

# GIẢI PHÁP PHÂN LOẠI CHỦ ĐỀ TỰ ĐỘNG CHO BẢN TIN THỜI SỰ TRUYỀN HÌNH BẰNG KỸ THUẬT HỌC MÁY

Nguyễn Tấn Phú, Lâm Thanh Toàn, Từ Thái Bảo

Trường Đại học Kỹ thuật - Công nghệ Cần Thơ

Email: ntanphu@ctu.edu.vn

## Thông tin chung:

Ngày nhận bài: 29.12.2023

Ngày nhận bài sửa: 08.4.2024

Ngày duyệt đăng: 08.4.2024

## Từ khóa:

Học máy, Phân loại video,  
Trích xuất đặc trưng, Xử lý  
ngôn ngữ tự nhiên

## TÓM TẮT

Phân loại video bằng học máy trở thành một lĩnh vực tiềm năng, giúp tự động nhận dạng và phân loại vào các danh mục tương ứng. Quá trình này bắt đầu bằng việc tiền xử lý dữ liệu video để trích xuất và chuyển đổi thông tin thành đặc trưng số học. Đặc biệt, các thuật toán học máy như KNN, SVM, CNN và PhoBERT được sử dụng để xử lý và phân tích nội dung video cũng như thông tin ngôn ngữ trong video. Trong thực nghiệm, dữ liệu được thu thập từ hệ thống lưu trữ nội bộ của Đài Phát thanh và Truyền hình thành phố Cần Thơ, với mỗi video có độ dài trung bình khoảng 3 phút. Các thuật toán đã được triển khai và đánh giá trên tập dữ liệu này để đo lường và so sánh hiệu suất. Kết quả của thuật toán PhoBERT, với độ chính xác đạt tới 98%. Từ kết quả cho thấy khả năng vượt trội của PhoBERT trong việc xử lý và nhận dạng nội dung video, tạo điều kiện thuận lợi cho việc phát triển hệ thống phân loại video tự động.

## 1. GIỚI THIỆU

Trong những năm qua, sự tiến bộ đột phá của công nghệ và sự lan tỏa mạnh mẽ của mạng xã hội đã biến Internet thành một nguồn thông tin đa dạng, từ sách, báo, hình ảnh đến video và âm nhạc. Trong đó, video đóng vai trò ngày càng quan trọng, tác động sâu rộng vào xã hội. Điều này làm cho việc phân loại video dựa trên nội dung trở nên cực kỳ quan trọng, đặc biệt trong lĩnh vực bản tin thời sự truyền hình, việc tự động phân loại chủ đề đóng vai trò then chốt để cung cấp thông tin chính xác và đáng tin cậy.

Công nghệ trí tuệ nhân tạo (AI) và xử lý ngôn ngữ tự nhiên (NLP) đã đạt được sự phát triển đáng kể, mở ra khả năng tự động hóa việc phân loại video. Ở Việt Nam, việc áp dụng AI trong lĩnh vực truyền thông không chỉ góp phần thúc đẩy sự tiến bộ của ngành mà còn tối ưu hóa quá trình sản xuất nội dung, nâng cao trải nghiệm người xem.

Trên phạm vi toàn cầu, việc ứng dụng AI để tự động phân loại chủ đề đã thu hút sự quan tâm đặc biệt, mang lại những tiến bộ đột phá cho ngành truyền thông.

## 2. PHƯƠNG PHÁP NGHIÊN CỨU

### 2.1. Các nghiên cứu có liên quan

Nghiên cứu của (Ahmed và cộng sự, 2020) đã giới thiệu một phương pháp tiên tiến để phát hiện tin giả thông qua kỹ thuật học máy, chủ yếu tập trung vào xử lý ngôn ngữ tự nhiên. Nghiên cứu này tập trung vào việc phát triển các thuật toán có khả năng phân tích nội dung văn bản để phân biệt giữa thông tin đáng tin cậy và thông tin không chính xác. Bằng cách kết hợp phương pháp rút trích đặc trưng TF-IDF (Term Frequency-Inverse Document Frequency) với các thuật toán Naïve Bayes, Passive Aggressive và SVM. Kết quả thực

nghiệm trên nhiều tập dữ liệu cho thấy, mô hình này đạt được độ chính xác lên đến 93%.

