

**Nhóm thực hiện: nhóm 3**

**Nhóm trưởng:**

* **Nguyễn Quốc Trung – Lớp: CNTT 1 K11 – Khoa: CNTT – Mã Sv: 20200372**

**Thành viên:**

* **Hoàng Đức Vinh – Lớp: CNTT1 K11 – Khoa: CNTT – Mã Sv: 20200129**
* **Nguyễn Hồng Quân – Lớp: CNTT1 K11 – Khoa: CNTT – Mã Sv: 20200125**
* **Phạm Anh Tuấn – Lớp: CNTT1 K11 – Khoa: CNTT – Mã Sv: 20200126**
* **Nguyễn Mạnh Tiến – Lớp: CNTT1 K11 – Khoa: CNTT – Mã Sv: 20200892**

**XÂY DỰNG HỆ THỐNG**

**QUẢN LÝ BÁN VÉ TÀU HOẢ**

**BÁO CÁO ĐỒ ÁN LẦN 2**

**ĐỒ ÁN CHUYÊN NGÀNH LẦN 2 CỦA SINH VIÊN**

**Năm học: 2022 - 2023**

**Giảng viên hướng dẫn: ThS. Trần Xuân Thanh**

**HÀ NỘI 04/2023**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ ĐÔNG Á**

------------**🕮**------------

**Giảng viên hướng dẫn: ThS. Trần Xuân Thanh**

**HÀ NỘI 04/2023**

**Nhóm thực hiện: nhóm 3**

**Nhóm trưởng:**

* **Nguyễn Quốc Trung – Lớp: CNTT 1 K11 – Khoa: CNTT – Mã Sv: 20200372**

**Thành viên:**

* **Hoàng Đức Vinh – Lớp: CNTT1 K11 – Khoa: CNTT – Mã Sv: 20200129**
* **Nguyễn Hồng Quân – Lớp: CNTT1 K11 – Khoa: CNTT – Mã Sv: 20200125**
* **Phạm Anh Tuấn – Lớp: CNTT1 K11 – Khoa: CNTT – Mã Sv: 20200126**
* **Nguyễn Mạnh Tiến – Lớp: CNTT1 K11 – Khoa: CNTT – Mã Sv: 20200892**

**XÂY DỰNG HỆ THỐNG**

**QUẢN LÝ BÁN VÉ TÀU HOẢ**

**BÁO CÁO ĐỒ ÁN LẦN 2**

**ĐỒ ÁN CHUYÊN NGÀNH LẦN CỦA SINH VIÊN**

**Năm 2022 - 2023**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ ĐÔNG Á**

------------**🕮**------------

**Mục lục**

[Lời nói đầu 1](#_Toc132057274)

[Chương I: Giới thiệu vấn đề nghiên cứu 2](#_Toc132057275)

[1.1. Tổng quan thực trạng nghiên cứu 2](#_Toc132057276)

[1.2. Bối cảnh nghiên cứu 2](#_Toc132057277)

[1.3. Mục tiêu, nội dung nghiên cứu 3](#_Toc132057278)

[1.4. Phương pháp nghiên cứu 3](#_Toc132057279)

[Chương II: Tổng quan tình hình nghiên cứu 5](#_Toc132057280)

[2.1. Các khái niệm cơ bản 5](#_Toc132057281)

[2.2. Thực trạng vấn đề nghiên cứu 5](#_Toc132057282)

[2.3. Các nghiên cứu của các nhà khoa học trên thế giới 6](#_Toc132057283)

[2.4. Các nghiên cứu đã được áp dụng ở Việt Nam 6](#_Toc132057284)

[Chương III: Phương pháp nghiên cứu 8](#_Toc132057285)

[**3.1.** **Phát triển giả thiết nghiên cứu** 8](#_Toc132057286)

[**3.2.** **Xây dựng mô hình** 8](#_Toc132057287)

[Chương IV: Kết quả và đánh giá 13](#_Toc132057288)

[4.1. Báo cáo kết quả 13](#_Toc132057289)

[4.2. Đánh giá nhận xét 14](#_Toc132057290)

[Chương 5: Kết luận và đánh giá 15](#_Toc132057291)

[5.1. Kết luận 15](#_Toc132057292)

[5.2. Hướng phát triển của đề tài 15](#_Toc132057293)

[Chương 6: Tài liệu tham khảo 16](#_Toc132057294)

**Danh mục các từ viết tắt**

|  |  |  |
| --- | --- | --- |
| **Từ viết tắt (Tiếng anh)** | **Nghĩa tiếng anh** | **Nghĩa tiếng việt** |
| AI | Advanced Encryption Standard | Trí tuệ nhân tạo |
| ML | Machine Learning | Học máy |
| DL | Deep Learning | Học sâu |
| Q&A | Question and Answer | Câu hỏi và trả lời |
| NLP | Natural Language Processin | Xử lý ngôn ngữ tự nhiên |

# **Lời nói đầu**

Trong những năm gần đây, Công nghệ thông tin và Internet đã được con người chú trọng phát triển với những dấu mốc quan trọng. Trải qua các cuộc cách mạng công nghiệp lớn, đặc biệt là cuộc cách mạng 4.0, nhu cầu của con người trong việc sử dụng mạng Internet đã tăng mạnh. Các dịch vụ và ứng dụng trên Internet cũng ngày càng phát triển kéo theo đó yêu cầu bảo mật thông tin trên mạng thông tin cũng được ưu tiên phát triển.

