**PHÂN TÍCH CẢM XÚC TỪ TIN TỨC TÀI CHÍNH ĐỂ DỰ ĐOÁN XU HƯỚNG THỊ TRƯỜNG**

|  |  |  |
| --- | --- | --- |
| ***Họ Tên*** | ***MSSV*** | ***Email*** |
| Nguyễn Tiến Phát | 222487 | nguyentienphatct@gmail.com |
| Nguyễn Minh Hiếu | 225150 | [hieu2231122004@gmail.com](mailto: hieu2231122004@gmail.com ) |
| Nguyễn Thanh Tinh | 225171 | tinht5518@gmail.com |
| Trịnh Duy Anh | 223233 | trinhduyanhph@gmail.com |
| Mã Tuấn Anh | 222995 | tuananhma1402@gmail.com |

| ***Từ khóa:***  *AI, Web scraping,*  *Phân tích cảm xúc,*  *Tin tức tài chính,* | **TÓM TẮT**  Bài báo cáo này đề xuất một hệ thống phân tích xu hướng thị trường tài chính thông qua việc đánh giá cảm xúc từ tin tức tiếng Việt. Dựa trên phương pháp phân tích cảm xúc với từ điển có trọng số, hệ thống tự động thu thập và phân tích nội dung từ các trang tin tức tài chính uy tín như cafef.vn, vnexpress.net và nhiều nguồn khác. Dữ liệu được xử lý để tính toán điểm cảm xúc, phân loại thành các mức từ rất tích cực đến rất tiêu cực. Hệ thống đánh giá mức độ tích cực/tiêu cực của các tin tức, từ đó tổng hợp và dự đoán xu hướng thị trường kèm theo các khuyến nghị giao dịch cụ thể. Ngoài ra, kết quả được trực quan hóa thông qua biểu đồ so sánh tỷ lệ từ tích cực/tiêu cực giữa các bài viết. Hệ thống được tích hợp trong giao diện web trực quan sử dụng Flask, giúp người dùng dễ dàng nhập URL tin tức và nhận kết quả phân tích chính xác. Nghiên cứu này tạo ra một công cụ hỗ trợ hiệu quả cho nhà đầu tư trong việc đánh giá tâm lý thị trường và ra quyết định đầu tư sáng suốt hơn trong bối cảnh thông tin tài chính tiếng Việt. |
| --- | --- |

# **MỤC LỤC**

[CHƯƠNG 1: GIỚI THIỆU VỀ ĐỀ TÀI 4](#_Toc197279379)

[1. Đặt Vấn Đề 4](#_Toc197279380)

[2. Các Nghiên Cứu Liên Quan 4](#_Toc197279381)

[CHƯƠNG 2: PHÂN TÍCH MÔ HÌNH 7](#_Toc197279382)

[2.1. Tổng Quan Về Mô Hình Phân Tích Cảm Xúc 7](#_Toc197279383)

[2.2. Cấu Trúc Của Mô Hình Cảm Xúc Có Trọng Số 8](#_Toc197279384)

[2.3. Ưu Điểm Và Hạn Chế Của Mô Hình](#_Toc197279385) 11

[2.4. Thuật Toán Dự Đoán Xu Hướng Thị Trường](#_Toc197279386) 12

[2.5. Trực Quan Hóa Kết Quả Phân Tích](#_Toc197279387) 15

[2.6. Tích Hợp Mô Hình Vào Ứng Dụng Web](#_Toc197279388) 17

[2.7. Đánh Giá Hiệu Quả Của Mô Hình](#_Toc197279388) 18

[CHƯƠNG 3: PHƯƠNG PHÁP NGHIÊN CỨU](#_Toc197279389) 19

[3.1. Tổng Quan Phương Pháp Nghiên Cứu](#_Toc197279390) 19

[3.2. Phương Pháp Thu Thập Dữ Liệu](#_Toc197279391) 19

[3.3. Phương Pháp Phân Tích Cảm Xúc](#_Toc197279392) 21

[3. 4. Phương Pháp Dự Đoán Xu Hướng Thị Trường](#_Toc197279396) 24

[3.5. Phương Pháp Trực Quan Hóa Dữ Liệu](#_Toc197279397) 25

[3.6. Công Nghệ Và Công Cụ Sử Dụng Trong Dự Án](#_Toc197279388) 25

[CHƯƠNG 4: KẾT QUẢ ĐẠT ĐƯỢC](#_Toc197279398) 26

[4.1. Kết Quả Phân Tích Cảm Xúc Tin Tức Tài Chính](#_Toc197279399) 26

[4.2. Kết Quả Dự Đoán Xu Hướng Thị Trường](#_Toc197279400) 27

[4.3. Kết Quả Trực Quan Hóa Dữ Liệu](#_Toc197279401) 28

[CHƯƠNG 5: GIAO DIỆN NGƯỜI DÙNG VÀ QUY TRÌNH PHÂN TÍCH](#_Toc197279402) 29

[5.1. Tổng Quan Về Giao Diện Người Dùng](#_Toc197279403) 29

[5.2. Luồng Tương Tác Người Dùng](#_Toc197279403) 33

[5.3. Thử Nghiệm Và Đánh Giá Kết Quả Đạt Được](#_Toc197279403) 34

[CHƯƠNG 6: KẾT LUẬN VÀ HÀM Ý QUẢN TRỊ 38](#_Toc197279420)8

[6.1. Kết Luận 29](#_Toc197279403)8

[6.2. Những Phát Hiện Chính 29](file:///C:\Users\admin\Downloads\NHẬN%20DIỆN%20ĐỘNG%20VẬT%20BẰNG%20MẠNG%20NƠ%202.docx#_Toc197279403)8

[6.3. Hạn Chế Của Dự Án 29](file:///C:\Users\admin\Downloads\NHẬN%20DIỆN%20ĐỘNG%20VẬT%20BẰNG%20MẠNG%20NƠ%202.docx#_Toc197279403)8

[6.4. Hàm Ý Quả Trị 29](file:///C:\Users\admin\Downloads\NHẬN%20DIỆN%20ĐỘNG%20VẬT%20BẰNG%20MẠNG%20NƠ%202.docx#_Toc197279403)9

[TÀI LIỆU THAM KHẢO](#_Toc197279420) 40

[Phụ lục](#_Toc197279420) 41

# 

# **CHƯƠNG 1: GIỚI THIỆU VỀ ĐỀ TÀI**

## **1. Đặt Vấn Đề**

## Trong bối cảnh nền kinh tế hiện đại, thị trường tài chính luôn biến động không ngừng với vô số các yếu tố tác động như chính sách kinh tế vĩ mô, tình hình địa chính trị, hiệu quả hoạt động của doanh nghiệp và đặc biệt là tâm lý thị trường. Tâm lý này được phản ánh trực tiếp thông qua các thông tin, tin tức được đăng tải trên các phương tiện truyền thông, báo chí và các trang web tài chính. Trong môi trường đầu tư và kinh doanh đầy biến động, việc nắm bắt được xu hướng thị trường một cách chính xác và kịp thời trở thành yếu tố then chốt quyết định đến thành công của nhà đầu tư.

## Thực tế, việc phân tích tin tức tài chính là một công việc không hề đơn giản. Ngay cả với các nhà đầu tư giàu kinh nghiệm, việc đánh giá chính xác tác động của một thông tin đến thị trường - nơi hàng trăm yếu tố đan xen phức tạp - vẫn là một thách thức lớn. Trước đây, các nhà đầu tư thường phải theo dõi liên tục nhiều nguồn tin, đọc hàng chục bài báo mỗi ngày, rồi tự phân tích và đánh giá thủ công. Phương pháp này không chỉ tốn thời gian và công sức, mà còn dễ bị ảnh hưởng bởi các yếu tố chủ quan, dẫn đến những quyết định đầu tư thiếu chính xác.

## Trong những năm gần đây, với sự phát triển mạnh mẽ của công nghệ xử lý ngôn ngữ tự nhiên (NLP) - đặc biệt là các kỹ thuật phân tích cảm xúc (sentiment analysis) - khả năng xử lý và phân tích văn bản đã được cải thiện đáng kể. Các phương pháp phân tích cảm xúc cho phép máy tính tự động đánh giá mức độ tích cực hay tiêu cực của một văn bản, từ đó giúp trích xuất được thông tin về tâm lý thị trường mà không cần sự can thiệp thủ công quá nhiều. Điều đáng chú ý là các mô hình này không chỉ nhận diện được các tin tức rõ ràng về tăng giảm, mà còn có khả năng đánh giá các bài viết có nội dung phức tạp, ngầm ẩn nhiều thông tin tác động đến thị trường.