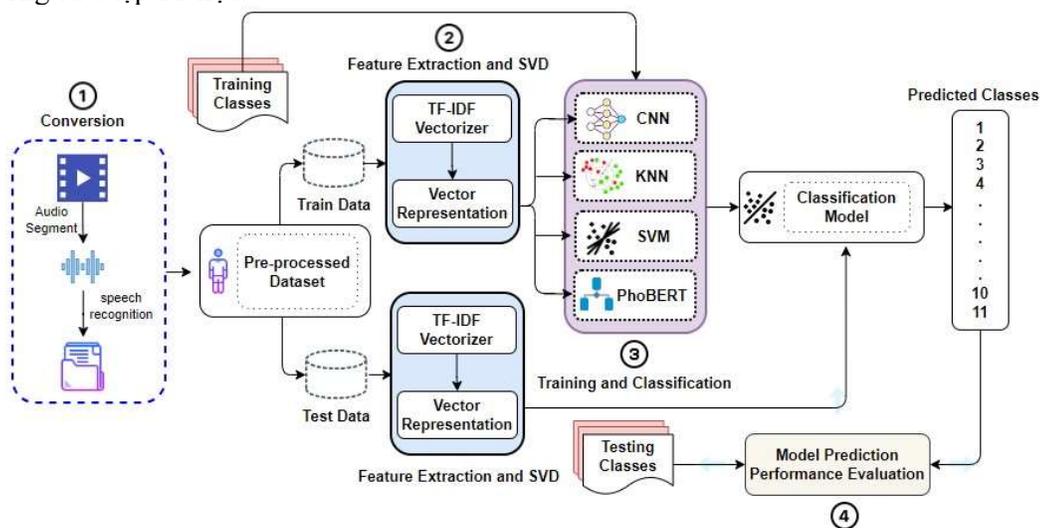
Trong lĩnh vực phân loại video, Gao (2021) đã tiến hành nghiên cứu chi tiết bằng việc áp dụng mô hình kiến trúc ResNet-v2. Tác giả đã đặt nền tảng nghiên cứu trên sự kết hợp và cải tiến của thuật toán Adam cùng thuật toán Gradient Descent, nhằm tối ưu hóa hiệu suất học. Kết quả thực nghiệm đã cho thấy thuật toán Adam cải tiến hiệu quả trong việc cập nhật trọng số mạng và đạt được sự hội tụ nhanh chóng. Đặc biệt, mô hình Inception-ResNet-v2 sau khi được cải tiến đã vượt trội so với các mô hình mạng nơron tích chập (CNN) thông thường, với tỷ lệ chính xác phân loại lên tới 91,47% trên tập dữ liệu video tin tức.

Luo (2021) triển khai phương pháp máy véc-tơ hỗ trợ (SVM) và các kỹ thuật học máy khác để phân loại văn bản. Sử dụng các phương pháp Naive Bayes, SVM, Logistic Regression, Logistic Regression Cross-Validation (LRCV) để phân loại trên 3 bộ dữ liệu khác nhau. Kết quả cho thấy rất khá tốt, SVM đạt mức đánh giá chính xác khoảng 90% trong cả 3 tập dữ liệu.

Dựa trên các công trình nghiên cứu hiện có, đã có nhiều phương pháp đề xuất để phân loại video dựa trên nội dung và tính chất của video. Mỗi phương pháp đều mang lại những ưu điểm và hạn chế riêng. Trong nghiên cứu này, chúng tôi đã sử dụng kỹ thuật học máy và phân tích thống kê, cùng với việc chuẩn hóa dữ liệu. Mục tiêu chính là phát triển một phương pháp phân loại video chính xác và hiệu quả hơn dựa trên nội dung và đặc điểm của video.