Trí tuệ nhân tạo, đặc biệt là xử lý ngôn ngữ tự nhiên hiện nay là một vấn đề đang được rất nhiều nước trên thế giới quan tâm, như gần đây nhất là có xu hướng là công cụ ChatGpt được quan tâm và ứng dụng trong rất nhiều lĩnh vực, tình huống; từ việc giải quyết các vấn đề đơn giản tới những việc yêu cầu phức tạp.

Bên cạnh thực tế là ChatGPT nói riêng và những sản phẩm của AI đang ngày càng trở thành hiện tượng toàn cầu thì trong bài Nghiên cứu khoa học dưới đây, nhóm chúng tôi muốn giới thiệu một phương pháp phân tích quan điểm, phân loại dựa quan điểm của các bình luận trên các nền tảng bình luận bài viết: “**Xây dựng hệ thống Chatbot nhằm phân tích quan điểm của bình luận tiếng anh bằng phương pháp phân tích cấu trúc cú pháp ngữ nghĩa của câu**”**.**

Do thời gian có hạn và kiến thức chuyên môn còn hạn chế nên bài nghiên cứu khoa học chưa được hoàn chỉnh. Rất mong nhận được sự góp ý, bổ sung và sửa đổi của thầy, cô giáo và các bạn giúp cho bài báo cáo được hoàn thiện hơn. Mọi ý kiến bổ sung, góp ý xin gửi về địa chỉ email: **maiyonaisu1102@gmail.com** hoặc số điện thoại **0822206919.**

# **Chương I: Giới thiệu vấn đề nghiên cứu**

* 1. **Tổng quan thực trạng nghiên cứu**

Với thực trạng là ngành công nghệ trí tuệ nhân tạo đang ngày càng trở thành những phần mà mọi người trên thế giới đang ngày càng quan tâm và dõi theo, năm 2023 là một năm đột phá với nhiều thành quả của trí tuệ nhân tạo, hơn nữa là các bài viết bài đánh giá về các công nghệ mới đang ngày một phổ biến.

Xử lý ngôn ngữ tự nhiên là một nhánh của Trí tuệ nhân tạo, tập trung vào việc nghiên cứu sự tương tác giữa máy tính và ngôn ngữ tự nhiên của con người, dưới dạng tiếng nói hoặc văn bản. Mục tiêu của lĩnh vực này là giúp máy tính hiểu và thực hiện hiệu quả những nhiệm vụ liên quan đến ngôn ngữ của con người như: tương tác giữa người và máy, cải thiện hiệu quả giao tiếp giữa con người với con người, hoặc đơn giản là nâng cao hiệu quả xử lý văn bản và lời nói.

* 1. **Bối cảnh nghiên cứu**

Ngày nay, nhiều người theo dõi các thông tin từ những phương tiện truyền thông để đánh giá xem có nên mua sản phẩm hoặc dịch vụ nào đó hay không. Việc phát hiện các bình luận rác rất khó vì các từ ngữ lừa đảo có thể được viết bởi các tổ chức, cá nhân với nhiều mioục đích khác nhau. Họ viết những đánh giá gây hiểu lầm này để đánh lừa người đọc lhoặc hệ thống nhận dạng tự động quảng bá sản phẩm của họ hoặc đánh giá thấp sản phẩm của đối thủ cạnh tranh. Trong nghiên cứu này, chúng tôi đề xuất một cách tiếp cận, đó là sử dụng Chatbot làm cơ sở tri thức để giải bài toán nhìn nhận và đánh giá ý kiến.

Nghiên cứu Chatbot đã tiến bộ đáng kể trong những năm qua. Nhiều doanh nghiệp đang nghiên cứu cách cải thiện hiệu suất, việc áp dụng và triển khai các công cụ này để giao tiếp với khách hàng hoặc các nhóm nội bộ thông qua phương tiện truyền thông xã hội. Bên cạnh đó, các doanh nghiệp cũng mong muốn quan tâm đến đánh giá chất lượng của khách hàng qua mạng xã hội về sản phẩm hiện có trên thị trường. Từ đó, họ lựa chọn một phương pháp mới để nâng cao chất lượng dịch vụ của mình. sản phẩm rồi gửi cho các cơ quan xuất bản để xuất bản dựa trên nhu cầu và đánh giá của xã hội. Mặc dù đã có nhiều nghiên cứu gần đây nhưng không phải tất cả chúng đều đề cập đến vấn đề đánh giá ý kiến trên hệ thống chatbot. Mục tiêu chính của nghiên cứu của bài viết này là để đánh giá bình luận của con người bằng tiếng Anh thông qua hệ thống chatbot. Các tài liệu của hệ thống được xử lý trước và quan điểm phù hợp để đưa ra các đánh giá quan điểm dựa trên các bình luận bằng tiếng Anh.

