

# Báo cáo Tổng kết Đề án Môn học: Web Mining Big Five Project

## 1. Giới thiệu Đề tài

Trong kỷ nguyên mạng xã hội, dấu chân kỹ thuật số (digital footprints) chứa đựng lượng thông tin khổng lồ về người dùng. Dự án "**Detection of Big Five Personality Traits from Social Media**" tập trung vào việc khai thác dữ liệu văn bản từ mạng xã hội (cụ thể là dataset PAN15) để dự đoán 5 nét tính cách cốt lõi (Mô hình OCEAN):

- \*\*O\*\*penness (Cởi mở)
- \*\*C\*\*onscientiousness (Tận tâm)
- \*\*E\*\*xtraversion (Hướng ngoại)
- \*\*A\*\*greeableness (Dễ chịu)
- \*\*S\*\*tability (Ôn định cảm xúc - Emotional Stability)

Dự án không chỉ dừng lại ở bài toán phân loại, mà còn mở rộng sang các ứng dụng thực tế như **Hệ thống gợi ý (Recommender System)** và **Giải thích dự đoán (Explainable AI)**.

## 2. Kiến trúc Hệ thống

Hệ thống được thiết kế theo mô hình Modular, đảm bảo tính mở rộng và dễ dàng bảo trì.

### Sơ đồ luồng dữ liệu (Data Flow):

Data Collection (PAN15) → Preprocessing (Cleaning/Splitting) → Feature Extraction (TF-IDF/Embeddings) → Model Training → Application (Streamlit)

### Các Module chính:

1. **Data Module:** Xử lý dữ liệu thô XML, trích xuất tweet và gán nhãn ground truth.
2. **Model Module:**
  - \*\*Baseline\*\*: Sử dụng TF-IDF kết hợp Ridge Regression (Multi-output). Đây là mô hình chính được sử dụng cho Demo vì tốc độ nhanh và hiệu quả tốt trên dữ liệu văn bản ngắn.
3. **Opinion Mining Module:** Tích hợp các mô hình Transformer (RoBERTa) để trích xuất cảm xúc (Sentiment/Emotion) làm đặc trưng bổ sung, giúp model hiểu sâu hơn về thái độ của người viết.

4. **Information Retrieval (IR) Module:** Sử dụng thuật toán BM25 để tìm kiếm các bài đăng đóng vai trò là "bằng chứng" (evidence) cho từng nhận định về tính cách.

5. **Recommender System (RecSys):** Gợi ý Hashtag dựa trên profile tính cách dự đoán.

### 3. Chức năng & Demo Ứng dụng

Ứng dụng Demo được xây dựng trên nền tảng **Streamlit**, cung cấp giao diện trực quan để người dùng tương tác với hệ thống.

