|  |  |
| --- | --- |
| HỌC VIỆN NÔNG NGHIỆP  VIỆT NAM  **KHOA CÔNG NGHỆ THÔNG TIN** | CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM  **Độc lập – Tự do – Hạnh phúc** |

 

**BÁO CÁO**

**TRÍ TUỆ NHÂN TẠO**

**ĐỀ TÀI:** **Hệ thống phân loại cảm xúc câu văn tiếng Việt**

|  |  |  |
| --- | --- | --- |
| Họ và tên | Lớp | Mã Sinh Viên |
| Bùi Tố Thành Hưng | K65CNPMA | 651288 |
| Phạm Quang Huy | K65CNTTA | 650706 |

|  |  |  |  |
| --- | --- | --- | --- |
| Môn học  Giảng viên | :  : | Trí tuệ nhân tạo   Nguyễn Thị Huyền |  |

HÀ NỘI - NĂM 2024

**BẢNG PHÂN CÔNG NHIỆM VỤ**

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Nhiệm vụ | Sinh viên thực hiện | |
| Họ tên | Mã sinh viên |
| 1 | * Phân tích từ & câu * Quản lý từ điển * Cấu hình & tích hợp | Bùi Tố Thành Hưng | 651288 |
| 2 | * Giao diện và trải nghiệm người dùng | Phạm Quang Huy | 650706 |

**MỤC LỤC**

[**CHƯƠNG I: MỞ ĐẦU** 4](#_Toc198521961)

[**1.** **Giới thiệu** 4](#_Toc198521962)

[**2.** **Mục đích và yêu cầu** 4](#_Toc198521963)

[2.1. Mục đích 4](#_Toc198521964)

[2.2. Yêu cầu 4](#_Toc198521965)

[**3** **Nội dung nghiên cứu** 5](#_Toc198521966)

[**4** **Phương pháp nghiên cứu** 5](#_Toc198521967)

[**CHƯƠNG II: TỔNG QUAN VỀ CÔNG NGHỆ SỬ DỤNG** 7](#_Toc198521968)

[**1.1.** **Tổng quan mô hình BERTweet:** 7](#_Toc198521969)

[1.2. Tổng quan về mô hình BERTweet**:** 7](#_Toc198521970)

[1.3. Ưu và nhược điểm của BERTweet: 7](#_Toc198521971)

[**2.** **Tổng quan về Streamlit** 8](#_Toc198521972)

[2.1. Tổng quan về framework Streamlit: 8](#_Toc198521973)

[2.2. Ưu và nhược điểm của framework Streamlit 8](#_Toc198521974)

[**3.** **Tổng quan về thư viện khác** 8](#_Toc198521975)

[**CHƯƠNG III: ĐẶC TẢ YÊU CẦU HỆ THỐNG** 12](#_Toc198521976)

[1. Giới thiệu 12](#_Toc198521977)

[2. Yêu cầu chức năng 12](#_Toc198521978)

[**CHƯƠNG IV: XÂY DỰNG CHƯƠNG TRÌNH** 14](#_Toc198521979)

[**KẾT LUẬN VÀ ĐỀ NGHỊ** 17](#_Toc198521980)

[**1.** **Kết luận** 17](#_Toc198521981)

[**2.** **Đề nghị** 17](#_Toc198521982)

[**TÀI LIỆU THAM KHẢO** 18](#_Toc198521983)