Dựa trên thực tế nhu cầu và điều kiện xã hội, phương pháp này nhằm mục đích phát triển nội dung chatbot dựa trên tương tác của người dùng các hành động, cho phép thực hiện quy trình theo chu kỳ và do con người giám sát với các bước sau để đánh giá bình luận bằng tiếng Anh. Đầu tiên, chúng tôi xử lý trước dữ liệu đầu vào bằng cách thu thập các bình luận trên mạng xã hội, và sau đó hệ thống của chúng tôi phân tích các nhận xét đó theo lượt xem xếp hạng cho từng chủ đề được đề cập. Cuối cùng, hệ thống của chúng tôi sẽ đưa ra kết quả xếp hạng và bình luận cho mỗi bình luận được nhập vào hệ thống.

* 1. **Mục tiêu, nội dung nghiên cứu**

Ngày nay, nhiều người theo dõi các thông tin từ những phương tiện truyền thông để đánh giá xem có nên mua sản phẩm hoặc dịch vụ nào đó hay không. Việc phát hiện các bình luận rác rất khó vì các từ ngữ lừa đảo có thể được viết bởi các tổ chức, cá nhân với nhiều mục đích khác nhau. Họ viết những đánh giá gây hiểu lầm này để đánh lừa người đọc hoặc hệ thống nhận dạng tự động quảng bá sản phẩm của họ hoặc đánh giá thấp sản phẩm của đối thủ cạnh tranh. Trong nghiên cứu này, chúng tôi đề xuất một cách tiếp cận, đó là sử dụng Chatbot làm cơ sở tri thức để giải bài toán nhìn nhận và đánh giá ý kiến.

Bài nghiên cứu khoa học này nhằm mục đích đề xuất một thuật toán phân tích, sử dụng phương pháp phân tích cú pháp ngữ nghĩa của từng thành phần câu, nội dung văn bản từ đó đưa ra những kết luận về quan điểm của văn bản đầu vào. Qua đó có thể ứng dụng được công cụ vào các sản phẩm của doanh nghiệp có nhu cầu phân tích tập dữ liệu khách hàng.

* 1. **Phương pháp nghiên cứu**

Trong bài nghiên cứu khoa học này, nhóm sẽ tìm hiểu khái niệm cơ bản, các hàm toán học được sử dụng và từng bước thuật toán được áp dụng để phân tích. Sau đó nhóm sẽ minh hoạ dựa trên ngôn ngữ Python để mô phong quá trình từ nhận vào dữ liệu sau đó phân tích ngữ nghĩa của văn bản nhằm đưa ra kết luận phân tích phù hợp.

# **Chương II: Tổng quan tình hình nghiên cứu**

* 1. **Các khái niệm cơ bản**

Trí tuệ nhân tạo và xử lý ngôn ngữ tự nhiên (NLP) được tích hợp với các thuật toán học máy đóng vai trò quan trọng vai trò quan trọng trong công nghệ và ứng dụng ngày nay. Chatbot sẽ tích lũy kinh nghiệm bằng cách học hỏi từ những kinh nghiệm trước đó và sử dụng các thuật toán khác nhau. Dữ liệu có thể được đào tạo cho chatbot, cho phép nó kiểm tra với cơ sở tri thức và đưa ra câu trả lời chính xác truy vấn của người dùng thông qua các ứng dụng phía máy khách.

* 1. **Thực trạng vấn đề nghiên cứu**

Thởi điểm hiện tại, chatbot đang được mở rộng sang hầu hết mọi lĩnh vực, chẳng hạn như, một chatbot giáo dục dành cho nền tảng Facebook Messenger. Tương tự, trong một nghiên cứu, phần mềm cũng sẽ đặt câu hỏi dựa trên câu trả lời trước đó của ứng viên, sử dụng một mô hình Xử lý ngôn ngữ tự nhiên (NLP). Sau cuộc phỏng vấn, phần mềm sẽ phân tích dữ liệu thu thập được để xác định ứng cử viên tốt nhất cho vị trí được cung cấp. Kết quả là, dự án Chatbot JARO nhằm mục đích đơn giản hóa quy trình tuyển dụng.

Mặt khác, một ứng dụng trò chuyện được xây dựng dựa trên công nghệ Chatbot, ứng dụng Chatbot y tế tự động giúp giảm chi phí chăm sóc sức khỏe trong khi cải thiện khả năng tiếp cận với kiến thức và dịch vụ y tế. Phương pháp này đã tạo ra một bot chẩn đoán trò chuyện với bệnh nhân về các câu hỏi và vấn đề y tế của họ để cung cấp một chẩn đoán cá nhân hóa dựa trên biểu hiện và hồ sơ được chẩn đoán của họ. Một phương pháp thực tế để xác định khả năng khám phá và các tính năng như ngôn ngữ, chủ đề và nền tảng của nhà phát triển. Hoặc đối với bệnh nhân, một khung hoạt động như một trợ lý ảo được tạo bằng thuật toán học máy. Nó có thể dự đoán các triệu chứng, giới thiệu bác sĩ và điều tra các phương pháp điều trị cho bệnh nhân bằng cách tương tác với họ – chăm sóc sức khỏe bệnh nhân hiệu quả.