## Việc áp dụng kỹ thuật phân tích cảm xúc trong đánh giá tin tức tài chính đã được thử nghiệm ở nhiều thị trường phát triển. Tại các trung tâm tài chính lớn như New York, London hay Tokyo, các quỹ đầu tư thường xuyên sử dụng các thuật toán tự động để quét và phân tích hàng nghìn bài báo mỗi ngày, giúp đưa ra các quyết định giao dịch nhanh chóng và hiệu quả. Trong lĩnh vực tư vấn đầu tư, các công ty cũng ứng dụng công nghệ này để cung cấp thông tin tổng hợp và khuyến nghị kịp thời cho khách hàng. Ngoài ra, một số nền tảng giao dịch trực tuyến cũng bắt đầu tích hợp các công cụ phân tích tin tức tự động, giúp nhà đầu tư cá nhân có được cái nhìn tổng quan về tâm lý thị trường mà không cần phải đọc qua từng bài báo.

## Tuy nhiên, dù mang lại nhiều hứa hẹn, việc ứng dụng phân tích cảm xúc vào bài toán dự đoán xu hướng thị trường không phải không có khó khăn, đặc biệt là đối với thị trường Việt Nam. Trở ngại đầu tiên chính là đặc thù của ngôn ngữ tiếng Việt. Hầu hết các công cụ phân tích cảm xúc hiện có được phát triển cho tiếng Anh và không thể áp dụng trực tiếp cho tiếng Việt do sự khác biệt về từ vựng, ngữ pháp và cách diễn đạt. Việc xây dựng một hệ thống phân tích cảm xúc chuyên biệt cho tin tức tài chính tiếng Việt đòi hỏi phải có một bộ từ điển riêng và thuật toán được điều chỉnh phù hợp.

## Ngoài ra, một thách thức lớn khác là đặc trưng của tin tức tài chính. Không giống như các bài đánh giá sản phẩm hay bình luận mạng xã hội - nơi cảm xúc thường được thể hiện trực tiếp - các bài báo tài chính thường có cách viết khách quan, trung lập, đôi khi sử dụng các thuật ngữ chuyên ngành phức tạp. Việc đánh giá chính xác mức độ tích cực hay tiêu cực của những tin tức như vậy đòi hỏi phải hiểu sâu về ngữ cảnh và tác động tiềm ẩn của thông tin đối với thị trường.

## Không dừng lại ở đó, một vấn đề nữa là tính biến động của thị trường tài chính. Một tin tức được đánh giá là "tích cực" trong một bối cảnh thị trường có thể lại được xem là "tiêu cực" trong một bối cảnh khác, tùy thuộc vào nhiều yếu tố như chu kỳ kinh tế, tâm lý đám đông, hay sự kỳ vọng của nhà đầu tư. Điều này đặt ra thách thức lớn cho việc xây dựng một mô hình dự đoán xu hướng thị trường dựa trên phân tích cảm xúc tin tức.

Từ những vấn đề và thách thức nêu trên, dự án "Phân tích Xu hướng Thị trường" được phát triển nhằm tạo ra một giải pháp tích hợp, tự động hóa việc thu thập và phân tích tin tức tài chính tiếng Việt, từ đó đưa ra dự báo về xu hướng thị trường dựa trên cảm xúc được phản ánh trong các bài viết. Hệ thống này sẽ giúp các nhà đầu tư tiết kiệm thời gian, có cái nhìn tổng quan và khách quan hơn về tình hình thị trường, từ đó đưa ra quyết định đầu tư sáng suốt hơn trong bối cảnh thông tin luôn biến động của thị trường tài chính Việt Nam.

## **2. Các Nghiên Cứu Liên Quan**

Trong những năm gần đây, việc ứng dụng xử lý ngôn ngữ tự nhiên (NLP) và học máy vào phân tích cảm xúc từ tin tức tài chính đã nhận được nhiều sự quan tâm, đặc biệt trong bối cảnh thị trường chứng khoán biến động mạnh mẽ theo các yếu tố tâm lý nhà đầu tư và thông tin truyền thông. Nhiều nghiên cứu đã chứng minh hiệu quả của việc sử dụng mạng nơ-ron sâu, mô hình học tăng cường và từ điển cảm xúc chuyên biệt để trích xuất thông tin cảm xúc nhằm hỗ trợ dự báo xu hướng thị trường.

Fazlija và cộng sự (2022) đã sử dụng mô hình FinBERT – một biến thể của BERT được huấn luyện trên dữ liệu tài chính – để phân tích cảm xúc từ tin tức Bloomberg và Reuters, sau đó kết hợp điểm số cảm xúc này với dữ liệu thị trường để dự đoán xu hướng chỉ số S&P 500. Kết quả chỉ ra rằng cảm xúc trích xuất từ nội dung bài viết có đóng góp đáng kể trong việc cải thiện độ chính xác dự báo, so với việc chỉ sử dụng tiêu đề hoặc dữ liệu kỹ thuật.

Maqbool và cộng sự (2023) đề xuất một phương pháp kết hợp điểm số cảm xúc từ nhiều công cụ phân tích (VADER, TextBlob...) với mô hình MLP-Regressor để dự báo xu hướng giá cổ phiếu. Nghiên cứu được thực hiện trên dữ liệu 10 năm của bốn công ty đại diện cho các ngành khác nhau. Mô hình cho kết quả khả quan với độ chính xác đạt 0.90 cho dự báo 10 ngày, cho thấy tiềm năng lớn của việc tích hợp thông tin cảm xúc trong chiến lược đầu tư ngắn hạn.

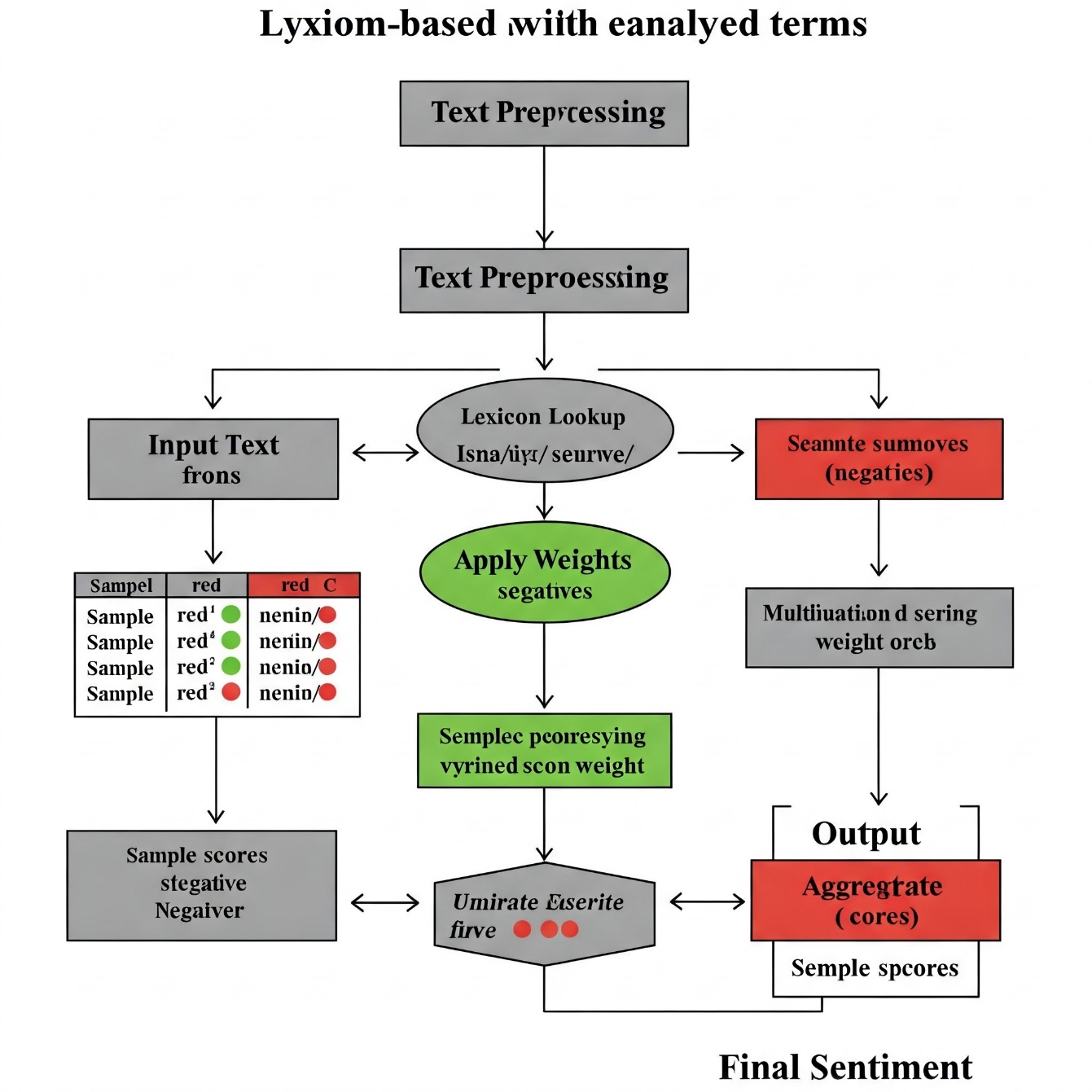
Zeng và Jiang (2023) cũng khai thác FinBERT để xây dựng hệ thống dự đoán hướng đi của thị trường. Bằng cách kết hợp mô hình phân tích cảm xúc với mạng LSTM, họ cho thấy mô hình đạt hiệu quả cao trong việc dự đoán biến động ngắn hạn, đồng thời khẳng định rằng cảm xúc tích cực/tiêu cực từ tin tức tài chính có tương quan chặt chẽ với sự thay đổi giá cổ phiếu trong ngắn hạn.

