ĐỀ CƯƠNG CHI TIẾT

Ngày thực hiện đề cương:02/05/2025

TÊN ĐỀ TÀI	Dự đoán sự hài lòng của khách hàng bằng các phương pháp học máy/học sâu (Predict the Customer Satisfaction)				
LĨNH VỰC CHUYÊN NGÀNH	Khoa Học Máy Tính				
LOẠI HÌNH NGHIÊN CỨU	Nghiên cứu mô tả				
NGƯỜI HƯỚNG DẪN	GV. Đỗ Như Tài				
SINH VIÊN THỰC HIỆN 1	Lê Hồng Sơn	3121410423			
SINH VIÊN THỰC HIỆN 2	Nguyễn Hoàng Thanh Phương	3122410329			
SINH VIÊN THỰC HIỆN 3	Đỗ Hữu Lộc	3123410201			
SINH VIÊN THỰC HIỆN 4	Văn Hoàng Như Ý	3122410493			

GIỚI THIỆU

CÂU HỎI GỢI Ý Lĩnh vực và nghiên cứu liên quan đã và đang ph vấn đề, và bài toán đặt ra cần giải quyết là gì?		Lĩnh vực và nghiên cứu liên quan đã và đang phát triển như thế nào? Các vấn đề, và bài toán đặt ra cần giải quyết là gì?
HƯỚNG DẪN Giới thiệu tổng quan về đề tài		Giới thiệu tổng quan về đề tài những vấn đề và lĩnh vực liên quan đến đề tài

Trong thời đại số hóa, sự hài lòng của khách hàng đóng vai trò quan trọng trong sự thành công của doanh nghiệp. Việc hiểu rõ mức độ hài lòng của khách hàng giúp các công ty cải thiện dịch vụ, tăng trải nghiệm người dùng và nâng cao lợi thế cạnh tranh. Tuy nhiên, việc thu thập và phân tích phản hồi của khách hàng theo cách truyền thống thường tốn nhiều thời gian và dễ bi sai sót.

Với sự phát triển mạnh mẽ của trí tuệ nhân tạo (AI) và học máy (Machine Learning), các doanh nghiệp có thể tận dụng các mô hình dự đoán để phân tích dữ liệu phản hồi khách hàng, từ đó dự báo mức độ hài lòng của họ. Điều này giúp các doanh nghiệp không chỉ hiểu được xu hướng của khách hàng mà còn có thể đưa ra các chiến lược phù hợp nhằm nâng cao chất lượng dịch vụ.

Mặc dù đã có nhiều nghiên cứu về dự đoán sự hài lòng của khách hàng, nhưng vẫn còn nhiều thách thức cần được giải quyết:

- Dữ liệu không đông nhất: Dữ liệu phản hồi khách hàng có thể không đồng nhất, chứa nhiều yếu tố cảm xúc khó đo lường [1].
- **Phương pháp truyền thống hạn chế:** Các phương pháp truyền thống không đủ hiệu quả trong việc khai thác các yếu tố tiềm ẩn ảnh hưởng đến sự hài lòng [2].
- Lựa chọn thuật toán phù hợp: Việc áp dụng mô hình học máy cần lựa chọn thuật toán phù

hợp để đạt độ chính xác cao nhất [3].

Dựa trên các vấn đề trên, nghiên cứu này hướng đến việc xây dựng một mô hình dự đoán sự hài lòng của khách hàng bằng cách sử dụng các thuật toán học máy. Cụ thể, nghiên cứu sẽ tập trung vào:

- **Thu thập và tiền xử lý dữ liệu:** Thu thập và tiền xử lý dữ liệu phản hồi khách hàng từ nhiều nguồn khác nhau [4].
- **Áp dụng thuật toán học máy:** Áp dụng các thuật toán học máy để phân tích và dự đoán mức độ hài lòng của khách hàng [5].
- Đánh giá hiệu suất mô hình: Đánh giá hiệu suất của các mô hình và đề xuất phương pháp tối ưu nhất [6].