Trường hợp khác nữa, hệ thống tạo chatbot phục vụ mục đích học thuật sử dụng NLP và ML mà các tổ chức giáo dục khác nhau có thể sử dụng. Có hai chế độ có sẵn: âm thanh và văn bản. Thay vì đặt cuộc trò chuyện vào d anh sách chờ, người dùng có thể tương tác với bot. cùng một câu hỏi được hỏi dưới nhiều hình thức khác nhau để kiểm tra độ chính xác. Do đó, kế hoạch là kết hợp phân loại mục đích và xử lý ngôn ngữ tự nhiên để tạo ra một tương tác giao diện người dùng và một chatbot. Mô hình nhằm mục đích nhận ra các truy vấn của người dùng và tạo truy vấn SPARQL.

Kỹ thuật học sâu (DL – Deep Learning) đã được sử dụng để phát triển trợ lý bài giảng video trực tuyến giúp cải thiện chất lượng dữ liệu Q&A bằng cách kết hợp nhiều chatbot từ nhiều góc độ khác nhau cho một video.

* 1. **Các nghiên cứu của các nhà khoa học trên thế giới**

Nghiên cứu về cấu trúc của một chatbot sử dụng trí tuệ nhân tạo (AI) nhằm mục đích đánh giá, chẩn đoán và đề xuất các biện pháp phòng ngừa và an toàn ngay lập tức các biện pháp đối với những bệnh nhân đã tiếp xúc với nCOV-19 và hoạt động như một ảo trợ lý để hỗ trợ đo lường mức độ nghiêm trọng của nhiễm trùng thông qua triệu chứng phân tích và kết nối với các cơ sở y tế có thẩm quyền khi nó di chuyển vào giai đoạn nặng.

Ngoài ra, một số nghiên cứu sử dụng Xử lý ngôn ngữ tự nhiên (NLP) và các kỹ thuật Deep Learning (DL) để phát triển một chatbot có thể tương tác trong các cuộc trò chuyện tương tác với khách tham quan trong ngày khai trương MPU, hoặc chatbot dựa trên AI để thu hút khách hàng.

Các nhà nghiên cứu ở trên thế giới thường đề cập tới việc dùng Chatbot nhà hàng, Chatbot sử dụng AI trong thị trường dịch vụ chăm sóc sức khỏe và AI đàm thoại trong nhà hàng, y tế hoặc giao tiếp các ngành nghề. Hơn nữa, trong triển lãm nghệ thuật, một nhóm nhà nghiên cứu đã trình bày một Giải pháp dựa trên NLP cho các tác nhân đàm thoại. Nhìn chung, các tác giả đã đề cập đến rất nhiều từ, tất cả đều đã tạo ra kết quả đáng kể, nhưng sử dụng NLP để phân tích và đánh giá ý kiến vẫn chưa được đề cập, và phân tích đánh giá vẫn cần cải thiện về mặt đánh giá nhận xét. Với những lợi ích rõ ràng của mình, Chatbot ngày càng trở nên phổ biến.

Với sự ra đời của công nghệ 4.0, con người đã tích hợp nhiều hơn vào thế giới ảo và thế giới thực. Do nhu cầu của xã hội ngày càng cao, đòi hỏi một lượng lớn thao tác, nhưng chọn một số lượng lớn các giá trị nhận xét cho bản thân là rất cần thiết.

* 1. **Các nghiên cứu đã được áp dụng ở Việt Nam**

Cũng có một số nhà nghiên cứu ở Việt Nam đề xuất một cách tiếp cận khác là xây dựng mô hình Chatbot cho quản lý bình luận tiếng Việt; tác giả cũng đã đề cập đến một số thay đổi đề xuất thuật toán và cải tiến đáng kể. Tuy nhiên, phương pháp này chỉ giới hạn trong phạm vi nghiên cứu này – nghiên cứu ngôn ngữ tiếng Việt và nó cũng có rất nhiều hạn chế.

**Chương III: Phương pháp nghiên cứu**

* 1. **Phát triển giả thiết nghiên cứu**

Phương pháp của chúng tôi dựa trên sự phân tích các đánh giá và ý kiến cá nhân về một hoặc nhiều sản phẩm được mở rộng trên mạng xã hội. Hơn nữa, một báo cáo chỉ ra rằng thiết kế chatbot chủ yếu là dựa trên thiết bị học quy tắc. Có ba bước để thực thi điều này chatbot. Bản ghi thô được xử lý trước khi bắt đầu. Kết quả xuất ra một bộ dữ liệu đã được xử lý. Sau đó quá trình phân tách xảy ra trong bước thứ hai.

Một số các kỹ thuật khác có thể được sử dụng, chẳng hạn như một số nhóm nghiên cứu sử dụng một trang web học thuật hỗ trợ việc tính toán, làm thế nào hệ thống nhanh chóng tính toán từ bắt nạt hoặc không còn. Còn nhiều kỹ thuật khác có thể được sử dụng với hệ thống Chatbot, bao gồm làm chủ thiết bị, làm mờ phán đoán, đối sánh mẫu và đánh giá tình cảm.

Gần đây, theo điều tra, có 74 bài báo gợi ý chatbot. Các nghiên cứu chủ yếu tập trung vào ứng dụng, phương pháp luận (phương pháp được sử dụng, cỡ mẫu, loại mẫu và nghiên cứu quốc gia) và thư mục (cơ quan xuất bản, trích dẫn và tiêu điểm).

