

AI Team Builder: Dự đoán MBTI từ Văn Bản bằng ML truyền thống và BERT

September 23, 2025

1. Abstract

Vấn đề. Bài toán cần dự đoán kiểu tính cách MBTI (16 loại, biểu diễn qua bốn trục IE, NS, TF, JP) trực tiếp từ văn bản người dùng viết, và chuyển hoá dự đoán thành gợi ý *vai trò phù hợp trong nhóm* để hỗ trợ hình thành đội nhóm hiệu quả.

Dữ liệu. Sử dụng bộ dữ liệu `mbti-type` (các tệp `data/mbti_1.csv`, `data/mbti_clean.csv`) gồm **8 675** mẫu tiếng Anh, mỗi mẫu là một người dùng với trường `posts` (tập hợp nhiều câu/đoạn) và nhãn MBTI. Bộ `mbti_clean.csv` bổ sung bốn cột nhãn nhị phân `mbti_IE`, `mbti_NS`, `mbti_TF`, `mbti_JP`. Tập kiểm thử điển hình có **1 735** mẫu (`data/test.csv`).

Cách tiếp cận. (i) Tuyến **ML truyền thống**: TF-IDF + bộ phân loại tuyến tính độc lập cho từng trục (mô hình được gói trong `scr_ml/models.py` và sử dụng bởi `demo/UI_ML/app_ml.py`). (ii) Tuyến **BERT**: tinh chỉnh `bert-base-uncased` với đầu ra 4 nhánh nhị phân, pooling kết hợp [CLS]+mean, dropout 0.4 và một tầng ẩn (`src/models.py`, suy luận trong `demo/backend/app.py`).

Kết quả. Trên tập kiểm thử 1735 mẫu: tuyến ML đạt Acc/F1 tốt trên từng trục (ví dụ: **IE** Acc 0.84, F1 0.78; **NS** Acc 0.89, F1 0.79; **TF** Acc 0.84, F1 0.84; **JP** Acc 0.81, F1 0.80). Tuyến BERT đạt kết quả cạnh tranh (ví dụ: **IE** Acc 0.764, Macro-F1 0.699; **NS** Acc 0.861, Macro-F1 0.712; **TF** Acc 0.749, Macro-F1 0.749). Hệ thống được đóng gói thành hai giao diện: Flask+HTML (BERT) và Streamlit (ML).

2. Introduction

Bối cảnh. Trong thực tế, việc ghép đội nhóm (teamwork), định hướng nghề nghiệp (career), hay gợi ý tương tác xã hội đều cần tín hiệu về phong cách làm việc/ưu tiên cá nhân.

Vai trò của MBTI. MBTI mô hình hoá sở thích nhận thức/hành vi qua 4 trục (IE, NS, TF, JP), giúp liên hệ thói quen làm việc và vai trò phù hợp trong nhóm (ví dụ ENTJ thiên về lãnh đạo, INTP mạnh về phân tích).

Mục tiêu. Xây dựng một *pipeline* từ văn bản → dự đoán 4 trục MBTI → ghép 16 kiểu MBTI → gợi ý vai trò team (“AI Team Builder”).

Điểm khác biệt. (i) Kết hợp **hai tuyến** (ML truyền thống và BERT) trên *cùng* dữ liệu, (ii) chuyển hoá đầu ra thành **vai trò trong nhóm** dùng bảng ánh xạ curated, (iii) cung cấp **demo thực thi** qua Flask/Streamlit để ứng dụng ngay.

3. Related Works

Ứng dụng MBTI phổ biến. MBTI đã được sử dụng rộng rãi trong nhiều lĩnh vực đời sống. Trong *tư vấn nghề nghiệp*, MBTI giúp cá nhân xác định điểm mạnh, phong cách làm việc để chọn ngành nghề phù hợp. Trong *xây dựng và quản trị đội nhóm*, MBTI hỗ trợ phân công vai trò (người lãnh đạo, người khởi xướng ý tưởng, người hỗ trợ chăm sóc, ...) nhằm tăng hiệu quả hợp tác. Trong *quan hệ cá nhân và hẹn hò*, nhiều ứng dụng sử dụng MBTI để gợi ý đối tượng phù hợp dựa trên tính cách bổ sung hoặc đồng điệu. Ngoài ra, MBTI còn được khai thác trong *giáo dục* (phân tích phong cách học tập) và *phát triển cá nhân* (self-coaching, kỹ năng mềm).