Câu hỏi nghiên cứu: Những yếu tố nào ảnh hưởng đến sự hài lòng của khách hàng? Mô hình học máy nào có thể đưa ra dự đoán chính xác nhất? Làm thế nào để tối ưu hóa mô hình dự đoán sự hài lòng của khách hàng?

GHI CHÚ

TÌNH HÌNH NGHIÊN CỨU TRONG VÀ NGOÀI NƯỚC

CÂU HỎI GỢI Ý	Lĩnh vực và nghiên cứu liên quan đã và đang phát triển như thế nào? Các vấn đề, và bài toán đặt ra cần giải quyết là gì?
HƯỚNG DẪN	Tìm hiểu các nghiên cứu đã công bố gần nhất (5 năm trở lại) về lĩnh vực liên quan trong và ngoài nước. Phân tích các kết quả đóng góp, nhận xét các hạn chế còn tồn tại.

Trên thế giới, các phương pháp học máy và học sâu đã được áp dụng rộng rãi để dự đoán sự hài lòng của khách hàng trong nhiều lĩnh vực khác nhau. Một nghiên cứu tiêu biểu là dự án phân tích sự hài lòng của hành khách sử dụng mô hình học máy có giám sát trên dữ liệu từ một hãng hàng không Mỹ. Mục tiêu của nghiên cứu là dự đoán sự hài lòng của hành khách dựa trên các yếu tố như khoảng cách bay, độ trễ khởi hành và hạng vé. Kết quả cho thấy các yếu tố như dịch vụ lên máy bay trực tuyến, Wi-Fi trên chuyến bay và giải trí trên chuyến bay có ảnh hưởng lớn đến sự hài lòng của hành khách. Đây là một minh chứng rõ ràng cho thấy học máy có thể hỗ trợ hiệu quả trong việc cải thiện trải nghiệm khách hàng,[11]

Một nghiên cứu khác so sánh giữa các mô hình học máy truyền thống như Random Forest và Support Vector Machines với các kỹ thuật học sâu như Multi-Layer Perceptrons trong việc dự đoán sự hài lòng của khách hàng thương mại điện tử. Kết quả chỉ ra rằng Random Forest có thể dự đoán sự hài lòng trong tương lai với độ chính xác 92%, vượt trội hơn so với học sâu. Đồng thời, nghiên cứu cũng xác định được các yếu tố chính ảnh hưởng đến sự hài lòng như thời gian giao hàng và độ chính xác của đơn hàng, giúp các doanh nghiệp có hướng điều chỉnh dịch vụ hiệu quả. [10]

Tại Việt Nam, việc áp dụng các phương pháp học máy trong đo lường và dự đoán sự hài lòng của khách hàng đã bắt đầu được quan tâm. Một nghiên cứu nổi bật của Thái Kim Phụng (2023) đã sử dụng các mô hình học máy như Neural Network (NN), Decision Tree (DT), Naïve Bayes (NB), Support Vector Machines (SVM), Logistic Regression (LR) và Random Forest (RF) để phân loại mức độ hài lòng của khách hàng dựa trên các bình luận trực tuyến về khách sạn trên trang Agoda.com. Kết quả nghiên cứu cho thấy Random Forest đạt hiệu suất phân loại tốt nhất, đồng thời cung cấp thông tin hữu ích giúp các doanh nghiệp khách sạn cải thiện dịch vụ và nâng cao trải nghiệm khách hàng. [12]

Các nghiên cứu hiện nay còn gặp nhiều hạn chế như chất lượng dữ liệu không đồng nhất, có thể chứa thông tin sai lệch hoặc thiên lệch, gây ảnh hưởng đến độ chính xác của mô hình dự đoán. Việc xử lý ngôn ngữ tự nhiên (NLP) cũng gặp khó khăn do các phản hồi của khách hàng thường mang tính cảm xúc, sử dụng ẩn dụ hoặc từ lóng. Ngoài ra, một số mô hình học sâu dễ bị overfitting khi dữ liệu huấn luyện không đủ đa dạng, làm giảm hiệu suất trên dữ liệu thực tế. Khả năng giải thích kết quả của các mô hình học sâu cũng còn hạn chế, gây khó khăn cho các doanh nghiệp trong việc hiểu và áp dụng kết quả dự đoán.