Mục tiêu chính là tiến hành một xem xét có hệ thống các bài báo nghiên cứu tạp chí chất lượng cao để tóm tắt hiện trạng nghiên cứu về chatbot để xác định vai trò của chúng trong kinh doanh kỹ thuật số sự biến đổi. Ngược lại, chúng tôi đã tiến hành nghiên cứu sâu để đánh giá quan điểm của các bài bình luận tiếng Anh, từ đó đánh giá và đưa ra những ý kiến hữu ích thông tin cho doanh nghiệp khi quảng bá sản phẩm trên mạng xã hội.

Nhìn chung, với kỹ thuật phân tích cú pháp câu và từ tiếng Anh mà nhóm chúng tôi áp dụng vào, chúng tôi đã cải thiện độ chính xác lên tới 78,53%.

* 1. **Xây dựng mô hình**

Trong phần này, chúng tôi tiến hành Xử lý ngôn ngữ tự nhiên (NLP) như sau:

* Thứ nhất: Ta phân tích câu và chia câu nhận xét thành từ
* Thứ hai: Chúng tôi sử dụng thẻ POS Penn Treebank được chia thành ba loại: tính từ, danh từ và động từ
* Thứ ba: Việc sử dụng SentiWordNet có thể được tinh chỉnh bằng cách sử dụng các thẻ này để nhìn vào ý nghĩa của các từ phù hợp với thẻ POS của họ.
* Cuối cùng, để đánh giá một bình luận, chúng tôi cũng tiến hành phân tích ngữ pháp. Từ đó, chúng tôi phân tích và dựa trên những từ thu thập được, xem xét ý kiến ​​của người khác trực tuyến trong xã hội để kết luận.

Chúng tôi phân tích và viết lại câu dựa trên các điều kiện được liệt kê như trong hình dưới đây.

Table

Description automatically generated

Hình 2.1.1 – Ký hiệu POS

Hệ thống của chúng tôi được thiết kế để sử dụng nhiều bình luận nhằm đánh giá chất lượng bình luận để phân tích và đánh giá tốt hơn bằng cách sử dụng các bình luận đã nhập vào hệ thống. Hơn nữa, hệ thống của chúng tôi cung cấp phương pháp đánh giá để chọn lọc những ý kiến có giá trị cho doanh nghiệp, từ đó mang đến những sản phẩm chất lượng cho người dùng. Ở đây có thể thấy rằng chúng ta sẽ viết lại câu đã phân tích cú pháp ở dạng TreeBank.

Diagram

Description automatically generatedHình 3.2.1 – TreeBank

Trên cơ sở đó, chúng tôi xây dựng một thuật toán để đánh giá các bình luận trên mạng xã hội để tìm ra giải pháp hợp lý cho phù hợp. Đánh giá các bình luận là một thách thức lớn đối với Chatbots. Trong phạm vi nghiên cứu của bài báo này, hệ thống của chúng tôi đánh giá dựa trên cơ sở phân tích cú pháp của câu sau đó đánh giá nội dung và cuối cùng là kết luận đánh giá của người bình luận. Tuy nhiên, vấn đề gặp phải là có nhiều ý kiến cần chỉnh sửa; ví dụ như bình luận ngắn, viết tắt, sai ngữ pháp,… Căn cứ vào nội dung này, chúng tôi đánh giá các bình luận để kết luận là tiêu cực hay tích cực thông qua việc đào tạo trước đó.

Về mặt tự nhiên, việc đánh giá dựa trên thuật toán được đề xuất để xác định tích cực hay tiêu cực. Các thuật toán dựa trên một tập dữ liệu đào tạo để đưa ra kết luận nhận xét nó là bình luận mang tính chất tích cực hay tiêu cực.

Mô hình đề xuất của chúng tôi. Trong phần này, chúng tôi đề xuất một cấu trúc hệ thống, trong đó hệ thống được chia thành nhiều phần chính, bao gồm Ontology và giai đoạn tiền xử lý.

Diagram

Description automatically generatedHình 3.2.2 – Mô hình hệ thống đề xuất.

Mô-đun này chịu trách nhiệm truy xuất cơ sở tri thức từ Ontology và trích xuất tất cả các thực thể từ nhánh tương ứng của sản phẩm. Đầu tiên, đầu vào của hệ thống là một câu, có nhiệm vụ nhận một câu để xử lý và phân tích xem câu đó là tích cực hay tiêu cực. Tiếp theo, dữ liệu đầu vào sẽ chuyển qua một Ontology dùng để tạo thực thể (Module xây dựng thực thể), hỗ trợ hệ thống đánh giá các từ mà hệ thống đã xử lý trước đó. Đây là nền tảng cho quá trình tìm kiếm. Ontology tìm kiếm và kết hợp.

Các mô-đun tiền xử lý sẽ nhận dạng các thực thể này, tính toán và lưu kết quả xác suất xác định tính tích cực hoặc tiêu cực của thực thể để hỗ trợ khớp các thực thể trong câu.

Giai đoạn tiền xử lý, mô-đun tiền xử lý phân tích nội dung và tiêu đề của đánh giá và tạo ra dữ liệu theo yêu cầu của mô hình phân loại. Đầu vào cần thiết cho mô hình là nội dung của đánh giá. Do đó, việc chuẩn hóa phải được hoàn thành trước khi tiến hành các bước xử lý khác để tạo nguồn dữ liệu chuẩn và tránh phân tích lỗi.