## 2.2. Mô hình đề xuất của hệ thống

Mô hình đề xuất cho hệ thống tự động phân loại chủ đề được biểu diễn trong Hình 1. Mô hình này được thiết kế với 4 giai đoạn: (1) Chuyển đổi và tiền xử lý dữ liệu; (2) Rút trích đặc trưng dữ liệu và giảm chiều dữ liệu; (3) Huấn luyện và phân loại bằng cách sử dụng KNN, SVM, CNN, và PhoBERT; (4) Đánh giá hiệu suất của các thuật toán tại giai đoạn (3) dựa trên các tiêu chí như Accuracy, Precision, Recall, F1-Score.



Hình 1. Mô hình đề xuất của hệ thống phân loại chủ đề tự động

Nguồn: Tác giả đề xuất (2024)



**Thuật toán: kNN**

**Input:** X: dữ liệu huấn luyện. Y: nhãn lớp của X, K: số lượng láng giềng gần nhất.

**Output:** Lớp của mẫu thực nghiệm  $x$ .

**Start**

**Classify** (X,Y,  $x$ )

1. *For each* sample  $x$  *do*

Tính khoảng cách:  $d(x, X) = \sqrt{\sum_{i=1}^n (x_i - X_i)^2}$

**End for**

2. Phân loại  $x$  vào lớp:  $C(x_i) = \operatorname{argmax}_k \sum_{X_j \in kNN} C(X_j, Y_k)$

**End**

**2.2.3. THUẬT TOÁN SVM**

SVM dựa trên lý thuyết học máy thống kê do (Cortes, C. và Vapnik, 1995) đề xuất. Ý tưởng của SVM là xây dựng một siêu phẳng tốt nhất có thể chia các điểm trên không gian này thành hai lớp riêng biệt tương ứng, lớp dương (+) và lớp âm (-). Phương trình mặt siêu phẳng có véc-tơ  $x$  trong không gian:  $w \cdot x + b = 0$ . Trong đó,  $w$  là véc-tơ trọng số,  $b$  là độ lệch (bias). Hướng và khoảng cách từ gốc tọa độ đến mặt siêu phẳng thay đổi khi  $w$  và  $b$  thay đổi. Bộ phân lớp SVM được định nghĩa như sau:

$$\text{class}(x) = \text{sign}(w \cdot x + b) \quad (6)$$

Trong đó:

$$\begin{cases} \text{class}(x) = -1, w \cdot x + b < 0 \\ \text{class}(x) = +1, w \cdot x + b > 0 \end{cases} \quad (7)$$

Gọi  $y_i$  mang giá trị +1 hoặc -1. Nếu  $y_i = +1$  thì  $x$  thuộc về lớp (+), ngược lại  $y_i = -1$  thì  $x$  thuộc lớp (-). Hai mặt siêu phẳng phân chia các mẫu thành hai phần được cho bởi các phương trình:  $w \cdot x + b = 1$  và  $w \cdot x + b = -1$ . Có thể tính khoảng cách giữa hai mặt siêu phẳng này là:  $\frac{2}{\|w\|}$ , để khoảng cách biên là lớn nhất cần tìm giá trị nhỏ nhất của  $\|w\|$  và ngăn không cho các

điểm dữ liệu rơi vào vùng bên trong biên, cần thêm điều kiện ràng buộc sau:

$$\begin{cases} w \cdot x_i + b \leq -1, \text{ với mẫu } (-) \\ w \cdot x_i + b \geq 1, \text{ với mẫu } (+) \end{cases} \quad (8)$$