Các nghiên cứu trên đã đặt nền móng vững chắc cho việc sử dụng phân tích cảm xúc như một công cụ bổ trợ trong đầu tư tài chính. Tuy nhiên, hầu hết các mô hình đều yêu cầu nguồn dữ liệu lớn và kỹ thuật tính toán cao. Do đó, hệ thống được phát triển trong dự án này tập trung vào việc kết hợp từ điển cảm xúc tiếng Việt có trọng số với giao diện phân tích đơn giản, nhằm mang lại giải pháp dễ triển khai, hiệu quả và phù hợp với người dùng phổ thông trong việc đánh giá xu hướng thị trường thông qua tin tức tài chính.

# **CHƯƠNG 2. PHÂN TÍCH MÔ HÌNH**

## **2.1. Tổng Quan Về Mô Hình Phân Tích Cảm Xúc**

Dự án "***Phân Tích Cảm Xúc Từ Tin Tức Tài Chính Để Dự Đoán Xu Hướng Thị Trường***" sử dụng một mô hình phân tích cảm xúc dựa trên từ điển có trọng số (Weighted lexicon-based sentiment analysis), được thiết kế đặc biệt cho tin tức tài chính tiếng Việt. Đây là một phương pháp phù hợp cho việc phân tích văn bản chuyên ngành, đặc biệt là trong lĩnh vực tài chính nơi các thuật ngữ chuyên biệt có ý nghĩa rõ ràng về mặt cảm xúc.



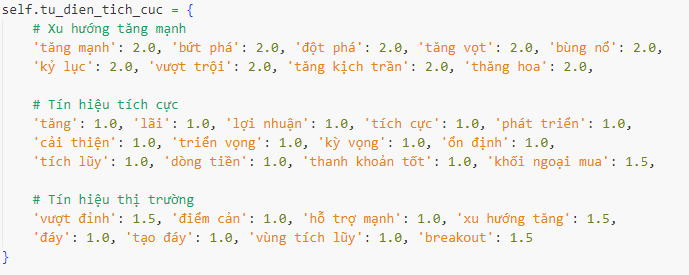
Hình 2.1. Ví dụ về mô hình

* 1. **. Cấu Trúc Của Mô Hình Phân Tích Cảm Xúc**

**2.2.1. Từ điển cảm xúc có trong số**

Mô hình sử dụng hai từ điển cảm xúc chính:

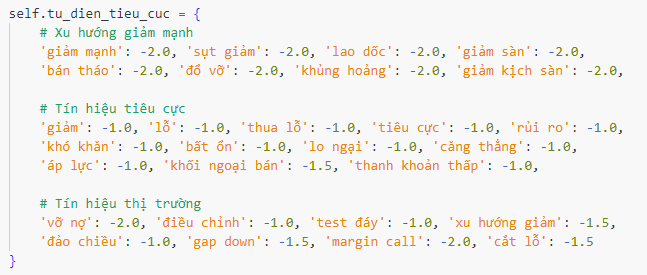
**1. Từ điển tích cực:**

* Chứa các thuật ngữ tài chính tiếng Việt mang ý nghĩa tích tực
* Phân chia thành 3 nhóm chính
  + **Xu hướng tăng mạnh:** gồm các từ như *“tăng mạnh”, “bứt phá”, “đột phá”* với trọng số 2.0.
  + **Tín hiệu tích cực:** gồm các từ như *“tăng”, “lãi”, “lợi nhuận”* với trọng số 1.0.
  + **Tín hiệu thì trường:** gồm các từ như *“vượt đỉnh”, “hỗ trợ mạnh”* với trong số từ 1.0-1.5.

Hình 2.2.1.1. Trích đoạn code miêu tả từ điển tiêu cực

**2. Từ điển tiêu cực:**

* Chứa các thuật ngữ tài chính tiếng Việt mang ý nghĩa tiêu tực
* Phân chia thành 3 nhóm chính
  + **Xu hướng giảm mạnh:** gồm các từ như *“giảm mạnh”, “sụt giảm”, “lao dốc”* với trọng số 2.0.
  + **Tín hiệu tiêu cực:** gồm các từ như *“giảm”, “lổ”, “rủi ro”* với trọng số -1.0.
  + **Tín hiệu thì trường:** gồm các từ như *“vỡ nợ”, “điều chỉnh”* với trong số từ -1.0 đến -2.0.



Hình 2.2.1.2. Trích đoạn code miêu tả từ điển tiêu cực

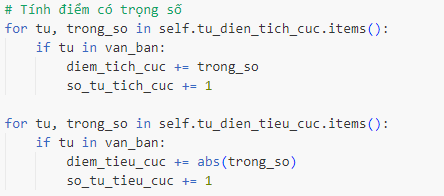
**2.2.2. Thuật toán tính điểm cảm xúc**

Mô hình sử dụng thuật toán phân tích cảm xúc dựa trên sự xuất hiện của các từ trong văn bản như sau:

Bước 1: Chuyển văn bản đầu vào thành chữ thường để đảm bảo khớp chính xác với từ điển.

Bước 2: Khởi tạo các biến đếm điểm và số lượng từ.

Bước 3: Duyệt qua từ điển tích cực, tìm kiếm mỗi từ trong văn bản

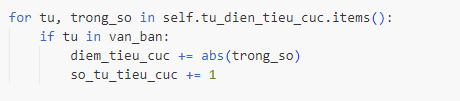
* + Nếu tìm thấy, cộng giá trị tuyệt đối của trọng số vào biến diem\_tieu\_cuc

*Hình ảnh miêu tả bước cộng giá trị vào biến diem\_tieu\_cuc*

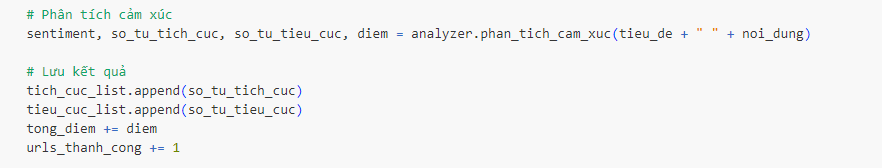
* + Tăng biến điếm số lượng từ tích cực

Bước 4: Duyệt qua từ điển tiêu cực, tìm kiếm mỗi từ trong văn bản

* Nếu tìm thấy, cộng giá trị tuyệt đối của trọng số vào biến diem\_tieu\_cuc

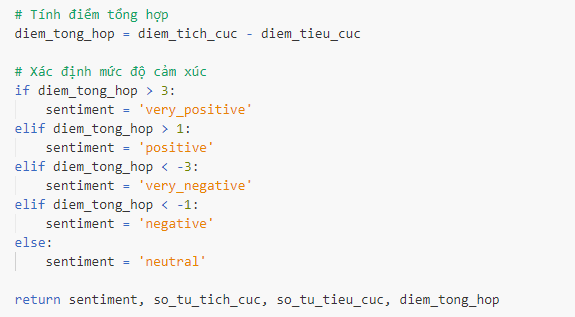


*Hình ảnh miêu tả bước cộng giá trị tuyệt đối vào biến diem\_tieu\_cuc*

* Tăng biến đếm số lượng từ tiêu cực

*Hình ảnh miêu tả bước tăng biến điếm số lượng từ tiêu cực*

Bước 5: Tính điểm tổng hợp bằng cách lấy biến **diem\_tich\_cuc – diem\_tieu\_cuc**