Các bài toán quan trọng đặt ra trong lĩnh vực này bao gồm việc cải thiện độ chính xác của mô hình dự đoán thông qua tối ưu hóa hoặc kết hợp các mô hình học máy tiên tiến. Phát triển các hệ thống phân tích cảm xúc hiệu quả hơn bằng công nghệ NLP hiện đại cũng là một yêu cầu cấp thiết để nắm bắt chính xác tâm lý khách hàng. Bên cạnh đó, cần xây dựng các phương pháp giải thích mô hình như SHAP hoặc LIME để giúp doanh nghiệp hiểu rõ hơn các yếu tố tác động đến sự hài lòng của khách hàng. Việc ứng dụng thực tế các hệ thống dự đoán này sẽ giúp doanh nghiệp cải thiện dịch vụ, tối ưu quy trình vận hành và nâng cao trải nghiệm khách hàng.

GHI CHÚ

MỤC TIÊU VÀ PHẠM VI NGHIÊN CỨU

CÂU HỎI GỢI Ý	Mục tiêu nghiên cứu chính của đề tài là gì? Phạm vi nghiên cứu là gì?
HƯỚNG DẪN	Đặt bài toán giải quyết và trình bày mục tiêu nghiên cứu chính của đề tài. Nêu phạm vi nghiên cứu của đề tài, bao gồm việc giới hạn phạm vi nghiên cứu và triển khai, các giả định ban đầu đối với nghiên cứu.

Muc tiêu của đề tài

Hiện nay trong kinh doanh thương mại, khách hàng là một yếu tố ảnh hưởng không nhỏ đến việc thành công của một doanh nghiệp. Vì vậy việc chăm sóc khách hàng là một điều hầu như bắt buộc nếu như muốn phát triển và mở rộng quy mô kinh doanh.

Vì vậy, việc doanh nghiệp chú trọng vào phân tích hành vi, phân tích sự hài lòng của khách hàng sẽ giúp doanh nghiệp đi được đúng hướng trong việc kinh doanh của mình.[7][8]

Với đề tài này, mục tiêu của đề tài bao gồm:

- Phân tích các yếu tố tác động đến sự hài lòng của khách hàng.
- Dư đoán sư hài lòng của khách hàng dưa trên các yếu tố, thông tin sẵn có.
- Tìm ra mô hình tối ưu nhất để đưa vào sử dụng

Nghiên cứu này giúp cho các doanh nghiệp có giải pháp, hướng đi đúng đắn khi nắm được trải nghiệm của người dùng trong lĩnh vực thương mại điện tử.

Giảm chi phí, nắm bắt được tâm lý, không tốn quá nhiều chi phí không cần thiết cho các sự kiện dành cho khách hàng

Pham vi nghiên cứu

Nghiên cứu này tập trung vào việc tìm hiểu, phân tích các yếu tố ảnh hưởng đến trải nghiệm của người dùng, khách hàng trong lĩnh vực thương mại điện tử.

Ngoài các phương pháp truyền thống (Đo lường CSAT, NPS,...) bài báo này tập trung vào nghiên cứu các phương pháp hiện đại hơn đó là Deep Learning và Machine learning.

Trong bài báo này, bộ dữ liệu sẽ được lấy từ cuộc thi **Predict the Customer Satisfaction - CSE**22 và tập dữ liệu được thu thập trong vòng 4 năm từ 2020 đến 2024.[9]

Nghiên cứu này chỉ giới hạn ở một sàn thương mại điện tử nhất định. Nó không thể bao quát, đại diện cho toàn bộ các sàn thương mại điện tử.

