

DOI:10.22144/ctu.jvn.2021.169

PHÂN LOẠI Ý KIẾN CÔNG CHÚNG VỀ SỰ KIỆN XÃ HỘI

Nguyễn Văn Toàn^{1*}, Trần Thanh Điền², Nguyễn Tí Hon³, Phạm Thanh Sơn⁴, Trần Lan Anh¹ và Lê Thanh Sang⁵

¹Phòng Tổ chức-Cán bộ, Trường Đại học Cần Thơ

²Ban Quản lý Dự án ODA, Trường Đại học Cần Thơ

³Khoa Phát triển Nông thôn, Trường Đại học Cần Thơ

⁴Phòng Tài Chính, Trường Đại học Cần Thơ

⁵Trung tâm Thông tin và Quản trị mạng, Trường Đại học Cần Thơ

*Người chịu trách nhiệm bài viết: Nguyễn Văn Toàn (email: nvtoan@ctu.edu.vn)

Thông tin chung:

Ngày nhận bài: 30/08/2021

Ngày nhận bài sửa: 01/11/2021

Ngày duyệt đăng: 25/12/2021

Title:

Classification of community opinion about social events

Từ khóa:

Phân tích dư luận xã hội, bình luận (comments); SVM; Neural Network

Keywords:

Analyzing comments, SVM, neural network

ABSTRACT

Public opinion on social media has been paid much attention in new era of information technology. It has crucial impact on daily life and the operations of organizations. For that reason, analyzing public opinion will help the organizations to appropriately adjust their behavior. The objective of this study is to evaluate public opinion by analyzing comments on social network (Facebook) related to Can Tho University; then to help the university make appropriate adjustment policies. First of all, the public opinions (comments) were collected, divided into identify consensus (same direction) or disagree (opposite) opinions to the university activities. Then, two machine learning models, SVM and Neural Network, were installed for training, running experiments and comparing the results. Experimental results show that the Neural Network machine learning model has a quite high evaluation result of 85%, and the SVM model obtains 83% of accuracy.

TÓM TẮT

Dư luận xã hội, đặc biệt là dư luận được thể hiện trên các trang mạng xã hội, là vấn đề đang được quan tâm hiện nay; nó có tác động mạnh mẽ đến đời sống hàng ngày của mọi người, đến hoạt động của các tổ chức, doanh nghiệp và cơ quan nhà nước, chính phủ. Do đó, phân tích dư luận xã hội có ý nghĩa rất quan trọng đối với sự điều chỉnh cách ứng xử của cá nhân và tổ chức. Mục tiêu của nghiên cứu này là đánh giá dư luận xã hội thông qua phân tích thông tin các bình luận trên các trang mạng xã hội (cụ thể là Facebook) liên quan đến Trường Đại học Cần Thơ; từ đó giúp nhà trường đưa ra các chính sách điều chỉnh cho phù hợp. Để thực hiện điều này, 5.848 dư luận (bình luận) có liên quan đến các hoạt động của nhà trường được thu thập, với sự trợ giúp của những người có chuyên môn phân loại các bình luận thành các ý kiến đồng thuận (cùng chiều) hoặc không đồng thuận (trái chiều). Sau đó, hai mô hình máy học là SVM (Support Vector Machine) và Neural Network được sử dụng để huấn luyện, đánh giá thực nghiệm và so sánh độ chính xác và tin cậy nhằm lựa chọn được mô hình phù hợp cho việc xây dựng công cụ tự phân loại các bình luận trên mạng xã hội. Kết quả thực nghiệm cho thấy mô hình máy học Neural Network có kết quả đánh giá khá cao là 85%, và mô hình dựa trên giải thuật SVM đạt 83%.

1. GIỚI THIỆU

Dur luận xã hội là tập hợp các ý kiến cá nhân của của nhiều người trong xã hội về một chủ đề nào đó (Phạm Chiên Khu, 2011) và tất cả các ý kiến đều được đối xử như nhau bất kể cá nhân có hiểu biết về vấn đề đang thảo luận hay không. Thông thường, dur luận xã hội có xu hướng nghiêng về ý kiến được nhiều người đồng thuận nhất (nghiêng về số đông). Các ý kiến thể hiện sự phán xét đánh giá, sự nhận định (bằng lời hoặc không bằng lời), phản ánh ý nghĩa của các thực tế, quá trình, hiện tượng, sự kiện đối với các tập thể, xã hội nói chung và thái độ công khai hoặc che đậy của các nhóm xã hội lớn nhỏ đối với các vấn đề của cuộc sống xã hội. Dur luận xã hội có ảnh hưởng đến các lợi ích chung của họ, có tầm quan trọng và có tính cấp bách, đòi hỏi phải có ý kiến nhận xét, đánh giá và đưa ra những phương hướng cụ thể. Đó có thể là vấn đề chính trị, kinh tế, văn hóa, xã hội, hay đạo đức (Nguyễn Văn Dương, 2021).