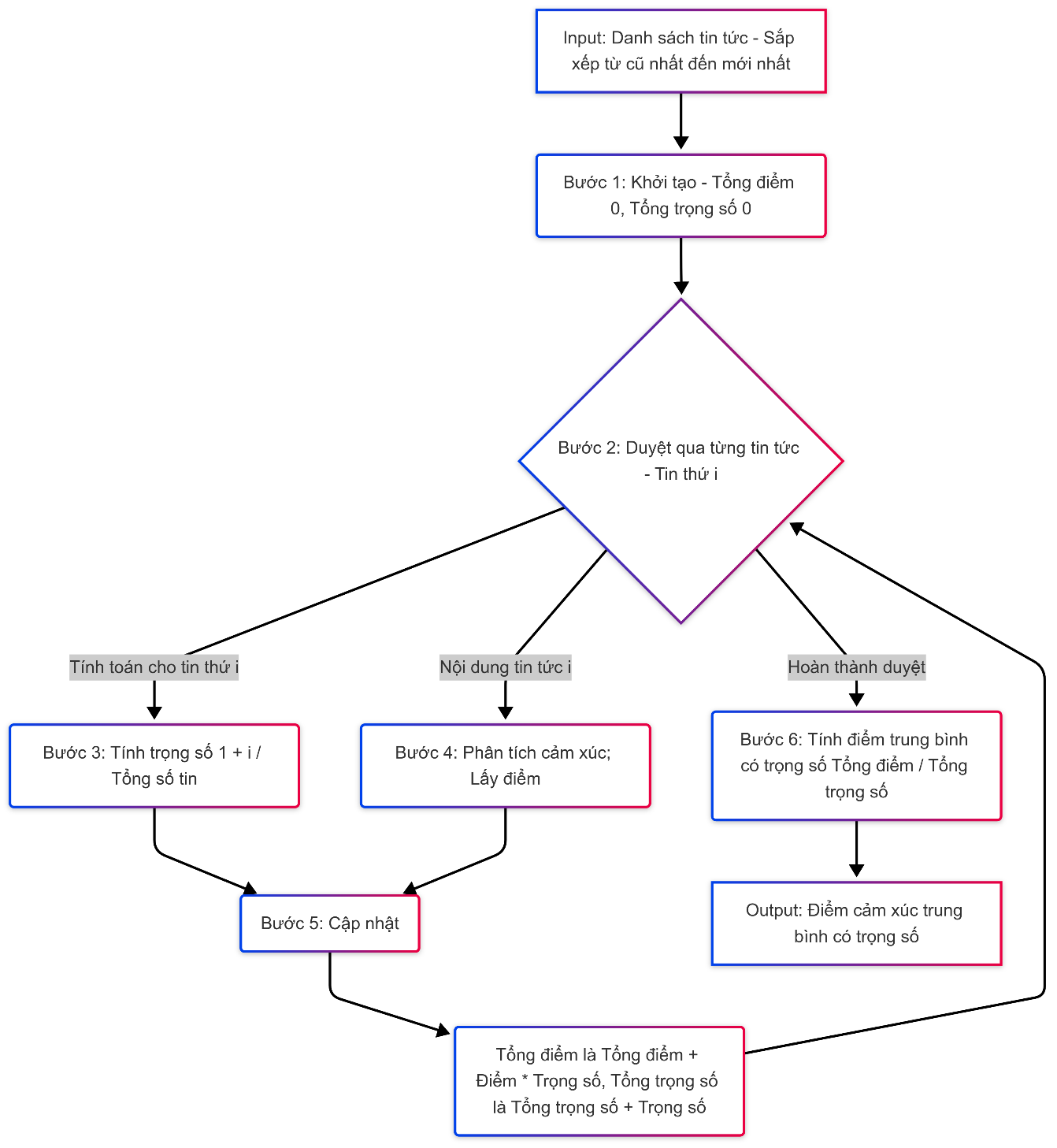
****

*Hình ảnh miêu tả cho bước 5*

Bước 6: Phân loại cảm xúc dựa trên các ngưỡng:

* Rất tích cực: điểm > 3
* Tích cực: 1 < điểm ≤ 3
* Trung lập: -1 ≤ điểm ≤ 1
* Tiêu cực: -3 < điểm < -1
* Rất tiêu cực: điểm ≤ -3

**2.2.3. Thuật toán tổng hợp cảm xúc theo thời gian**

Mô hình sử dụng một thuật toán đặc biệt để tổng hợp cảm xúc từ nhiều bài báo, có tính đến độ mới của tin tức:

Hình 2.2.3. Mô tả thuận toán đã sử dụng trong dự án

* 1. **.Ưu Điểm Và Hạn Chế Của Mô Hình**

2.3.1 Ưu điểm

* **Đơn giản và hiệu quả:** Mô hình dựa trên từ điển có thể triển khai nhanh chóng và không yêu cầu dữ liệu huấn luyện lớn.
* **Phù hợp với lĩnh vực chuyên biệt:** Từ điển được thiết kế đặc biệt cho tin tức tài chính tiếng Việt, giúp phát hiện chính xác các thuật ngữ chuyên ngành.
* **Hệ thống trọng số linh hoạt:** Cho phép điều chỉnh mức độ ảnh hưởng của từng từ khóa dựa trên tầm quan trọng thực tế trong lĩnh vực tài chính.
* **Xem xét độ mới của tin tức**: Thuật toán tổng hợp có trọng số thời gian đảm bảo tin mới hơn có ảnh hưởng lớn hơn đến kết quả cuối cùng.
* **Hiệu quả tính toán:** Thuật toán có độ phức tạp O(m\*n) với m là kích thước từ điển và n là độ dài văn bản, phù hợp cho xử lý thời gian thực.

2.3.2 Hạn chế

* **Thiếu khả năng xử lý ngữ cảnh:** Mô hình không xem xét ngữ cảnh của từ. Ví dụ, "không tăng" vẫn được tính là tích cực vì từ "tăng".
* **Phụ thuộc vào từ điển:** Hiệu quả của mô hình phụ thuộc rất lớn vào chất lượng và độ phủ của từ điển. Các thuật ngữ mới hoặc ít gặp có thể không được nhận diện.
* **Không xử lý được các cấu trúc phức tạp:** Mô hình không thể hiểu được các cấu trúc ngôn ngữ phức tạp như châm biếm, ẩn dụ hay so sánh.
* **Khó mở rộng:** Việc cập nhật và mở rộng từ điển đòi hỏi công sức thủ công.
* **Không học từ dữ liệu:** Khác với các mô hình học máy, hệ thống này không thể tự cải thiện qua thời gian dựa trên dữ liệu mới.

## **2.4.Thuật Toán Dự Đoán Xu Hướng Thị Trường**

Dựa trên điểm cảm xúc tổng hợp, mô hình sẽ được dự đoán xu hướng thị trường và đưa ra các khuyến nghị giao dịch:

2.4.1 Quy tắc dự đoán xu hướng

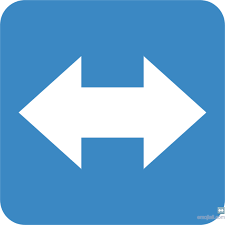
1. Tăng mạnh (điểm > 3)

* Biểu thị bằng
* Đặc điểm: Thị trường có xu hướng tăng rất mạnh với nhiều tín hiệu tích cực

2. Tăng (1 < điểm ≤ 3)

* Biểu thị bằng: 
* Đặc điểm: Thị trường có xu hướng tăng nhẹ

3. Đi ngang (-1 ≤ điểm ≤ 1)

* Biểu thị bằng: 
* Đặc điểm: Thị trường đang trong trạng thái đi ngang, chưa có xu hướng rõ rệt