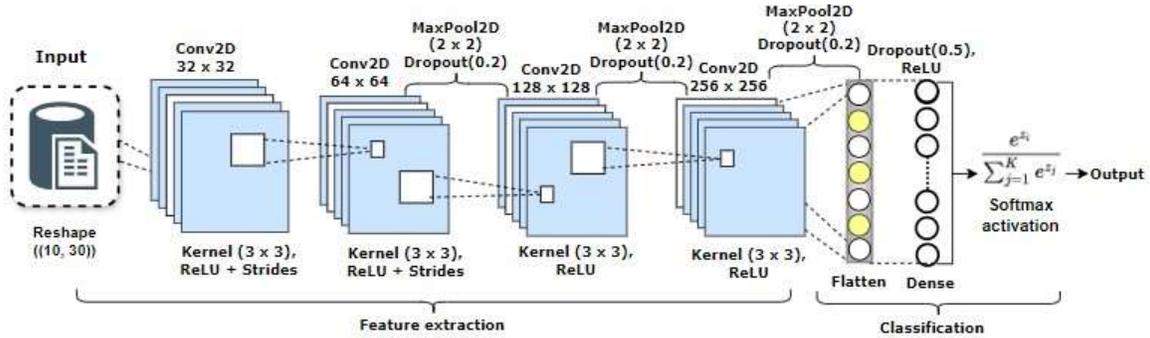
Ta có thể viết lại:  $y_i(w \cdot x_i + b) \geq 1$ , với  $i \in (1, n)$ . Khi đó, việc tìm siêu phẳng  $h$  tương đương giải bài toán tìm  $\operatorname{Min}\|w\|$  với  $w$  và  $b$  thỏa điều kiện sau:  $\forall i \in (1, n): y_i(w \cdot x_i + b) \geq 1$  và sử dụng hàm nhân RBF thực sự tăng đáng kể về hiệu năng của mô hình.

$$k(x, z) = \exp(-\gamma \|x - z\|_2^2), \gamma > 0 \quad (9)$$

**2.2.4. MẠNG NORON TÍCH CHẬP (CNN)**

Trong mô hình mạng nơron tích chập (CNN) (Krizhevsky và cộng sự, 2017) được đề xuất với kiến trúc mạng trong Hình 2, có 4 lớp tích chập (Conv2D) và 3 lớp MaxPool2D, được xếp chồng lên nhau trong mạng.

Lớp tích chập đầu tiên (Conv2D<sub>1</sub>) trong mô hình CNN mà chúng tôi đề xuất là lớp đầu vào liên kết trực tiếp với lớp tích chập tiếp theo (Conv2D<sub>2</sub>). Lớp (Conv2D<sub>1</sub>) sử dụng 32 bộ lọc, mỗi bộ lọc với kernel có kích thước là (3 x 3), sử dụng stride là 1 và áp dụng hàm kích hoạt ReLU.



**Hình 2. Mô hình kiến trúc mạng CNN của hệ thống phân loại**

*Nguồn: Tác giả đề xuất (2024)*

Lớp tích chập thứ 2 (Conv2D<sub>2</sub>) của mô hình bao gồm 64 bộ lọc với kernel kích thước (3 x 3), sử dụng stride là 1 và hàm kích hoạt ReLU. Tiếp theo, sau lớp (Conv2D<sub>2</sub>) là lớp MaxPool2D (MP<sub>1</sub>) với kích thước (2 x 2) nhằm lựa chọn đặc trưng quan trọng. Để giảm hiện tượng overfitting, chúng tôi đã thêm lớp Dropout (DR<sub>1</sub>) với tỷ lệ 20% (dựa trên việc điều chỉnh mô hình cho từng giá trị khác nhau), giúp loại bỏ 20% neuron trong quá trình huấn luyện.

Lớp tích chập thứ 3 (Conv2D<sub>3</sub>) sử dụng 128 bộ lọc với kernel kích thước (3 x 3), sau đó áp dụng lớp MaxPool2D (MP<sub>2</sub>) với kích thước (2 x 2) và lớp Dropout (DR<sub>2</sub>) với tỷ lệ 20%. Còn lớp tích chập thứ 4 (Conv2D<sub>4</sub>) bao gồm 256 bộ lọc với kernel (3 x 3), được tiếp tục bởi lớp MaxPool2D (MP<sub>3</sub>) (2 x 2) và Dropout (DR<sub>3</sub>) tỷ lệ 20%.

Lớp Flatten (FL) được sử dụng trong giai đoạn này để chuyển đổi dữ liệu ma trận hai chiều thành vector, theo sau là lớp Dense (Dense<sub>1</sub>) với hàm kích hoạt ReLU và Dropout (DR<sub>4</sub>) 50% để thu được đầu ra cuối cùng sẽ được xử lý cho các lớp tiếp theo.

