Tạp chí
Kinh tế và Dự báo

Economy & Forecast Review

8/2024
Số 15

VIỆN CHIẾN LƯỢC PHÁT TRIỂN - BỘ KẾ HOẠCH VÀ ĐẦU TƯ

ISSN 1859-4972

Hệ sinh thái khởi nghiệp tại Việt Nam và một số yếu tố cấu thành quan trọng

Số 15 tháng 08/2024 (881) - Năm thứ 57

Kinh tế và Dự báo

ECONOMY AND FORECAST REVIEW

CƠ QUAN
CỦA VIỆN CHIẾN LƯỢC PHÁT TRIỂN,
BỘ KẾ HOẠCH VÀ ĐẦU TƯ

Tổng Biên tập
ĐỖ THỊ PHƯƠNG LAN

Phó Tổng Biên tập
TRẦN THỊ THANH HÀ
PHÙNG THỊ PHƯƠNG ANH

Hội đồng Biên tập
GS. TS. PHẠM HỒNG CHƯƠNG
GS. TS. PHẠM BẢO DƯƠNG
GS. TS. TRẦN THỌ ĐẠT
PGS. TS. LÊ XUÂN ĐÌNH
TS. VƯƠNG QUÂN HOÀNG
GS. TS. NGÔ THẮNG LỢI
PGS. TS. TRẦN TRỌNG NGUYÊN
PGS. TS. BÙI HUY NHƯỢNG
TS. TRẦN HỒNG QUANG
TS. CAO VIẾT SINH
PGS. TS. NGUYỄN HỒNG SƠN
GS. TS. SỬ ĐÌNH THÀNH

Tòa soạn và trị sự
65 Văn Miếu - Đống Đa - Hà Nội
Tel: 080.43174 / 080.44474
Fax: 024.3747.3357
Email: kinhtedubao@mpi.gov.vn
Tạp chí điện tử
http://kinhtevadubao.vn

Quảng cáo và phát hành
Tel: 080.4474 / 0983 720 868
Qua mạng lưới Bưu điện Việt Nam

Giấy phép xuất bản: 477/GP-BTTTT
In tại: Công ty CP in Công đoàn Việt Nam

Giá 39.000 đồng

MỤC LỤC

CHIẾN LƯỢC - CHÍNH SÁCH

NGHIÊN CỨU - TRAO ĐỔI

Phân lớp bình luận của du khách về dịch vụ khách sạn ở tỉnh Thừa Thiên Huế dựa trên các thuật toán học máy

DƯƠNG THỊ HẢI PHƯƠNG*
HỒ QUỐC DŨNG**

Tóm tắt

Mục tiêu của nghiên cứu là ứng dụng các thuật toán phân lớp học máy để dự đoán cảm xúc của du khách về dịch vụ khách sạn ở tỉnh Thừa Thiên Huế nhằm đánh giá các thuật toán học máy trong việc dự đoán cảm xúc, cũng như giúp các khách sạn có thêm cơ sở để hoạch định các chiến lược trong quản trị quan hệ khách hàng. Bằng cách sử dụng kết hợp 2 thư viện hỗ trợ thu thập dữ liệu trong Python là Selenium và Scrapy, một bộ dữ liệu gồm 22,557 dòng về các bình luận và đánh giá của du khách đối với các khách sạn ở tỉnh Thừa Thiên Huế đã được thu thập. Dữ liệu sau khi được tiền xử lý với các kỹ thuật khác nhau đã được đưa vào huấn luyện bởi 3 mô hình học máy là: KNN, SVM và Naïve Bayes. Kết quả cho thấy, SVM cho kết quả dự đoán cảm xúc tốt hơn so với các mô hình còn lại.

