Bộ Giáo dục và Đào tạo

**Trường Đại Học Sài Gòn**

Khoa Công Nghệ Thông Tin

*Bộ môn PPNCKH*

ĐỀ CƯƠNG CHI TIẾT

Ngày thực hiện đề cương:19/03/2025

|  |  |  |
| --- | --- | --- |
| **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 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 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 bị 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 đồng 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. |
| **Mục 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. * Dự đoán sự hài lòng của khách hàng dựa 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   * **Phạm 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** | |

**TÀI LIỆU THAM KHẢO**

|  |  |
| --- | --- |
| **CÂU HỎI GỢI Ý** | **Các tài liệu tham khảo thuộc lĩnh vực nghiên cứu liên quan là gì?** |
| **HƯỚNG DẪN** | [[1] Azizi, M., & Djouhri, L. *Predicting customer satisfaction using machine learning techniques*.](https://dspace.univ-ouargla.dz/jspui/bitstream/123456789/33505/1/AZIZI-DJOUHRI.pdf)  [[2] Research Team. *Predicting customer satisfaction: An approach based on machine learning*.](https://www.researchgate.net/publication/343874839_Incidence_and_risk_factors_for_postictal_delirium_in_patients_after_electroconvulsive_therapy_in_China)  [[3] C Research Group. *Predicting customer satisfaction for distribution companies using machine learning*.](https://ui.adsabs.harvard.edu/abs/2021IJESM..15..743C/abstract)  [[4] Author(s) Unknown. *Predicting E-commerce customer satisfaction: Traditional machine learning versus deep learning*.](https://www.sciencedirect.com/science/article/abs/pii/S0969698924001619.)  [[5] Lao Research Team. *Performance of machine learning models to predict customer satisfaction scores for Lao National Convention Center*.](https://www.researchgate.net/publication/384674264_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*.](https://repository.rit.edu/cgi/viewcontent.cgi?article=12513&context=theses)  *[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 approaches  [11] assenger 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 luận trực tuyến |
|  | |
|  | |



|  |  |  |
| --- | --- | --- |
| **NGƯỜI THỰC HIỆN ĐỀ CƯƠNG** | | |
| **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 |

|  |  |  |
| --- | --- | --- |
| **XÁC NHẬN CỦA GIÁO VIÊN HƯỚNG DẪN** | | |
| **NGƯỜI HƯỚNG DẪN** |  | |
| **Ý KIẾN** |  | |
|  |  |  |