#### 3.1. Phân tích Tính cách (Personality Analysis)

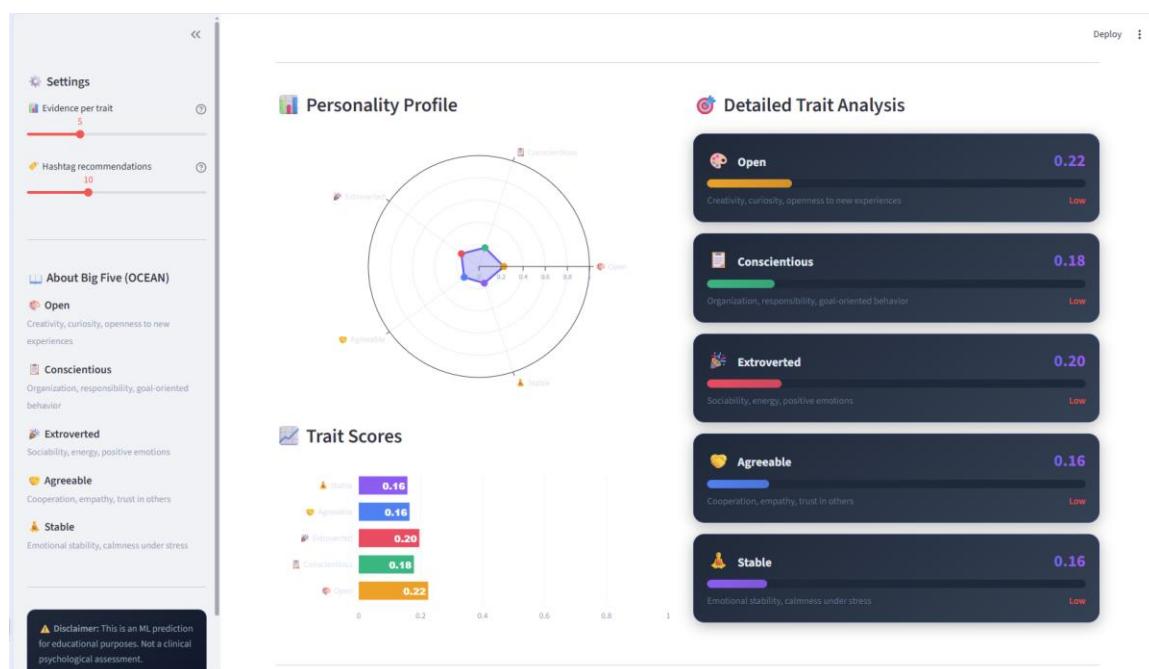
Người dùng nhập văn bản (các bài đăng trên mạng xã hội) hoặc upload file. Hệ thống sẽ:

- Tiền xử lý văn bản (xóa URL, chuẩn hóa).
- Dự đoán điểm số cho 5 nét tính cách (thang 0.0 - 1.0).
- Hiển thị kết quả dưới dạng biểu đồ Radar và biểu đồ cột trực quan.

#### Giao diện hiển thị:

- \*\*Radar Chart ("Your Profile")\*\*: Cho cái nhìn tổng quan về "hình dáng" tính cách.
- \*\*Trait Cards\*\*: Hiển thị chi tiết từng chỉ số kèm nhãn (High/Moderate/Low).

\*\*Hình 1\*\*: Giao diện chính sau khi phân tích, hiển thị biểu đồ Radar và điểm số tính cách.



### 3.2. Truy xuất Bằng chứng (Evidence Retrieval)

Để tăng tính minh bạch (Explainability), hệ thống trích xuất lại các bài đăng của người dùng có đóng góp lớn nhất vào dự đoán.

- Ví dụ: Nếu người dùng được dự đoán là "High Extroversion", hệ thống sẽ chỉ ra các tweet chứa từ khóa liên quan đến tiệc tùng, giao tiếp xã hội mà họ đã viết.
- Công nghệ: Sử dụng \*\*BM25\*\* với bộ query đặc thù cho từng nét tính cách.

\*\*Hình 2\*\*: Phản hiển thị "Evidence by Trait", minh chứng cho các dự đoán.

The screenshot displays two main sections: "Evidence by Trait" on the left and "Recommended Hashtags" on the right.

**Evidence by Trait:** Posts most relevant to each personality trait:

- OPEN (score: 0.22):**
  - "I love exploring new art galleries and finding inspiration in abstract paintings 🎨"
  - Relevance: 0.000
- CONSCIENTIOUS (score: 0.18):**
  - "Just started learning a new language, the culture is so fascinating!"
  - Relevance: 0.000
  - "Thinking about the universe and our place in it keeps me up at night 🌌"
  - Relevance: 0.000

**Recommended Hashtags:** Personalized hashtag suggestions based on your personality:

- #creativity #travel #love #in #art #me
- #searchsolutions2014 #nowplaying #teiputsch #enjoywe

**Recommendation Details:**

| Hashtag              | Score |
|----------------------|-------|
| #creativity          | 2.637 |
| #travel              | 2.569 |
| #love                | 2.513 |
| #in                  | 2.436 |
| #art                 | 2.620 |
| #me                  | 2.407 |
| #searchsolutions2014 | 0.447 |
| #nowplaying          | 0.497 |
| #teiputsch           | 0.418 |
| #enjoywe             | 0.489 |

### 3.3. Gợi ý Hashtag (Hashtag Recommendation)

Dựa trên profile tính cách vừa phân tích, hệ thống gợi ý các hashtag phù hợp mà người dùng có thể quan tâm.