Có rất nhiều phương pháp để trích xuất các từ từ một văn bản. Chúng tôi đã chọn cách mô hình n-gram của unigram, kết hợp với mô hình gắn thẻ POS, cho nghiên cứu này. Công cụ gắn thẻ POS của Đại học Stanford (Stanford POS Tagger) có cơ sở dữ liệu tương đối lớn và đã được sử dụng rộng rãi trong nghiên cứu xử lý ngôn ngữ. Chúng tôi chọn công cụ này cho mô-đun tách từ vì độ chính xác và hiệu suất xử lý cao của nó. Để làm điều này, chúng tôi đã thực hiện mô phỏng sau:

Đầu tiên, bộ dữ liệu ở đây bao gồm các Bình luận được thu thập từ các nguồn trên Internet như diễn đàn, mạng xã hội, website,… Sau đó, thông qua Cấu trúc ngữ pháp, gán nhãn POS cho từng thành phần của câu thông qua mô hình gán nhãn POS (máy gắn nhãn POS của Đại học Stanford có cơ sở dữ liệu sẵn có khá lớn và đã được ứng dụng rộng rãi trong các khóa học nghiên cứu về xử lý ngôn ngữ tự nhiên). Chúng tôi chọn công cụ này trong xử lý phân từ trong câu và phân tích cấu trúc các thành phần trong câu vì nó có độ chính xác tương đối cao và đã được ứng dụng nhiều nơi trong việc phân tích tính chất xử lý ngôn ngữ.

Thứ hai, việc xử lý các câu sau khi chúng đã được tách thành các từ sẽ khớp với WordLists và Classes được sử dụng để phân loại loại nào câu thuộc câu. Word là một danh sách chứa POS-tagging cho mỗi từ trong câu, và classes là một danh sách chứa kiểu phân loại của câu ở đây là khẳng định. Hệ thống của chúng tôi chia thành năm loại câu: câu đơn, câu điều kiện, câu so sánh hơn, câu ghép và câu đặc biệt. Tùy về loại câu, hệ thống có các cách xử lý từ ngữ khác nhau trong câu để có kết quả chính xác. Cụ thể, sau khi được phân hủy thành từ và được gắn nhãn thông qua mô hình POS, câu sẽ được sắp xếp thành một mảng để duyệt từng từ trong câu, từ đó hệ thống sẽ đánh giá cho dù mỗi từ là trong đó. Khả năng từ đó có một ký tự như vậy là bao nhiêu phụ thuộc vào từ đó là tích cực hay tiêu cực. Sau khi đánh giá từng từ trong câu, hệ thống sẽ phân loại nhận xét tích cực hoặc tiêu cực.

Kết quả: Tính năng cuối cùng sẽ quyết định bình luận thuộc nhóm nào thuộc về, tích cực, tiêu cực, hoặc trung tính. Mô hình phân loại có năm tính năng phân loại cơ bản loại câu gồm câu đơn, câu điều kiện, câu so sánh, câu ghép, câu đặc biệt. Ngoài ra, đối với mỗi loại trên, trong câu đều có một phủ định. Một mặt, trong hệ thống chatbot, có một phần quan trọng không kém một phần của đào tạo học máy.

Text lemmatizati là một quá trình viết thường từng từ và loại bỏ từ trùng lặp bình luận. Bổ đề hóa là đưa các từ về định dạng ban đầu của chúng bằng cách sử dụng bảng tra cứu từ vựng thực tế của các từ trong câu. Có thể có những quy tắc xử lý bình luận mà chúng ta chưa từng thấy trước đây.

Tạo mô hình đào tạo. Bước 01: Lớp 1, 128 nơ ron, bỏ học (0,01). Lớp 2, 64 tế bào thần kinh, tỷ lệ bỏ tiếp nhận học (0,01). Bước 02: Biên dịch mạng với các tham số để huấn luyện mạng để đánh giá sự tối ưu. Bước 03: Tỷ lệ học tập = 0,001.

# **Chương IV: Kết quả và đánh giá**

* 1. **Báo cáo kết quả**

Cài đặt thử nghiệm. Trong phần này, chúng ta tiến hành cài đặt và Bài kiểm tra. Thử nghiệm được viết bằng Python. Dữ liệu chúng tôi thu thập là các sản phẩm, bao gồm máy tính (531 câu), bộ định tuyến (879 câu) và Loa (689 câu) để kiểm tra.

Table

Description automatically generatedKhi làm việc với dữ liệu văn bản, chúng ta phải thực hiện nhiều bước tiền xử lý khác nhau trên dữ liệu trước khi xây dựng mô hình học máy hoặc học sâu. Chúng ta phải áp dụng các hoạt động khác nhau để xử lý trước dữ liệu dựa trên các yêu cầu. Tokenization là điều cơ bản nhất và đầu tiên bạn có thể làm với dữ liệu văn bản. Tokenizing đang chia một văn bản lớn thành các phần nhỏ, chẳng hạn như các từ. Chúng tôi lặp lại các mẫu, mã hóa câu bằng hàm nltk.word tokenized() và nối thêm từng từ trong danh sách từ. Chúng tôi cũng tạo một danh sách lớp cho các thẻ của chúng tôi. Bây giờ chúng ta sẽ bổ sung từ vựng cho từng từ và loại bỏ bất kỳ từ trùng lặp nào khỏi danh sách.