Trong thời đại công nghệ thông tin hiện nay, việc sử dụng các phương tiện truyền thông xã hội giúp cho việc chia sẻ thông tin, thể hiện quan điểm cá nhân trở nên dễ dàng, nhanh chóng và rộng khắp. Các thông tin mạng xã hội có tác động rất lớn đến mọi hoạt động của cá nhân và tổ chức (cả tích cực và tiêu cực). Để bắt kịp xu thế nhằm hạn chế những ảnh hưởng tiêu cực đồng thời phát huy những điểm mạnh của mạng xã hội, các tổ chức cần phải phân tích và đánh giá các thông tin từ mạng xã hội liên quan đến hoạt động của tổ chức mình. Điều quan trọng là phải có giải pháp thu thập và phân tích thông tin tự động (phân loại được các nhận xét: tích cực hay tiêu cực). Việc thu thập và phân tích thông tin có ý nghĩa rất quan trọng đối với các nhà quản lý. Đặc biệt, khi đưa ra một chính sách hoặc một quy định cụ thể thì họ cũng muốn biết xem phản ứng của công chúng về chính sách hay quy định đó như thế nào từ đó có thể điều chỉnh cho phù hợp.

Phân tích dur luận xã hội là thực hiện xác định đối tượng dur luận xã hội và chủ thể dur luận xã hội thông qua việc nghiên cứu, nắm bắt dur luận xã hội nhằm đánh giá tác động của dur luận về ý thức pháp luật. Việc nghiên cứu, tìm hiểu về dur luận có ý nghĩa lớn về mặt vật chất và tinh thần của con người, là phương tiện quan trọng và mạnh mẽ để con người phát huy quyền làm chủ và tính tích cực của công dân đối với đất nước trong đó. **Đối tượng của dur luận xã hội** không phải là mọi thực tế xã hội nói chung mà chỉ là những vấn đề được cộng đồng xã hội quan tâm tới vì nó có liên quan tới các nhu cầu lợi ích về vật chất hay về tinh thần của họ. **Chủ thể**

của dur luận là cộng đồng người hay nhóm người mang dur luận xã hội. Nắm bắt dur luận xã hội là khả năng dự báo chính xác tình hình tâm trạng, tư tưởng, thái độ, nguyện vọng của các tầng lớp xã hội trước các sự kiện, hiện tượng, vấn đề xã hội. Các cơ quan lãnh đạo, quản lý đất nước khó có thể ban hành được các chủ trương, quyết sách sát thực, có sức sống, có tính khả thi nếu không nắm chắc được tâm trạng, tư tưởng của đối tượng có liên quan đến các chủ trương, quyết sách đó. Làm tốt công tác nắm bắt dur luận xã hội ở cơ sở là điều kiện vô cùng quan trọng, một mặt, các cá nhân trong tổ chức phát huy quyền làm chủ và mở rộng nền dân chủ xã hội chủ nghĩa; mặt khác, giúp cho cấp uỷ và các tổ chức công đoàn, đoàn thanh niên các cấp nắm bắt ý nguyện của viên chức, người lao động và sinh viên, để có những đối sách, hoàn thiện các chủ trương, đường lối, chính sách một cách thích hợp. Đó cũng chính là điều kiện tiên quyết để có thể sử dụng sức mạnh dur luận quần chúng vào việc quản lý xã hội. Các nội dung của dur luận xã hội được thu thập thường quan tâm đến: (i) thái độ (đồng tình hay không đồng tình; ủng hộ hay không ủng hộ,...), tâm trạng (phân khởi, chán nản hay thờ ơ,...) đối với những chủ trương, chính sách mới ra của Đảng, Nhà nước và cấp uỷ, chính quyền các cấp; (ii) dur luận của viên chức, người lao động (VC, NLĐ) về việc tín nhiệm, hay không tín nhiệm đối với các cán bộ lãnh đạo, đối với bộ máy điều hành quản lý, đặc biệt là khi có những vụ việc bất thường liên quan đến việc thực hiện các chế độ chính sách, đến mối quan hệ lao động trong cơ quan đơn vị; (iii) tâm trạng và mức độ ổn định của người học đối với điều kiện học tập, môi trường sinh sống và các sự kiện khác có liên quan đến quá trình trước, trong và sau khi học tại trường.

Để phân loại được dur luận xã hội thông qua việc nắm bắt thông tin trên các trang mạng xã hội thì chúng ta cần phải phân loại, sắp xếp thông tin dur luận theo từng nội dung, phạm trù nhất định để xử lý, ví dụ: (1) dur luận về các sự kiện, vấn đề chính trị; (2) dur luận về các sự kiện, vấn đề tôn giáo, dân tộc; (3) dur luận về các sự kiện, vấn đề môi trường, sinh thái; (4) dur luận về các vấn đề kinh tế; (5) dur luận xã hội về các vấn đề quốc tế,... Trong nghiên cứu này, dur luận được phân loại dựa theo các bình luận của công chúng ủng hộ hay phản đối hoặc ý kiến khác về một nội dung có liên quan đến Trường Đại học Cần Thơ và đưa ra tỷ lệ của các nhận định đó.