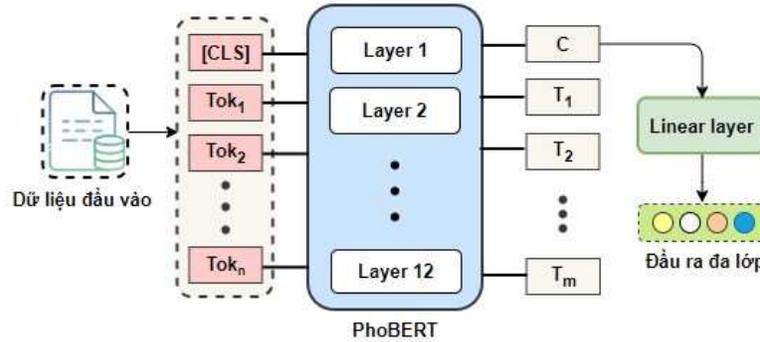
Lớp Dense thứ 2, 3 (Dense<sub>2</sub>, Dense<sub>3</sub>) nhận đầu ra của lớp Dense đầu tiên với hàm kích

hoạt ReLU, lớp này kết nối sử dụng hàm kích hoạt Softmax để phân phối xác suất cho từng chủ đề.

### 2.2.5. PHOBERT

PhoBERT (Dat và cộng sự, 2020) là một mô hình tiếng Việt dựa trên kiến trúc được cải tiến từ RoBERT so với BERT. Mô hình này được huấn luyện chuyên biệt cho Tiếng Việt, sử dụng khoảng 20GB dữ liệu từ bộ văn bản Wikipedia và các trang tin tức Tiếng Việt khác, nhằm giải quyết hiệu quả các thách thức trong xử lý ngôn ngữ tự nhiên cho Tiếng Việt.

Mô hình PhoBERT bao gồm hai phiên bản: PhoBERT-Base với 12 lớp và PhoBERT-Large với 24 lớp. Đầu vào của PhoBERT cần điều chỉnh để hoàn thiện tính tương thích với nhiệm vụ phân loại. Dựa trên kiến trúc của PhoBERT, văn bản đầu vào của mô hình cần được chuyển đổi thành chuỗi token và được bổ sung thêm hai token quan trọng là [CLS] và [SEP] để đánh dấu cuối câu. Trong nhiệm vụ phân loại, trạng thái ẩn tương ứng với token [CLS] là đại diện cho toàn bộ câu, được sử dụng để thực hiện các nhiệm vụ phân loại, khác với vectơ trạng thái ẩn liên quan đến token biểu diễn từ thông thường.



**Hình 3. Kiến trúc mô hình PhoBERT của hệ thống phân loại**  
 Nguồn: Tác giả điều chỉnh, đề xuất (2024) và tham khảo (Dat và cộng sự, 2020)

Do đó, khi đưa vào mô hình một câu trong quá trình huấn luyện, nhận được đầu ra là một vector trạng thái ẩn tương ứng với token đó. Lớp bổ sung thêm vào mô hình bao gồm các nơron tuyến tính chưa được huấn luyện, có kích thước là [kích thước vector trạng thái ẩn, số chủ đề], có nghĩa là đầu ra của PhoBERT khi kết hợp với lớp phân loại là một vector chứa hai số, đại diện cho điểm số sử dụng làm cơ sở cho quá trình phân loại câu.

Trong thử nghiệm của bài báo này, chúng tôi sử dụng mô hình PhoBERT-Base và tinh chỉnh các tham số được sử dụng trong mô hình như: *max\_length*; *pad\_to\_max\_length* =

*true* (tự động thêm khoảng đệm vào phía sau [SEP]); *learning\_rate* (*adam*); *epochs*; *batch\_size*; *cross\_entropy*.

### 3. KẾT QUẢ VÀ THẢO LUẬN

Dữ liệu được thu thập từ hệ thống lưu trữ nội bộ của Đài Phát thanh và Truyền hình thành phố Cần Thơ, với mỗi đoạn video có thời lượng trung bình là 3 phút. Tổng cộng, bộ dữ liệu bao gồm 14.503 mẫu, được phân thành 11 chủ đề chính được mô tả chi tiết trong Bảng 1. Quá trình thực nghiệm mô hình được thực hiện dựa trên máy tính cá nhân, với cấu hình như sau: Intel® Core™ i5- 7200U CPU @ 2.50GHz 2.70GHz , RAM 8GB.