Từ khóa: *học máy, phân lớp, khách sạn Huế, bình luận của du khách, phân tích cảm xúc*

Summary

This study intends to apply machine learning classification algorithms to predict tourists' emotions about hotel services in Thua Thien Hue province, to evaluate machine learning algorithms in predicting emotions, and to help hotels have more basis in planning strategies for customer relationship management. Using a combination of two data collection support libraries in Python, Selenium, and Scrapy, a dataset of 22,557 rows of comments and reviews of tourists for hotels in Thua Thien Hue province was collected. After being preprocessed with different techniques, the data was trained by 3 machine learning models: KNN, SVM, and Naïve Bayes. The results showed that SVM gave better emotion prediction than the remaining models.

Keywords: *machine learning, classification, Hue hotels, tourists' comments, sentiment analysis*

ĐẶT VẤN ĐỀ

Ngành kinh doanh khách sạn đóng vai trò quan trọng trong việc thúc đẩy phát triển du lịch và kinh tế địa phương ở tỉnh Thừa Thiên Huế, đồng thời góp phần vào việc tạo ra cơ hội việc làm, nâng cao hình ảnh và vị thế của địa phương trên thị trường du lịch quốc tế. Trong những năm qua, Tỉnh đã và đang tập trung đẩy mạnh đầu tư kinh doanh khách sạn theo nhiều hướng khác nhau để cung cấp dịch vụ tốt nhất cho du khách nhằm thu hút ngày càng nhiều du khách đến Thừa Thiên Huế. Hầu hết các khách sạn đều cung cấp dịch vụ đặt phòng qua website riêng của khách sạn. Các website này cung cấp các xếp hạng trực tuyến và phản hồi của du khách để giúp khách hàng có thêm cơ sở đưa ra các quyết định đặt phòng. Tuy nhiên, cũng tồn tại sự nghi ngờ về độ tin cậy của những xếp hạng này và để tăng sự hài lòng của du khách, cần có thêm sự tương tác với các du khách khác, cũng như các nhà cung cấp dịch vụ của bên thứ ba. Đánh giá của du khách trên các website của bên thứ ba mang tính khách quan hơn và cung cấp cái nhìn sâu sắc hơn về trải nghiệm thực tế của các du khách khác khi lưu trú tại khách sạn (Nadeem Akhtar và cộng sự, 2017). Do đó, các nguồn thông tin, như: xếp hạng và đánh giá của du khách trên TripAdvisor, Booking, Agoda…, ngày càng trở thành nguồn thông tin chính cho các du khách trong việc lựa chọn một nơi lưu trú phù hợp; đồng thời, các nguồn thông tin này cũng giúp các khách sạn nhanh chóng nắm bắt được tâm lý và nhu cầu khách hàng, hiểu được điểm mạnh và điểm yếu trong các dịch vụ của mình để từ đó có thể đưa ra những cải tiến phù hợp hơn với khách hàng, mang lại lợi nhuận cao hơn.

Trên thế giới đã xuất hiện các nghiên cứu đề cập đến một số phương pháp và kỹ thuật phân tích ý kiến và cảm xúc của du khách thông qua các bình luận. Hầu hết các phương pháp này đều dựa trên việc sử dụng các thuật toán học máy để phân lớp cảm xúc của du khách. Các phương pháp học máy không chỉ có thể tìm hiểu tính phân cực cảm xúc của các từ khóa gây ảnh hưởng, mà còn có thể xem xét tính phân cực của các từ khóa

* TS., Trường Đại học Kinh tế - Đại học Huế | Email: dthphuong@hce.edu.vn
** TS., Khoa Kỹ thuật và Công nghệ - Đại học Huế | Email: hoquocdung@gmail.com
Ngày nhận bài: 25/5/2024; Ngày phản biện: 20/6/2024; Ngày duyệt đăng: 05/8/2024

HÌNH 1: MA TRẬN NHẦM LẪN

		Prediction	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

Nguồn: Han và cộng sự (2012)