Phân loại văn bản với nhiều từ khóa là vấn đề cơ bản trong lĩnh vực khai thác dữ liệu văn bản, giúp tiết kiệm thời gian trong việc tìm kiếm, tổng hợp thông tin và quản lý dữ liệu (Liu et al., 2009;

Birmingham et al., 2010). Phân loại văn bản được áp dụng trong các lĩnh vực khác nhau như truy xuất thông tin, lọc văn bản, bộ tổng hợp tin tức tự động và thư viện điện tử.

Đối với Trường Đại học Cần Thơ, việc nắm bắt được phản ứng của sinh viên, viên chức và những người có liên quan về một quy định hoặc quy chế mà Trường Đại học Cần Thơ ban hành là điều rất cần thiết. Việc sử dụng các công cụ hỗ trợ tìm kiếm có thể giúp theo dõi tin tức một cách hệ thống, đầy đủ, chi tiết về một sự kiện cụ thể và các luồng dư luận xã hội khác nhau về sự kiện đó. Các ý kiến nhận xét, bình luận từ các trang báo mạng chính thức và phi chính thức đều có thể chứa đựng những luồng dư luận xã hội nhất định. Từ đó, lãnh đạo Nhà trường có những phân tích, đánh giá và đưa ra điều chỉnh cho phù hợp với thực tiễn. Để phân tích được dư luận xã hội thông qua việc nắm bắt thông tin trên các trang mạng xã hội đặc biệt là Facebook (Facebook 2021) thì chúng ta cần phải phân loại, sắp xếp thông tin dư luận theo từng nội dung, phạm trù nhất định để xử lý. Tuy nhiên, với lượng thông tin được chia sẻ trên các phương tiện xã hội rất lớn và tăng lên từng ngày như hiện nay vượt khả năng theo dõi, giám sát của cá nhân. Vì những lý do nêu trên ứng dụng được xây dựng để phân loại nhận định của công chúng (nhận định ủng hộ và nhận định phản đối) liên quan đến các hoạt động của Trường Đại học Cần Thơ.

Có rất nhiều công trình dựa trên các kỹ thuật phân loại cổ điển như: Support Vector Machine (SVM), Naïve Bayes,... Gần đây, Vateekul and Koomsubha (2016) của Trường Đại học Chulalongkorn, Bangkok, Thailand áp dụng hai kỹ thuật học sâu: Bộ nhớ dài hạn (Long Short Term Memory - LSTM) và Mạng thần kinh chuyển đổi năng động (Dynamic Convolutional Neural Network - DCNN) để phân loại tình cảm của dữ liệu Twitter Thái Lan. Ngoài ra, Troussas et al. (2013) đã phân tích cảm xúc có thể hỗ trợ việc học ngôn ngữ, bằng cách kích thích quá trình giáo dục và kết quả thử nghiệm trên Naive Bayes Classifier.

Hai mô hình máy học SVM và Neural network được nghiên cứu và chọn mô hình có kết quả đánh giá cao hơn để xây dựng ứng dụng phân loại ý kiến nhận định của công chúng: đồng thuận hoặc không đồng thuận hoặc ý kiến khác vì các lý do: (1) So với các phương pháp máy học khác, SVM rất mạnh trong việc nhận dạng các mẫu tinh tế trong các bộ dữ liệu phức tạp (Aruna & Rajagopalan, 2011); (2) phân loại là một trong những lĩnh vực nghiên cứu và

ứng dụng tích cực nhất của mạng neuron (Zhang, 2000).

2. PHƯƠNG PHÁP NGHIÊN CỨU

Đề tài tiếp cận phương pháp nghiên cứu: Nghiên cứu lý thuyết - thực nghiệm - phân tích dữ liệu. Trong đó, phương pháp nghiên cứu lý thuyết thì dựa vào ý kiến của các chuyên gia để xác định loại ý kiến của công chúng bằng cách đọc nội dung và đọc từng bình luận kết hợp với các biểu tượng kèm theo đưa ra nhận định đồng ý, không đồng ý hoặc ý kiến khác. Phương pháp thực nghiệm thì huấn luyện mô hình máy học, so sánh các bộ tham số để chọn bộ tham số tối ưu của mô hình. Phương pháp phân tích dữ liệu dựa vào ứng dụng được xây dựng sử dụng tham số mô hình máy học đã huấn luyện được để phân tích dữ liệu thu thập được và đưa ra xu hướng dư luận xã hội về hoạt động của nhà trường.

2.1. Cài đặt thực nghiệm

Quy trình cài đặt thực nghiệm được thực hiện bao gồm các bước sau: Dữ liệu được thu thập từ các trang facebook của Trường Đại học Cần Thơ được tiền xử lý, loại bỏ html, sau đó gán nhãn phân loại, mỗi bình luận sẽ được gán duy nhất một nhãn là đồng thuận, không đồng thuận hoặc ý kiến khác. Mô hình huấn luyện phân lớp được thực hiện dựa trên tập gán nhãn để phân lớp và đưa ra kết quả phân loại bình luận. Mô hình phân lớp đã huấn luyện được sử dụng để phân lớp cho các bình luận mới.

