**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**TRƯỜNG ĐẠI HỌC PHENIKAA**

****

**BÁO CÁO BÀI TẬP LỚN**

**TÍCH HỢP VÀ PHÂN TÍCH DỮ LIỆU LỚN**

**Nhóm: 4**

**TÊN ĐỀ TÀI  
SENTIMENT ANALYSIS TIN TỨC CHỨNG KHOÁN**

**Giảng viên hướng dẫn: Đặng Thị Thùy An**

**02/2025**

# LỜI CẢM ƠN

Lời đầu tiên, chúng em xin gửi lời cảm ơn chân thành đến Trường Đại Học Phenikaa đã tạo cơ hội cho chúng em được học tập tại một ngôi trường có cơ sở vật chất hàng đầu, với chất lượng giảng dạy vô cùng chuyên nghiệp.

Trong suốt thời gian học tập và rèn luyện tại Trường Đại Học Phenikaa cho đến nay, chúng em đã nhận được rất nhiều sự quan tâm, giúp đỡ của quý Thầy cô và bạn bè.

Chúng em gửi lời cảm ơn đặc biệt đến thầy Đặng Thị Thúy An , giảng viên hướng dẫn đề tài môn học này của chúng em trong quá trình làm project giữa kỳ . Nhờ sự giúp đỡ nhiệt tình, nhẫn nại dành thời gian hướng dẫn, chỉ bảo, của thầy trong suốt quá trình nên chúng em có kết quả tốt nhất.

Với điều kiện thời gian cũng như kinh nghiệm còn hạn chế, bài báo cáo này không thể tránh được những thiếu sót. Chúng em rất mong nhận được sự chỉ bảo, đóng góp ý kiến của các quý thầy cô để chúng em có điều kiện bổ sung, nâng cao ý thức của mình, phục vụ tốt hơn công tác thực tế sau này.

Chúng em xin chân thành cảm ơn!

Hà Nội ,27 tháng 02 năm 2025

Sinh viên thực hiện

**Hoàng Văn Phú**

**NHẬN XÉT**

Chữ ký Giáo viên hướng dẫn

MỤC LỤC

[LỜI CẢM ƠN 1](#_Toc191561077)

[CHƯƠNG 1. GIỚI THIỆU (Introduction) 1](#_Toc191561078)

[1.1 Mục tiêu 1](#_Toc191561079)

[1.2 Tổng quan 1](#_Toc191561080)

[CHƯƠNG 2. CƠ SỞ LÝ THUYẾT 3](#_Toc191561081)

[2.1. Phân tích cảm xúc (Sentiment Analysis) 3](#_Toc191561082)

[**2.2.** **Ảnh hưởng của tin tức đến thị trường chứng khoán** 5](#_Toc191561083)

[CHƯƠNG 3: KIẾN TRÚC HỆ THỐNG (System Architecture) 7](#_Toc191561084)

[3.1 Cấu hình dữ liệu 7](#_Toc191561085)

[3.2 Tổng quan kiến trúc chương trình 8](#_Toc191561086)

[3.3 Cấu trúc database 8](#_Toc191561087)

[3.4 Cấu hình và triển khai ứng dụng 9](#_Toc191561088)

[CHƯƠNG 4. KẾT QUẢ VÀ ĐÁNH GIÁ ( Results and reviews) 12](#_Toc191561089)

[**4.1 Các kết quả đạt được** 12](#_Toc191561090)

[**4.2 Hạn chế và cải tiến** 12](#_Toc191561091)

[CHƯƠNG 5. KẾT LUẬN (Conclusion) 14](#_Toc191561092)

[DANH MỤC TÀI LIỆU THAM KHẢO 15](#_Toc191561093)

# CHƯƠNG 1. GIỚI THIỆU (Introduction)

## 1.1 Mục tiêu

Trong bối cảnh thị trường chứng khoán biến động liên tục, thông tin từ báo chí đóng vai trò quan trọng trong việc định hình tâm lý nhà đầu tư và xu hướng thị trường. Những tin tức tích cực có thể thúc đẩy niềm tin và gia tăng giao dịch mua vào, trong khi tin tức tiêu cực có thể gây hoang mang và kích thích bán tháo. Do đó, việc phân tích cảm xúc từ các bài báo tài chính giúp nhà đầu tư có cái nhìn tổng quan hơn về tình hình thị trường và ra quyết định phù hợp.