Cách dự đoán MBTI thường gặp. Truyền thống nhất là qua *trắc nghiệm tự báo cáo*, nơi người dùng trả lời một danh sách câu hỏi tình huống (ví dụ: thích làm việc một mình hay trong nhóm, ra quyết định dựa trên lý trí hay cảm xúc). Dựa vào tổng hợp các lựa chọn, hệ thống ánh xạ sang 4 cặp đối cực (I/E, N/S, T/F, J/P) và suy ra kiểu MBTI. Ngoài ra, một số nghiên cứu và ứng dụng hiện đại khai thác *ngôn ngữ tự nhiên* (bài đăng, bình luận, nhật ký, hồ sơ cá nhân) để suy luận MBTI, với giả định rằng cách dùng từ và cách diễn đạt phản ánh đặc điểm tính cách.

Điểm tương đồng/khác biệt. Đề tài này cũng dựa trên ngôn ngữ làm tín hiệu chính, giống với hướng nghiên cứu gần đây. Tuy nhiên, điểm khác biệt quan trọng là: (i) triển khai **song song** hai hướng mô hình — ML truyền thống (TF-IDF + classifier) và BERT fine-tuning — để so sánh hiệu quả; (ii) không chỉ dừng lại ở phân loại MBTI, mà còn **chuyển hoá đầu ra thành gợi ý vai trò trong nhóm** thông qua một bảng ánh xạ được thiết kế, nhằm phục vụ trực tiếp cho bài toán xây dựng đội nhóm (AI Team Builder).

4. Methods

Phân rã bài toán.

1. Tiền xử lý văn bản (`mbti_clean.csv` cung cấp nhãn nhị phân cho 4 trục).
2. Huấn luyện tuyến ML & tuyến BERT.
3. Suy luận: dự đoán 4 trục → ghép thành 16 MBTI.
4. **AI Team Builder**: ánh xạ MBTI → vai trò trong nhóm.

Mô hình.

- **ML**: TF-IDF → bộ phân loại tuyến tính theo từng trục (độc lập IE/NS/TF/JP). Suy luận được hiện thực trong `demo/UI_ML/app_ml.py`.
- **BERT**: tinh chỉnh `bert-base-uncased`, pooling [CLS]+mean, dropout 0.4, có tầng ẩn; huấn luyện trong `notebooks/BERT_MBTI.ipynb` (GPU Kaggle), suy luận trong `demo/backend/app.py`.

AI Team Builder. Đầu vào: văn bản tự do. Đầu ra: (i) MBTI 4 ký tự, (ii) *vai trò gợi ý* (ví dụ: ENTJ → Leader; INTP → Analyst). Bảng ánh xạ được định nghĩa trong mã nguồn của hai tuyến suy luận (Flask & Streamlit).

5. Experiments & Evaluation

Dataset & Split

Bộ `mbti-type` gồm **8 675** mẫu; mỗi mẫu chứa trường `posts` (nhiều câu/đoạn) và một nhãn MBTI 4 ký tự. Ta tách nhãn thành bốn trục nhị phân IE, NS, TF, JP.

Tuyến ML: dữ liệu được chia **80/20 (train/test)**. Tập test có đúng **1 735** mẫu, trùng khớp báo cáo trong `classification_reports_ml.txt`.

Tuyến BERT: trong notebook, dữ liệu được chia **70/20/10 (train/valid/test)** với *stratify theo type*: (1) tách train (70%) và temp (30%), (2) tách temp thành validation (20% tổng) và test (10% tổng). File `test.csv` sau đó dùng cho đánh giá cuối cùng.

Đáng chú ý, dù đã tách thành 4 cột, **mất cân bằng lớp** (*class imbalance*) vẫn rõ rệt: IE (1335 vs. 400), NS (1496 vs. 239), TF (796 vs. 939), JP (686 vs. 1049). Điều này ảnh hưởng đến F1 của BERT dù đã áp dụng `pos_weight` trong `BCEWithLogitsLoss`.

Modeling Setups

ML (baseline): TF-IDF trên văn bản đầu vào + các bộ phân loại tuyến tính độc lập cho bốn trục (triển khai trong `app_ml.py`).