HÌNH 2: QUY TRÌNH NGHIÊN CỨU

Nguồn: Nhóm tác giả xây dựng

HÌNH 3: BÌNH LUẬN VÀ ĐÁNH GIÁ CỦA DU KHÁCH TRÊN TRIPADVISOR

HÌNH 4: MINH HỌA BỘ DỮ LIỆU SAU KHI GÁN NHÃN CHO BIẾN MỤC TIÊU

	Review	Rating	Sentiment
11624	We spent 3 days at this gem of a hotel. The mo...	5,0	pos
1954	Welcome drinks were served upon our arrival.Ir...	4,0	pos
19565	Been travelling Vietnam for one week now. Been...	5,0	pos
11036	The room here was amazing value, super spaciou...	4,0	pos
8156	I have travelled for one month in several hote...	5,0	pos
16708	We stayed at Vedana Lagoon for 3 nights as a p...	3,0	neu
13176	I had a great stay. Clean room and good breakf...	5,0	pos
22408	I cannot rate this hotel highly enough. I am c...	5,0	pos
18429	The rooms were lovely and big with a huge bath...	4,0	pos
11977	We only stayed here one night while travelling...	4,0	pos

tùy ý khác và tần suất xuất hiện của từ (Cambria và cộng sự, 2017) và một lượng lớn nghiên cứu của các tác giả khác nhau đã áp dụng các thuật toán học máy trên các bộ dữ liệu về khách sạn (Moro và cộng sự, 2017).

Ở Huế cũng đã xuất hiện một số nghiên cứu liên quan đến lĩnh vực khách sạn (Trần Thị Thu Hiền, 2019; Hoàng Bá Lộc và Hoàng Trọng Hùng, 2022). Phần lớn các nghiên cứu này sử dụng các kỹ thuật phân tích dữ liệu được hỗ trợ trong các phần mềm viết sẵn, như: SPSS, STATA…, để nghiên cứu ảnh hưởng của chất lượng dịch vụ khách sạn đến sự hài lòng của khách hàng, về chiến lược tiếp thị và quản lý khách hàng trong ngành khách sạn… Tuy nhiên, chưa thấy sự xuất hiện của các nghiên cứu liên quan đến việc phân tích, đánh giá các bình luận của du khách. Điều này cho thấy vẫn còn khoảng trống trong nghiên cứu về vấn đề này. Vì vậy, nghiên cứu này sử dụng các thuật toán học máy để phân lớp các bình luận của du khách về dịch vụ khách sạn ở Thừa Thiên Huế nhằm đánh giá hiệu suất của các thuật toán học máy trong phân lớp bình luận của du khách, cũng như giúp các khách sạn hiểu rõ hơn cảm nhận của du khách về các dịch vụ được cung cấp; từ đó, giúp các khách sạn có thêm cơ sở để hoạch định các chiến lược nhằm mục đích nâng cao chất lượng, cải thiện hình ảnh, giữ chân khách hàng cũng như thu hút khách hàng mới *(Bài viết sử dụng cách viết số thập phân theo chuẩn quốc tế).*

CƠ SỞ LÝ THUYẾT VÀ PHƯƠNG PHÁP NGHIÊN CỨU

Cơ sở lý thuyết

Phân tích cảm xúc

Theo Liu (2012), phân tích cảm xúc là một lĩnh vực nghiên cứu thực hiện việc phân tích ý kiến, đánh giá tình cảm, cảm xúc của một người đối với các thực thể (sản phẩm, dịch vụ, tổ chức, sự kiện…) và các thuộc tính của các thực thể. Phân tích cảm xúc có thể phân lớp tính phân cực của văn bản trong câu hoặc tài liệu để tìm hiểu xem ý kiến trong câu hoặc tài liệu là tích cực, tiêu cực hay trung lập. Đối với dịch vụ lưu trú, phân tích cảm xúc có thể được sử dụng để xác định mức độ hài lòng của du khách về các dịch vụ lưu trú nhằm tạo ra E-WOM tích cực.

Phân tích cảm xúc được thực hiện theo 2 hướng tiếp cận: Tiếp cận dựa trên từ điển và Tiếp cận học máy. Mỗi cách tiếp cận đều có những ưu và nhược điểm riêng. Tiếp cận dựa trên từ điển thực hiện đơn giản và được trợ giúp bởi các phần mềm sẵn có, như: GI hay DICTION. Tiếp cận học máy phức tạp và tốn kém thời gian hơn. Tuy nhiên, tiếp cận học máy mang lại tỷ lệ chính xác cao hơn tiếp cận dựa trên từ điển.

