**TRƯỜNG ĐẠI HỌC SÀI GÒN**

**KHOA CÔNG NGHỆ THÔNG TIN**

****

**ĐỀ CƯƠNG**

**ĐỀ TÀI: Phân tích đánh giá sản phẩm trên nền tảng thương mại điện tử bằng Machine Learning.**

**NHÓM 30:**

3124720066 - Nguyễn Gia Thành

3123410151 - Nguyễn Dương Khang

3121560024 - Nguyễn Quang Dương

**Giảng viên hướng dẫn**: Đỗ Như Tài

**TPHCM Tháng 4/2025**

**MỤC LỤC**

[**I. Thông tin chung**](#_21kctc3l5rm4) **2**

[**II. Đặt vấn đề**](#_i7gj1xxw65vx) **2**

[1. Lý do chọn đề tài](#_xbuk5224codg) 2

[2. Tính cấp thiết](#_mip3u4l730b7) 3

[3. Tổng quan nghiên cứu](#_5wbvmmtg77bk) 3

[4. Mục tiêu nghiên cứu](#_352rbjn0njlw) 3

[5. Câu hỏi/Giả thuyết nghiên cứu](#_aszm56v507sg) 3

[**III. Đối tượng và phạm vi nghiên cứu**](#_q52g1dz3m2yw) **4**

[**IV. Phương pháp nghiên cứu**](#_nhtx66c7d8yl) **4**

[**V. Dự kiến kết quả và ý nghĩa nghiên cứu**](#_r09meiem3w3) **5**

[**VI. Kế hoạch thực hiện**](#_2avs0o6455ir) **5**

[**VII. Tài liệu tham khảo**](#_mkohbefejl85) **6**

[**VIII. Phụ lục .**](#_kvp2tcf1o0n2) **6**

**NỘI DUNG**

### **I. Thông tin chung**

* **Tên đề tài**: Phân tích đánh giá sản phẩm trên nền tảng thương mại điện tử bằng Machine Learning.
* **Nhóm thực hiện**:

Nhóm 30:

3124720066 - Nguyễn Gia Thành

3123410151 - Nguyễn Dương Khang

3121560024 - Nguyễn Quang Dương

* **Chủ nhiệm đề tài**: GV. Đỗ Như Tài
* **Cơ sở thực hiện**: Khoa Công nghệ Thông tin, Trường Đại học Sài Gòn
* **Thời gian dự kiến**: 2 tháng (Từ tháng 3 đến tháng 5 năm 2025)

### **II. Đặt vấn đề**

#### **1. Lý do chọn đề tài**

Trong kỷ nguyên số, người tiêu dùng ngày càng có xu hướng tham khảo các đánh giá trực tuyến trước khi quyết định mua hàng. Những đánh giá này phản ánh cảm nhận thực tế của người dùng, đồng thời cũng là nguồn dữ liệu quan trọng giúp cải thiện sản phẩm và dịch vụ. Do đó, việc phân tích đánh giá sản phẩm bằng kỹ thuật xử lý ngôn ngữ tự nhiên và học máy là cần thiết.

#### **2. Tính cấp thiết**

Giúp người tiêu dùng có góc nhìn khách quan hơn khi mua sắm trực tuyến.

Hỗ trợ doanh nghiệp đánh giá mức độ hài lòng của khách hàng và cải tiến sản phẩm.

Đóng góp ứng dụng thực tế cho lĩnh vực trí tuệ nhân tạo trong phân tích dữ liệu văn bản.

#### **3. Tổng quan nghiên cứu**

Nghiên cứu các công trình liên quan đến phân tích cảm xúc (sentiment analysis).

Các phương pháp sử dụng NLP kết hợp mô hình học máy như Naive Bayes, SVM, Logistic Regression, LSTM, BERT,...

Ứng dụng NumPy, Matplotlib, Pandas, Scikit-learn, Keras/TensorFlow, Nltk, Imbalanced-learn trong xử lý dữ liệu đánh giá.

#### **4. Mục tiêu nghiên cứu**

Phân tích và đánh giá cảm xúc từ các bình luận sản phẩm trên nền tảng thương mại điện tử dựa vào tệp dữ liệu Amazon Food Reviews ( Kaggle )[1].

Xây dựng mô hình dự đoán cảm xúc (tích cực, tiêu cực, trung tính) từ văn bản đánh giá.

#### **5. Câu hỏi/Giả thuyết nghiên cứu**

Các đặc trưng ngôn ngữ nào thường xuất hiện trong đánh giá tích cực/tiêu cực/trung tính ?

Mô hình học máy nào có thể cho ra hiệu quả tốt nhất trong việc phân loại đánh giá sản phẩm ?

Có thể thay thế mô hình học máy tốt nhát sau nghiên cứu bằng mô hình học sâu để giải quyết vấn đề hay không và có đạt được hiệu quả như mô hình học máy ?

### **III. Đối tượng và phạm vi nghiên cứu**

**Đối tượng nghiên cứu**: Các đánh giá sản phẩm được người dùng đăng tải trên nền tảng thương mại điện tử Amazon được tổng hợp trên Kaggle.

**Phạm vi nghiên cứu**:

**Không gian**: Dữ liệu từ Kaggle..  
**Thời gian**: Các đánh giá trong năm 2012.

**Khung lý thuyết/Mô hình nghiên cứu**:

Mô hình phân tích cảm xúc sử dụng pipeline xử lý văn bản (tiền xử lý → trích đặc trưng → mô hình học máy).

### **IV. Phương pháp nghiên cứu**

**Thiết kế nghiên cứu**: Mô hình dự đoán cảm xúc dựa trên dữ liệu văn bản.

**Phương pháp thu thập dữ liệu**: Dùng dữ liệu Kaggle đánh giá sản phẩm (bằng Python hoặc sử dụng nguồn dữ liệu công khai).