BERT: tinh chỉnh `bert-base-uncased` với 4 nhánh nhị phân, pooling [CLS] + mean, dropout=0.4, có tầng ẩn. Chuỗi được cắt ở `max_length=256`, batch size = 8, huấn luyện 40 epoch với AdamW + linear scheduler và warmup=10%. Checkpoint tốt nhất (`mbti_best.pt`) được chọn theo Macro-F1 trên validation.

Metrics

Dùng **Accuracy** và **Macro-F1**. Macro-F1 ưu tiên để phản ánh hiệu năng trong điều kiện mất cân bằng.

Kết quả định lượng

Trục	IE	NS	TF	JP
ML (Acc)	0.840	0.887	0.844	0.809
ML (F1)	0.779	0.791	0.844	0.799
BERT (Acc)	0.764	0.861	0.749	0.730
BERT (Macro-F1)	0.699	0.712	0.749	0.704

Nhận xét. BERT đạt tốt ở trục **NS** (Acc 0.861), nhưng **không vượt** ML ở IE, TF, JP. ML ổn định và cao hơn ở nhiều trục, đặc biệt TF.

Vì sao BERT chưa vượt ML như kỳ vọng?

Ban đầu kỳ vọng BERT sẽ vượt ML nhờ biểu diễn ngữ cảnh. Tuy nhiên:

- **Imbalance:** Tách 4 cột không khử được lệch lớp (IE, NS mất cân bằng mạnh). Với ngưỡng 0.5 cố định, BERT dễ thiên vị lớp đa số.
- **Dữ liệu nhỏ và nhiều:** 8.6k mẫu là ít cho BERT; `posts` chứa slang/emoji/link, khi truncation 256 có thể bỏ mất thông tin.
- **Độc lập 4 trục:** BERT huấn luyện 4 nhánh tách biệt, chưa tận dụng quan hệ NS-TF, JP-IE; trong khi ML ít bị ảnh hưởng cấu trúc này.
- **Config & threshold:** Chưa tối ưu sâu hyperparameters; chưa tuning ngưỡng riêng theo trục → Macro-F1 chưa tối ưu.

Error Analysis

Lỗi phổ biến (từ `error_samples.csv`): IE: văn bản ngắn/emoji/link, tín hiệu mờ; NS: nhầm giữa mô tả sự kiện và suy luận trừu tượng; TF/JP: câu đa dụng (vừa lý trí vừa cảm xúc; vừa ngăn nắp vừa linh hoạt).

Đối chiếu kỳ vọng & Hướng cải thiện

Kết quả cho thấy BERT chỉ nhỉnh ở NS, nhưng thua hoặc tiệm cận ML ở các trục khác.

Khuyến nghị: - Dùng *class weights/focal loss*, hoặc sampling cân bằng. - Tuning ngưỡng riêng cho từng trục (tối ưu Macro-F1). - Tăng `max_length` hoặc dùng long-sequence BERT. -

Dùng *multi-task head* hoặc mô hình 16 lớp với ràng buộc cấu trúc để tận dụng quan hệ giữa các trục.

6. Limitation

- **Hạn chế về dữ liệu.** Bộ dữ liệu chỉ gồm $\approx 8,675$ mẫu tiếng Anh trực tuyến, được thu thập từ diễn đàn và mạng xã hội. Điều này dẫn đến hai điểm yếu: (i) chưa đủ lớn để khai thác hết tiềm năng của BERT, (ii) văn bản chứa nhiều nhiễu (slang, emoji, link) và không phản ánh hoàn toàn ngôn ngữ viết/chuyên nghiệp.
- **Mất cân bằng lớp.** Một số trục MBTI có chênh lệch lớn (ví dụ: NS với 1496 vs. 239). Dù đã dùng `pos_weight`, hiện tượng này vẫn làm giảm Macro-F1 của BERT.
- **Giả định nhân MBTI.** Nghiên cứu giả định rằng nhân MBTI gốc là đúng và ổn định. Thực tế, MBTI chỉ là một mô hình tham chiếu, còn gây tranh luận về tính khoa học và độ tin cậy. Việc dự đoán MBTI chính xác không đồng nghĩa với việc phản ánh đầy đủ tính cách cá nhân.
- **Tổng quát hoá.** Mô hình có thể không hoạt động tốt trên văn bản cực ngắn, văn bản mang tính ẩn dụ/mỉa mai, hoặc khi áp dụng cho ngôn ngữ khác (ví dụ: tiếng Việt).
- **Ràng buộc đầu ra.** Việc huấn luyện bốn trục độc lập có thể tạo tổ hợp MBTI hiếm hoặc không nhất quán với phân bố thực tế, vì không tận dụng ràng buộc liên trục (ví dụ: sự đồng biến giữa NS và TF).