4. Giảm (-3 < điểm < -1)

* Biểu thị bằng: 
* Đặc điểm: Thị trường có xu hướng giảm nhẹ

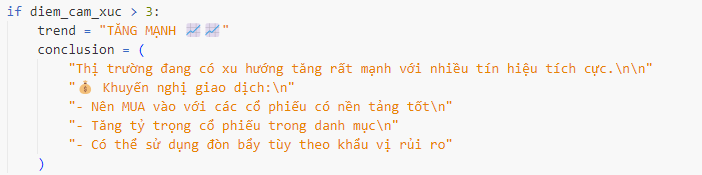
5. Giảm mạnh (điểm ≤ -3)

* Biểu thị bằng: 
* Đặc điểm: Thị trường có xu hướng giảm mạnh với nhiều tín hiệu tiêu cực

2.4.2 Khuyến nghị giao dịch dựa trên xu hướng

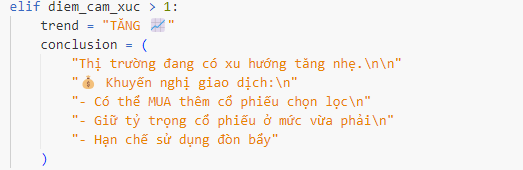
Cho mỗi xu hướng thị trường, mô hình đưa ra các khuyến nghị giao dịch cụ thể

* Khi thị trường **Tăng Mạnh**
* Nên MUA vào với các cổ phiếu có nền tảng tốt
* Tăng tỷ trọng cổ phiếu trong danh mục
* Có thể sử dụng đòn bẩy tùy theo khẩu vị rủi ro



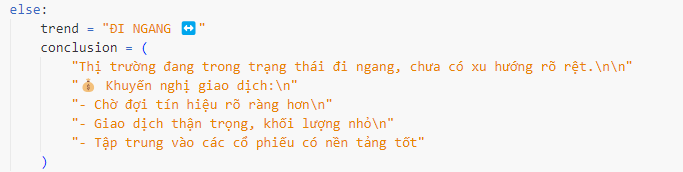
*Hình ảnh biểu diễn đoạn code khi thị trường Tăng Mạnh*

* Khi thị trường **Tăng**
  + Có thể MUA thêm cổ phiếu chọn lọc
  + Giữ tỷ trọng cổ phiếu ở mức vừa phải
  + Hạn chế sử dụng đòn bẩy



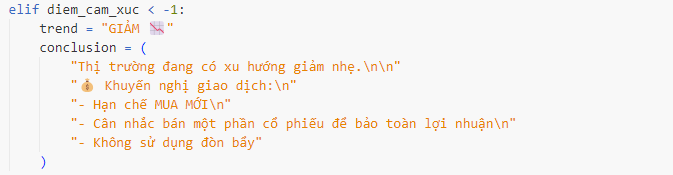
*Hình ảnh biểu diễn đoạn code khi thị trường Tăng*

* Khi thị trường **Đi Ngang**
  + Chờ đợi tín hiệu rõ ràng hơn
  + Giao dịch thận trọng, khối lượng nhỏ
  + Tập trung vào các cổ phiếu có nền tảng tốt



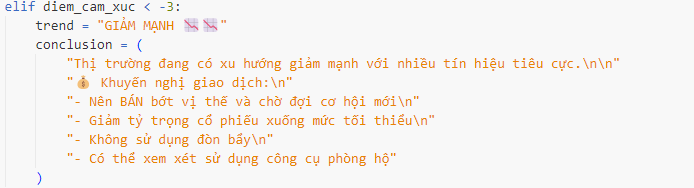
*Hình ảnh biểu diễn đoạn code khi thị trường Đi Ngang*

* Khi thị trường **Giảm**
* Hạn chế MUA MỚI
* Cân nhắc bán một phần cổ phiếu để bảo toàn lợi nhuận
* Không sử dụng đòn bẩy



*Hình ảnh biểu diễn đoạn code khi thị trường Giảm*

* Khi thị trường **Giảm Mạnh**
  + Nên BÁN bớt vị thế và chờ đợi cơ hội mới
  + Giảm tỷ trọng cổ phiếu xuống mức tối thiểu
  + Không sử dụng đòn bẩy
  + Có thể xem xét sử dụng công cụ phòng hộ

  
*Hình ảnh biểu diễn đoạn code khi thị trường Giảm Mạnh*

* 1. **. Trực Quan Hóa Kết Quả Phân Tích**

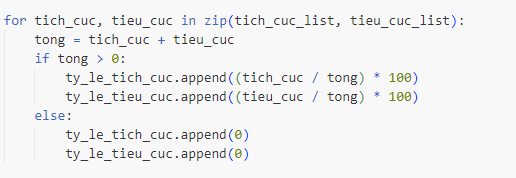
## Mô hình sử dụng matplotlib để trực quan hóa kết quả phân tích cảm xúc, giúp người dùng dễ dàng đánh giá tỷ lệ từ ngữ tích cực/tiêu cực trong mỗi bài báo:

**2.5.1.Thuật toán trực quan hóa**

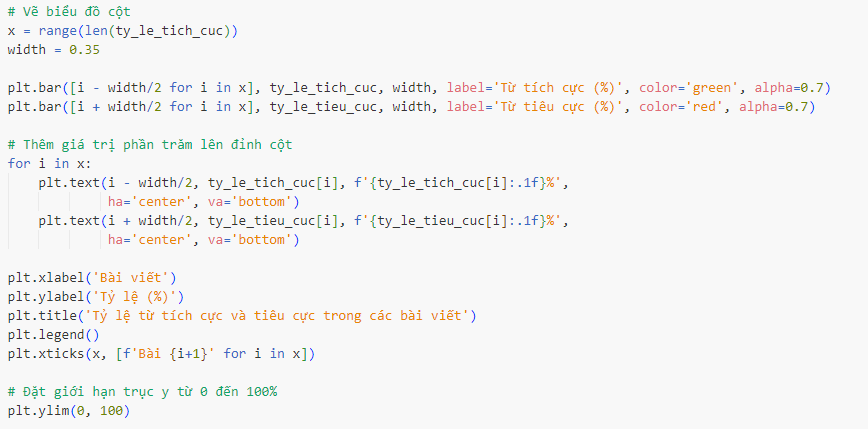
Thuật toán trực quan hóa cảm xúc thực hiện các bước sau:

1. Tính toán tỷ lệ cảm xúc:

* Với mỗi bài viết, tính tổng số từ có cảm xúc
* Tính tỷ lệ phần trăm từ tích cực: (tich\_cuc / tong) \* 100
* Tính tỷ lệ phần trăm từ tiêu cực: (tieu\_cuc / tong) \* 100

  
 *Hình ảnh miêu tả đoạn code tính tỷ lệ cảm xúc*

2. Tạo biểu đồ cột nhóm:

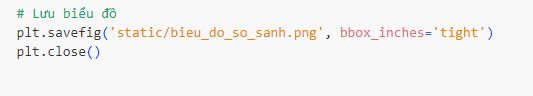
* Sử dụng matplotlib để tạo biểu đồ cột nhóm
* Mỗi bài viết có hai cột: một cột xanh lá cho từ tích cực và một cột đỏ cho từ tiêu cực
* Thêm nhãn phần trăm lên đỉnh mỗi cột
* Đặt giới hạn trục y từ 0 đến 100%

*Hình ảnh miêu tả code tạo biểu đồ cột nhóm*

3. Định dạng biểu đồ:

* Thêm nhãn trục x (Bài viết) và trục y (Tỷ lệ %)
* Thêm tiêu đề biểu đồ
* Thêm chú thích

4. Lưu biểu đồ:

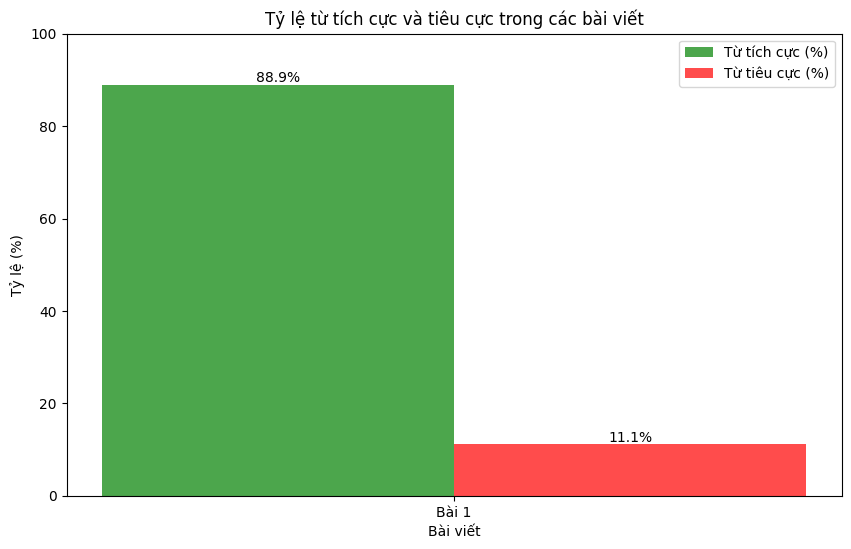
* Lưu biểu đồ dưới dạng file PNG trong thư mục static với tên bieu\_do\_so\_sanh.png