2.2. Xử lý số liệu

Tập dữ liệu thu thập về xử lý (1) tách nội dung và mã của nội dung thành một tập dữ liệu về nội dung; (2) tách các bình luận có gán mã nội dung thành một tập dữ liệu riêng.

Bảng 1. Số liệu thu thập từ các trang facebook của Trường

Nguồn thu thập dữ liệu	Số mẫu tin nội dung	Số mẫu tin bình luận
Trang ctuDHCT	293	4.161
Trang ctu.Youh	278	4.232

Bảng 2. Tập dữ liệu được gán nhãn

Đồng thuận	Không đồng thuận	Ý kiến khác	Tổng
2.497	1.054	2.297	5.848

* Chú thích: đồng thuận (đồng ý, đồng tình); không đồng thuận (không đồng ý, không đồng tình)

Dữ liệu sau đó được vector hóa theo mô hình túi từ (Bag of word) để làm dữ liệu đầu vào cho hai mô hình máy học.

Một số quy tắc cơ bản khi gán nhãn sau: (1) đối với biểu tượng (icon) thể hiện sự cười tươi, rất hạnh phúc, ngạc nhiên, nháy mắt, trái tim, thiên thần, hôn, bình tĩnh như Bảng 3 thể hiện sự đồng ý. Đồng thời cũng phân biệt một số biểu tượng không đồng ý như biểu tượng buồn, khóc, bực điên, đang giận, ác quỷ,... Trường hợp không biết hoặc lưỡng lự như đang suy nghĩ, mơ hồ,... thì đưa vào lớp ý kiến khác. Phần gán nhãn biểu tượng được cụ thể hóa theo nguyên tắc ở Bảng 3. (2) Đối với một số từ không theo quy tắc Tiếng Việt nhưng khi đọc vào vẫn hiểu nghĩa là đồng ý ví dụ: làm đi, tới luôn, ok, thank, yes, yeah,.. hoặc kêu gọi người khác vào đọc nội dung đó. Trường hợp 'tag' người khác vào nhưng không viết gì thêm thì sẽ đưa vào ý kiến khác. Loại

bỏ một số câu trùng hoặc các bình luận lặp đi lặp lại nhằm mục đích kêu gọi người khác vào hoặc đề cập đến vấn đề khác không liên quan đến nội dung. Nếu trong một bình luận vừa có phần chữ vừa có biểu tượng thì ưu tiên đọc phần chữ để gán nhãn trước và tiếp theo sẽ gán phần biểu tượng

Ví dụ đối với nội dung: “[phần 2] chung kết olympic các môn lý luận chính trị và tư tưởng hồ chí minh trong sinh viên đại học cần thơ năm 2018”

- Ý kiến đồng thuận:
-
- Dạng văn bản: “và đây là kiến thức sử địa haha”
- Dạng icons: 👍 ,
- Ý kiến không đồng thuận:
- Dạng icons: 😬😬😬