**Bảng 1. Số lượng mẫu dữ liệu và các chủ đề được tổng hợp và thu thập**

STT	Chủ đề	Huấn luyện	Kiểm tra	Tổng số
1	Chính trị - Xã hội	1.165	292	1.457
2	Dự báo thời tiết	1.055	264	1.319
3	Kinh tế	1.046	262	1.308
4	Môi trường	933	233	1.166
5	Nông nghiệp	1.123	281	1.404
6	Pháp luật	1.094	273	1.367
7	Sức khỏe	1.050	262	1.312
8	Thể giới	1.084	271	1.355

9	Thể thao	1.050	263	1.313
10	Văn hóa	1.097	274	1.371
11	Giáo dục	905	226	1.131
<b>Tổng cộng:</b>		<b>11.602</b>	<b>2.901</b>	<b>14.503</b>

*Nguồn: Đài Phát thanh và Truyền hình thành phố Cần Thơ*

### 3.1. Kết quả thực nghiệm

Kết quả thực nghiệm cho thấy hiệu quả phân loại của các thuật toán là tương đối tốt. Trong đó, kNN thực nghiệm với nhiều tham số k trong ứng {1, 2, 3, 4, 5, 6, 7}, và p trong ứng {1, 2} kết hợp với GridSearchCV tự động kiểm tra tìm ra giá trị tối ưu của k và hàm khoảng cách nhằm tăng hiệu suất mô hình.

Kết quả thu được từ mô hình với độ chính xác 83% (k = 5, p = 1).

Tiếp theo, SVM sử dụng GridSearchCV áp dụng cho các bộ thông số khác nhau của mô hình được thiết lập, qua đó tìm được bộ thông số tối ưu nhất của hàm kernel = RBF, C = 10, Gamma = 1 đạt được kết quả cao nhất trong các trường hợp thực nghiệm với độ chính xác lên đến 91%.

**Bảng 2. Tổng hợp kết quả thực nghiệm của kNN, SVM**

STT	Chủ đề	kNN				SVM			
		Pr	Re	F1	Acc	Pr	Re	F1	Acc
1	Chính trị - Xã hội	86	72	78		92	89	91	
2	Dự báo thời tiết	98	99	98		99	100	99	
3	Kinh tế	87	85	86		90	91	91	
4	Môi trường	82	78	80		87	90	89	
5	Nông nghiệp	89	84	86		94	92	93	
6	Pháp luật	81	81	81	83	86	90	88	91
7	Sức khỏe	87	81	84		90	88	89	
8	Thể giới	78	81	79		92	88	90	
9	Thể thao	85	93	89		96	97	96	
10	Văn hóa	61	73	66		83	82	83	
11	Giáo dục	79	85	82		89	90	89	

\* Pr = Precision (%); Re = Recall (%); F1 = F1 Score (%); Acc = Accuracy (%)

*Nguồn: Kết quả xử lý số liệu thực nghiệm kNN, SVM năm 2024.*

Trong mô hình CNN dựa trên các giá trị của tham số như: Batch\_Size, Optimizer {RMSprop, Adam, Nadam, SGD}, Epoch tăng dần và điều chỉnh các thông số trong mô hình đã thay đổi đáng kể về tốc độ học, tỉ lệ lỗi, độ chính xác của mô hình. Kết quả thực nghiệm cho thấy giá trị Batch\_Size = 32, Optimizer = SGD, Epoch = 23 cho kết quả với độ chính xác là 89%.

Mô hình PhoBERT để tìm kiếm siêu tham số, nhóm tác giả điều chỉnh kết hợp với siêu tham số tối ưu hóa Adam được đề xuất từ tác giả (Nguyen, D. Q và cộng sự, 2020), learning\_rate {2e-5, 3e-5, 5e-5}, batch\_size {128, 256}, Epoch. Ở giai đoạn thử nghiệm, các tiêu chí đánh giá mô hình trên tập dữ liệu kiểm tra đạt được kết quả mong đợi với độ chính xác 98% ứng với learning\_rate = 5e-5, batch\_size = 256, epoch = 5.