Dữ liệu không bao gồm sự ảnh hưởng từ các yếu tố bên ngoài như (dịch bệnh, covid, chiến tranh) Tập trung vào các mô hình học máy/ học sâu từ cơ bản đến phức tạp để có thể đưa ra sự so sánh

khách quan nhất như: CNN, ANN, Logistic Regression, k-Nearest Neighbors,...[8]

GHI CHÚ

CƠ SỞ LÝ THUYẾT

CÂU HỎI GỢI Ý	Khái niệm về sự hài lòng của khách hàng, những yếu tố ảnh hưởng? Tổng quan học máy/học sâu, các phương pháp học máy/học sâu, phương pháp nào ứng dụng vào dự đoán.
HƯỚNG DẪN	Đọc các lý thuyết về sự hài lòng của khách hàng và các yếu tố tác động đến nó. Nghiên cứu các thuật toán học máy như Random Forest, SVM và các mô hình học sâu như ANN, CNN . Tìm hiểu sự khác biệt giữa học máy và học sâu trong việc phân tích và dự đoán.

Sự hài lòng của khách hàng

Sự hài lòng của khách hàng là sự đánh giá dựa trên kinh nghiệm sử dụng một dịch vụ hoặc sản phẩm cụ thể trong một thời gian. Sự hài lòng của khách hàng là kim chỉ nam cho các doanh nghiệp khi triển khai marketing cho thấy sự tin tưởng và hài lòng của khách hàng khi trải nghiệm các dịch vụ và sản phẩm marketing có liên quan đến mức độ gắn bó với thương hiệu và lợi thế cạnh tranh so với đối thủ, từ đó giúp doanh nghiệp cải thiện [13].

Học máy và học sâu trong phân tích sự hài lòng

Học máy (Machine Learning): Sử dụng các thuật toán như Random Forest, XGBoost, và LightGBM để phân loại và dự đoán sự hài lòng dựa trên các đặc trưng đầu vào.

Học sâu (Deep Learning): Áp dụng các mô hình như Mạng Nơ-ron Tích chập (CNN) để xử lý dữ liệu phức tạp, đặc biệt hiệu quả với dữ liệu hình ảnh hoặc chuỗi thời gian.

GHI CHÚ

MÔ HÌNH NGHIÊN CỨU

CÂU HỎI GỢI Ý	Các biến độc lập và phụ thuộc trong mô hình là gì? Giả thuyết nghiên cứu của đề tài là gì? Mô hình học máy và học sâu nào bạn sẽ áp dụng?
HƯỚNG DẪN	Xác định các yếu tố quan trọng như giá trị sản phẩm, chất lượng dịch vụ, v.v. sẽ được sử dụng trong mô hình nghiên cứu. Lựa chọn các giả thuyết phù hợp với các yếu tố nghiên cứu, ví dụ: "Chất lượng dịch vụ cao sẽ làm tăng sự hài lòng của khách hàng." Nghiên cứu các mô hình học máy và học sâu phù hợp với dữ liệu và bài toán.

Biến độc lập và phụ thuộc

Biến phụ thuộc: customer_experience với 3 mức độ: "Good", "Neutral", "Bad".

Biến độc lập: Bao gồm các đặc trưng như thời gian giao hàng, giá trị đơn hàng, phương thức thanh toán, và các đặc trưng khác liên quan đến đơn hàng và khách hàng.

Giả thuyết nghiên cứu

H1: Thời gian giao hàng nhanh hơn dẫn đến sự hài lòng cao hơn.

H2: Giá trị đơn hàng cao có thể liên quan đến sự hài lòng thấp hơn nếu không đáp ứng kỳ vọng.

H3: Sử dụng các phương thức thanh toán linh hoạt tăng sự hài lòng của khách hàng.

Mô hình áp dụng

Hoc máy: Random Forest, LightGBM.

Học sâu: Mạng Nơ-ron Tích chập (CNN) cho dữ liệu hình ảnh hoặc chuỗi thời gian nếu có.

Các yếu tố quan trọng trong mô hình

Giá trị sản phẩm: Tổng giá trị của các mặt hàng trong đơn hàng.

Chất lượng dịch vụ: Được phản ánh qua thời gian giao hàng và phản hồi từ khách hàng.

Phương thức thanh toán: Sự đa dạng và linh hoạt trong thanh toán có thể ảnh hưởng đến trải nghiệm khách hàng.