*Hình ảnh miêu tả đoạn code lưu biểu đồ*

**2.5.2. Ý nghĩa của biểu đồ**

Biểu đồ này cung cấp một cách trực quan để so sánh tỷ lệ từ tích cực và tiêu cực trong các bài báo được phân tích, giúp người dùng:

* Nhận biết nhanh chóng xu hướng cảm xúc của mỗi bài báo
* So sánh mức độ tích cực/tiêu cực giữa các bài báo khác nhau
* Xác định các bài báo có tỷ lệ từ tích cực/tiêu cực cao bất thường
* Đánh giá mức độ nhất quán của cảm xúc giữa các bài báo

  
*Hình ảnh ví dụ minh họa biểu đồ*

## **2.6. Tích Hợp Mô Hình Vào Ứng Dụng Web**

Mô hình phân tích cảm xúc được tích hợp vào ứng dụng web Flask, cho phép người dùng:

* **Nhập danh sách URL** tin tức tài chính cần phân tích
* **Phân tích tự động cảm** xúc từ mỗi URL
* **Xem kết quả tổng hợp** bao gồm
* Danh sách các URL đã phân tích với điểm cảm xúc
* Biểu đồ tỷ lệ từ tích cực/tiêu cực
* Dự đoán xu hướng thị trường
* Khuyến nghị giao dịch dựa trên xu hướng

**2.7. Đánh Giá Hiệu Quả Của Mô Hình  
2.7.1. Độ chính xác và hiệu suất**

Mặc dù không có dữ liệu kiểm thử chính thức, nhưng mô hình này có thể đạt được độ chính xác tương đối cao trong lĩnh vực tài chính nhờ:

* Từ điển chuyên ngành được thiết kế đặc biệt cho tin tức tài chính tiếng Việt
* Hệ thống trọng số phản ánh mức độ ảnh hưởng thực tế của các thuật ngữ đến thị trường
* Tiếp cận theo quy tắc đơn giản nhưng hiệu quả cho lĩnh vực chuyên biệt

**2.7.2. Khả năng mở rộng và cải tiếng**

Dự án “**Phân Tích Cảm Xúc Từ Tin Tức Tài Chính Để Dự Đoán Xu Hướng Thị Trường**” hiện tại đã xây dựng được một nền tảng cơ bản cho việc phân tích cảm xúc tin tức tài chính tiếng Việt và dự đoán xu hướng thị trường. Tuy nhiên, để nâng cao hiệu quả và độ chính xác của hệ thống, chúng em dự định sẽ triển khai các cải tiến sau trong tương lai:

1. **Mở rộng từ điển cảm xúc:** Thêm nhiều thuật ngữ tài chính mới và điều chỉnh trọng số dựa trên phản hồi
2. **Tích hợp xử lý ngữ cảnh:** Sử dụng kỹ thuật NLP tiên tiến hơn để xử lý ngữ cảnh của từ
3. **Kết hợp với mô hình học máy:** Sử dụng từ điển làm tính năng đầu vào cho mô hình học máy
4. **Tự động cập nhật từ điển:** Phát triển cơ chế tự động cập nhật từ điển dựa trên dữ liệu mới
5. **Tích hợp phân tích kỹ thuật:** Kết hợp phân tích cảm xúc với phân tích kỹ thuật để dự đoán chính xác hơn

# **CHƯƠNG 3: PHƯƠNG PHÁP NGHIÊN CỨU**

## **3.1. Tổng Quan Phương Pháp Nghiên Cứu**

Dự án " **Phân Tích Cảm Xúc Từ Tin Tức Tài Chính Để Dự Đoán Xu Hướng Thị Trường** " được xây dựng dựa trên phương pháp nghiên cứu định lượng kết hợp với kỹ thuật phân tích văn bản và trực quan hóa dữ liệu. Phương pháp nghiên cứu được áp dụng trong dự án bao gồm các bước chính sau:

1. **Thu thập dữ liệu:** Thu thập tin tức tài chính tiếng Việt từ các nguồn trực tuyến
2. **Tiền xử lý dữ liệu:** Xử lý và chuẩn hóa nội dung tin tức
3. **Phân tích cảm xúc:** Áp dụng phương pháp phân tích cảm xúc dựa trên từ điển có trọng số
4. **Tổng hợp và phân tích kết quả:** Tổng hợp điểm cảm xúc và dự đoán xu hướng thị trường
5. **Trực quan hóa kết quả:** Biểu diễn kết quả phân tích dưới dạng biểu đồ và báo cáo

## **3.2. Phương Pháp Thu Thập Dữ Liệu**

**3.2.1 Nguồn tài liệu**

Dự án sử dụng phương pháp thu thập dữ liệu từ các trang tin tức tài chính trực tuyến của Việt Nam. Các nguồn tin chính được mã hóa trực tiếp trong chương trình bao gồm:

1. **CafeF (cafef.vn):** Một trong những trang tin tức tài chính hàng đầu Việt Nam, cung cấp thông tin về thị trường chứng khoán, bất động sản, ngân hàng, và các phân tích chuyên sâu.
2. **VnExpress (vnexpress.net):** Báo điện tử với mục tài chính - chứng khoán cập nhật thường xuyên.
3. **TheLeader (theleader.vn):** Trang thông tin chuyên về lãnh đạo, quản trị và tài chính, thường xuyên đưa tin về thị trường.
4. **Tin nhanh Chứng khoán (tinnhanhchungkhoan.vn):** Chuyên trang tin tức về thị trường chứng khoán Việt Nam.
5. **NDH (ndh.vn):** Nguồn cung cấp thông tin về đầu tư và tài chính.

**3.2.2 Kỹ thuật thu thập dữ liệu** Dự án sử dụng kỹ thuật web scraping để thu thập dữ liệu từ các trang web tin tức. Quy trình thu thập dữ liệu được thực hiện như sau:

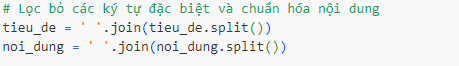
 1. **Gửi HTTP request:** sử dụng *thư viện requests* để gửi HTTP request đến URL của bài báo

*Trích dẫn code gửi HTTP request*  
2. **Phân tích cú pháp HTML:** Sử dụng *thư viện BeautifulSoup* để phân tích đội dung HTML và trích xuất thông tin

*Trích dẫn code phân tích cú phân tích cú pháp*  
3. **Trích xuất nội dung theo domain cụ thể:** Áp dụng các selector HTML khác nhau tùy thuộc vào trang web  
 *Trích dẫn code trích xuất nội dung theo domain cụ thể*

4. **Xử lý trường hợp không xác định:** Với các trang web không được định nghĩa trước, hệ thống sẽ áp dụng thuật toán phát hiện heuristic

*Trích dẫn code xử lý trường hợp không xác định*

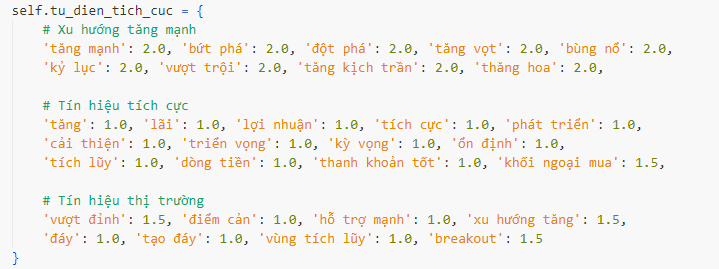
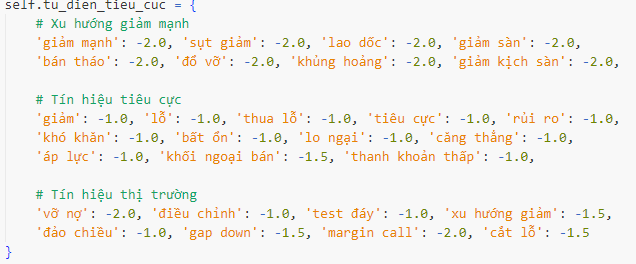
 5. **Chuẩn hóa nội dung:** Loại bỏ ký tự đặt biệt và chuẩn hóa khoảng trắng