# **CHƯƠNG I: MỞ ĐẦU**

1. **Giới thiệu**

* Trong bối cảnh phát triển nhanh chóng của công nghệ thông tin và truyền thông, việc phân tích cảm xúc trong văn bản trở thành một lĩnh vực quan trọng, đặc biệt là trong việc hiểu và tương tác với người dùng trên các nền tảng mạng xã hội. Dự án "Phân tích Cảm xúc Tiếng Việt với BERTweet" được thực hiện nhằm mục tiêu phát triển một hệ thống phân tích cảm xúc hiệu quả cho văn bản tiếng Việt, sử dụng mô hình BERTweet - một biến thể của mô hình BERT được tối ưu hóa cho ngôn ngữ tiếng Việt.
* Hệ thống không chỉ giúp xác định các loại cảm xúc như tích cực, tiêu cực và trung lập trong văn bản mà còn cung cấp các điểm số cụ thể cho từng loại cảm xúc, từ đó hỗ trợ người dùng trong việc hiểu rõ hơn về tâm trạng và ý kiến của cộng đồng. Với giao diện thân thiện và khả năng lưu trữ lịch sử phân tích, dự án hứa hẹn sẽ mang lại những giá trị thiết thực cho các doanh nghiệp, nhà nghiên cứu và người dùng cá nhân trong việc khai thác thông tin từ văn bản tiếng Việt. Thông qua việc áp dụng công nghệ học sâu và xử lý ngôn ngữ tự nhiên, dự án không chỉ góp phần nâng cao khả năng phân tích cảm xúc mà còn mở ra hướng đi mới cho các nghiên cứu và ứng dụng trong lĩnh vực này.

1. **Mục đích và yêu cầu** 
   1. Mục đích

Đề tài "Hệ thống phân loại cảm xúc câu văn tiếng Việt" được thực hiện nhằm xây dựng một hệ thống có khả năng tự động nhận diện và phân loại cảm xúc (ví dụ: tích cực, tiêu cực, trung lập) trong các câu văn tiếng Việt. Mục đích cụ thể bao gồm:

* Ứng dụng các kỹ thuật của Trí tuệ Nhân tạo, đặc biệt là Xử lý ngôn ngữ tự nhiên (NLP) và Học máy (Machine Learning), vào việc phân tích cảm xúc văn bản tiếng Việt.
* Tạo ra một mô hình có thể áp dụng trong thực tế, chẳng hạn như đánh giá phản hồi khách hàng, phân tích bình luận mạng xã hội, hoặc hỗ trợ chatbot hiểu cảm xúc người dùng.
* Nâng cao khả năng nghiên cứu và thực hành các thuật toán học máy, xử lý ngôn ngữ tự nhiên, và tiền xử lý văn bản tiếng Việt.
* Góp phần giải quyết bài toán phân tích cảm xúc cho ngôn ngữ tiếng Việt – một lĩnh vực vẫn còn hạn chế về tài nguyên và công cụ so với các ngôn ngữ phổ biến khác như tiếng Anh.
  1. Yêu cầu

Để đạt được mục đích trên, hệ thống cần đáp ứng các yêu cầu sau:

* Thu thập và xử lý dữ liệu: Có tập dữ liệu huấn luyện và kiểm tra phù hợp, bao gồm các câu văn tiếng Việt đã được gán nhãn cảm xúc rõ ràng.
* Tiền xử lý văn bản tiếng Việt: Loại bỏ ký tự đặc biệt, chuẩn hóa câu chữ, tách từ (word segmentation), loại bỏ stop words, v.v.
* Lựa chọn mô hình học máy hoặc học sâu: Sử dụng các mô hình như Naive Bayes, SVM, hoặc các mô hình mạng nơ-ron như LSTM, BERT tiếng Việt,...
* Huấn luyện và đánh giá mô hình: Đánh giá độ chính xác, độ nhạy, độ đặc hiệu và F1-score để đo lường hiệu quả mô hình.
* Xây dựng giao diện đơn giản (nếu có): Cho phép người dùng nhập câu văn tiếng Việt và nhận kết quả phân loại cảm xúc tương ứng.
* Khả năng mở rộng: Mở rộng phân loại từ 2 nhãn (tích cực/tiêu cực) sang nhiều loại cảm xúc hơn (vui, buồn, giận dữ, sợ hãi, ngạc nhiên, v.v.)