Mục tiêu của dự án này là xây dựng một hệ thống tự động phân tích cảm xúc từ tin tức chứng khoán, nhằm:

* Thu thập tin tức từ nhiều nguồn báo tài chính (Bloomberg, Reuters, CafeF, Vietstock, v.v.).
* Xác định số lượng tin tức liên quan đến thị trường chứng khoán trong một khoảng thời gian nhất định.
* Phân loại tin tức thành các nhóm tích cực, tiêu cực, hoặc trung lập dựa trên nội dung bài báo.
* Xây dựng hệ thống trực quan hóa dữ liệu để hiển thị xu hướng cảm xúc của thị trường theo từng ngày, tuần hoặc tháng.
* Cung cấp API hoặc dashboard hỗ trợ nhà đầu tư theo dõi tình hình tin tức và xu hướng cảm xúc thị trường.

## 1.2 Tổng quan

*1.2.1 Tầm quan trọng của phân tích cảm xúc trong thị trường chứng khoán*

Trong lĩnh vực tài chính, tin tức đóng vai trò quan trọng trong việc ảnh hưởng đến quyết định của nhà đầu tư. Những sự kiện lớn như báo cáo tài chính, chính sách tiền tệ, hoặc biến động kinh tế vĩ mô có thể làm thay đổi giá cổ phiếu trong thời gian ngắn. Phân tích cảm xúc từ các bài báo tài chính giúp đánh giá mức độ lạc quan hay bi quan của thị trường, từ đó hỗ trợ các nhà đầu tư đưa ra quyết định chính xác hơn.

Các nghiên cứu đã chỉ ra rằng tin tức có thể tác động đến tâm lý đám đông và làm thay đổi xu hướng mua/bán cổ phiếu. Ví dụ:

* Tin tức tích cực (doanh thu tăng trưởng mạnh, hợp tác chiến lược) có thể thúc đẩy giá cổ phiếu tăng.
* Tin tức tiêu cực (khủng hoảng kinh tế, scandal doanh nghiệp) có thể khiến giá cổ phiếu giảm mạnh.

*1.2.2 Giới thiệu về phân tích cảm xúc (Sentiment Analysis)*

Phân tích cảm xúc (Sentiment Analysis) là một nhánh của Xử lý ngôn ngữ tự nhiên (NLP), nhằm xác định thái độ, quan điểm hoặc cảm xúc trong văn bản. Trong bài toán phân tích tin tức chứng khoán, mô hình sẽ phân loại bài báo thành tích cực, tiêu cực hoặc trung lập dựa trên nội dung.

Có ba phương pháp phổ biến để thực hiện phân tích cảm xúc:

* Phương pháp dựa trên từ điển (Lexicon-based): Dựa vào danh sách từ vựng cảm xúc để đánh giá mức độ tích cực hoặc tiêu cực của một văn bản.
* Phương pháp học máy (Machine Learning-based): Sử dụng các thuật toán như Naïve Bayes, SVM để huấn luyện mô hình phân loại cảm xúc.
* Phương pháp học sâu (Deep Learning-based): Ứng dụng mạng nơ-ron như LSTM, Transformer để phân tích và dự đoán cảm xúc với độ chính xác cao.

1.2.3 Các nghiên cứu liên quan

Một số nghiên cứu trước đây đã chứng minh mối quan hệ giữa phân tích cảm xúc và thị trường tài chính:

* Bollen et al. (2011): Sử dụng dữ liệu từ Twitter để dự đoán biến động của chỉ số Dow Jones.
* Tetlock (2007): Chứng minh rằng tin tức tiêu cực có thể làm giảm giá cổ phiếu trong ngắn hạn.
* Loughran & McDonald (2011): Phát triển từ điển tài chính để phân tích cảm xúc trong báo cáo doanh nghiệp

# CHƯƠNG 2. CƠ SỞ LÝ THUYẾT

## Phân tích cảm xúc (Sentiment Analysis)

*2.1.1 Khái niệm về phân tích cảm xúc*

Phân tích cảm xúc (Sentiment Analysis) là một nhánh quan trọng của Xử lý ngôn ngữ tự nhiên (NLP - Natural Language Processing), nhằm xác định và phân loại thái độ, quan điểm hoặc cảm xúc thể hiện trong văn bản. Trong bài toán phân tích tin tức chứng khoán, mục tiêu của phân tích cảm xúc là xác định xem một bài báo có nội dung tích cực, tiêu cực hay trung lập, từ đó đánh giá tác động của tin tức đến tâm lý thị trường và biến động giá cổ phiếu.

