

PHÂN TÍCH TÌNH CẢM NGƯỜI DÙNG TIKTOK ĐỂ KINH DOANH: SỬ DỤNG PHƯƠNG PHÁP LONG SHORT-TERM MEMORY

● ĐỖ HOÀNG NAM - PHẠM ĐÌNH TÀI

TÓM TẮT:

Mục đích của nghiên cứu này là sử dụng phân tích tình cảm để đánh giá tiềm năng kinh doanh, giải trí trên TikTok. Dữ liệu từ các bài đánh giá trên Google Play, App Store được phân tích bằng thuật toán long short-term memory (LSTM). Nghiên cứu này cho thấy triển vọng kinh doanh, giải trí, định hướng quảng cáo của các công ty, người dùng trên TikTok, đồng thời thấy được những khó khăn, thách thức ảnh hưởng đến dự đoán tình cảm.

Từ khóa: cửa hàng TikTok, kinh doanh trực tuyến, phân tích tình cảm, thương mại điện tử, thuật toán LSTM.

1. Đặt vấn đề

Khi thiết bị thông minh phổ biến người dùng mạng xã hội tăng nhanh chóng, đặc biệt các lĩnh vực thương mại điện tử, giải trí, quảng cáo, người dùng có nhiều độ tuổi khác nhau nên tình cảm/cảm xúc trước các nội dung trên nền tảng mạng xã hội TikTok cũng khác nhau. Đặc điểm quan trọng khi người dùng sử dụng mạng xã hội kéo dài sẽ khuyến khích họ mua hàng trực tuyến, tiếp nhận thông tin quảng cáo giới thiệu sản phẩm hay dịch vụ. Ứng dụng mạng xã hội TikTok được tải xuống nhiều nhất trên App Store và Google Play, mang đến cho người dùng trải nghiệm.

Một trong những ứng dụng mạng xã hội xếp hạng thứ 5 trên toàn thế giới năm 2023 là TikTok [1]. Ứng dụng này đã cách mạng hóa cách chúng ta xem các video ngắn, cạnh tranh trực tiếp với Facebook. Những ai thích những video ngắn mang tính giải trí và hài hước có thể đã quen thuộc với ứng dụng này, đặc biệt là giới trẻ. Tuy nhiên, không phải ai cũng quan tâm đến TikTok, một số người

dùng khen ngợi nó, trong khi những người khác chỉ trích nó. Mục tiêu chính của việc thực hiện phân tích tình cảm người dùng trên ứng dụng TikTok là:

- ❖ Để đánh giá cảm xúc của người dùng về ứng dụng, bao gồm đánh giá và nhận xét của người dùng.

- ❖ Để hiểu rõ hơn về cảm nhận của người dùng về nội dung trên ứng dụng, liệu họ có trải nghiệm tích cực hay tiêu cực cũng như những điều họ thích hay không thích.

- ❖ Để các nhà phát triển sử dụng nó để cải thiện chức năng của ứng dụng, giải quyết các vấn đề và đáp ứng các mối quan tâm của người dùng.

- ❖ Dành cho người kinh doanh điều chỉnh chiến lược tiếp thị của họ trên TikTok.

Trong thời đại kỹ thuật số, mạng xã hội (MXH) là nơi mọi người có thể chia sẻ cảm xúc/tình cảm về nhiều vấn đề với nhiều chủ đề khác nhau, từ tổng quát đến cá nhân. Người kinh doanh có thể tìm hiểu về xu hướng, lắng nghe ý kiến của khách hàng và hiểu họ muốn gì từ một sản phẩm. Phân tích, trích

xuất ý kiến của công chúng hoặc người tiêu dùng. Với hàng loạt công cụ, phương pháp và công nghệ có sẵn trực tuyến, MXH thúc đẩy giao tiếp kỹ thuật số mở. Phân tích dữ liệu MXH bao gồm việc thu thập và trích xuất dữ liệu hiện có, cấu trúc dữ liệu theo các mẫu cụ thể và sau đó đưa ra dự đoán. [2]

Cửa hàng TikTok là một nền tảng mua sắm trực tuyến kết hợp thương mại điện tử với trải nghiệm trên MXH, cho phép người sáng tạo, thương hiệu và người bán quảng cáo và bán các mặt hàng trực tiếp thông qua ứng dụng TikTok. Các video trong nguồn cấp dữ liệu, phát trực tiếp và trang giới thiệu sản phẩm đều có sẵn cho các hoạt động bán hàng.