Phân lớp K láng giềng gần KNN

KNN (K-Nearest Neighbors) là một trong những thuật toán học có giám sát đơn giản nhất. KNN đi tìm đầu ra của một điểm dữ liệu mới chỉ dựa trên thông tin của K điểm dữ liệu gần nhất trong tập huấn luyện (Agrawal, 2014).

Với KNN, việc tính khoảng cách từ điểm kiểm tra đến dữ liệu huấn luyện sẽ quyết định độ chính xác của lớp mà nó thuộc về, do đó, việc quyết định sử dụng loại khoảng cách nào đóng vai trò quan trọng. Có nhiều loại khoảng cách khác nhau tùy vào bài toán, nhưng khoảng cách được sử dụng nhiều nhất là khoảng cách Euclid (Vũ Hữu Tiệp, 2020).

Phân lớp Naïve Bayes

Naïve Bayes là thuật toán phân lớp dựa trên định lý Bayes về lý thuyết xác suất để đưa ra các phán đoán, cũng như phân lớp dữ liệu dựa trên các dữ liệu được quan sát và thống kê. Đây là thuật toán thuộc nhóm học có giám sát và là một trong những thuật toán được ứng dụng nhiều trong lĩnh vực học máy. Naïve Bayes có 3 mô hình thường được sử dụng là: Gaussian Naïve Bayes, Multinomial Naïve Bayes và Bernoulli Naïve.

Phân lớp máy vector hỗ trợ SVM

Máy vector hỗ trợ SVM là một trong những thuật toán phân lớp phổ biến và hiệu quả. Thuật toán SVM nhằm mục đích tìm Siêu phẳng cận biên tối đa (MMH) bằng cách sử dụng các vectơ hỗ trợ và lề. MMH là siêu phẳng tốt nhất với khoảng cách lề lớn nhất được

sử dụng để phân tách dữ liệu một cách tối đa và chính xác cho mỗi lớp. Lề có thể được định nghĩa là khoảng cách ngắn nhất của siêu phẳng đến một phía của lề giống như khoảng cách siêu phẳng đến phía bên kia của lề, miễn là cả hai lề đều ở vị trí song song với siêu phẳng (Han và cộng sự, 2012).

Đánh giá hiệu suất của mô hình

Có nhiều phương pháp để đánh giá hiệu suất của một mô hình phân lớp. Nghiên cứu này sử dụng các chỉ tiêu thường được sử dụng là: Accuracy, Precision, Recall và F1-score. Các chỉ tiêu này được xác định dựa trên ma trận nhầm lẫn (Hình 1).

Phương pháp nghiên cứu

Quy trình nghiên cứu bắt đầu bằng việc thu thập dữ liệu. Tiếp theo là giai đoạn gán nhãn dữ liệu để áp dụng các thuật toán học có giám sát. Kế đến, dữ liệu về các bình luận sẽ được tiền xử lý nhằm đáp ứng yêu cầu về dữ liệu trước khi tiến hành vector hóa dữ liệu để đưa vào huấn luyện các mô hình phân lớp cảm xúc. Việc huấn luyện các mô hình phân lớp cảm xúc được thực hiện bằng các thuật toán phân lớp trong học máy. Cuối cùng, việc phân tích và đánh giá kết quả phân lớp được thực hiện bằng cách sử dụng các chỉ số Accuracy, Precision, Recall và F1-Score. Các bước trong quy trình này chủ yếu được thực hiện trên cơ sở sử dụng ngôn ngữ lập trình Python (Hình 2).