*2.1.2 Ứng dụng của phân tích cảm xúc trong tài chính*

Phân tích cảm xúc đã được ứng dụng rộng rãi trong lĩnh vực tài chính, đặc biệt là để hỗ trợ ra quyết định đầu tư. Một số ứng dụng phổ biến bao gồm:

* Đánh giá tác động của tin tức tài chính: Xác định xem tin tức có ảnh hưởng tích cực hay tiêu cực đến cổ phiếu hoặc thị trường chung.
* Dự đoán xu hướng thị trường: Phân tích dữ liệu cảm xúc từ các nguồn tin tức, báo cáo tài chính hoặc mạng xã hội để dự báo biến động thị trường.
* Tự động hóa giao dịch chứng khoán: Kết hợp phân tích cảm xúc với các thuật toán giao dịch tự động (Algorithmic Trading) để tối ưu hóa chiến lược đầu tư.

*2.1.3 Các phương pháp phân tích cảm xúc*

Có ba phương pháp chính để thực hiện phân tích cảm xúc:

* Phương pháp dựa trên từ điển cảm xúc (Lexicon-based Approach)
  + Nguyên lý: Sử dụng danh sách từ vựng cảm xúc (sentiment lexicon) chứa các từ có sắc thái tích cực hoặc tiêu cực để đánh giá văn bản.
  + Ưu điểm: Dễ triển khai, không cần dữ liệu huấn luyện lớn.
  + Nhược điểm: Khó xử lý các ngữ cảnh phức tạp, từ đồng nghĩa, hoặc câu có nghĩa đảo ngược.
  + Ví dụ: Từ điển tài chính của Loughran & McDonald được thiết kế để phân tích báo cáo tài chính.
* Phương pháp học máy (Machine Learning-based Approach)
  + Nguyên lý: Sử dụng các thuật toán học máy như Naïve Bayes, Support Vector Machine (SVM) hoặc Random Forest để phân loại cảm xúc dựa trên dữ liệu huấn luyện.
  + Ưu điểm: Độ chính xác cao hơn so với phương pháp từ điển, có thể học từ dữ liệu thực tế.
  + Nhược điểm: Cần tập dữ liệu huấn luyện lớn, khó giải thích kết quả mô hình.
  + Ví dụ: Huấn luyện mô hình SVM trên tập dữ liệu tin tức tài chính để phân loại cảm xúc của bài báo.
* Phương pháp học sâu (Deep Learning-based Approach)
  + Nguyên lý: Sử dụng các mô hình mạng nơ-ron như LSTM, BiLSTM, Transformer (BERT, FinBERT) để phân tích và dự đoán cảm xúc từ văn bản.
  + Ưu điểm: Hiệu suất cao, có thể xử lý các ngữ cảnh phức tạp và mối quan hệ giữa các từ trong câu.
  + Nhược điểm: Đòi hỏi tài nguyên tính toán lớn, cần nhiều dữ liệu huấn luyện.
  + Ví dụ: Sử dụng FinBERT – một biến thể của BERT được huấn luyện đặc biệt trên dữ liệu tài chính – để phân tích cảm xúc của bài báo chứng khoán.

*2.1.4 So sánh các phương pháp*

*Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, số

Nội dung do AI tạo ra có thể không chính xác.*

* 1. **Ảnh hưởng của tin tức đến thị trường chứng khoán**

*2.2.1 Vai trò của tin tức trong thị trường chứng khoán*

Tin tức là một trong những yếu tố quan trọng ảnh hưởng đến thị trường chứng khoán. Các bài báo tài chính, thông báo doanh nghiệp, báo cáo kinh tế và sự kiện chính trị đều có thể tác động đến tâm lý nhà đầu tư, dẫn đến biến động giá cổ phiếu.

* Tin tức tích cực (ví dụ: lợi nhuận tăng trưởng, hợp tác chiến lược, chính sách hỗ trợ) có thể thúc đẩy nhà đầu tư mua vào, làm giá cổ phiếu tăng.
* Tin tức tiêu cực (ví dụ: khủng hoảng kinh tế, scandal doanh nghiệp, thay đổi chính sách bất lợi) có thể làm gia tăng sự lo lắng và dẫn đến xu hướng bán tháo cổ phiếu.
* Tin tức trung lập (ví dụ: báo cáo kinh tế ổn định, thông tin không ảnh hưởng mạnh) thường không tác động nhiều đến giá cổ phiếu.