1. **Nội dung nghiên cứu**

* Mục tiêu nghiên cứu: Phát triển một hệ thống phân tích cảm xúc cho văn bản tiếng Việt, sử dụng mô hình BERTweet để cải thiện độ chính xác và hiệu suất trong việc nhận diện cảm xúc
* Tập dữ liệu: Xây dựng và sử dụng một tập dữ liệu phong phú, bao gồm các văn bản tiếng Việt từ nhiều nguồn khác nhau như mạng xã hội, bài viết, và bình luận, để huấn luyện và kiểm tra mô hình.
* Phân tích từ và câu: Nghiên cứu cách mà các từ và cấu trúc câu ảnh hưởng đến cảm xúc tổng thể của văn bản. Điều này bao gồm việc phân tích ngữ nghĩa và ngữ pháp để hiểu rõ hơn về cách mà cảm xúc được thể hiện trong ngôn ngữ.
* Phát triển từ điển cảm xúc: Tạo ra một từ điển cảm xúc chi tiết, phân loại các từ theo cảm xúc tích cực, tiêu cực và trung tính, nhằm hỗ trợ cho quá trình phân tích.
* Đánh giá mô hình: Thực hiện các bài kiểm tra và đánh giá để xác định độ chính xác của mô hình trong việc phân tích cảm xúc, bao gồm việc so sánh với các phương pháp phân tích cảm xúc khác.
* Thống kê và báo cáo: Thu thập và phân tích dữ liệu từ các lần phân tích cảm xúc, xuất báo cáo để trình bày kết quả và rút ra các kết luận từ nghiên cứu.

1. **Phương pháp nghiên cứu**

* Nghiên cứu lý thuyết: Tìm hiểu các phương pháp phân tích cảm xúc hiện có, đặc biệt là trong ngữ cảnh tiếng Việt. Nghiên cứu các mô hình học sâu, đặc biệt là BERT và các biến thể của nó như BERTweet, để hiểu cách chúng hoạt động và ứng dụng trong phân tích ngữ nghĩa.
* Tạo và quản lý từ điển: Sử dụng các nguồn tài liệu để xây dựng từ điển cảm xúc, bao gồm từ đơn, cụm từ, thành ngữ, từ tăng cường, giảm nhẹ và phủ định. Từ điển này sẽ được lưu trữ dưới dạng file JSON để dễ dàng cập nhật và truy xuất.
* Sử dụng mô hình BERTweet: Nạp mô hình BERTweet đã được huấn luyện trước để phân tích cảm xúc câu. Mô hình này sẽ được điều chỉnh để phù hợp với ngữ cảnh tiếng Việt và các quy tắc từ điển đã xây dựng.
* Tiền xử lý văn bản: Chuẩn hóa văn bản đầu vào, tách từ và nhận diện các dấu câu, emoji.
* Phân tích cảm xúc: Sử dụng các quy tắc từ điển để xác định cảm xúc của từng từ, cụm từ và toàn bộ câu. Kết hợp giữa phân tích từ và phân tích câu để đạt được kết quả chính xác hơn.
* Đánh giá kết quả: So sánh kết quả phân tích với các dữ liệu thực tế hoặc các phương pháp phân tích khác để đánh giá độ chính xác.
* Điều chỉnh mô hình: Dựa trên phản hồi và kết quả đánh giá, điều chỉnh các tham số của mô hình và từ điển để cải thiện độ chính xác.
* Triển khai ứng dụng: Xây dựng giao diện web cho phép người dùng nhập văn bản và nhận kết quả phân tích cảm xúc.
* Thu thập phản hồi: Nhận phản hồi từ người dùng để cải thiện và mở rộng các chức năng của ứng dụng.

# **CHƯƠNG II: TỔNG QUAN VỀ CÔNG NGHỆ SỬ DỤNG**

* 1. **Tổng quan mô hình BERTweet:**
  2. Tổng quan về mô hình BERTweet**:**
* BERTweet là một biến thể của mô hình BERT (Bidirectional Encoder Representations from Transformers) được tối ưu hóa cho ngôn ngữ tiếng Việt. Mô hình này sử dụng kiến trúc Transformer để hiểu ngữ nghĩa của văn bản trong ngữ cảnh, cho phép phân tích cảm xúc một cách chính xác hơn.
* Xử lý ngôn ngữ tự nhiên (NLP): Công nghệ NLP được áp dụng để xử lý và phân tích văn bản tiếng Việt. Các kỹ thuật như tokenization (phân tách từ), stemming (chuẩn hóa từ), và lemmatization (chuyển đổi từ về dạng gốc) được sử dụng để chuẩn bị dữ liệu cho mô hình.
* Học sâu (Deep Learning): Dự án sử dụng các phương pháp học sâu để huấn luyện mô hình BERTweet trên tập dữ liệu cảm xúc. Các thuật toán tối ưu hóa như Adam hoặc SGD (Stochastic Gradient Descent) được áp dụng để cải thiện hiệu suất của mô hình.
* Thư viện và công cụ: Sử dụng các thư viện Python như transformers từ Hugging Face để triển khai mô hình BERTweet, cùng với các thư viện như pandas và numpy để xử lý dữ liệu. Ngoài ra, Flask hoặc FastAPI có thể được sử dụng để xây dựng giao diện web cho ứng dụng phân tích cảm xúc.
* Quản lý dữ liệu: Dữ liệu được lưu trữ và quản lý thông qua các file JSON hoặc cơ sở dữ liệu, cho phép dễ dàng truy xuất và cập nhật thông tin từ điển cảm xúc và lịch sử phân tích.
* Giao diện người dùng: Dự án có thể bao gồm một giao diện người dùng thân thiện, cho phép người dùng nhập văn bản và nhận kết quả phân tích cảm xúc một cách nhanh chóng và dễ dàng
  1. Ưu và nhược điểm của BERTweet:

1. Ưu điểm:

* **Độ chính xác cao:** Sử dụng mô hình BERTweet giúp cải thiện độ chính xác trong việc phân tích cảm xúc so với các phương pháp truyền thống.
* **Khả năng xử lý ngữ nghĩa:** Mô hình BERT có khả năng hiểu ngữ cảnh và ngữ nghĩa của từ, giúp phân tích cảm xúc một cách tinh vi hơn.
* **Tính mở rộng:** Dự án có thể dễ dàng mở rộng để bao gồm nhiều ngôn ngữ khác hoặc các loại dữ liệu khác nhau.
* **Giao diện thân thiện:** Nếu sử dụng Streamlit hoặc các công cụ tương tự, người dùng có thể dễ dàng tương tác với ứng dụng mà không cần kiến thức lập trình.

1. Nhược điểm:

* **Yêu cầu tài nguyên tính toán cao:** Mô hình BERT và các biến thể của nó thường yêu cầu phần cứng mạnh mẽ để huấn luyện và triển khai.
* **Khó khăn trong việc thu thập dữ liệu:** Việc xây dựng một tập dữ liệu phong phú và đa dạng có thể gặp khó khăn, đặc biệt là trong việc thu thập dữ liệu từ các nguồn khác nhau.
* **Thời gian huấn luyện lâu:** Huấn luyện mô hình BERT có thể mất nhiều thời gian, đặc biệt khi làm việc với tập dữ liệu lớn.
* **Cần bảo trì và cập nhật:** Mô hình và từ điển cảm xúc cần được bảo trì và cập nhật thường xuyên để đảm bảo tính chính xác và phù hợp với ngữ cảnh hiện tại.

1. **Tổng quan về Streamlit**
   1. Tổng quan về framework Streamlit:

Mô tả: Streamlit là một framework mã nguồn mở cho phép xây dựng ứng dụng web tương tác cho machine learning và data science một cách nhanh chóng và dễ dàng.

* 1. Ưu và nhược điểm của framework Streamlit

1. Ưu điểm

* Dễ sử dụng: Chỉ cần viết mã Python, không cần kiến thức về HTML hay JavaScript.
* Tích hợp tốt với các thư viện Python như NumPy, Pandas, Matplotlib, và Plotly.
* Cung cấp giao diện người dùng trực quan và tương tác.

1. Nhược điểm

* Không phù hợp cho các ứng dụng phức tạp yêu cầu nhiều tính năng tùy chỉnh.
* Hiệu suất có thể không tốt bằng các framework web truyền thống khi xử lý lượng người dùng lớn.

1. **Tổng quan về thư viện khác**
   1. Tổng quan về thư viện NumPy:

**NumPy** Thư viện cho tính toán khoa học với Python, cung cấp các cấu trúc dữ liệu như mảng n-dimensional.

1. Ưu điểm:

* Hiệu suất cao trong các phép toán số học và ma trận.
* Là nền tảng cho nhiều thư viện khoa học khác trong Python.

