **crucial**.
* Helped boost **fairness macro F1 from 0.46 to 0.52 ie. 6%** , not just raw accuracy.
* Results show **models are learning to handle rare personality types** better.

**🤔 Self-Reflection:**

* Feeling neutral but proud — and that’s valid.
* NLP with real-world text is **messy, noisy, and rarely perfect**.
* Even small improvements mean your model is working harder and smarter.
* Even small improvements mean your model is working harder and smarter.

🔁 Will keep revising key concepts like F1-scores, class weighting, and evaluation techniques to stay confident for future interviews.