*2.2.2 Các loại tin tức ảnh hưởng đến chứng khoán*

Có nhiều loại tin tức khác nhau có thể ảnh hưởng đến thị trường chứng khoán, bao gồm:

* Tin tức vĩ mô (Macro News)
  + Chính sách tiền tệ (lãi suất, tỷ giá hối đoái, chính sách của ngân hàng trung ương).
  + Dữ liệu kinh tế (GDP, tỷ lệ thất nghiệp, chỉ số lạm phát).
  + Sự kiện chính trị (bầu cử, chiến tranh thương mại, xung đột địa chính trị).
* Tin tức vi mô (Micro News)
  + Báo cáo tài chính doanh nghiệp (doanh thu, lợi nhuận, nợ).
  + Thay đổi trong ban lãnh đạo công ty (CEO mới, scandal nội bộ).
  + Hợp tác chiến lược, mua bán & sáp nhập (M&A).
* Tin tức đột biến (Breaking News)
  + Đại dịch, thiên tai, khủng hoảng tài chính.
  + Tin đồn thị trường, các vụ bê bối doanh nghiệp lớn.

*2.2.3 Nghiên cứu về tác động của tin tức lên giá cổ phiếu*

Nhiều nghiên cứu đã chứng minh rằng tin tức có ảnh hưởng trực tiếp đến biến động giá cổ phiếu:

* Tetlock (2007): Chứng minh rằng tin tức tiêu cực trong các bài báo tài chính có thể làm giảm lợi nhuận thị trường trong ngắn hạn.
* Bollen et al. (2011): Sử dụng dữ liệu Twitter để dự đoán xu hướng chứng khoán, cho thấy mối quan hệ chặt chẽ giữa cảm xúc trên mạng xã hội và biến động thị trường.
* Loughran & McDonald (2011): Phát triển từ điển cảm xúc tài chính giúp đánh giá chính xác hơn tác động của tin tức đến giá cổ phiếu.

# CHƯƠNG 3: KIẾN TRÚC HỆ THỐNG (System Architecture)

## Cấu hình dữ liệu

Trong dự án này, dữ liệu được thu thập từ chuyên mục Chứng khoán trên trang báo điện tử VnExpress (https://vnexpress.net/kinh-doanh/chung-khoan). Quá trình thu thập được thực hiện thông qua web scraping bằng thư viện BeautifulSoup4 trong Python. Sau đó, dữ liệu được chuẩn hóa và đưa vào trong MongoDB. Sau đó tiến hành train thông qua spark và hiển thị kết quả thông qua jupyter và các dự đoán thông qua ứng dụng web xây dựng bằng Flask.

Số lượng dữ liệu thu thập được:

* Tổng cộng 300 bài báo khác nhau.
* Dung lượng dữ liệu đạt 600 KB.
* Lưu trữ trên MongoDB để dễ dàng truy vấn và xử lý.

A diagram of a data flow

AI-generated content may be incorrect.

* 1. Tổng quan kiến trúc chương trình

Chương trình được triển khai với Docker, với một Spark chạy ở localmode, ba collection của mongodb để lưu trữ ( tin gốc, tin được làm chuẩn và kết quả dự đoán) xử lý dữ liệu được triển khai trên jupyter lab và các kết quả dự đoán được triển khai ở một ứng dụng Flask:

A screenshot of a computer program

AI-generated content may be incorrect.

* 1. Cấu trúc database

|  |  |
| --- | --- |
| ID | Khóa của database, đánh giấu các dữ liệu. Các dữ liệu sẽ không bị trùng nhau. |
| title | Tiêu đề của bài báo. Cần lưu lại để sau này triển khai trên ứng dụng web |
| url | Link của bài báo đó. Cần lưu lại để sau này triển khai trên ứng dụng web |
| content | Nội dung của bài báo, nội dung dùng để tiến hành train model và kiểm thử model |
| sentiment | Nhãn của bài báo, mỗi bài báo gồm ba nhãn khác nhau (neutral, positive, negative) |

* 1. Cấu hình và triển khai ứng dụng

Đầu tiên cần phải cấu hình cho ứng dụng với Docker thông qua việc tạo một Dockerfile và docker-compose.yml. Sau khi cấu hình xong, thì tiến hành tạo Container đó thông qua lệnh docker compose build. Sử dụng docker compose up -d để khởi tạo container. Kết quả sau khi hoàn thiện bước này, các công cụ cần thiết được triển khai (Jupyter, Flask và MongoDB), có thể xem qua docker desktop:

A close-up of a white box

AI-generated content may be incorrect.