Bảng 3. Các biểu tượng đặc trưng để gán nhãn

Nội dung	Biểu tượng
Các biểu tượng được gán nhãn đồng thuận	😊👍👏🙌🙏🙇🙄🙆🙇🙈🙉🙊🙋🙌🙍🙎🙏🙐🙑🙒🙓🙔🙕🙖🙗🙘🙙🙚🙛🙜🙝🙞🙟🙠🙡🙢🙣🙤🙥🙦🙧🙨🙩🙪🙫🙬🙭🙮🙯🙰🙱🙲🙳🙴🙵🙶🙷🙸🙹🙺🙻🙼🙽🙾🙿🚶🚷🚸🚹🚺🚻🚼🚽🚾🚿🛀🛁🛂🛃🛄🛅🛆🛇🛈🛉🛊🛋🛌🛍🛎🛏🛐🛑🛒🛓🛔🛕🛖🛗🛘🛙🛚🛛🛜🛝🛞🛟🛠🛡🛢🛣🛤🛥🛦🛧🛨🛩🛪🛫🛬🛭🛮🛯🛰🛱🛲🛳🛴🛵🛶🛷🛸🛹🛺🛻🛼🛽🛾🛿🚲🚴🚵🚶🚷🚸🚹🚺🚻🚼🚽🚾🚿🛀🛁🛂🛃🛄🛅🛆🛇🛈🛉🛊🛋🛌🛍🛎🛏🛐🛑🛒🛓🛔🛕🛖🛗🛘🛙🛚🛛🛜🛝🛞🛟🛠🛡🛢🛣🛤🛥🛦🛧🛨🛩🛪🛫🛬🛭🛮🛯🛰🛱🛲🛳🛴🛵🛶🛷🛸🛹🛺🛻🛼🛽🛾🛿
Các biểu tượng được gán nhãn không đồng thuận	😡😠😡😢😣😤😥😦😧😨😩😪😫😬😭😮😯😰😱😲😳😴😵😶😷😸😹😺😻😼😽😾😿👎👏🙄🙆🙇🙈🙉🙊🙋🙌🙍🙎🙏🙐🙑🙒🙓🙔🙕🙖🙗🙘🙙🙚🙛🙜🙝🙞🙟🙠🙡🙢🙣🙤🙥🙦🙧🙨🙩🙪🙫🙬🙭🙮🙯🙰🙱🙲🙳🙴🙵🙶🙷🙸🙹🙺🙻🙼🙽🙾🙿🚶🚷🚸🚹🚺🚻🚼🚽🚾🚿🛀🛁🛂🛃🛄🛅🛆🛇🛈🛉🛊🛋🛌🛍🛎🛏🛐🛑🛒🛓🛔🛕🛖🛗🛘🛙🛚🛛🛜🛝🛞🛟🛠🛡🛢🛣🛤🛥🛦🛧🛨🛩🛪🛫🛬🛭🛮🛯🛰🛱🛲🛳🛴🛵🛶🛷🛸🛹🛺🛻🛼🛽🛾🛿
Các biểu tượng gán nhãn ý kiến khác	😏😐😑😒😓😔😕😖😗😘😙😚😛😜😝😞😟😠😡😢😣😤😥😦😧😨😩😪😫😬😭😮😯😰😱😲😳😴😵😶😷😸😹😺😻😼😽😾😿👎👏🙄🙆🙇🙈🙉🙊🙋🙌🙍🙎🙏🙐🙑🙒🙓🙔🙕🙖🙗🙘🙙🙚🙛🙜🙝🙞🙟🙠🙡🙢🙣🙤🙥🙦🙧🙨🙩🙪🙫🙬🙭🙮🙯🙰🙱🙲🙳🙴🙵🙶🙷🙸🙹🙺🙻🙼🙽🙾🙿🚶🚷🚸🚹🚺🚻🚼🚽🚾🚿🛀🛁🛂🛃🛄🛅🛆🛇🛈🛉🛊🛋🛌🛍🛎🛏🛐🛑🛒🛓🛔🛕🛖🛗🛘🛙🛚🛛🛜🛝🛞🛟🛠🛡🛢🛣🛤🛥🛦🛧🛨🛩🛪🛫🛬🛭🛮🛯🛰🛱🛲🛳🛴🛵🛶🛷🛸🛹🛺🛻🛼🛽🛾🛿

3. KẾT QUẢ VÀ THẢO LUẬN

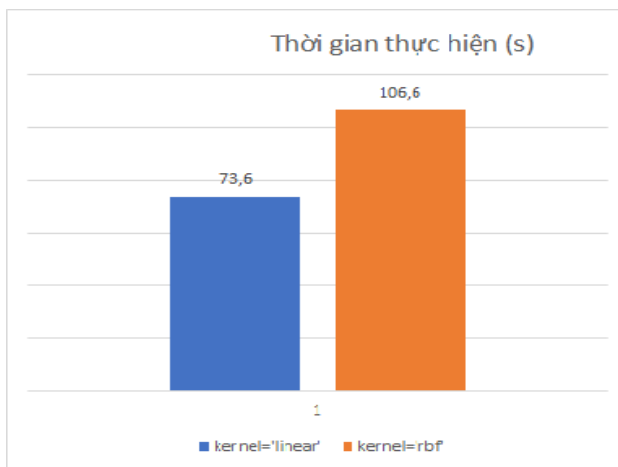
3.1. Các tham số thực nghiệm với mô hình máy học vector hỗ trợ (SVM)

Bằng thực nghiệm cho thấy khi sử dụng hàm nhân (kernel) với hai giá trị là 'linear' và 'rbf' kết hợp tham số c được thay đổi trong tập giá trị {0.01; 0.025; 0.5; 10; 100; 1e³ 5e³} và giá trị Gamma được

thay đổi khi chọn được tham số tối ưu và được kết hợp thay đổi {0.001; 0.01; 1; 2.5}. Các thông số khác có giá trị mặc định sẽ cho kết quả khả quan.

3.2. Đánh giá thời gian thực hiện với mô hình máy học vector hỗ trợ (SVM)

Khi huấn luyện mô hình SVM với hàm nhân bằng 'linear' có thời gian ngắn hơn khoảng 20s so với sử dụng hàm nhân bằng 'rbf' như Hình 1.

