**AI ML Internship Log**

# Day 1 - Project Understanding & Clarity

# Date - 09 June 2025

# Team Role - Member

# Project Title - Personality Prediction from Social Media

## **1. What is the Project About?**

* The project focuses on predicting a person's personality based on their interactions on various social media platforms.
* The prediction is done by analyzing various texts, posts, and interactions done on social media.
* It is a real-world **Natural Language Processing (NLP)** problem involving text classification.
* The project will work on text data such as tweets, facebook posts, Instagram captions, or Reddit content.
* The aim is to classify users into personality types based on what they write.

**2. Why Personality Prediction from Social Media?**

**Social Media is a Mirror of Personality**

* In today’s world, digital expressions often outweigh real-life expressions. People tend to express themselves more through tweets, posts, captions on social media.
* These expressions reflect their **real-life behaviour** and their **psychological traits.**
* Instead of taking long personality tests, we can analyze their natural writing to understand them better.
* **Highly Useful in Real World Applications**
* **For Example:** Identifying early signs of depression, anxiety, or stress through social media.

**3. How will Personality Prediction Usually Be Done?**

**(General Workflow)**

**Step 1: Data Collection**

* Collect social media texts with personality traits (eg. tweets, posts etc)

**Step 2: Data Preprocessing**

* Clean the text: Remove URLs, emojis, mentions.
* Convert to lowercase, remove stop words etc.

**Step 3: Feature Extraction**

* Convert text to Numbers using techniques like:

Bag of Word

TF-IDF

Word Embeddings (Word2Vec, BERT)

**Step 4: Model Training**

* Train ML Models
* E.g., Logistic Regression, SVM, Random Forest on labeled data.
* Understanding the relationship between writing style and personality traits.

**Step 5: Prediction & Evaluation**

* Test models on new data and test its accuracy etc.

**4. What Kind of Data will be Used?**

* Each data point includes:

**Text:** Social Media content (posts, tweets etc).

**Label:** Personality traits like Introvert, Extrovert,

High Openness, etc.

* **Data Format Example:**

|  |  |
| --- | --- |
| **Text** | **Personality** |
| I love trying new experiences. | High Openness |

**5. What are the challenges in this project?**

* **Human Personality** is complex to analyze by just looking at social media texts.
* **Social media content is noisy -** full of slang, emojis, abbreviations, sarcasm etc.
* People sometimes don’t express their true personality online as some posts are edited, exaggerated or made for attention.
* **Strong NLP skills are required** as it involves data cleaning, feature extraction etc.

**6 . My Challenges in understanding of project**

* I was initially confused that each text must be linked with a personality label in the data collection step.

Then I got to know the meaning of this statement that to train a model we provide text with also the correct personality trait it reflects.

This helps the model to develop or find a relation between input(text) and output(personality).

**For Example:**

|  |  |
| --- | --- |
| **Text** | **Personality Label** |
| I love trying to be adventure | Extrovert |
| I prefer quiet time and alone | Introvert |
| I am grateful to God | High agreeableness |

I learned that NLP-based personality prediction is a supervised learning task, where models are trained using labeled data — known text paired with the personality traits it represents.

**The model learns:**

If someone writes like this, their personality is like that.

Then it uses that learning to predict personality on new text ie. unlabeled data.

**7. Conclusion**

* Today, I developed a clear understanding of my project “***Personality Prediction from Social Media.***” I learned that this task falls under **Natural Language Processing (NLP)** and is a **supervised learning problem**, where a labeled dataset is used to train a model to find patterns between input (text) and output (personality). I also recognized the challenges involved in NLP, such as handling noisy data and capturing subtle human traits. Overall, I now feel more confident and prepared to dive deeper into the technical aspects of the project.

This foundational understanding will help me move confidently into coding and data analysis in the next phase of the project.

**8. References**

* **ChatGPT Guidance**

<https://chatgpt.com/share/6846b9f9-203c-8007-b527-dd4ab5145450>

* **Personality Prediction Model for Social Media Using Machine Learning Techniques.**

[**https://www.sciencedirect.com/science/article/abs/pii/S0045790622001446#:~:text=Statistical%20information%20about%20the%20human,on%20big%20five%20personality%20traits**](https://www.sciencedirect.com/science/article/abs/pii/S0045790622001446#:~:text=Statistical%20information%20about%20the%20human,on%20big%20five%20personality%20traits)**.**

## 

## 