*3.3.1 Thu thập dữ liệu và chuẩn hóa.*

Đầu tiên cần tiến hành thu thập dữ liệu và crawl chúng để lưu trữ vào các collection của mongodb. Đầu tiên, thu thập các trang web cung cấp, cập nhật các bài báo về tin tức chứng khoán, nguồn Việt Nam thì có VnExpress và CafeF, nước ngoài thì có các tài liệu như Bloomberg,... Sau khi xác định được các bài báo cần thiết thì tiến hành crawl hàng loạt các bài báo đó và lưu trữ vào trong MongoDB.

Tiếp theo sử dụng các công cụ để chuẩn hóa dữ liệu như bỏ các từ vô nghĩa, sau đó tiến hành lowercase bỏ các dấu không sử dụng như (, ), -,...

*3.3.2 Chuẩn bị dữ liệu huấn luyện*

Sau khi dữ liệu đã được gắn nhãn, cần chuẩn bị dữ liệu theo định dạng phù hợp để huấn luyện mô hình:

* Biểu diễn văn bản bằng các kỹ thuật tiền xử lý bằng TF-IDF (Term Frequency - Inverse Document Frequency)
* Chia tập dữ liệu thành tập huấn luyện (80%) và tập kiểm tra (20%).

*3.3.3 Lựa chọn và huấn luyện mô hình*

Cấu hình mô hình sử dụng phương pháp LogisticRegression của pyspark. Dưới đây là cấu hình cho mô hình huấn luyện:

|  |
| --- |
| # Tokenization  tokenizer = Tokenizer(inputCol="clean\_content", outputCol="words")  # Loại bỏ stop words  remover=StopWordsRemover(inputCol="words",outputCol="filtered\_words")  # Tạo feature vector  countVectorizer = CountVectorizer(inputCol="filtered\_words",  outputCol="raw\_features",  minDF=2.0)  # TF-IDF  idf = IDF(inputCol="raw\_features", outputCol="features")  # Label Indexing  labelIndexer = StringIndexer(inputCol="sentiment",  outputCol="label")  # Mô hình Logistic Regression  lr = LogisticRegression(maxIter=20, regParam=0.3, elasticNetParam=0)  # Tạo pipeline  pipeline = Pipeline(stages=[  tokenizer,  remover,  countVectorizer,  idf,  labelIndexer,  lr  ]) |

Các bước huấn luyện bao gồm:

* Chuyển đổi dữ liệu thành dạng số hóa phù hợp với mô hình.
* Chạy huấn luyện với tập dữ liệu đã chia.
* Điều chỉnh siêu tham số để tối ưu hóa hiệu suất.
* Lưu mô hình để sử dụng trong bước dự đoán.

*3.3.4 Đánh giá mô hình*

Sau khi huấn luyện, mô hình được đánh giá bằng cách sử dụng tập kiểm tra. Các chỉ số đánh giá bao gồm:

* Accuracy (Độ chính xác): Phần trăm bài báo được phân loại đúng.
* Precision (Độ chính xác theo lớp): Tỉ lệ bài báo được mô hình dự đoán là tích cực/tiêu cực thực sự thuộc về lớp đó.
* Recall (Độ bao phủ): Mô hình có thể phát hiện được bao nhiêu bài báo thực sự tích cực/tiêu cực.
* F1-score: Trung bình điều hòa giữa Precision và Recall, giúp đánh giá mô hình một cách toàn diện.

# CHƯƠNG 4. KẾT QUẢ VÀ ĐÁNH GIÁ ( Results and reviews)

**4.1 Các kết quả đạt được**

Kết quả đánh giá của mô hình: Chỉ số đánh giá hồi quy

A graph of a graph showing a number of different levels

AI-generated content may be incorrect.

Chỉ số đánh giá phân loại:

A number on a white background

AI-generated content may be incorrect.

Đánh giá: mô hình đạt độ chính xác ở mức chấp nhận được, có thể gây sai sót.

