

ĐỀ CƯƠNG ĐỒ ÁN TỐT NGHIỆP

Họ và tên sinh viên: Nguyễn Hải Đăng

Mã sinh viên : 1571020059

Lớp: CNTT 1505

Địa chỉ : Lào Cai

E-mail : nguyendang2467@gmail.com

Ngành : Công nghệ thông tin

Tên đề tài: Phân tích và dự đoán tỷ lệ thất nghiệp kết hợp dữ liệu kinh tế và xã hội

Giảng viên hướng dẫn: ThS. Trần Thị Thanh Nhân

NỘI DUNG ĐỀ CƯƠNG

1. Tính cấp thiết của đề tài/ Lý do chọn đề tài:

- Tỷ lệ thất nghiệp là một chỉ số quan trọng phản ánh tình trạng kinh tế và xã hội của một quốc gia. Việc phân tích và dự đoán tỷ lệ thất nghiệp giúp các nhà hoạch định chính sách đưa ra các biện pháp phù hợp để kiểm soát thị trường lao động, giảm thiểu rủi ro kinh tế và thúc đẩy tăng trưởng bền vững.

- Trong bối cảnh công nghệ phát triển nhanh chóng và thị trường việc làm có nhiều biến động, việc sử dụng các mô hình dự báo hiện đại kết hợp dữ liệu kinh tế - xã hội giúp nâng cao độ chính xác trong dự đoán thất nghiệp. Đặc biệt, ứng dụng công nghệ trí tuệ nhân tạo vào phân tích dữ liệu lớn (big data) mang lại những giá trị quan trọng trong việc hỗ trợ chính sách lao động và việc làm.

2. Mục đích nghiên cứu:

- Phân tích ảnh hưởng của các yếu tố kinh tế - xã hội đến tỷ lệ thất nghiệp.
- Ứng dụng các mô hình học máy để dự báo tỷ lệ thất nghiệp với độ chính xác cao.
- Phát triển giao diện trực quan giúp người dùng theo dõi và phân tích dữ liệu.

3. Phạm vi nghiên cứu:

- **Không gian:** Nghiên cứu trên phạm vi quốc tế hoặc tập trung vào một số nền kinh tế lớn.
- **Thời gian:** Sử dụng dữ liệu trong giai đoạn 2023 - 2024 để đảm bảo tính cập nhật.
- Dữ liệu kinh tế và xã hội từ các nguồn chính thống

4. Phương pháp nghiên cứu:

- Phân tích dữ liệu: Xử lý dữ liệu, trực quan hóa, mối tương quan giữa các yếu tố
- Phương pháp phân tích dữ liệu:
 - **Xây dựng mô hình:**
 - Các mô hình thống kê (Linear Regression, ARIMA)
 - Các mô hình Machine Learning (Random Forest, XGBoost)
 - Mô hình Deep Learning (LSTM, GRU)
 - Trực quan hóa dữ liệu bằng Python (Matplotlib, Seaborn).
- **Phát triển ứng dụng web:**
 - Ứng dụng công nghệ Flask để xây dựng backend.
 - Sử dụng Streamlit cho giao diện người dùng (frontend).
 - Triển khai mô hình dự báo lên nền tảng web để dễ dàng truy cập và sử dụng.

5. Kết cấu của đồ án:

CHƯƠNG 1: TỔNG QUAN VỀ THẤT NGHIỆP VÀ DỰ BÁO DỮ LIỆU KINH TẾ

1.1 Tổng quan về thất nghiệp

1.2 Các yếu tố ảnh hưởng đến tỷ lệ thất nghiệp

1.3 Các phương pháp dự báo dữ liệu kinh tế

CHƯƠNG 2: XÂY DỰNG MÔ HÌNH AI DỰ ĐOÁN TỶ LỆ THẤT NGHIỆP

2.1. Quy trình xây dựng mô hình AI

2.1.1. Tiền xử lý dữ liệu

2.1.2. Trích xuất và lựa chọn đặc trưng

2.2 Xây dựng mô hình AI dự đoán thất nghiệp

2.2.1 Mô hình Machine Learning (Linear Regression, Random Forest, XGBoost)

2.2.2 Mô hình Deep Learning (LSTM, GRU)

2.3 Đánh giá hiệu suất mô hình

2.3.1 Các chỉ số đánh giá mô hình (MSE, RMSE, R^2)

2.3.2 Tối ưu hóa mô hình

CHƯƠNG 3: XÂY DỰNG HỆ THỐNG DỰ ĐOÁN VÀ GIAO DIỆN NGƯỜI DÙNG

3.1 Yêu cầu hệ thống

3.1.1 Chức năng chính của hệ thống

3.1.2 Kiến trúc hệ thống và công nghệ sử dụng

3.1.3 Mô hình triển khai API và giao diện

3.2 Xây dựng API dự đoán

- 3.2.1 Triển khai mô hình AI lên server
- 3.2.2 Xây dựng API kết nối giữa AI và giao diện
- 3.2.3 Kiểm thử và tối ưu API