**Bảng 3. Tổng hợp kết quả thực nghiệm của CNN, PhoBERT**

STT	Chủ đề	CNN				PhoBERT			
		Pr	Re	F1	Acc	Pr	Re	F1	Acc
1	Chính trị - Xã hội	91	88	90		99	97	98	
2	Dự báo thời tiết	99	100	99		100	100	100	
3	Kinh tế	88	91	90		98	98	98	
4	Môi trường	84	89	87		97	96	97	
5	Nông nghiệp	90	92	91		99	99	99	
6	Pháp luật	84	92	88	89	96	95	96	98
7	Sức khỏe	85	85	85		98	98	98	
8	Thể giới	90	86	88		96	98	97	
9	Thể thao	92	96	94		100	99	99	
10	Văn hóa	88	72	79		95	96	95	
11	Giáo dục	88	88	88		98	98	98	

\*Pr = Precision (%); Re = Recall (%); F1 = F1 Score (%); Acc = Accuracy (%)

*Nguồn: Kết quả xử lý số liệu thực nghiệm CNN, PhoBERT năm 2024.*

### 3.2. So sánh kết quả thực nghiệm

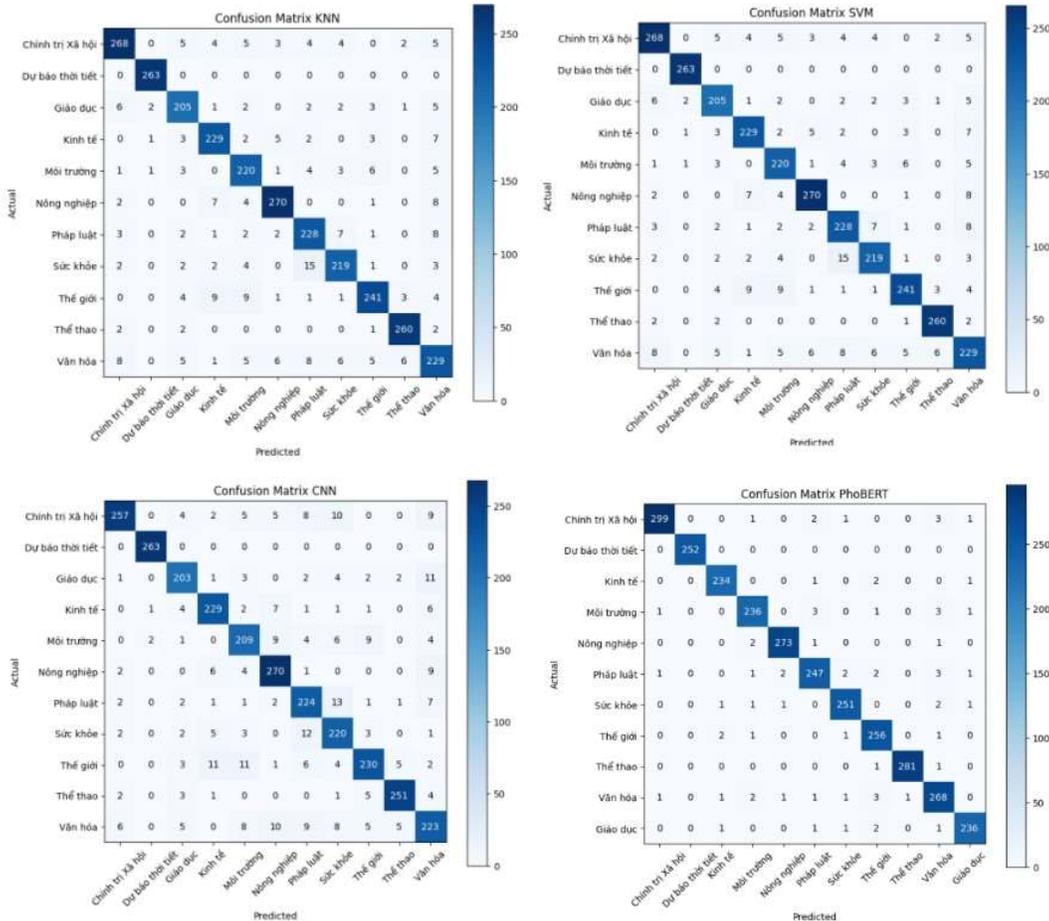
Dựa trên kết quả thực nghiệm từ các mô hình và kết hợp với kỹ thuật rút trích đặc trưng TF-IDF, giảm chiều dữ liệu SVD được mô tả so sánh giữa các tiêu chí