- Logic: Kết hợp giữa sự tương đồng nội dung (Content-based) và sự tương thích tính cách ("Personality-aware").
- Ví dụ: Người có tính "Openness" cao có thể được gợi ý các hashtag về #art, #travel, #creativity.

\*\*Hình 3\*\*: Danh sách Hashtag được gợi ý dựa trên tính cách.

The screenshot displays a user interface for analyzing personality traits. On the left, under 'Evidence by Trait', there are two sections: 'OPEN' (score: 0.22) and 'CONSCIENTIOUS' (score: 0.18). Each section contains three dark blue cards with text snippets and relevance scores. On the right, under 'Recommended Hashtags', personalized suggestions are shown in purple rounded rectangles, followed by a table titled 'Recommendation Details' with a red border.

| Hashtag              | Score |
|----------------------|-------|
| #creativity          | 2.637 |
| #travel              | 2.569 |
| #love                | 2.513 |
| #in                  | 2.436 |
| #art                 | 2.620 |
| #me                  | 2.407 |
| #searchsolutions2014 | 0.447 |
| #nowplaying          | 0.497 |
| #teiputsch           | 0.418 |
| #enjoywe             | 0.489 |

### 3.4. Giao diện Người dùng (UI/UX)

- Thiết kế hiện đại với Dark Mode, màu gradient và animations.
- Bố cục chia tab rõ ràng: "Analyze Text", "Upload File", "How It Works".
- Các chỉ số thống kê nhanh: Số lượng bài viết, Tổng số từ, Tính cách nổi trội nhất.

## 4. Công nghệ Sử dụng

- Ngôn ngữ**: Python 3.10+
- Machine Learning**: Scikit-learn (Ridge, TF-IDF), PyTorch, Transformers (HuggingFace).
- Web Framework**: Streamlit.
- Visualization**: Plotly Interactive Charts.
- Data Processing**: Pandas, Numpy.
- IR/Search**: Rank-BM25.

## 5. Kết luận

Dự án đã xây dựng thành công một pipeline trọn vẹn từ xử lý dữ liệu đến ứng dụng cuối cùng. Hệ thống không chỉ đạt được mục tiêu dự đoán tính cách mà còn cung cấp giá trị gia tăng thông qua việc giải thích kết quả và ứng dụng vào bài toán gợi ý nội dung. Đây

là một minh chứng rõ ràng cho tiềm năng ứng dụng của Web Mining và NLP trong việc thấu hiểu người dùng mạng xã hội.

**Note:**

1. Ở phần Experiments hiện tại tôi thấy 4.1 là Experimental Setup thì thì theo tôi 4.2 mình sẽ đưa các hình ảnh về training curve vào đây rồi 4.3 mới là Personality Prediction Results ...
2. Ngoài ra thì các logs cũng sẽ phải được ghi trong nội dung báo cáo (Có thể đưa logs ra 1 dạng file csv rồi upload lên Driver xong ghim link vào báo cáo chẳng hạn) => Để chứng minh training curve không phải mình bịa
3. Trong phần Prediction thì tôi nghĩ nên thêm cả ảnh mà mình chạy sản phẩm lên, có input đầu vào là chuỗi twitt của người dùng và dùng mô hình để phân tích và cho ra được output có định dạng như mình đã mô tả trong phần checkpoint gồm có: Score của từng loại tính cách, có k tweet truy hồi để chứng minh cho dự đoán, có hashtag dự đoán , ...
4. Do là nhóm mình làm sản phẩm nên tôi nghĩ cũng nên thêm 1 phần Tech Stack (Như nội dung phần 4 trong doc mà ông list ra)