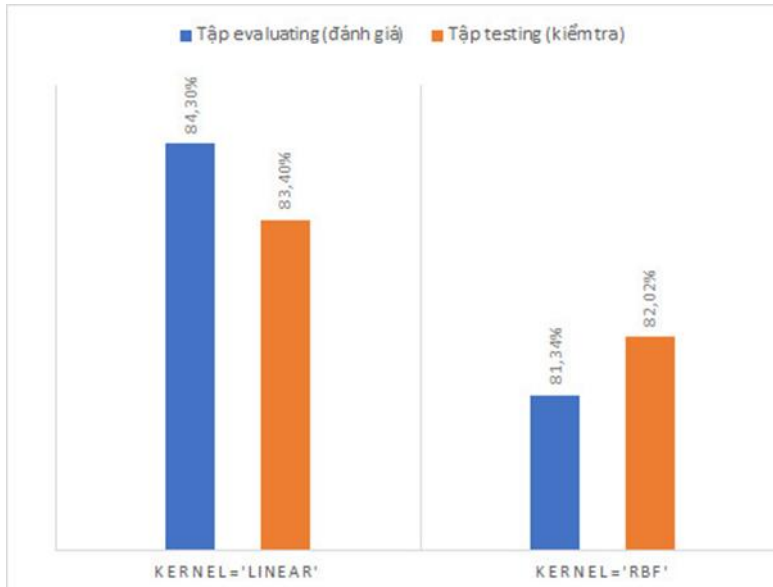


Hình 1. So sánh thời gian huấn luyện mô hình SVM bằng thay đổi hàm nhân

3.3. Đánh giá độ chính xác các mô hình SVM

Kết quả đánh giá độ chính xác (evaluate) theo Hình 2, cho thấy độ chính xác thay đổi khi thay đổi

hàm nhân “linear” cao hơn hàm nhân “rbf” (85% so với 82%).



Hình 2. So sánh độ chính xác của mô hình SVM bằng cách thay đổi hàm nhân

3.4. Đánh giá và kiểm tra thời gian thực hiện với mô hình Neural Network

Sử dụng mô hình Neural Network với bộ giải pháp (solve) gồm ('adam', 'sgd', 'lbfgd') và chỉ số tế bào lớp ẩn thay đổi từ 08 đến 512 theo cấp 2^n (n =

3, 4, 5, 6), lấy ngẫu nhiên (random) là 100, hàm kích hoạt ReLU mặc định. Kết quả huấn luyện với thời gian thực hiện (giây), kết quả kiểm tra trên tập đánh giá; kết quả kiểm tra trên tập test của ba bộ giải pháp trên được thể hiện ở Bảng 4, Bảng 5 và Bảng 6 như sau:

Bảng 4. Kết quả thực nghiệm huấn luyện mô hình Neural Network sử dụng bộ giải pháp solve='sgd'

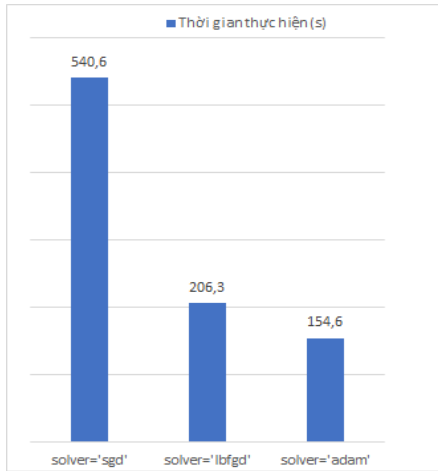
Số lớp ẩn (hidden layer size)	Kết quả đạt trên tập đánh giá (%)	Kết quả đạt trên tập kiểm tra (%)	Thời gian thực hiện (s)
8	71,10	72,90	80,4
16	72,80	73,10	82,4
32	73,10	75,00	100,7
64	81,30	84,90	89,3
128	81,80	84,80	115,3
256	72,40	73,60	271,3
512	72,60	73,40	540,6

Bảng 5. Kết quả thực nghiệm huấn luyện mô hình Neural Network sử dụng bộ giải pháp solve='lbfgd'

Số lớp ẩn (hidden layer size)	Kết quả đạt trên tập đánh giá (%)	Kết quả đạt trên tập kiểm tra (%)	Thời gian thực hiện (s)
8	82,50	85,60	95,5
16	82,30	85,80	90,3
32	81,60	86,00	93,6
64	82,53	86,30	131,6
128	82,02	85,53	219,9
256	82,51	85,96	205,0
512	82,53	85,96	206,3

Bảng 6. Kết quả thực nghiệm huấn luyện mô hình Neural Network sử dụng bộ giải pháp solve='adam'

Số lớp ẩn (hidden layer size)	Kết quả đạt trên tập đánh giá (%)	Kết quả đạt trên tập kiểm tra (%)	Thời gian thực hiện (s)
8	81,30	85,60	161,9
16	82,00	86,30	162,7
32	80,10	85,00	110,8
64	81,30	84,90	89,3
128	81,80	84,80	115,3
256	80,90	87,20	129,6
512	82,40	86,50	154,6