**Phương pháp xử lý và phân tích dữ liệu**:

Tiền xử lý văn bản: loại bỏ stopword, chuẩn hóa văn bản, tokenization.

Trích đặc trưng: TF-IDF, Word Embedding.

Mô hình: Naive Bayes[2], Logistic Regression[3], Random Forest[4], LSTM[5],...

Đánh giá mô hình: độ chính xác, precision, recall, F1-score.

**Công cụ/phần mềm sử dụng**:

Python (Pandas, NumPy, scikit-learn, Keras/TensorFlow, Nltk, Imbalanced-learn)

Jupyter Notebook, VS Code

Matplotlib/Seaborn để trực quan hóa dữ liệu

### **V. Dự kiến kết quả và ý nghĩa nghiên cứu**

**Kết quả mong đợi**:

Xây dựng mô hình phân tích cảm xúc từ đánh giá sản phẩm.

Trực quan hóa dữ liệu và đưa ra nhận xét từ kết quả mô hình.

**Ý nghĩa lý luận/thực tiễn**:

Là tài liệu tham khảo cho sinh viên và nghiên cứu viên về NLP.

Gợi ý hướng phát triển chatbot hỗ trợ khách hàng, phân tích thị trường.

**Những đóng góp của nghiên cứu**:

Ứng dụng kỹ thuật NLP và học máy vào thực tế.

Tạo cơ sở cho các nghiên cứu nâng cao như phân tích chủ đề, phát hiện gian lận trong đánh giá,…

### **VI. Kế hoạch thực hiện và phân công.**

Kế hoạch thực hiện :

|  |  |
| --- | --- |
| **Thời gian** | **Nội dung công việc** |
| Tuần 1 - 2 | Xác định đề tài, tìm hiểu tổng quan lý thuyết |
| Tuần 3 - 7 | Thu thập và tiền xử lý dữ liệu |
| Tuần 8 - 11 | Xây dựng và huấn luyện mô hình |
| Tuần 11 - 13 | Đánh giá mô hình, điều chỉnh |
| Tuần 14 | Tổng hợp kết quả, viết báo cáo |
| Tuần 14 | Chuẩn bị slide trình bày |
| Tuần 15 | Báo cáo đề tài |

### Bảng phân công :

|  |  |  |
| --- | --- | --- |
| **STT** | **Họ và Tên** | **Nhiệm vụ** |
| 1 | Nguyễn Dương Khang & Nguyễn Quang Dương | Thu thập và tiền xử lý dữ liệu |
| 2 | Cả nhóm | Phân tích khám phá dữ liệu (EDA) và xây dựng mô hình |
| 3 | Nguyễn Gia Thành | Cải tiến mô hình |
| 4 | Nguyễn Gia Thành | Viết báo cáo và đánh giá mô hình |
| 5 | Cả nhóm | Làm powerpoint và poster. |
| 6 | Nguyễn Dương Khang & Nguyễn Quang Dương | Làm video thuyết trình và tạo UI |

Bảng đánh giá đóng góp :

|  |  |
| --- | --- |
| **Họ và Tên** | **Đánh giá (%)** |
| Nguyễn Gia Thành | 100% |
| Nguyễn Dương Khang | 100% |
| Nguyến Quang Dương | 100% |

### **VII. Tài liệu tham khảo**

[1] “SNAP: Web data: Amazon Fine Foods reviews.” Accessed: May 18, 2025. [Online]. Available: https://snap.stanford.edu/data/web-FineFoods.html

[2] “(PDF) Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter,” in *ResearchGate*, doi: 10.1109/ICIC47613.2019.8985884.

[3] “(PDF) Sentimental Analysis using Logistic Regression.” Accessed: May 02, 2025. [Online]. Available: https://www.researchgate.net/publication/353514662\_Sentimental\_Analysis\_using\_Logistic\_Regression?utm\_source=chatgpt.com

[4] T. Nguyen, “Topic Modelling and Sentiment Analysis of Customer Reviews for B2C E-commerce Platforms in Vietnam: A Comparative Study of Lazada, Shopee, Tiki, and Sendo,” *Sr. Indep. Study Theses*, Jan. 2023, [Online]. Available: https://openworks.wooster.edu/independentstudy/10623

[5] U. B. Mahadevaswamy and P. Swathi, “Sentiment Analysis using Bidirectional LSTM Network,” *Procedia Comput. Sci.*, vol. 218, pp. 45–56, Jan. 2023, doi: 10.1016/j.procs.2022.12.400.

[6] “Natural Language Processing with Python: Analyzing Text with the Natural ... - Steven Bird, Ewan Klein, Edward Loper - Google Sách.” Accessed: May 02, 2025. [Online]. Available: https://books.google.com.vn/books?hl=vi&lr=&id=KGIbfiiP1i4C&oi=fnd&pg=PR5&dq=Natural+Language+Processing+with+Python+Bird&ots=Y5Clx8JCM2&sig=imoLAVJUFztkzfUP6DU42qPXs\_A&redir\_esc=y#v=onepage&q=Natural%20Language%20Processing%20with%20Python%20Bird&f=false

[7] J. Brownlee, “Machine Learning Mastery With Python”.

[8] J. Brownlee, A. Tam, and Z. M. Chng, “Deep Learning with Python, Second Edition”.

[9] B. Pang and L. Lee, “Opinion Mining and Sentiment Analysis,” *Found. Trends® Inf. Retr.*, vol. 2, no. 1–2, pp. 1–135, Jul. 2008, doi: 10.1561/1500000011.