GHI CHÚ

PHƯƠNG PHÁP NGHIÊN CỨU

CÂU HỎI GỢI Ý	Thu thập dữ liệu từ đâu? Làm thế nào để phân tích dữ liệu và kiểm tra mô hình? Phương pháp đánh giá mô hình của đề tài là gì?
HƯỚNG DẪN	Quyết định nguồn dữ liệu: khảo sát khách hàng, dữ liệu từ hệ thống, v.v Nghiên cứu phương pháp phân tích dữ liệu như xử lý dữ liệu thô, làm sạch dữ liệu. Tìm hiểu về các chỉ số đánh giá mô hình như Accuracy, Precision, Recall, F1-score.

Thu thập dữ liệu

Dữ liệu được cung cấp bởi cuộc thi trên Kaggle, bao gồm thông tin về đơn hàng, khách hàng, sản phẩm, và phản hồi từ khách hàng.

Phân tích và kiểm tra mô hình

- Tiền xử lý dữ liệu: Làm sạch dữ liệu, xử lý giá trị thiếu, và mã hóa các biến phân loại.
- Phân tích dữ liệu: Khám phá dữ liệu để hiểu rõ mối quan hệ giữa các biến.
- Huấn luyện mô hình: Sử dụng tập huấn luyện để đào tạo các mô hình học máy và học sâu.
- Kiểm tra mô hình: Đánh giá hiệu suất của mô hình trên tập kiểm tra.

Phương pháp đánh giá mô hình

- Accuracy: Tỷ lệ dự đoán đúng trên tổng số dự đoán.
- **Precision**: Tỷ lệ dự đoán đúng trong số các dự đoán dương tính.
- **Recall**: Tỷ lệ dự đoán đúng trong số các trường hợp dương tính thực sự.
- **F1-score**: Trung bình điều hòa của Precision và Recall, đặc biệt hữu ích khi dữ liệu mất cân bằng.

GHI CHÚ

KẾT QUẢ DỰ KIẾN VÀ PHÂN TÍCH

CÂU HỎI GỌI Ý	Kết quả dự đoán của mô hình như thế nào? Làm thế nào để so sánh các mô hình học máy và học sâu trong nghiên cứu của đề tài ?
HƯỚNG DẪN	Tìm hiểu cách đánh giá hiệu quả của mô hình thông qua các chỉ số cụ thể. So sánh kết quả của mô hình học máy và học sâu về độ chính xác và hiệu quả.

Nghiên cứu đã triển khai và đánh giá hai mô hình học máy phổ biến: Random Forest và LightGBM, trong nhiệm vụ phân loại đa lớp nhằm dự đoán mức độ hài lòng của khách hàng dựa trên các đặc trưng hành vi và giao dịch. Kết quả thực nghiệm cho thấy:

• LightGBM là mô hình có hiệu năng tốt hơn so với Random Forest với các chỉ số:

Accuracy: 67.31%F1 Score: 67.41%

- Khả năng cân bằng tốt giữa Precision và Recall trên cả ba lớp phân loại.

- Mô hình thể hiện tính ổn định và khả năng phân biệt khá rõ giữa các nhóm khách hàng.

• Random Forest mặc dù là mô hình mạnh mẽ về tính khái quát, nhưng trong tập dữ liệu cụ thể này lại cho kết quả thấp hơn:

Accuracy: 60.43%F1 Score: 60.54%

- Có sự mất cân đối trong độ chính xác giữa các lớp, đặc biệt ở lớp 1.

Việc so sánh giữa hai mô hình cho thấy LightGBM vượt trội về khả năng tổng quát hóa và xử lý các mối quan hệ phức tạp trong dữ liệu. Đây là minh chứng rõ ràng cho việc lựa chọn đúng mô hình phù hợp với tính chất dữ liêu sẽ mang lai hiệu quả cao hơn.