Hình 3. So sánh thời gian thực hiện của mô hình Neural Network

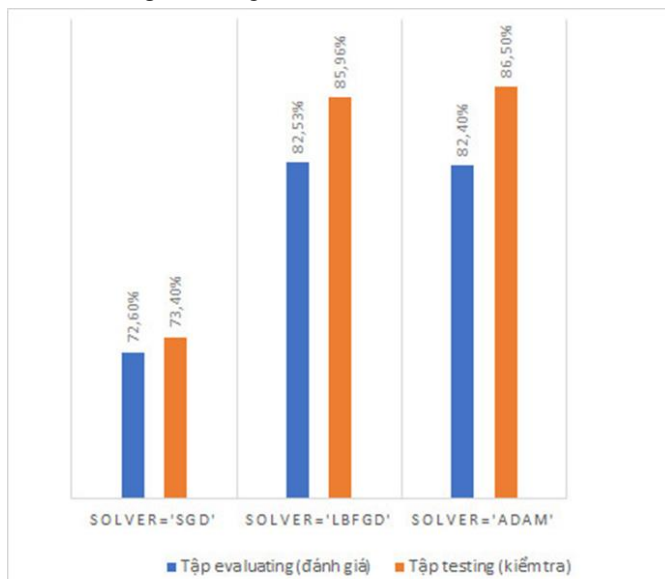
Theo thống kê thực nghiệm ở các Bảng 4, 5, 6 cho thấy mỗi bộ giải pháp (solver) khi chỉ số lớp ẩn càng cao thì thời gian thực hiện càng cao đồng thời

kết quả đánh giá trên tập kiểm tra và tập đánh giá cũng đạt tỷ lệ cao. Đề tài chọn chỉ số lớp ẩn 512 để thực hiện mô hình.

Theo Hình 4 thì dễ dàng nhìn thấy khi huấn luyện mô hình Neural Network với chỉ số lớp ẩn là 512 với bộ giải pháp (solver='adam') thì sẽ ít tốn thời gian nhất (khoảng 154 s). Trong khi thời gian thực hiện dài nhất (hơn 540,6 s) với bộ giải pháp là (solver='sgd').

3.5. Đánh giá độ chính xác các mô hình Neural Network

Khi đánh giá độ chính xác theo hình 5 ta thấy với bộ giải pháp solver='adam' và solver='lbfgd' đạt kết quả khá cao (khoảng 83% đối với tập đánh giá và 86% đối với tập kiểm tra). Trong khi bộ giải pháp solver='sgd' thì kết quả không khả quan (dưới 75%). Đề tài chọn mô hình Neural Network với bộ giải pháp (solver=adam) và chỉ số lớp ẩn (hidden_layer_size=512) để viết ứng dụng phân tích bình luận.



Hình 4. Đánh giá độ chính xác của mô hình Neural Network

3.6. Kết quả phân tích bình luận trên tập dữ liệu thực nghiệm khi thực hiện mô hình

Mỗi bài viết được thu thập về được thống kê gồm: ngày đăng, nội dung, số lượng bình luận, trong đó có số đồng thuận, số không đồng thuận và ý kiến khác với các tỷ lệ thống kê tương ứng. Người quản

trị có thể quan tâm đến những nội dung có số bình luận nhiều để hiểu rõ hơn hoặc quan tâm đến những nội dung có nhiều số không đồng thuận. Từ đó có thể điều chỉnh các chính sách cho phù hợp.

Ví dụ Bài viết số: 112 được thể hiện như Bảng 7.

Bảng 7. Kết quả phân tích bài viết số 112 của mô hình

Ngày đăng	Nội dung	Số lượng bình luận	Số đồng thuận	Số không đồng thuận	Số ý kiến khác
2018-11-19 14:32:32	“hiển máu cứu người một nghĩa cử cao đẹp. 😊 hiển máu tình nguyện 24/11/2018 (thứ bảy tuần này) - thời gian: 7h00 - 11h00 ngày 24/11/2018 - địa điểm: sảnh văn phòng đoàn trường - khu II Đại học Cần Thơ - link đăng ký vui lòng xem thông tin trước khi đăng ký để hiển máu an toàn bạn nhé”	18	4 22,22%	6 33,33%	8 44,44%

3.7. Kết quả kiểm tra phân loại bình luận

Mỗi bình luận tự động thu thập về sẽ có kết quả thống kê gồm: Mã bình luận (Comment code), Ngày đăng (comment date), tên người bình luận (Name of commenter), thông tin người bình luận, nội dung bình luận (Comment text), liên kết bình luận và đặc biệt là loại của bình luận (Comment kind) được thực hiện số 1, 2, 3 tương ứng với 1 là đồng thuận, 2 là không đồng thuận và 3 là ý kiến khác dựa theo mô hình máy học.

Ví dụ người quản trị muốn biết thông tin của tất cả các bình luận liên quan đến bài viết số 92; Ngày đăng: 2020-10-08T07:32:25.000Z; có nội dung là hướng dẫn đổi mật khẩu tài khoản hộp thư điện tử. Bài viết này có 10 bình luận (trong đó: có 03 đồng thuận, 03 không đồng thuận và 04 ý kiến khác) như sau:

Comment date: 2020-10-08T09:52:15.000Z

Name of commenter: Trần Đức Nam

Comment text: Ngày 12 mình mới đăng nhập vào hệ thống được phải không ạ

Comment kind: 2

Comment date: 2020-10-11T16:04:50.000Z

Name of commenter: Huỳnh Nhật Tâm

Comment text: Cho em hỏi có ai học CTCLC k ạ? Em học CTCLC thì ngày 26 mới lên trường theotkb phải k ạ? Hay ngày 12 vẫn lên trường ạ?