KẾT QUẢ NGHIÊN CỨU VÀ THẢO LUẬN

Kết quả nghiên cứu

Thu thập dữ liệu

Nghiên cứu sử dụng số liệu được thu thập từ website Tripadvisor (www.tripadvisor.com). Đây là trang web chứa thông tin về các khách sạn, nhà hàng và các địa điểm tham quan, cũng như phản hồi của du khách khi trải nghiệm dịch vụ tại những địa điểm này. Tripadvisor cung cấp nhiều thông tin liên quan đến khách sạn; tuy nhiên, trong phạm vi của nghiên cứu này, với mục đích chính là phân lớp cảm xúc của du khách dựa trên các bình luận, nghiên cứu chỉ tập trung thu thập dữ liệu về nội dung bình luận (Review) và đánh giá tương ứng (Rating) của du khách (Hình 3). Bằng cách sử dụng kết hợp 2 thư viện hỗ trợ thu thập dữ liệu trong Python là: Selenium và Scrapy, một bộ dữ liệu về các bình luận và đánh giá của du khách đối với các khách sạn ở tỉnh Thừa Thiên Huế đã được thu thập. Bộ dữ liệu gồm 22,557 dòng với 2 trường dữ liệu.

Gán nhãn dữ liệu

Nghiên cứu áp dụng phân lớp cảm xúc dựa theo điểm đánh giá (Rating) của du khách để gán nhãn theo quy tắc sau: nếu Rating > 3, thì gán nhãn tích cực (pos); nếu Rating = 3, thì gán nhãn trung lập (neu) và Rating < 3, thì gán nhãn tiêu cực (neg). Hình 4 minh hoạ 10 dòng ngẫu nhiên trong bộ dữ liệu sau khi gán nhãn cho biến Cảm xúc của du khách (Sentiment).

Tiền xử lý dữ liệu

Để đảm bảo cho việc huấn luyện mô hình, nghiên cứu tiến hành tiền xử lý dữ liệu. Ngoài việc xóa các

HÌNH 5: MINH HỌA BỘ DỮ LIỆU SAU TIỀN XỬ LÝ

Review	Rating	Sentiment	Rev
fantastic room excellent staff and fabulous ex...	5.0	pos	fantast room excel staff fabul excurs outsid c...
we stayed at sanna la residence hue during th...	5.0	pos	stay sanna la resid hue tet holiday group 13...
im just back from two nights at the sanna la ...	5.0	pos	im back two night sanna la resid delight stay...
we stayed here for four nights in april 2021 o...	5.0	pos	stay four night april 2021 obvious covid pande...
we spent a few days exploring hu and one were s...	5.0	pos	spend day explor hu happ eventh hotel comfor...
we were pleased with the room compared to othe...	4.0	pos	pleas room compar other price vang help staff
we were met by the owner in the street he conv...	4.0	pos	met owner street convinc u choos hotel disappo...
stayed for two nights and chose the penthouse ...	4.0	pos	stay two night choos penthous option fun birth ...
guide book recommendation we got a big airy roo...	3.0	neu	guidebook recommend get big airi room top floo...
we stayed at binh duong 3 for three nights in ...	4.0	pos	stay binh duong 3 three night sale decent a...

HÌNH 6: MÃ LỆNH VECTOR HÓA VÀ CÂN BẰNG BỘ DỮ LIỆU

```
### create an object for tfidfvectorizer
from sklearn.feature_extraction.text import TfidfVectorizer
tfidfconverter = TfidfVectorizer(max_features=500, min_df=0.01, max_df=0.9)
tfidf = tfidfconverter.fit_transform(df['Rev']).toarray()

### make train and test set
X_train, X_test, y_train, y_test = train_test_split(tfidf,df['Sentiment'],test_size=0.2,random_state=42)

# use SMOTE to balance the dataset
from imblearn.over_sampling import SMOTE
oversample = SMOTE()
X_train, y_train = oversample.fit_resample(X_train, y_train)
```

HÌNH 7: MÃ LỆNH HUẤN LUYỆN MÔ HÌNH PHÂN LỚP BẰNG THUẬT TOÁN KNN

```
# kNN modeling:

from sklearn.neighbors import KNeighborsClassifier

kNNmodel=KNeighborsClassifier(n_neighbors=3)

kNNmodel.fit(X_train, y_train)
```