Hình 4.1.1 – Kết quả thử nghiệm

* 1. **Đánh giá nhận xét**

**Table

Description automatically generated**

Hình 4.2.1 – Bảng so sánh phương pháp đề xuất với các phương pháp khác

Kết quả cho thấy, phương pháp đề xuất luôn chiếm tỷ lệ cao hơn so với các phương pháp còn lại khi xét trên 736 câu. Phương pháp ACCLE cho kết quả thấp khi cho kết quả trung bình 50,27%, trong khi phương pháp ITEAI cho kết quả từ 76,22% trở lên. Phương pháp đề xuất đạt mức thấp nhất là 78,53%

**Chương 5: Kết luận và đánh giá**

**5.1. Kết luận**

Chúng tôi đã xây dựng một mô hình Chatbot để giải quyết một số vấn đề đơn giản các câu, chẳng hạn như câu đơn và câu so sánh với câu điều kiện và câu ghép, đáng tin cậy, nhưng đối với câu đáng nhớ vì cần có thêm thời gian để giải quyết vấn đề.

Bài báo đã xây dựng mô hình đánh giá tự động khai phá quan điểm qua hệ thống Chatbot. Khái niệm này được phát triển để đáp ứng các vấn đề hiện tại mà các doanh nghiệp phải đối mặt khi mạng xã hội phát triển, nhưng các giá trị chất lượng vẫn còn hạn chế. Hàng loạt rà soát tài liệu để đảm bảo tính thống nhất trong mọi công việc và chatbot được xác định là mô hình tốt nhất đáp ứng yêu cầu.

Nghiên cứu về Chatbot thu hút các kết nối để tìm hiểu thêm về các công nghệ nhất thời mới nổi và các thuật toán tương thích như Trí tuệ nhân tạo, Học máy, Python và Xử lý ngôn ngữ tự nhiên (NLP). Kết quả cho thấy phương pháp chúng tôi đề xuất đạt tới 78,53%.

* 1. **Hướng phát triển của đề tài**

Chúng tôi nghiên cứu về đề tài này với mong muốn là đề xuất một mô hình mới để tối ưu hơn trong xử lý ngôn ngữ tự nhiên, cụ thể là giúp xử lý phân loại các bình luận và có thể được ứng dụng vào các doanh nghiệp trong quá trình xử lý thông tin dựa vào việc thu thập dữ liệu người dùng trong thực tế.

Và chúng tôi cũng mong muốn là các nhà nghiên cứu khác sau này có thể sử dụng mô hình này và tham khảo được trong quá trình nghiên cứu, sáng tạo ra các mô hình cao cấp, tiên tiến và tối ưu hơn.

# **Chương 6: Tài liệu tham khảo**

1. Scholz T., Conrad S. Opinion mining in newspaper articles by entropy-based word connections. Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. Seattle, Washington, USA: Association for Computational Linguistics, 2013. pp. 1828–1839.

2. Feldman R. Techniques and applications for sentiment analysis. Communications of the ACM. 2013. vol. 56. pp. 82–89.

3. Mauldin M. Chatbot. Available at: https://en.wikipedia. org/wiki/Chatbot. (accessed: 10.11.2022).

4. Ganesan M., Deepika C., HarievashiniB., Krithikha A., Lokhratchana B. A survey on chatbots using artificial intelligence. International Conference on System, Computation, Automation and Networking (ICSCAN). 2020. pp. 1–5.

5. Singh S., Kaur M., Tanwar P., Sharma S. Design and development of conversational chatbot for covid-19 using nlp: an ai application. 6th International Conference on Computing Methodologies and Communication (ICCMC). 2022. pp. 1654–1658.

6. Liu F., Yang X., Wang Y. An interactive chatbot for university open day. IEEE 13th International Conference on Software Engineering and Service Science (ICSESS). 2022. pp. 214–217.

7. Rawat B., Bist A., Rahardja U., Aini Q., Ayu Sanjaya Y. Recent deep learning based nlp techniques for chatbot development: An exhaustive survey. 10th International Conference on Cyber and IT Service Management (CITSM). 2022. pp. 1–4.

8. Garg R., Riya R., Thakur S., Tyagi N., Basha K., Vij D., Sodhi G. Nlp based chatbot for multiple restaurants. 10th International Conference on System Modeling Advancement in Research Trends (SMART). 2021. pp. 439–443.

9. Wahal A., Aggarwal M., Poongodi T. Iot based chatbots using nlp and svm algorithms. 3rd International Conference on Intelligent Engineering and Management (ICIEM). 2022. pp. 484–489.

10. Adam M., Wessel M., Benlian A. AI-based chatbots in customer service and their effects on user compliance. Electronic Markets. 2021. vol. 31. no. 2. pp. 427–445.

11. Ceuca L., Rednic A., Chifu E. Safer museum guide interaction during a pandemic and further using nlp in human interactive museum visits: Museum guide chatbot. IEEE 17th International Conference on Intelligent Computer Communication and Processing (ICCP). 2021. pp. 313–318.

12. Smutny P., Schreiberova P. Chatbots for learning: A review of educational chatbots for the facebook messenger. Computers Education. 2020. vol. 151. p. 103862.