1. Nhược điểm:

* Cần có kiến thức cơ bản về toán học để sử dụng hiệu quả.
  1. [Tổng quan về thư viện Pandas](#_CHƯƠNG_II:_TỔNG)

**Pandas** Thư viện mạnh mẽ cho việc xử lý và phân tích dữ liệu, cho phép thao tác với dữ liệu dạng bảng (DataFrame).

1. Ưu điểm:

* Cung cấp nhiều chức năng để xử lý dữ liệu, như lọc, nhóm, và thống kê.
* Dễ dàng tích hợp với các thư viện khác như NumPy và Matplotlib.

1. Nhược điểm:

* Có thể tiêu tốn bộ nhớ khi làm việc với tập dữ liệu lớn.
  1. Tổng quan về thư viện Transformers

Thư viện từ Hugging Face cung cấp các mô hình pre-trained cho nhiều tác vụ xử lý ngôn ngữ tự nhiên, bao gồm BERT, BERTweet, GPT, và nhiều mô hình khác.

* + 1. Ưu điểm:
* Dễ dàng sử dụng với API thân thiện.
* Hỗ trợ nhiều ngôn ngữ và tác vụ khác nhau.
* Cung cấp các mô hình đã được huấn luyện sẵn, giúp tiết kiệm thời gian và tài nguyên.
  + 1. Nhược điểm:
* Yêu cầu tài nguyên tính toán cao khi huấn luyện các mô hình lớn.
  1. Tổng quan về thư viện **PyTorch**

PyTorch là một thư viện mã nguồn mở cho học sâu (deep learning) được phát triển bởi Facebook's AI Research lab (FAIR). Nó được thiết kế để dễ sử dụng và linh hoạt, cho phép các nhà nghiên cứu và lập trình viên xây dựng và triển khai các mô hình học máy một cách nhanh chóng và hiệu quả.

1. Ưu điểm:

* Tính linh hoạt: PyTorch cho phép người dùng dễ dàng thay đổi cấu trúc của mô hình trong quá trình huấn luyện, điều này rất hữu ích cho việc thử nghiệm và phát triển các mô hình mới.
* Tính năng động: PyTorch sử dụng một hệ thống tính toán động (dynamic computation graph), cho phép người dùng xây dựng và thay đổi đồ thị tính toán trong thời gian thực. Điều này giúp dễ dàng gỡ lỗi và kiểm tra các mô hình.
* Hỗ trợ GPU: PyTorch hỗ trợ tính toán trên GPU, giúp tăng tốc độ huấn luyện mô hình một cách đáng kể. Người dùng có thể dễ dàng chuyển đổi giữa CPU và GPU chỉ bằng cách thay đổi một vài dòng mã.
* Thư viện phong phú: PyTorch cung cấp nhiều thư viện con như torchvision cho xử lý hình ảnh, torchaudio cho xử lý âm thanh, và torchtext cho xử lý văn bản, giúp mở rộng khả năng của thư viện.

1. Nhược điểm:

* Hiệu suất: Mặc dù PyTorch có hiệu suất tốt, nhưng trong một số trường hợp, các thư viện khác như TensorFlow có thể tối ưu hóa tốt hơn cho các mô hình lớn và phức tạp.
* Thiếu tính năng sản xuất: PyTorch chủ yếu được thiết kế cho nghiên cứu và phát triển, nên một số tính năng cho sản xuất (như triển khai mô hình) có thể không mạnh mẽ bằng TensorFlow.
* Tài liệu và hướng dẫn: Mặc dù PyTorch có tài liệu phong phú, nhưng một số người dùng có thể cảm thấy tài liệu chưa đầy đủ hoặc không rõ ràng cho một số tính năng nâng cao.
* Quản lý mô hình: Việc quản lý và triển khai mô hình trong môi trường sản xuất có thể phức tạp hơn so với một số thư viện khác, đặc biệt là khi cần tích hợp với các hệ thống khác.
* Khó khăn trong việc chuyển giao kiến thức: Nếu một nhóm đã quen với TensorFlow, việc chuyển sang PyTorch có thể gặp khó khăn do sự khác biệt trong cách tiếp cận và cú pháp.