HÌNH 8: MÃ LỆNH HUẤN LUYỆN MÔ HÌNH PHÂN LỚP BẰNG THUẬT TOÁN NAÏVE BAYES

```
# Support Vector Machine modeling

from sklearn import svm

SVMmodel = svm.SVC(kernel='linear') # Linear Kernel

SVMmodel.fit(X_train, y_train)
```

HÌNH 9: MÃ LỆNH HUẤN LUYỆN MÔ HÌNH PHÂN LỚP BẰNG THUẬT TOÁN SVM

```
# Support Vector Machine modeling

from sklearn import svm

SVMmodel = svm.SVC(kernel='linear') # Linear Kernel

SVMmodel.fit(X_train, y_train)
```

dòng chứa dữ liệu trống, quá trình tiền xử lý dữ liệu chủ yếu tập trung vào việc chuẩn hóa các bình luận. Quá trình này bao gồm các thủ tục: chuyển đổi các ký tự sang dạng in thường, xóa các ký tự đặc biệt, tách từ, xóa các từ dừng, loại bỏ các hậu tố của từ. Các thủ tục này được thực hiện bởi các phương thức xử lý văn bản, như: lower (), sub (), tokenize (), stem ()… trong thư viện NLTK của Python. Bộ dữ liệu sau khi chuẩn hóa và làm sạch gồm 21,929 dòng (Hình 5).

Vector hóa dữ liệu

Sau khi được tiền xử lý, các lời bình luận được chuyển thành các vector số để chuẩn bị dữ liệu cho việc huấn luyện mô hình. Mặc khác, do bộ dữ liệu không cân bằng (các bình luận tích cực chiếm đa số với 90% tổng số bình luận so với 6% bình luận trung

BẢNG: KẾT QUẢ ĐÁNH GIÁ HIỆU SUẤT CÁC MÔ HÌNH

Thuật toán	Precision	Recall	F1-score	Accuracy
KNN	89.72%	89.67%	89.70%	89.74%
Naïve Bayes	81.75%	78.59%	80.14%	82.14%
SVM	90.86%	89.92%	90.39%	91.53%

lập và 4% bình luận tiêu cực), nên thuật toán SMOTE được sử dụng lấy mẫu thêm cho các lớp thiểu số, đưa bộ dữ liệu về cân bằng (Hình 6).

Huấn luyện mô hình phân lớp cảm xúc

Nghiên cứu thực hiện phân lớp cảm xúc của du khách trên cơ sở sử dụng 3 thuật toán: phân lớp KNN (Hình 7), Naïve Bayes (Hình 8) và SVM (Hình 9). Tất cả các bước trong giai đoạn huấn luyện các mô hình phân lớp cảm xúc đều sử dụng thư viện Sklearn của Python.

Đánh giá hiệu suất mô hình

Kết quả đánh giá hiệu suất các mô hình phân lớp cảm xúc dựa vào bình luận của du khách được thể hiện như trong Bảng.

Thảo luận

Kết quả thực nghiệm cho thấy, hiệu quả phân lớp của 3 giải thuật: KNN, Naïve Bayes và SVM là tương đối tốt. Trong đó, giải thuật SVM có kết quả phân lớp tốt hơn với độ chính xác > 91%. Đây sẽ là một cơ sở rất tốt cho việc xây dựng hệ thống tự động phân lớp bình luận của du khách, góp phần giúp cho quá trình phân tích cảm xúc của du khách của các nhà quản lý và hoạch định chiến lược của các khách sạn được nhanh và chính xác hơn.