**4.2 Hạn chế và cải tiến**

**Hạn chế:**

* Lượng dữ liệu dùng để huấn luyện còn ít, chưa đủ đa dạng để mô hình học được hết các sắc thái ngữ nghĩa và cách diễn đạt trong tiếng Việt.
* Tiếng Việt có nhiều từ khóa đặc biệt, từ viết tắt, tiếng lóng, dấu câu, và ngữ cảnh khác nhau, gây khó khăn trong việc xử lý và phân loại chính xác.
* Mô hình có thể chưa tối ưu hóa đầy đủ do chưa thực hiện quá trình tinh chỉnh (fine-tuning) trên tập dữ liệu lớn hơn hoặc chưa thử nghiệm nhiều kỹ thuật xử lý trước dữ liệu (preprocessing).
* Thiếu cơ chế xử lý các trường hợp dữ liệu không rõ ràng (ambiguous data), dẫn đến sai số khi mô hình phân tích cảm xúc.
* Một số phương pháp xử lý ngôn ngữ tự nhiên hiện tại chưa tận dụng hết ngữ cảnh trong câu, khiến việc xác định cảm xúc gặp khó khăn.

**Cải tiến:**

* Cập nhật và mở rộng tập dữ liệu huấn luyện bằng cách thu thập thêm dữ liệu từ nhiều nguồn khác nhau như mạng xã hội, diễn đàn, bình luận, đánh giá sản phẩm, và bài báo.
* Sử dụng các phương pháp tiền xử lý dữ liệu nâng cao như chuẩn hóa văn bản (text normalization), loại bỏ nhiễu (noise removal), và xử lý chính tả để cải thiện độ chính xác.
* Thử nghiệm các mô hình mạnh hơn như BERT, PhoBERT, hoặc các mô hình học sâu khác chuyên biệt cho tiếng Việt để nâng cao hiệu suất phân loại cảm xúc.
* Sử dụng kỹ thuật tăng cường dữ liệu (data augmentation) để tạo thêm nhiều mẫu huấn luyện, giúp mô hình học tốt hơn.
* Áp dụng kỹ thuật điều chỉnh tham số (hyperparameter tuning) để tối ưu hóa hiệu suất mô hình.
* Cải thiện khả năng nhận diện ngữ cảnh bằng cách sử dụng mô hình có khả năng xử lý câu dài hơn và hiểu rõ hơn về mối quan hệ giữa các từ.
* Thực hiện đánh giá mô hình trên nhiều bộ dữ liệu khác nhau để kiểm tra tính tổng quát và khả năng ứng dụng thực tế của mô hình.

# CHƯƠNG 5. KẾT LUẬN (Conclusion)

Trong dự án này, nhóm đã xây dựng một hệ thống phân tích cảm xúc tin tức chứng khoán nhằm hỗ trợ nhà đầu tư đánh giá tác động của tin tức đến thị trường tài chính. Hệ thống thu thập dữ liệu từ các nguồn báo uy tín như VnExpress, CafeF, Vietstock, sau đó sử dụng Xử lý ngôn ngữ tự nhiên (NLP) để xác định cảm xúc của bài báo (tích cực, tiêu cực hoặc trung lập). Kết quả thử nghiệm cho thấy hệ thống có thể tự động phân loại tin tức với độ chính xác cao, giúp nhà đầu tư nhanh chóng nắm bắt xu hướng thị trường. Việc trực quan hóa dữ liệu qua dashboard cũng mang lại cái nhìn tổng quan về diễn biến cảm xúc của thị trường chứng khoán theo thời gian. Hệ thống này có thể ứng dụng trong nhiều lĩnh vực tài chính, từ hỗ trợ nhà đầu tư cá nhân, phân tích thị trường cho công ty chứng khoán đến tích hợp vào hệ thống giao dịch tự động. Trong tương lai, nhóm sẽ tiếp tục mở rộng tập dữ liệu, cải thiện độ chính xác của mô hình bằng cách sử dụng các mô hình học sâu như FinBERT, LSTM, Transformer, đồng thời nghiên cứu mối tương quan giữa tin tức và biến động giá cổ phiếu. Ngoài ra, hệ thống sẽ được phát triển thành một API hoặc ứng dụng web để người dùng có thể tra cứu xu hướng tin tức chứng khoán một cách thuận tiện. Dự án này là một bước quan trọng trong việc ứng dụng AI và phân tích dữ liệu vào tài chính, giúp nhà đầu tư ra quyết định thông minh hơn dựa trên dữ liệu tin tức.

# DANH MỤC TÀI LIỆU THAM KHẢO

<https://vinbigdata.com/camera-ai/xu-ly-hinh-anh-trong-python-tu-thuat-toan-den-cong-cu.html>

https://vnexpress.net/

<http://dulieu.tailieuhoctap.vn/books/cong-nghe-thong-tin/the-loai-khac/file_goc_769882.pdf>

<https://phamdinhkhanh.github.io/2020/01/06/ImagePreprocessing.html>