3.3 Xây dựng giao diện người dùng

- 3.3.1 Thiết kế giao diện web/app
- 3.3.2 Chức năng nhập dữ liệu và hiển thị kết quả dự đoán
- 3.3.3 Kiểm thử giao diện và tối ưu trải nghiệm người dùng

CHƯƠNG 4: KẾT QUẢ, ĐÁNH GIÁ VÀ HƯỚNG PHÁT TRIỂN

- 4.1. Kết quả thực nghiệm và đánh giá hiệu suất hệ thống
- 4.2. Hướng phát triển trong tương lai
- 4.3. Kết luận

TÀI LIỆU THAM KHẢO

6. Kế hoạch chi tiết thực hiện đồ án tốt nghiệp:

Giai đoạn 1: Nghiên cứu tổng quan và thu thập dữ liệu (10/01/2025 - 23/01/2025)

- **Tuần 1 (10/01 - 16/01/2025):**
 - Nghiên cứu lý thuyết về thất nghiệp và các yếu tố kinh tế - xã hội.
 - Tìm hiểu các mô hình dự báo: Hồi quy tuyến tính, Random Forest, Mạng nơ-ron nhân tạo (ANN).
- **Tuần 2 (17/01 - 23/01/2025):**
 - Thu thập dữ liệu từ:
 - *Unemployment Rate Dataset (Dữ liệu thất nghiệp).*
 - *World Economic Indicators Dataset (Chỉ số kinh tế toàn cầu).*
 - *LinkedIn Job Postings (2023 - 2024).*
 - Tiền xử lý dữ liệu: Loại bỏ dữ liệu trùng, xử lý giá trị thiếu.

Giai đoạn 2: Phân tích dữ liệu và thử nghiệm mô hình (24/01/2025 - 13/02/2025)

- **Tuần 3 (24/01 - 30/01/2025):**
 - Phân tích dữ liệu bằng Python (Pandas, Seaborn, Matplotlib).
 - Trực quan hóa dữ liệu để xác định xu hướng thất nghiệp.
- **Tuần 4 - 5 (31/01 - 13/02/2025):**
 - Triển khai thử nghiệm các mô hình dự đoán
 - Đánh giá sơ bộ hiệu suất mô hình bằng các chỉ số: MSE, RMSE, R^2 .

Giai đoạn 3: Huấn luyện và tối ưu mô hình (14/02/2025 - 27/02/2025)

- **Tuần 6 (14/02 - 20/02/2025):**

- Cải thiện dữ liệu đầu vào và xử lý outliers.
- Huấn luyện lại các mô hình với tập dữ liệu đầy đủ hơn.
- **Tuần 7 (21/02 - 27/02/2025):**
 - So sánh hiệu suất các mô hình và chọn mô hình tốt nhất.
 - Điều chỉnh tham số (*hyperparameter tuning*) để tăng độ chính xác.

Giai đoạn 4: Xây dựng ứng dụng web và tích hợp mô hình (28/02/2025 - 27/03/2025)

- **Tuần 8 - 9 (28/02 - 13/03/2025):**
 - Xây dựng backend với Flask/Python.
 - Phát triển frontend với Streamlit.
- **Tuần 10 (14/03 - 27/03/2025):**
 - Tích hợp mô hình dự báo vào ứng dụng web.
 - Kiểm thử ứng dụng với dữ liệu thực tế.

Giai đoạn 5: Hoàn thiện báo cáo và bảo vệ đồ án (28/03/2025 - 02/06/2025)

- **Tuần 11 - 12 (28/03 - 10/04/2025)**
 - Viết báo cáo đồ án theo bố cục chuẩn.
 - Chuẩn bị bài thuyết trình.
- **Tuần 13 - 14 (11/04 - 24/04/2025):**
 - Kiểm tra và chỉnh sửa nội dung báo cáo.
 - Tối ưu lại ứng dụng web dựa trên phản hồi (nếu có).
- **Tuần 15 - 16 (25/04 - 08/05/2025):**
 - Nộp báo cáo chính thức (hạn cuối: 27/05/2025).
 - Thực hiện thử nghiệm thuyết trình.
- **Tuần 17 - 19 (09/05 - 02/06/2025):**
 - Chuẩn bị và tham gia buổi bảo vệ đồ án

Hà Nội, ngày 24 tháng 2 năm 2025

Ý kiến giảng viên hướng dẫn

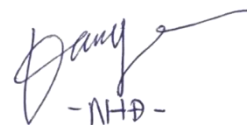
(Ký và ghi rõ họ tên)



ThS. Trần Thị Thanh Nhàn

Sinh viên thực hiện

(Ký và ghi rõ họ tên)



Nguyễn Hải Đăng