Comment kind: 3

Comment date: 2020-10-09T03:44:39.000Z

Name of commenter: Yến Milam

Comment text: Trường cho e hỏi đăng kí nhập học đùm được không ạ?

Comment kind: 3

Comment date: 2020-10-08T13:02:01.000Z

Name of commenter: Di Sản

Comment text: Mình k đổi được không ạ?

Comment kind: 2

Comment date: 2020-10-15T04:41:45.000Z

Name of commenter: Ngọc Nguyễn

Comment text: Ví dụ mình đổi mk rồi nhưng lỡ quên mk mới thì s ạ .Tại gmail này ko đổi mk như bình thường được.

Comment kind: 2

Comment date: 2020-10-12T03:16:00.000Z

Name of commenter: Huy Hiệp

Comment text: Phạm Sang chia sẻ này được nè

Comment kind: 1

Comment date: 2020-10-11T11:52:01.000Z

Name of commenter: Dk Kieu

Comment text:

Comment kind: 3

Comment date: 2020-10-08T13:15:21.000Z

Name of commenter: Duy Quang

Comment text: Quý Phú Cho đổi kia:((

Comment kind: 1

Comment date: 2020-10-11T11:25:41.000Z

Name of commenter: Nguyễn Thẩm

Comment text: .

Comment kind: 3

Comment date: 2020-10-09T08:21:24.000Z

Name of commenter: An Duy

Comment text: M.n ơi, cho mình hỏi, học bạ công chứng ở trường hay công chứng ở UB xã nơi mình sống v m.n

Comment kind: 1

4. KẾT LUẬN VÀ ĐỀ XUẤT

4.1. Kết luận

Đề tài đã cơ bản đáp ứng được các mục tiêu đề ra đó là phân loại dư luận dựa theo các bình luận về một nội dung bài viết có liên quan đến Trường Đại học Cần Thơ. Cụ thể, hai mô hình máy học là SVM và Neural Network được dùng để huấn luyện cho việc phân loại bình luận và chọn mô hình có kết quả cao nhất.

Chọn mô hình máy học Neural Network với bộ giải pháp (solver='adam') và chỉ số lớp ẩn (hidden_layer=512) có kết quả đánh giá khá cao trên 85% để xây dựng ứng dụng phân tích tỷ lệ đồng thuận, không đồng thuận hay ý kiến khác về nội dung một bài viết cụ thể được đăng trên các trang Facebook của Trường.

Thông qua mô hình của đề tài, các nhà quản lý có thể tiếp cận và phân tích được luồng thông tin liên quan đến một vấn đề cụ thể, từ đó đưa ra các

chính sách hoặc điều chỉnh chính sách về vấn đề đó cho phù hợp.

TÀI LIỆU THAM KHẢO

- Birmingham, A., & Smeaton, A. F. (2010, October). Classifying sentiment in microblogs: Is brevity an advantage?. In *Proceedings of the 19th ACM international conference on Information and knowledge management* (pp. 1833-1836).
- Nguyễn Văn Dương. (30.01.2021). *Dư luận xã hội là gì? Ảnh hưởng của dư luận xã hội đối với ý thức pháp luật*. <https://luatduonggia.vn/du-luan-xa-hoi/>
- Phạm Chiến Khu. (2021). *Cơ sở lý luận và thực tiễn của công tác nghiên cứu dư luận xã hội*. <https://www.tuyengiao.vn/ngghien-cuu/ngghiep-vu-cong-tac-tuyen-giao/co-so-ly-luan-va-thuc-tien-cua-cong-tac-ngghien-cuu-du-luan-xa-hoi-32565>
- Liu, F., Pennell, D., Liu, F., & Liu, Y. (2009, June). Unsupervised approaches for automatic keyword extraction using meeting transcripts. In *Proceedings of human language technologies: The 2009 annual conference of the North American chapter of the association for computational linguistics* (pp. 620-628).
- Vateekul, P., & Koomsubha, T. (2016, July). A study of sentiment analysis using deep learning techniques on Thai Twitter data. In *2016 13th International joint conference on computer science and software engineering (JCSSE)* (pp. 1-6). IEEE.
- Troussas, C., Virvou, M., Espinosa, K. J., Llaguno, K., & Caro, J. (2013, July). Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning. In *IISA 2013* (pp. 1-6). IEEE.
- Aruna, S., & Rajagopalan, S. P. (2011). A novel SVM based CSSFFS feature selection algorithm for detecting breast cancer. *International journal of computer applications*, 31(8), 0975-8887.
- Zhang, G. P. (2000). Neural networks for classification: a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 30(4), 451-462.