GHI CHÚ

KẾT LUẬN

CÂU HỎI GỢI Ý	Những kết quả chính mà nghiên cứu mang lại là gì?
HƯỚNG DẪN	Tóm tắt các kết quả đạt được và mức độ đạt được mục tiêu. Đề xuất hướng nghiên cứu mới hoặc cải tiến mô hình để tối ưu hóa kết quả.

Tổng kết kết quả đạt được

• LightGBM là mô hình có hiệu năng tốt hơn so với Random Forest với các chỉ số:

Nghiên cứu đã triển khai và đánh giá hai mô hình học máy phổ biến: Random Forest và LightGBM, trong nhiệm vụ phân loại đa lớp nhằm dự đoán mức độ hài lòng của khách hàng dựa trên các đặc trưng hành vi và giao dịch. Kết quả thực nghiệm cho thấy:

Accuracy: 67.31%F1 Score: 67.41%

- Khả năng cân bằng tốt giữa Precision và Recall trên cả ba lớp phân loại.
- Mô hình thể hiện tính ổn định và khả năng phân biệt khá rõ giữa các nhóm khách hàng.
- Random Forest mặc dù là mô hình mạnh mẽ về tính khái quát, nhưng trong tập dữ liệu cụ thể này lại cho kết quả thấp hơn:

Accuracy: 60.43%F1 Score: 60.54%

Có sự mất cân đối trong độ chính xác giữa các lớp, đặc biệt ở lớp 1.

Việc so sánh giữa hai mô hình cho thấy LightGBM vượt trội về khả năng tổng quát hóa và xử lý các mối quan hệ phức tạp trong dữ liệu. Đây là minh chứng rõ ràng cho việc lựa chọn đúng mô hình phù hợp với tính chất dữ liệu sẽ mang lại hiệu quả cao hơn.

Mức độ đạt được mục tiêu đề tài

Mục tiêu ban đầu của nghiên cứu là xây dựng một hệ thống dự đoán mức độ hài lòng khách hàng hiệu quả nhằm hỗ trợ doanh nghiệp trong việc cá nhân hóa dịch vụ và tối ưu hóa chăm sóc khách hàng. Qua quá trình triển khai mô hình, xử lý dữ liệu, huấn luyện và đánh giá, nhóm nghiên cứu đã:

- Chứng minh được khả năng áp dụng mô hình học máy hiện đại để giải quyết bài toán thực tiễn.
- Hiểu sâu hơn về tầm quan trọng của tiền xử lý dữ liệu, chọn đặc trưng và điều chỉnh siêu tham số.
 - Đề xuất mô hình LightGBM như một giải pháp tiềm năng có thể được triển khai trong thực tế.

GHI CHÚ

TÀI LIỆU THAM KHẢO

CÂU HỎI GỢI Ý	Các tài liệu tham khảo thuộc lĩnh vực nghiên cứu liên quan là gì?
	[1] Azizi, M., & Djouhri, L. Predicting customer satisfaction using machine learning techniques.
	[2] Research Team. Predicting customer satisfaction: An approach based on machine learning.
	[3] C Research Group. Predicting customer satisfaction for distribution companies using machine learning.
	[4] Author(s) Unknown. Predicting E-commerce customer satisfaction: Traditional machine learning versus deep learning.
HƯỚNG DẪN	[5] Lao Research Team. Performance of machine learning models to predict customer satisfaction scores for Lao National Convention Center.
	[6] RIT Research Team. Predicting & optimizing airlines customer satisfaction using machine learning techniques.
	[7] Can machine learning techniques predict customer dissatisfaction? A feasibility study for the automotive industry
	[8] OPTIMIZING E-COMMERCE PRICING STRATEGIES
	[9] Enhancing the Prediction of User Satisfaction with Metaverse Service Through Machine Learning
	[10] Predicting E-commerce customer satisfaction: Traditional machine learning vs. deep learning
	<u>approaches</u>
	[11] Passenger satisfaction analysis using supervised machine learning model
	[12] Ứng dụng phương pháp máy học trong đo lường sự hài lòng của khách hàng dựa trên các bình luân trưc tuyến
	[13] Predictive model for customer satisfaction analytics in E-commerce sector using machine learning and deep learning.