7. Discussion

Bài học chính.

- TF-IDF + tuyến tính là một *baseline* mạnh, dễ huấn luyện và ổn định hơn kỳ vọng.
- BERT nhạy cảm với lựa chọn pooling và hyperparameters; cấu hình [CLS]+mean với dropout 0.4 cho kết quả cân bằng nhất trong thí nghiệm.
- Quá trình suy luận cần hiệu chỉnh ngưỡng riêng cho từng trục để tối ưu Macro-F1, thay vì giữ mặc định 0.5.

Cải thiện tương lai. Có thể mở rộng bằng cách: thu thập dữ liệu đa miền/đa ngôn ngữ; áp dụng *prompt-tuning*, *adapter* hoặc *long-sequence models*; huấn luyện một mô hình 16 lớp trực tiếp thay vì 4 nhánh độc lập, từ đó tận dụng ràng buộc cấu trúc.

Triển khai thực tế. Một hệ thống AI Team Builder cần thêm: (i) dữ liệu từ chat/email, (ii) biểu mẫu tự đánh giá ngắn để bổ trợ MBTI, (iii) cơ chế giải thích kết quả (ví dụ hiển thị top-k câu quyết định). Ngoài ra nên hiển thị độ tin cậy dự đoán để người dùng cân nhắc.

Câu hỏi thảo luận và quan điểm.

- *Nên đa dạng hay đồng nhất MBTI trong đội?* Theo em, một đội nhóm hiệu quả nên có sự **đa dạng** MBTI có kiểm soát: đa dạng để bổ sung điểm mạnh, nhưng vẫn cần tối thiểu một vài điểm chung để đảm bảo giao tiếp trôi chảy.
- *Ngôn ngữ trực tuyến có phản ánh đúng tính cách thực không?* Không hoàn toàn. Người dùng có thể thể hiện khác biệt giữa online và offline. Để cải thiện, cần bổ sung dữ liệu thực tế (ví dụ: phản hồi từ đồng nghiệp, hiệu suất công việc) để cân đối.
- *Nếu dự đoán sai MBTI, rủi ro gì?* Rủi ro là gán nhãn sai vai trò, tạo định kiến hoặc ảnh hưởng động lực cá nhân. Để giảm thiểu, hệ thống nên: (i) hiển thị độ tin cậy, (ii) cho phép phản biện kết quả, (iii) kết hợp nhiều nguồn dữ liệu khác ngoài văn bản.

8. Conclusion

Bài viết trình bày pipeline dự đoán MBTI từ văn bản theo hai tuyến (ML và BERT) trên bộ `mbti-type` 8675 mẫu, cùng cơ chế AI Team Builder ánh xạ sang vai trò nhóm. Các kết quả

Acc/Macro-F1 cạnh tranh trên nhiều trục và có thể ứng dụng ngay qua hai giao diện demo.

Ý nghĩa thực tiễn. Hỗ trợ tạo đội nhóm, gợi ý vai trò và giao tiếp trong môi trường học tập/làm việc.

Hướng mở. Trong tương lai, em kỳ vọng mở rộng sang đa ngôn ngữ, huấn luyện trên dữ liệu giàu ngữ cảnh hơn và triển khai đánh giá người dùng thực địa.

Tái lập thí nghiệm. Notebook `BERT_MBTI.ipynb` **phải chạy trên Kaggle** (GPU T4/P100), **input** là tập `mbti-type` (`data/mbti_1.csv/mbti_clean.csv`). Tuyến ML có thể chạy trên CPU; hai ứng dụng suy luận: Flask (`demo/backend`) và Streamlit (`demo/UI_ML`).

Acknowledgement

Trong quá trình soạn thảo và trình bày báo cáo, em có sử dụng công cụ ChatGPT (OpenAI) để hỗ trợ về gợi ý cách diễn đạt, cấu trúc văn bản và định dạng LaTeX. Tất cả nội dung phân tích, kết quả thực nghiệm và số liệu kỹ thuật được em trực tiếp thực hiện, kiểm chứng và chỉnh sửa lại để đảm bảo tính chính xác.