**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**TRƯỜNG ĐẠI HỌC NGOẠI NGỮ - TIN HỌC TP.HCM**

**KHOA CÔNG NGHỆ THÔNG TIN**

**KHÓA LUẬN TỐT NGHIỆP**

**TÌM HIỂU CƠ CHẾ ATTENTION**

**VÀ ÁP DỤNG CHO BÀI TOÁN**

**PHÂN TÍCH CẢM XÚC**

**GIẢNG VIÊN HƯỚNG DẪN: ThS. Trần Khải Thiện**

**SINH VIÊN THỰC HIỆN:**

**Nhan Kim Thành – 15DH110281**

**Trần Đức Khang – 15DH110240**

**TP.HỒ CHÍ MINH - 2019**

LỜI CẢM ƠN

🙞🕮🙜

Để hoàn thành khóa luận đề tài nghiên cứu khoa học này, bên cạnh sự nỗ lực của bản thân đã vận dụng và tìm tòi học hỏi, chúng em đã nhận được sự giúp đỡ, hỗ trợ từ nhiều cơ quan, tổ chức, và cá nhân. Nghiên cứu khoa học được hoàn thành dựa trên nhiều tham khảo, kinh nghiệm từ các kết quả nghiên cứu liên quan, các bài báo chuyên ngành của nhiều tác giả ở các tổ chức nghiên cứu khác.

Với tình cảm chân thành, chúng em xin dành lời cảm ơn chân thành đối với các giáo viên trong trường Đại học Ngoại Ngữ Tin Học Thành phố Hồ Chí Minh, đặc biệt là giáo viên hướng dẫn trực tiếp của nhóm nghiên cứu – Thạc Sĩ Trần Khải Thiện. Cảm ơn thầy đã dành nhiều thời gian cũng như công sức và truyền đạt nhiều kiến thức trong việc theo dõi và hướng dẫn nhóm hoàn thành đề tài nghiên cứu.

Do vẫn còn giới hạn về kiến thức, thời gian và khả năng luận lý của nhóm nên không thể tránh khỏi nhiều thiếu sót, kính mong sự chỉ dẫn của các giáo viên cũng như đóng góp ý kiến từ những người quan tâm đề tài của nhóm, để có thể tiếp thu và cải thiện hơn.

Xin chân thành cảm ơn!

MỤC LỤC

[Chương 1. NỘI DUNG 1](#_Toc10069488)

[Chương 2. BÀI TOÁN SENTIMENT ANALYSIS 2](#_Toc10069489)

[Chương 3. CÁC HƯỚNG TIẾP CẬN 3](#_Toc10069490)

[Chương 4. THỰC NGHIỆM 4](#_Toc10069491)

[Chương 5. ĐÁNH GIÁ 5](#_Toc10069492)

[KẾT LUẬN 7](#_Toc10069493)

[TÀI LIỆU THAM KHẢO 8](#_Toc10069494)

**DANH SÁCH BẢNG**

(Nếu có)

**DANH SÁCH HÌNH**

(Nếu có)

# GIỚI THIỆU CHUNG

1. **Giới thiệu tổng quan về trí tuệ nhân tạo**

Hiện nay, mọi thứ đang trên đà hiện đại hóa, mọi vấn đề đều được xử lý bởi công nghệ nói chung, trong đó nổi trội nhất là mảng trí tuệ nhân tạo nói riêng. Trí tuệ nhân tạo, gọi là AI (Artificial Intelligence) ra đời nhằm thay thế con người trong việc xử lý các vấn đề một cách tự động, hiệu quả cao và chính xác hơn.

Thuật ngữ trí tuệ nhân tạo được sử dụng bởi một nhóm các nhà khóa học máy tính tại hội nghị Dartmouth vào năm 1956, những thập kỷ sau đó, lĩnh vực này đã có những bước phát triển mạnh mẽ vượt bật, và được dự đoán sẽ là chìa khóa mở ra một tương lai mới đối với nền văn minh nhân loại.

Trí tuệ nhân tạo hiện đang được áp dụng trong các mọi lĩnh vực như sinh học, khoa học, giảng dạy, an ninh,… Những ông lớn trong ngành công nghệ cũng áp dụng chúng vào ứng dụng để cải thiện và quản lý. Ví dụ như khả năng nhận dạng khuôn mặt của Facebook hay trình phân loại hình ảnh của Pinterest,…

Vậy trí tuệ nhân tạo bắt đầu từ đâu? Trí tuệ nhân tạo được đào tạo từ một lượng lớn dữ liệu và thực hiện việc học hỏi, đưa ra quyết định hoặc dự đoán về vấn đề liên quan bằng các thuật toán. Việc đào tạo này gọi là học máy (Machine Learning). Học máy cung cấp những kỹ thuật, phương pháp để áp dụng vào thực hiện các khả năng của máy móc, một trong các lĩnh vực ứng dụng tốt nhất cho máy học trong nhiều năm qua là *thị giác máy tính* (computer vision), *khả năng phát hiện hình ảnh* (image detection) hay *xử lý ngôn ngữ tự nhiên* (natural language processing).

Phương pháp tiếp cận tiếp theo là khả năng khiến AI tự đào tạo bản thân, phương pháp này được gọi là Học sâu (Deep Learning). Deep learning tập trung xử lý các vấn đề liên quan đến mạng thần kinh, được dựa trên mạng thần kinh sinh học của con người, nhằm nâng cao trí tuệ nhân tạo trong các bài toán về nhận diện giọng nói, khuôn mặt hay xử lý ngôn ngữ tự nhiên. Trong vài năm qua, học sâu đã trở thành lĩnh vực nổi nhất trong khoa học máy tính, như khả năng dịch tự dộng, nhận thức sự vật, nhận diện giọng nói, phân tích cảm xúc...

1. **Giới thiệu về bài toán Phân tích cảm xúc:**

Bài toán về xử lý văn bản (text classification) xuất hiện khá nhiều trong các ứng dụng về xử lý ngôn ngữ tự nhiên (NLP – Natural Language Processing) được sử dụng nhiều trong kinh doanh và được thực hiện bằng các kỹ thuật học máy, học sâu. Mục đích chính của việc phân loại văn bản được chia thành nhiều mục như:

* Tìm hiểu về cảm xúc của người dùng thông qua các trang mạng xã hội.
* Xác định và phân loại các loại thư rác (spam email).
* Phân loại các bài báo thành các chủ đề xác định.
* Ứng dụng vào tạo chatbot.
* Tóm tắt văn bản.
* Dịch đoạn văn.

Về đề tài này, chúng ta sẽ tìm hiểu sâu bài toán xử lý văn bản trong vấn đề phân loại cảm xúc người dùng (sentiment analysis) thông qua các đề tài khác nhau. Sentiment analysis – phân tích cảm xúc của đối tượng từ văn bản, là một trong các thách thức đối với bài toán về xử lý ngôn ngữ tự nhiên. Qua những bình luận của người dùng, ta sẽ phân loại các thái độ cảm xúc như giận hờn, vui vẻ, chê bai, buồn bã, trung tính… để tập trung định hướng phát triển sản phẩm hoặc cải thiện ứng dụng. Ở đây, ta chủ yếu xét về hai cảm xúc chính, đó là tích cực (positive) và tiêu cực (negative).

# BÀI TOÁN SENTIMENT ANALYSIS

# CÁC HƯỚNG TIẾP CẬN

# THỰC NGHIỆM

# ĐÁNH GIÁ

# KẾT LUẬN

1. Những vấn đề đã làm được

2. Những vấn đề chưa làm được

3. Hướng phát triển

# TÀI LIỆU THAM KHẢO