| Trường Đại học Nguyễn Tất Thành  **Khoa Công nghệ thông tin**  🙜🙜🙝🙝 | CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM  **Độc lập – Tự do – Hạnh phúc**  🙜🙜🙝🙝 |
| --- | --- |

**NHIỆM VỤ KHÓA LUẬN TỐT NGHIỆP**

*(Sinh viên phải đóng tờ này vào báo cáo)*

Họ và tên: Nguyễn Lê Quốc BảoMSSV: 2100004053

Chuyên ngành: Trí tuệ nhân tạo Lớp: 21DTH1C Email: 2100004053@nttu.edu.vn SĐT:0326032296 Tên đề tài: Phát triển hệ thống khuyến nghị sản phẩm dựa trên Collaborative Filtering

Giáo viên hướng dẫn: ThS. Phạm Đình Tài

Thời gian thực hiện:11/2/2025 đến 8/5/2025

Nhiệm vụ/nội dung (mô tả chi tiết nội dung, yêu cầu, phương pháp… ) :

**MÔ TẢ ĐỀ TÀI:**

**Tên đề tài**: **Phát hiện Ngôn ngữ Tiếng Việt bằng Mô hình Học Sâu.**

Phát triển hệ thống khuyến nghị dựa trên Collaborative Filtering, một phương pháp phổ biến dựa trên sự tương đồng giữa người dùng hoặc sản phẩm. Đề tài sẽ sử dụng các thuật toán như User-based Collaborative Filtering hoặc Item-based Collaborative Filtering để khuyến nghị sản phẩm dựa trên lịch sử người dùng hoặc người dùng tương tự. Sinh viên sẽ làm việc với dữ liệu mua sắm và tìm hiểu cách tối ưu hóa mô hình để đưa ra các đề xuất cá nhân hóa hơn.

**NỘI DUNG VÀ PHƯƠNG PHÁP:**

* Nghiên cứu các kiến trúc và mô hình nền tản của lĩnh vực xử lý ngôn ngữ tự nhiên(rnn, lstm, tranformer, bert, …).
* Nghiên cứu và tìm hiểu về các bài toán Collaborative Filtering ưu và nhược điểm của các loại biến thể, và sự phát triển, ứng dụng của nó trong ngành thời trang.
* Tìm kiếm và sử dụng các thuật toán tokenize, embeding tối ưu cho bài toán đề xuất(word2vec, embeding nhúng giá trị, dùng hàm log để mã hóa dữ liệu,…).
* Tìm kiếm, giả lập, xủ lý dữ liệu về dạng thích hợp đê training(viết ra các logic hợp lý, quy luật để tạo dữ liệu ảo một cách chân thật nhất).
* Nghiên cứu về kiến trúc micro-service cho backend và các thư viện như spring boot, flask, react, và sql server để xây dựng ứng dụng web cho mô hình.
* Nghiên cứu về nghiệp vụ người dùng để thiết kế chức năng hợp lý cho trang web
* Nghiên cứu về tối ưu truy vấn sql để tối ưu hiệu năng.
* Nghiên cứu về cách bảo mật của web tránh rò rĩ thông tin và tói ưu hiệu năng server.
* Học về tư duy thiết kế, các phương pháp thiết kế để thiết kế giao diện font end tối ưu và hiệu quả.
* Tìm kiếm các hàm loss tối ưu, các độ đo thích hợp và các phương pháp tránh overfit cho mô hình, vì đây là một dạng mô hình rất dễ bị overfit.

**YÊU CẦU:**

* Xây dựng được một tập dữ liệu có ít nhất 500000 mẫu và đáp ứng nhiều tiêu chí như: đa dạng, không trùng lập, có nhiễu, có logic đủ phức tạp và hợp lý
* Xây dựng được một mô hình ít nhất đạt mức trên 70% cho độ đo f1, acc,…
* Ứng dụng những công nghệ mới vào cơ chế Collaborative Filtering cũ
* Có một sự đột biến nhất định trong kiến trúc và phải man lại hiệu quả cao
* Các dữ liệu phải được xử lý hợp lý và phân bổ tối ưu nhất với từng đặc trưng để tối đa hóa sự liên quan của các đặc trưng.
* Phải xây dựng được một hệ thống ứng dụng web hoàn chỉnh tối ưu hiệu năng, dể bảo trì, mở rộng và bảo mật cao.

**Nội dung và yêu cầu đã được thông qua Bộ môn.**

*TP.HCM, ngày .. tháng .. năm 2023*

| **TRƯỞNG BỘ MÔN**  *(Ký và ghi rõ họ tên)* | **GIÁO VIÊN HƯỚNG DẪN**  *(Ký và ghi rõ họ tên)* |
| --- | --- |

**ThS. Trần Châu Thanh Thiện ThS. Phạm Đình Tài**