Accuracy, Precision, Recall, F1-Score trong Bảng 4, Hình 4 thể hiện độ chính xác và hiệu suất phân loại như sau:

**Bảng 4. Bảng tổng hợp so sánh kết quả thực nghiệm**

Tiêu chí đánh giá	Mô hình			
	kNN	SVM	CNN	PhoBERT
Accuracy (%)	83	91	89	98

*Nguồn: Kết quả xử lý số liệu thực nghiệm tổng hợp năm 2024.*



**Hình 4. Ma trận thể hiện hiệu suất phân loại của các mô hình**

*Nguồn: Kết quả số liệu phân lớp của các mô hình năm 2024.*

So sánh kết quả cho thấy được các thuật toán có độ chính xác khi phân lớp có thể xem là tương đối tốt. Nhưng PhoBERT cho kết quả tốt hơn trên tập dữ liệu.

#### 4. KẾT LUẬN

Trong nghiên cứu này, chúng tôi giới thiệu một giải pháp phân loại chủ đề tự động

cho bản tin thời sự truyền hình, sử dụng các thuật toán học máy kết hợp với phương pháp trích đặc trưng dữ liệu TF-IDF, và SVD giảm chiều dữ liệu để tối ưu hóa tính toán nhanh và hiệu quả. Kết quả thực nghiệm thu được từ KNN, SVM, CNN, và PhoBERT cho thấy các kỹ thuật học máy có thể dễ dàng áp dụng vào các bài toán phân loại. So sánh hiệu suất giữa

các mô hình cho thấy phương pháp PhoBERT đạt được kết quả tốt nhất.

Trong thời gian tới, chúng tôi dự kiến sẽ phát triển, và cải tiến thuật toán, cũng như sử dụng phương pháp lai giữa các thuật toán để tối ưu hóa tốc độ quá trình huấn luyện và phân loại.

#### Tài liệu tham khảo

Ahmed, S., Hinkelmann, K., and Corradini, F. (2022), “Development of fake news model using machine learning through natural language processing”. arXiv preprint arXiv:2201.07489.

Cortes, C., and Vapnik, V. (1995), “Support-vector networks”, Machine learning, 20, pp. 273-297.

Dat, Q. N., and Anh, T. N. (2020), “PhoBERT: Pre-trained language models for Vietnamese”. arXiv preprint arXiv:2003.00744.

Gao, Y. (2021), “News Video Classification Model Based on ResNet-2 and Transfer Learning”, Security and Communication Networks, pp. 1-9.

Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003). “KNN model-based approach in classification”, In On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings, Springer Berlin Heidelberg, pp.986-996.

Luo, X. (2021), “Efficient English text classification using selected machine learning techniques”, Alexandria Engineering Journal, 60(3), pp. 3401-3409.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017), “ImageNet classification with deep convolutional neural networks”, Communications of the ACM, 60(6), pp. 84-90.

## AUTOMATIC TOPIC CLASSIFICATION SOLUTION FOR TELEVISION NEWS USING MACHINE LEARNING TECHNIQUES

### ABSTRACT

*Video classification using machine learning has become a promising field, aiding in the automatic recognition and categorization into corresponding groups. This process begins with preprocessing video data to extract and convert information into numerical features. Specifically, machine learning algorithms such as KNN, SVM, CNN, and PhoBERT are employed to process and analyze the video content as well as language information within the video. In the experiment, data was collected from the internal storage system of the Can Tho City Radio and Television Station, with each video averaging about 3 minutes in length. These algorithms were deployed and evaluated on this dataset to measure and compare performance. The results of the PhoBERT algorithm achieved an accuracy rate of up to 98%. These results demonstrate the outstanding capability of PhoBERT in processing and recognizing video content, paving the way for the development of an automatic video classification system.*

**Keywords:** Feature Extraction, Machine Learning, Natural Language Processing, Video Classification