2. Phương pháp nghiên cứu

Nghiên cứu này mô tả một đối tượng cần nghiên cứu và kết quả phân tích tình cảm (SA) cũng như các bài kiểm tra sẽ được thực hiện. Phương pháp nghiên cứu được áp dụng là phương pháp tiếp cận kỹ thuật định lượng, với thuật toán LSTM.

2.1. Hướng tiếp cận nghiên cứu

Dựa trên quy trình trong Hình 1, nghiên cứu bắt đầu bằng việc thu thập dữ liệu bằng kỹ thuật quét web bằng Google-Play-Scraper. Dữ liệu thu thập được sau đó được lọc bằng từ khóa TikTok shop (cửa hàng TikTok) để có được dữ liệu phù hợp với nhu cầu của nghiên cứu này. Dữ liệu sẽ được gắn nhãn để phân biệt giữa các danh mục. (Hình 1)

Sau khi dán nhãn, dữ liệu sẽ chuyển sang giai đoạn tiền xử lý để làm sạch dữ liệu. Bước này rất quan trọng, vì nó chuyển đổi dữ liệu thô thành định dạng phù hợp để phân tích. Phương thức word2vec sẽ được sử dụng để trích xuất các tính năng từ bộ dữ liệu được xử lý trước. Word2vec hoạt động dựa trên nguyên tắc deep learning (học sâu) biểu diễn các từ dưới dạng vectơ; hơn nữa, việc sử dụng word2vec để trích xuất đặc điểm thường mang lại độ chính xác phân loại cao hơn. Sau đó, thuật toán

LSTM sẽ được sử dụng để đánh giá tình cảm. Phương pháp LSTM [3] được tin dùng, vì nó bảo tồn thông tin trên các chuỗi dài hơn và giải quyết vấn đề độ dốc biến mất do có cơ chế lưu data trên các node. Mỗi từ, được biểu diễn dưới dạng một vectơ (vectơ từ), được đưa vào đơn vị LSTM một cách tuần tự, theo thứ tự của tệp đánh giá. Sau khi nhận được kết quả LSTM, họ sẽ được đánh giá bằng cách sử dụng Ma trận kết hợp để xác định độ chính xác, độ chính xác, khả năng thu hồi và điểm F1. Sau khi dữ liệu được thu thập chính xác, sẽ được phân tích để đưa ra kết luận.

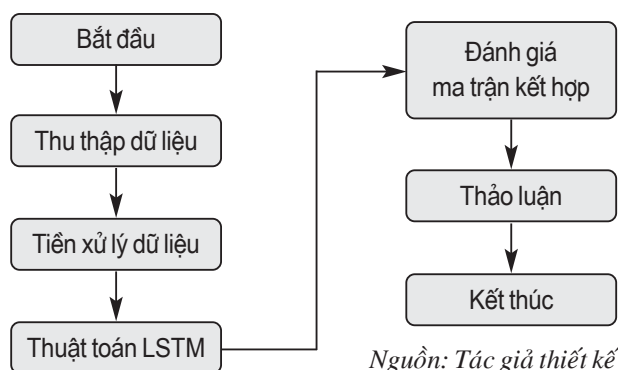
2.2. Phương pháp thu thập dữ liệu

Trong bài báo này, chúng tôi sử dụng kỹ thuật thu thập dữ liệu bằng từ khóa “TikTok shop” để thu thập dữ liệu. Dữ liệu được thu thập bằng Python, được hỗ trợ bởi thư viện Google-Play-Scraper. Bước này là phần quan trọng vì nếu không có dữ liệu liên quan, nó có thể dẫn đến phân tích không chính xác. Tập dữ liệu mà chúng tôi sẽ sử dụng là tập hợp các bài đánh giá TikTok được tải xuống từ Kaggle [4] dựa trên xếp hạng trên Cửa hàng Google Play. Sử dụng Python để thực hiện phân tích cảm xúc khi đánh giá TikTok để xem mọi người phản ứng thế nào với ứng dụng này. Sự sẵn có của các mô-đun Python khác nhau để phân tích mang lại cho chúng tôi lợi thế đáng kể và tăng tốc đáng kể quá trình nghiên cứu.