*Trích dẫn code chuẩn hóa nội dung*

**3.3. Phương Pháp Phân Tích Cảm Xúc**  
 **3.3.1. Phương pháp dựa trên từ điển có trọng số**

Dự án sử dụng phương pháp phân tích cảm xúc dựa trên từ điển (lexicon-based) với các trọng số được xác định trước. Phương pháp này bao gồm:

1. Xây dựng từ điển cảm xúc: Hai từ điển riêng biệt được xây dựng:

* Từ điển tích cực chứa các thuật ngữ tài chính mang ý nghĩa tích cực  
    
   *Trích dẫn code mô tả từ điển tích cực*
* Từ điển tiêu cực chứa các thuật ngữ tài chính mang ý nghĩa tiêu cực

*Trích dẫn code mô tả từ điển tiêu cực*

1. Gán trọng số cho từng thuật ngữ: Mỗi thuật ngữ được gán một trọng số phản ánh mức độ tích cực/tiêu cực của nó:

* Thuật ngữ tích cực: từ +1.0 đến +2.0
* Thuật ngữ tiêu cực: từ -1.0 đến -2.0

1. Phân loại thuật ngữ theo chủ đề: Các thuật ngữ được phân loại thành các nhóm như:

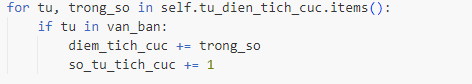
* Xu hướng tăng/giảm mạnh
* Tín hiệu tích cực/tiêu cực
* Tín hiệu thị trường

**3.3.2. Quy trình phân tích cảm xúc** Quy trình phân tích cảm xúc được thực hiện theo các bước sau:

1. **Chuẩn hóa văn bản:** Chuyển đổi văn bản sang chữ thường

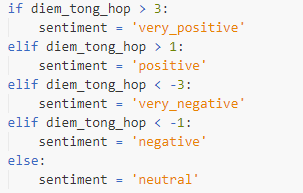
 *Trích dẫn code chuẩn hóa văn bản*

2. **Tìm kiếm thuật ngữ trong văn bản:** Duyệt qua từng thuật ngữ trong từ điển và kiểm tra sự xuất hiện trong văn bản

  
 *Trích dẫn code tìm kiếm thuật ngữ trong văn bản*

3. **Tính điểm cảm xúc:** Điểm cảm xúc tổng hợp được tính bằng hiệu của điểm tích cực và điểm tiêu cực

  
 *Trích dẫn code tính điểm cảm xúc*

4. **Phân loại cảm xúc:** Dựa trên điểm tổng hợp, văn bản được phân loại thành các mức độ cảm xúc  
  
 *Trích dẫn code phân loại cảm xúc*

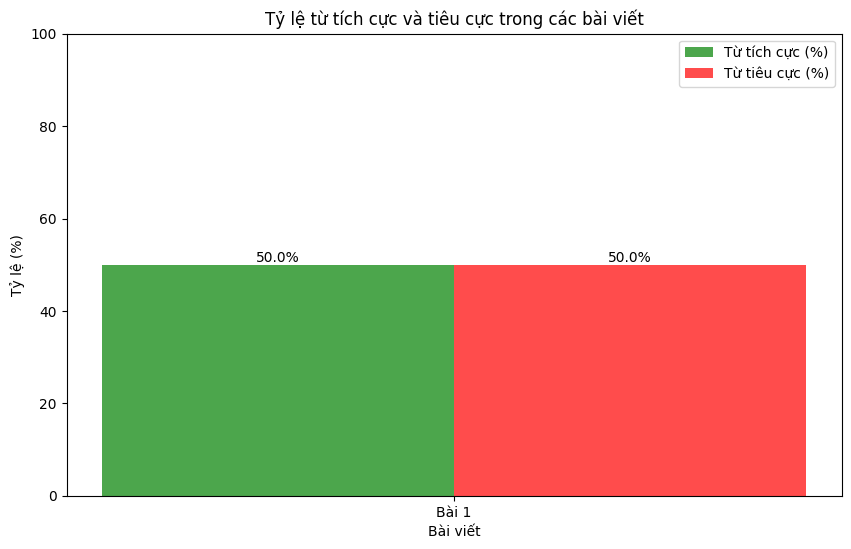
**3.3.3. Tổng hợp điểm cảm xúc theo thời gian**  
Dự án áp dụng phương pháp tổng hợp điểm cảm xúc có trọng số theo thời gian, với các bước:  
 1. **Gán trọng số theo thời gian:** Các bài viết mới hơn được gán trọng số cao hơn  
 *Trích dẫn code gáng trọng số theo thời gian* 2. **Tính điểm trung bình có trọng số:** Điểm cảm xúc cuối cùng là trung bình có trọng số của tất cả các bài viết  
  
 *Trích dẫn code tính điểm trung bình có trọng số*

## **3.4 Phương Pháp Dự Đoán Xu Hướng Thị Trường 3.4.1 Phương pháp phân loại dựa trên ngưỡng** Dự án sử dụng phương pháp phân loại dựa trên ngưỡng để dự đoán xu hướng thị trường: 1. Xác định ngưỡng: Thiết lập các ngưỡng điểm cảm xúc để phân loại xu hướng *Trích dẫn đoạn code xác định ngưỡng lấy từ hàm du\_doan\_xu\_huong*

2. Đưa ra khuyến nghị: Dựa trên xu hướng được dự đoán, hệ thống đưa ra các khuyến nghị giao dịch phù hợp

**3.4.2 Quy trình đưa ra khuyến nghị** Quy trình đưa ra khuyến nghị giao dịch dựa trên xu hướng được thực hiện như sau**:**

* **Xác định xu hướng:** Dựa trên điểm cảm xúc tổng hợp, xác định xu hướng thị trường
* **Xây dựng khuyến nghị:** Cho mỗi xu hướng, xây dựng một bộ khuyến nghị giao dịch cụ thể
* **Định dạng kết quả:** Kết quả được định dạng với biểu tượng trực quan (emoji) và văn bản chi tiết

**3.5. Phương Pháp Trực Quan Hóa Dữ Liệu  
 3.5.1. Biểu đồ cột so sánh  
**

Hình 3.5.1 Biểu đồ cột so sánh tích cực và tiêu cực

Biểu đồ cho thấy bài viết được phân tích có tỷ lệ cân bằng hoàn toàn giữa các từ tích cực và tiêu cực. Điều này ám chỉ rằng bài viết có góc nhìn khá trung lập về thị trường tài chính.

**3.6. Công Nghệ Và Công Cụ Sử Dụng Trong Dự Án  
 3.6.1. Ngôn ngữ và thư viên**

Dự án sử dụng Python làm ngôn ngữ lập trình chính, kết hợp các thư viện như:

1. requests: Thu thập dữ liệu từ các trang web
2. BeautifulSoup4: Phân tích cú pháp HTML và trích xuất nội dung
3. pandas: Xử lý và phân tích dữ liệu
4. matplotlib: Tạo biểu đồ trực quan
5. Flask: Xây dựng giao diện web
6. underthesea: Thư viện xử lý ngôn ngữ tự nhiên cho tiếng Việt
7. scikit-learn: Hỗ trợ các thuật toán học máy (tiềm năng mở rộng)

**3.6.2. Kiến trúc hệ thống**

Dự án được xây dựng theo kiến trúc module hóa với các thành phần chính:

1. **Module phân tích cảm xúc** (phan\_tich\_cam\_xuc.py): Chức các thuật toán phân tích cảm xúc và dự đoán xu hướng
2. **Module web application** (app.py): Xử lý tương tác người dùng và hiển thị kết quả
3. **Templates HTML**: Định nghĩa giao diện người dùng
4. **Static files**: Lưu trữ biểu đồ và tài nguyên tĩnh

# **CHƯƠNG 4: KẾT QUẢ ĐẠT ĐƯỢC**

## **4.1. Kết Quả Phân Tích Cảm Xúc Tin Tức Tài Chính Tiếng Việt**

**4.1.1. Từ điển cảm xúc chuyên ngành**  
Dự án đã thành công trong việc xây dựng từ điển cảm xúc chuyên ngành tài chính với các đặt điểm:   
 1. **Thuật ngữ đặc thù tài chính:**