13. Purohit J., Bagwe A., Mehta R., Mangaonkar O., George E. Natural language processing based jaro-the interviewing chatbot. 3rd International Conference on Computing Methodologies and Communication (ICCMC). 2019. pp. 134–136.

14. Srivastava P., Singh N. Automatized medical chatbot (medibot). International Conference on Power Electronics IoT Applications in Renewable Energy and its Control (PARC). 2020. pp. 351–354.

15. Polignano M., Narducci F., Iovine A., Musto C., De Gemmis M., Semeraro G., Healthassistantbot: A personal health assistant for the italian language. IEEE Access. 2020. vol. 8. pp. 107479–107497.

16. Kumari S., Naikwadi Z., Akole A., Darshankar P. Enhancing college chat bot assistant with the help of richer human computer interaction and speech recognition. International Conference on Electronics and Sustainable Communication Systems (ICESC). 2020. pp. 427–433.

17. Ait-Mlouk A., Jiang L. Kbot: A knowledge graph based chatbot for natural language understanding over linked data. IEEE Access. 2020. vol. 8. pp. 149220–149230.

18. Makhkamova O., Lee K.-H., Do K., Kim D. Deep learning-based multi-chatbot broker for qa improvement of video tutoring assistant. IEEE International Conference on Big Data and Smart Computing (BigComp). 2020. pp. 221–224.

19. Nguyen Huu P., Do Manh C., Nguyen Trong H. Proposing chatbot model for managing comments in vietnam,” in Industrial Networks and Intelligent Systems. Eds. N.-S. Vo, V.-P. Hoang, Q.-T. Vien. Cham: Springer International Publishing, 2021. pp. 287–297.

20. Sinha S., Basak S., Dey Y., Mondal A. An educational chatbot for answering queries. Emerging technology in modelling and graphics. Springer, 2020. pp. 55–60.

21. Selvi V., Saranya S., Chidida K., Abarna R. Chatbot and bullyfree chat. IEEE International Conference on System, Computation, Automation and Networking (ICSCAN). 2019. pp. 1–5.

22. Miklosik A., Evans N., Qureshi A. The use of chatbots in digital business transformation: A systematic literature review. IEEE Access. 2021. vol. 9. pp. 106530–106539.

23. Daniel G., Cabot J., Deruelle L., Derras M. Xatkit: A multimodal low-code chatbot development framework,” IEEE Access, 2020. vol. 8. pp. 15332–15346.

24. Santos G., De Andrade G., Silva G., Duarte F., Costa J., De Sousa R. A conversationdriven approach for chatbot management. IEEE Access. 2022. vol. 10. pp. 8474–8486.

25. Medeiros L., Bosse T. Gerritsen C. Can a chatbot comfort humans? studying the impact of a supportive chatbot on users’ self-perceived stress. IEEE Transactions on Human Machine Systems. 2022. vol. 52. no. 3. pp. 343–353.

26. Abdellatif A., Badran K., Costa D., Shihab E. A comparison of natural language understanding platforms for chatbots in software engineering. IEEE Transactions on Software Engineering. 2022. vol. 48. no. 8. pp. 3087–3102.

27. Zhang L., Yang Y., Zhou J., Chen C., He L. Retrieval-polished response generation for chatbot,” IEEE Access. 2020. vol. 8. pp. 123882–123890.

28. Ren R., Perez-Soler S., Castro J., Dieste O., Acuna S. Using the socio chatbot for uml modeling: A second family of experiments on usability in academic settings. IEEE Access. 2022. vol. 10. pp. 130542–130562.

29. The stanford natural language processing (nlp). Available at: https://nlp.stanford.edu/software/tagger.shtml. (accessed: 02.11.2022).

30. Nicolov N., Salvetti F., Ivanova S. Sentiment analysis: Does coreference matter. AISB 2008 convention communication, interaction and social intelligence. 2008. vol. 1. p. 37.

31. Kanayama H., Nasukawa T., Watanabe H. Deeper sentiment analysis using machine translation technology. COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics. 2004. pp. 494–500.

32. Hu M., Liu B. Mining and summarizing customer reviews. Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. 2004. pp. 168–177.

33. Ko M.-C., Lin Z.-H. Cardbot: A chatbot for business card management. Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion, ser. IUI ’18 Companion. New York, NY, USA: Association for Computing Machinery, 2018. Available at: https://doi.org/10.1145/3180308.3180313. (accessed: 02.11.2022).

34. English tense system. Available at: https://www.englishclub.com/grammar/verbtensessystem.htm. (accessed: 10.11.2022).

35. Narayanan R., Liu B., Choudhary A. Sentiment analysis of conditional sentences. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing. Singapore: Association for Computational Linguistics. 2009. pp. 180–189.

36. Liu Q., Gao Z., Liu B., Zhang Y. Automated rule selection for aspect extraction in opinion mining,” in Proceedings of the 24th International Conference on Artificial Intelligence, ser. IJCAI’15. AAAI Press, 2015. p. 1291–1297.

37. Kasthuri E., Balaji S. A chatbot for changing lifestyle in education. Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV). 2021. pp. 1317–1322.

38. Dharani M., Jyostna J., Sucharitha E., Likitha R., Manne S. Interactive transport enquiry with ai chatbot. 4th International Conference on Intelligent Computing and Control Systems (ICICCS). 2020. pp. 1271–1276.