2.3. Tiền xử lý dữ liệu

Tiền xử lý dữ liệu là một bước quan trọng trong phân tích tình cảm. Nó liên quan đến việc làm sạch và chuẩn bị dữ liệu để phân tích nhằm đảm bảo tính chính xác và hiệu quả của kết quả phân tích cảm tính. Chúng tôi sẽ sử dụng một số thư viện được khởi tạo vào thời điểm này. Các kỹ thuật tiền xử lý bao gồm loại bỏ các ký tự không mong muốn, chẳng hạn như dấu câu và chuyển đổi tất cả văn bản sang chữ thường để giúp quá trình phân tích dễ dàng hơn. Một bước quan trọng khác trong quá trình tiền xử lý dữ liệu là loại bỏ các từ dừng, là những từ phổ biến không cần thiết trong việc xác định cảm xúc của văn bản. Các từ dừng có thể bao gồm các từ như “the”, “is” và “and”. Việc loại bỏ những từ này có thể giúp giảm tiếng ồn và cải thiện độ chính xác của phân tích tình cảm. Các kỹ thuật tiền xử lý khác bao gồm mã thông báo, chia văn bản thành các từ hoặc cụm từ riêng lẻ và từ gốc hoặc từ vệt, bao gồm việc giảm các từ về dạng cơ sở của chúng để tính đến các biến thể chính tả và cách sử dụng từ.

Hình 1: Hướng tiếp cận



Nguồn: Tác giả thiết kế

2.4. Thuật toán LSTM

Giai đoạn đầu tiên liên quan đến việc hiển thị dữ liệu tiền xử lý từ giai đoạn trước. Sau đó kiểm tra xem mỗi nhãn có bao nhiêu dữ liệu: âm có 674 và dương có 555 với kiểu int64. Hơn nữa, dữ liệu cho thấy có bao nhiêu dữ liệu không rỗng trong mỗi hàng; nếu số lượng nhỏ hơn thì hàng chứa giá trị null trên đối tượng Khung dữ liệu từ thư viện Pandas. Mỗi hàng chứa hai cột được gắn nhãn như sau: "nội dung" và "giá trị". Cả hai cột đều có kiểu dữ liệu đối tượng (chuỗi). Hơn nữa, để tránh trang bị quá mức, dữ liệu được phân chia dựa trên nhãn. Nhãn phủ định là một ví dụ, với 539 trường hợp để huấn luyện và 135 trường hợp để kiểm tra. Đối với nhãn tích cực, có 444 trường hợp được chỉ định để huấn luyện và 111 trường hợp để kiểm tra. Sau khi phân tách, các tập dữ liệu được kết hợp lại cho mục đích huấn luyện và kiểm tra. Dữ liệu huấn luyện tổng hợp bao gồm 539 nhãn âm và 444 nhãn dương. Ngược lại, dữ liệu thử nghiệm hợp nhất bao gồm 135 nhãn âm tính và 111 nhãn dương tính. Bước tiếp theo liên quan đến việc tạo mô hình Word2Vec sau khi hợp nhất dữ liệu. Mô hình này được đào tạo về dữ liệu huấn luyện để tạo ra các vectơ biểu diễn từ bằng cách sử dụng các tham số như kích thước nhúng, kích thước cửa sổ và phương pháp học. Văn bản được vector hóa bằng TensorFlow, kết hợp với bộ mã thông báo để chuyển đổi thành một chuỗi số nguyên. Trình tự này sau đó được chuyển đổi thành mảng NumPy 2D. Phương thức chuỗi pad chuyển đổi danh sách thành mảng NumPy 2D này, trong khi văn bản thành chuỗi chuyển đổi từng văn bản trong nội dung văn bản thành một chuỗi số nguyên. (Hình 2)

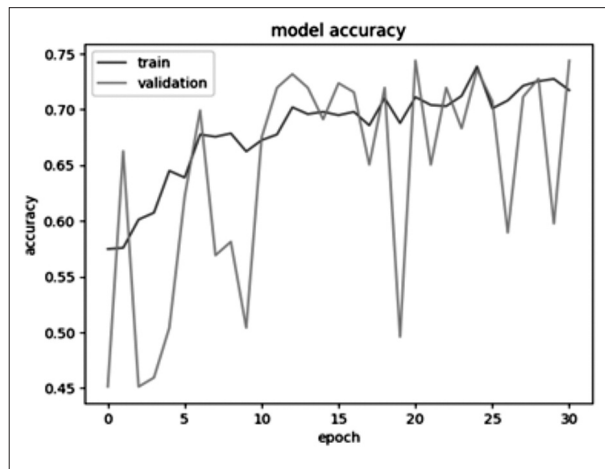
2.5. Đánh giá ma trận kết hợp

Kết quả đánh giá ma trận nhầm lẫn từ mô hình LSTM trên Hình 3. Giá trị chính xác của dữ liệu huấn luyện và xác nhận là 74% dựa trên quy trình được đề cập ở trên. Ma trận nhầm lẫn được tạo ra có 4 góc phần tư khác nhau:

- Tiêu cực sai (Góc phần tư trên cùng bên trái)
- Tích cực giả (Góc phần tư trên cùng bên phải)
- Tiêu cực thật sự (Góc phần tư dưới cùng bên trái)
- Tích cực thật sự (Góc phần tư dưới cùng bên phải).