# **CHƯƠNG III: ĐẶC TẢ YÊU CẦU HỆ THỐNG**

1. Giới thiệu

* Mục tiêu: Phát triển một hệ thống phân tích cảm xúc cho văn bản tiếng Việt sử dụng mô hình BERTweet.
* Phạm vi: Hệ thống sẽ cho phép người dùng nhập văn bản và nhận phân tích cảm xúc, bao gồm các loại cảm xúc như tích cực, tiêu cực và trung lập.

1. Yêu cầu chức năng

* Nhập văn bản: Hệ thống phải cho phép người dùng nhập văn bản để phân tích.
* Phân tích cảm xúc: Hệ thống phải sử dụng mô hình BERTweet để phân tích cảm xúc của văn bản và trả về kết quả.
* Hiển thị kết quả: Kết quả phân tích phải được hiển thị rõ ràng, bao gồm:
* Điểm số cho từng loại cảm xúc (tích cực, tiêu cực, trung lập).
* Cảm xúc chính và độ tin cậy của phân tích.
* Quản lý từ điển: Hệ thống phải có khả năng nạp và quản lý từ điển cảm xúc từ file JSON.
* Lưu lịch sử phân tích: Hệ thống phải lưu lại lịch sử phân tích để người dùng có thể xem lại.
* Xuất báo cáo: Hệ thống phải cho phép xuất kết quả phân tích ra file DOCX hoặc định dạng khác.

1. Yêu cầu phi chức năng

* Hiệu suất: Hệ thống phải có khả năng phân tích văn bản trong thời gian thực, với thời gian phản hồi không quá 2 giây cho mỗi văn bản.
* Khả năng mở rộng: Hệ thống phải có khả năng mở rộng để hỗ trợ nhiều ngôn ngữ khác trong tương lai.
* Bảo mật: Dữ liệu người dùng phải được bảo mật và không được lưu trữ mà không có sự đồng ý của họ.
* Tính khả dụng: Hệ thống phải có tỷ lệ thời gian hoạt động (uptime) tối thiểu 99% trong một tháng.
* Giao diện người dùng: Giao diện phải thân thiện, dễ sử dụng và hỗ trợ trên các thiết bị di động và máy tính để bàn.

1. Yêu cầu phần cứng và phần mềm (thực hiện chạy trên máy tính cá nhân)

* Phần cứng:
* CPU: Tối thiểu 4 nhân.
* RAM: Tối thiểu 8GB.
* GPU: Tối thiểu 4GB VRAM (nếu sử dụng GPU cho huấn luyện và phân tích).
* Phần mềm:
* Hệ điều hành: Windows, macOS hoặc Linux.
* Python 3.6 trở lên.
* Các thư viện cần thiết: PyTorch, Transformers, Flask (hoặc Streamlit), Pandas, NumPy.

1. Giới hạn

* Hệ thống chỉ hỗ trợ phân tích văn bản tiếng Việt.
* Các mô hình phân tích cảm xúc có thể không chính xác 100% và cần được cải thiện theo thời gian.

**CHƯƠNG IV: XÂY DỰNG CHƯƠNG TRÌNH**

* 1. Kết quả thực hiện được:
* Phân tích cảm xúc: Hệ thống có khả năng phân tích cảm xúc của văn bản tiếng Việt, xác định các loại cảm xúc như tích cực, tiêu cực và trung lập.
* Giao diện người dùng: Cung cấp giao diện thân thiện cho người dùng để nhập văn bản và nhận kết quả phân tích.
* Lưu lịch sử phân tích: Hệ thống lưu lại lịch sử phân tích để người dùng có thể xem lại các kết quả trước đó.
* Xuất báo cáo: Cho phép xuất kết quả phân tích ra file DOCX hoặc định dạng khác để dễ dàng chia sẻ và trình bày.
  1. Môi trường cài đặt:
* Về phần cứng:
* Hệ điều hành: Windows, macOS hoặc Linux.
* Python: Phiên bản 3.6 trở lên.
* CPU: Tối thiểu 4 nhân.
* RAM: Tối thiểu 8GB.
* GPU: Tối thiểu 4GB VRAM (nếu sử dụng GPU cho huấn luyện và phân tích).
* Yêu cầu về thư viện cần thiết:
* Torch: Thư viện PyTorch cho học sâu
* Transformers: Thư viện từ Hugging Face cho các mô hình ngôn ngữ.
* numpy: Thư viện cho các phép toán số học.
* Flask hoặc Streamlit: Thư viện để xây dựng giao diện web.
* pandas: Thư viện để xử lý dữ liệu.
* json: Thư viện để xử lý dữ liệu JSON.
  1. Vận hành hệ thống :
* **Nhập văn bản**: Người dùng có thể nhập văn bản vào giao diện và nhấn nút phân tích để nhận kết quả.
* **Xem kết quả**: Kết quả phân tích sẽ được hiển thị trên giao diện, bao gồm điểm số cho từng loại cảm xúc và cảm xúc chính.
* **Lưu và xuất báo cáo**: Người dùng có thể lưu lại lịch sử phân tích và xuất báo cáo theo yêu cầu.
  1. Các giao diện chính:

A screenshot of a computer

AI-generated content may be incorrect.

Hình 1: Giao diện Phân tích cảm xúc

A screenshot of a video

AI-generated content may be incorrect.

Hình 2 : Giao diện Huấn luyện mô hình

A blue line on a black background

AI-generated content may be incorrect.

Hình 3: Giao diện Thống kê và báo cáo

# **KẾT LUẬN VÀ ĐỀ NGHỊ**

1. **Kết luận**

* Dự án "Phân tích Cảm xúc Tiếng Việt với BERTweet" đã thành công trong việc phát triển một hệ thống phân tích cảm xúc hiệu quả cho văn bản tiếng Việt. Hệ thống sử dụng mô hình BERTweet, một biến thể của BERT được tối ưu hóa cho ngôn ngữ tiếng Việt, cho phép phân tích cảm xúc với độ chính xác cao. Qua quá trình phát triển, chúng tôi đã xây dựng được một giao diện người dùng thân thiện, cho phép người dùng dễ dàng nhập văn bản và nhận kết quả phân tích một cách nhanh chóng và trực quan.
* Hệ thống không chỉ cung cấp các điểm số cho từng loại cảm xúc (tích cực, tiêu cực, trung lập) mà còn cho phép lưu lại lịch sử phân tích và xuất báo cáo dưới dạng file DOCX. Điều này giúp người dùng có thể theo dõi và chia sẻ kết quả phân tích một cách dễ dàng.

1. **Đề nghị**

* Được tiếp tục bổ sung, phát triển và hoàn thiện các chức năng còn còn hạn chế.
* Tích hợp thêm tính năng: Phát triển thêm các tính năng như phân tích cảm xúc theo ngữ cảnh, phân tích cảm xúc theo thời gian, hoặc tích hợp với các nền tảng mạng xã hội để thu thập dữ liệu phân tích.
* Tối ưu hóa hiệu suất: Nghiên cứu và áp dụng các kỹ thuật tối ưu hóa để giảm thời gian phản hồi của hệ thống, đặc biệt khi xử lý các văn bản dài hoặc khi có nhiều người dùng cùng truy cập.
* Đào tạo người dùng: Tổ chức các buổi đào tạo hoặc hội thảo để hướng dẫn người dùng cách sử dụng hệ thống hiệu quả, cũng như cách hiểu và áp dụng kết quả phân tích vào thực tiễn.
* Phản hồi từ người dùng: Thiết lập một kênh để thu thập phản hồi từ người dùng về hệ thống, từ đó có thể điều chỉnh và cải tiến hệ thống theo nhu cầu thực tế.

# **TÀI LIỆU THAM KHẢO**

* Tài liệu về BERT và các mô hình ngôn ngữ:
* BERT: [Pre-training of Deep Bidirectional Transformers for Language Understanding](https://arxiv.org/abs/1810.04805) (Bài báo gốc về BERT, có thể tìm bản dịch tiếng Việt hoặc tóm tắt trên các trang học thuật).
* BERTweet: [A pre-trained language model for English Tweets](https://arxiv.org/abs/2005.10200) (Bài báo về BERTweet, có thể tìm thông tin tiếng Việt trên các diễn đàn học thuật).
* Tài liệu về Học sâu và PyTorch
* [Học sâu với PyTorch](https://docs.pytorch.org/tutorials/) (Tài liệu chính thức của PyTorch, có thể tìm thấy các hướng dẫn và ví dụ).