KẾT LUẬN

Trên cơ sở áp dụng kết hợp 2 thư viện hỗ trợ thu thập dữ liệu trong Python là Selenium và Scrapy, một bộ dữ liệu gồm 22,557 dòng về các bình luận và đánh giá của du khách đối với các khách sạn ở tỉnh Thừa Thiên Huế đã được thu thập. Dữ liệu sau khi được tiền xử lý với các kỹ thuật khác nhau đã được đưa vào huấn luyện bởi 3 mô hình học máy là: K láng giềng gần KNN, máy vector hỗ trợ SVM và Naïve Bayes để phân lớp cảm xúc của du khách. Kết quả cho thấy, mô hình SVM cho kết quả phân lớp tốt nhất so với các mô hình còn lại xét trên các chỉ số quan trọng, như: Accuracy, F1-score, Recall và Precision. Kết quả này về cơ bản phù hợp với một số nghiên cứu trước, đồng thời cũng cho thấy tiềm năng của các thuật toán học máy ứng dụng trong phân tích cảm xúc dựa trên những bình luận của du khách về dịch vụ khách sạn ở tỉnh Thừa Thiên Huế nhằm hiểu rõ hơn cảm nhận của du khách về các dịch vụ được cung cấp; từ đó, giúp các khách sạn có thêm cơ sở để hoạch định các chiến lược phát triển bền vững nhằm mục đích nâng cao chất lượng, cải thiện hình ảnh, giữ chân khách hàng cũng như thu hút khách hàng mới (Vuong và Nguyen, 2024). Về mặt lý luận, kết quả nghiên cứu cũng cung cấp thêm một cơ sở để xây dựng các chương trình ứng dụng phân tích và dự đoán cảm xúc của du khách nói riêng và khách hàng nói chung dựa trên các lời bình luận, đánh giá.❑

TÀI LIỆU THAM KHẢO

1. Agrawal, R. (2014), K-Nearest Neighborn for Uncertain Data, *International Journal of Computer Applications,* 105(11), 13-16.

2. Cambria E., D. Das, S. Bandyopadhyay, and A. Feraco (2017), Affective computing and sentiment analysis, *A Practical Guide to Sentiment Analysis,* 5, https://doi.org/ 10.1007/978-3-319-55394-8_1.

3. Han J., Kamber M., and Pei J. (2012), *Data Mining: concepts and techniques,* 3rd Edition, Morgan Kaufmann Publishers, Waltham.

4. Hoàng Bá Lộc, Hoàng Trọng Hùng (2022), Các nhân tố thúc đẩy hành vi hướng đến môi trường tại nơi làm việc của nhân viên dịch vụ khách sạn trên địa TP. Huế, *Tạp chí Khoa học Đại học Huế: Kinh tế và Phát triển*, 131(5C), 63-82.

5. Liu B. (2012), Sentiment Analysis and Opinion Mining, *Synthesis Lectures on Human Language Technologies*, 5(1), 1-167.

6. Miao F., Zhang P., Jin L., and Wu H. (2018), *Chinese News Text Classification Based on Machine learning algorithm*, In 2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), 48-51.

7. Moro S., P. Rita, and J. Coelho (2017), Stripping customers' feedback on hotels through data mining: The case of Las Vegas Strip, *Tourism Management Perspectives,* 23, 41-52, doi:10.1016/j.tmp.2017.04.003.

8. Nadeem Akhtara, Nashez Zubaira, Abhishek Kumara, and Tameem Ahmada (2017), Aspect based Sentiment Oriented Summarization of Hotel Reviews, *Procedia Computer Science,* 115, 563-57.

9. Thủ tướng Chính phủ (2023), *Quyết định số 1745/QĐ-TTg, ngày 30/12/2023 phê duyệt Quy hoạch tỉnh Thừa Thiên Huế thời kỳ 2021-2030, tầm nhìn đến năm 2050.*

10. Trần Thị Thu Hiền (2019), Chất lượng nguồn nhân lực tại các khách sạn 3 sao trên địa bàn tỉnh Thừa Thiên Huế - Góc nhìn từ nhà quản lý, *Tạp chí Khoa học Đại học Huế: Khoa học Xã hội Nhân văn*, 128(6D), 87-100.

11. Vũ Hữu Tiệp (2020), *Machine learning cơ bản*, truy cập từ https://machinelearningcoban.com/ebook.

12. Vuong, Q. H., Nguyen, M. H. (2024), *Better Economics for the Earth: A Lesson from Quantum and Information Theories*, AISDL.