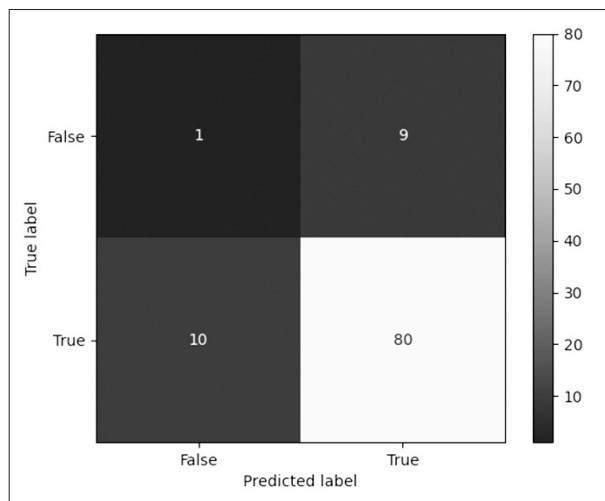
Đúng có nghĩa là các giá trị được dự đoán chính xác, Sai có nghĩa là đã xảy ra lỗi hoặc dự đoán sai.

Hình 2: Kết quả



Nguồn: Tác giả train mô hình

Hình 3: Ma trận nhầm lẫn (Confusion Matrix)



Nguồn: Tác giả thực nghiệm

3. Kết luận

Kết quả nghiên cứu từ ma trận kết hợp chỉ ra rằng thuật toán LSTM khi ghép với word2vec đạt độ chính xác. Dựa trên thông tin được cung cấp, triển vọng kinh doanh của cửa hàng TikTok có thể gặp nhiều thách thức. Tình cảm tiêu cực do người dùng thể hiện và độ chính xác tương đối thấp của mô hình phân tích tình cảm cho thấy các lĩnh vực tiềm năng cần cải thiện cho doanh nghiệp. Tuy nhiên, trong khi phân tích triển vọng kinh doanh, việc kiểm tra các yếu tố khác như mức độ hài lòng của khách hàng, nhu cầu thị trường, cạnh tranh, chất lượng sản phẩm, giá cả, giữ chân khách hàng

và kỹ thuật tiếp thị rất quan trọng. Chúng tôi cũng khuyến nghị thực hiện phân tích bổ sung, thu thập thêm dữ liệu và tích hợp những phát hiện này với

thông tin chi tiết liên quan đến hoạt động kinh doanh để có được kiến thức sâu sắc hơn về tiềm năng của cửa hàng TikTok ■

TÀI LIỆU THAM KHẢO:

1. Truc Hoang (2023). Tổng quan về thị trường TikTok 2023. Truy cập tại <https://megadigital.ai/vi/blog/tong-quan-xu-huong-thi-truong-tiktok/>
2. S. Sharma and A. Jain, (2020). Role of sentiment analysis in social media security and analytics. Wiley Interdisciplinary Reviews,10(5). <https://doi.org/10.1002/widm.1366>
3. W. Widayat, (2021). Analisis Sentimen Movie Review menggunakan Word2Vec dan metode LSTM Deep Learning. Jurnal Media Informatika Budidarma, 5(3), 1018-1026,
4. S. K. GANESH. "TikTok - Google Play Store Review." <https://www.kaggle.com/datasets/shivkumarganesh/tiktok-google-play-store-review>

Ngày nhận bài: 20/01/2024

Ngày phản biện đánh giá và sửa chữa: 6/02/2024

Ngày chấp nhận đăng bài: 20/02/2024

Thông tin tác giả:

ThS. ĐỖ HOÀNG NAM¹

ThS. PHẠM ĐÌNH TÀI¹

¹Trường Đại học Nguyễn Tất Thành

SENTIMENT ANALYSIS OF TIKTOK USERS FOR BUSINESS: USING LONG SHORT-TERM MEMORY METHODS

● Master. **DO HOANG NAM¹**

● Master. **PHAM DINH TAI¹**

¹ Nguyen Tat Thanh University

ABSTRACT:

In this study, emotional analysis was used to evaluate TikTok's business and entertainment potential. Long Short-Term Memory (LSTM) algorithms were employed to analyze data from reviews on Google Play and the App Store. This study showed the business, entertainment, and advertising prospects of companies and users on TikTok and also pointed out the difficulties and challenges that affect sentiment prediction.

Keywords: TikTok store, online business, entiment analysis, e-commerce, LSTM algorithms.