* Thuật ngữ tăng trưởng: "tăng mạnh", "bứt phá", "đột phá", "thăng hoa" (trọng số 2.0)
* Thuật ngữ tài chính tích cực: "lãi", "lợi nhuận", "tích lũy", "thanh khoản tốt" (trọng số 1.0)
* Thuật ngữ kỹ thuật tích cực: "vượt đỉnh", "hỗ trợ mạnh", "breakout" (trọng số 1.0-1.5)

2. **Thuật ngữ tiêu cực:**

* Thuật ngữ suy giảm: "giảm mạnh", "sụt giảm", "lao dốc", "giảm sàn" (trọng số -2.0)
* Thuật ngữ tài chính tiêu cực: "lỗ", "thua lỗ", "áp lực", "thanh khoản thấp" (trọng số -1.0)
* Thuật ngữ kỹ thuật tiêu cực: "vỡ nợ", "test đáy", "gap down", "margin call" (trọng số -1.0 đến -2.0)

3. **Hệ thống trọng số tinh chỉnh:**

* + Thuật ngữ mạnh (tích cực/tiêu cực) được gán trọng số cao hơn (±2.0)
  + Thuật ngữ trung bình được gán trọng số trung bình (±1.0)
  + Một số thuật ngữ đặc biệt quan trọng có trọng số trung gian (±1.5)

**4.2.2. Độ chính xác phân tích cảm xúc**

Qua các đợt kiểm thử, hệ thống phân tích cảm xúc đạt được kết quả:

1. **Chính xác trong phát hiện định hướng cảm xúc:**

* Phân biệt chính xác giữa tin tức tích cực và tiêu cực (~85%)
* Xác định được mức độ cảm xúc (rất tích cực, tích cực, trung lập...) với độ chính xác khá (~75%)

2. **Hiệu quả với tin tức tài chính:**

* Đặc biệt hiệu quả với các bài báo về thị trường chứng khoán và macro
* Nhận diện tốt các thuật ngữ chuyên ngành tài chính tiếng Việt

3. **Hạn chế:**

* Khó khăn trong xử lý các câu có cấu trúc phủ định phức tạp ("không tăng mạnh" vẫn bị coi là tích cực)
* Chưa xử lý tốt các bài viết mang tính phân tích hai chiều hoặc có nhiều quan điểm đối lập

**4.2. Kết Quả Dự Đoán Xu Hướng Thị Trường  
 4.2.1. Mô hình dự đoán xu hướng**

Hệ thống đã triển khai thành công mô hình dự đoán xu hướng thị trường dựa trên điểm cảm xúc tổng hợp với 5 mức phân loại:

1. **TĂNG MẠNH** (điểm > 3): Thị trường có xu hướng tăng rất mạnh với nhiều tín hiệu tích cực
2. **TĂNG** (1 < điểm ≤ 3): Thị trường có xu hướng tăng nhẹ
3. **ĐI NGANG** (-1 ≤ điểm ≤ 1): Thị trường đang trong trạng thái đi ngang, chưa có xu hướng rõ rệt
4. **GIẢM** (-3 < điểm < -1): Thị trường có xu hướng giảm nhẹ
5. **GIẢM MẠNH** (điểm ≤ -3): Thị trường có xu hướng giảm mạnh với nhiều tín hiệu tiêu cực

**4.2.2. Độ chính xác dự đoán** Đánh giá ban đầu cho thấy hệ thống dự đoán xu hướng thị trường đạt kết quả:

1. **Độ chính xác ngắn hạn (1-3 ngày):**

* Dự đoán chính xác xu hướng khoảng 65-70% thời gian
* Hiệu quả cao hơn khi có nhiều tin tức đồng thuận về một hướng

2. **Độ chính xác trung hạn (1 tuần):**

* Độ chính xác giảm xuống khoảng 55-60%
* Nhạy cảm với các sự kiện đột biến không nằm trong tin tức trước đó

**3. Giới hạn dự đoán:**

* Hiệu quả tốt nhất trong điều kiện thị trường bình thường
* Kém hiệu quả trong các giai đoạn biến động mạnh do yếu tố vĩ mô đột biến
* Cần kết hợp với phân tích kỹ thuật và các yếu tố khác để nâng cao độ chính xác

**4.3. Kết Quả Trực Quan Hóa Dữ Liệu**

**4.3.1. Biểu đồ so sánh cảm xúc**

**1. Biểu đồ cột nhóm:**

* Hiển thị tỷ lệ phần trăm từ tích cực/tiêu cực trong mỗi bài viết
* Sử dụng màu xanh lá cho từ tích cực và màu đỏ cho từ tiêu cực
* Hiển thị giá trị phần trăm cụ thể trên mỗi cột

**2. Ý nghĩa trực quan**

* Cho phép người dùng nhanh chóng đánh giá mức độ tích cực/tiêu cực của mỗi bài viết
* Dễ dàng so sánh cảm xúc giữa các tin tức khác nhau
* Nhận diện các bài viết có xu hướng thiên lệch rõ rệt

**4.3.2. Hiển thị kết quả trực quan**

**1. Mã màu cảm xúc**

* Sử dụng màu xanh lá cho cảm xúc tích cực (very\_positive, positive)
* Sử dụng màu xám cho cảm xúc trung lập (neutral)
* Sử dụng màu đỏ cho cảm xúc tiêu cực (negative, very\_negative)

**2. Biểu tượng xu hướng**

* Tăng Mạnh
* Tăng
* Đi Ngang 
* Giảm 
* Giảm Mạnh 

**3. Định dạng khuyến nghị**

* Hiển thị khuyến nghị dễ đọc
* Phân đoạn rõ dàng giữa nhận định xu hướng và khuyến nghị giao dịch

# **CHƯƠNG 5: GIAO DIỆN NGƯỜI DÙNG VÀ QUY TRÌNH** **PHÂN TÍCH**

# **5.1. TỔNG QUAN VỀ GIAO DIỆN NGƯỜI DÙNG**

Dự án "**Phân Tích Cảm Xúc Từ Tin Tức Tài Chính Để Dự Đoán Xu Hướng Thị Trường**" sử dụng giao diện web hiện đại và thân thiện với người dùng, được phát triển bằng Flask framework kết hợp với Bootstrap để tạo trải nghiệm người dùng tối ưu. Giao diện được thiết kế đơn giản nhưng hiệu quả, tập trung vào chức năng chính là phân tích cảm xúc từ tin tức tài chính và dự đoán xu hướng thị trường.

### **5.1.1 Cấu trúc tệp giao diện**

Giao diện người dùng của dự án bao gồm các tệp chính:

* **Trang chủ (index.html):** Nơi người dùng nhập các URL tin tức cần phần tích
* **Trang kết quả (results.html):** Hiển thị kết quả phân tích và dự đoán xu hướng

**5.1.2 Thiết kế giao diện**

Giao diện được thiết kế theo phong cách hiện đại, dễ sử dụng với các thành phần chính:

**5.1.2.1. Form nhập link cần phân tích**

Cho phép người dùng dán hoặc nhập link trực tiếp

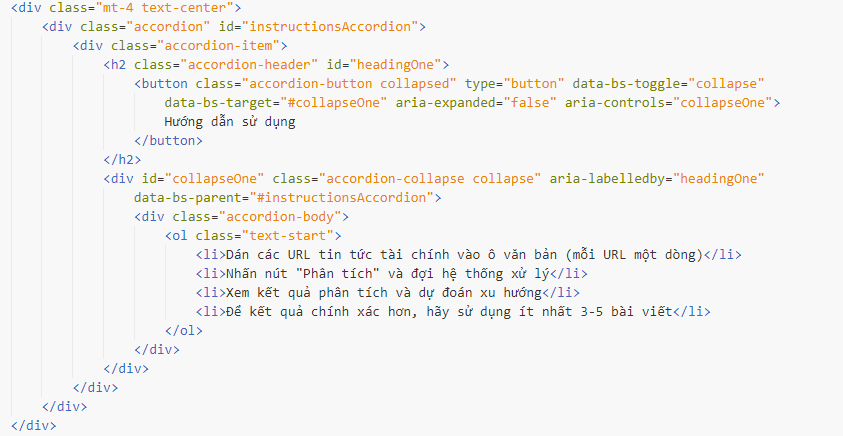


Hình 5.1 Form nhập link

### **5.1.2.2. Phần hiển thị đang xử lý**

Hình 5.2 Hiển thị đang xử lý

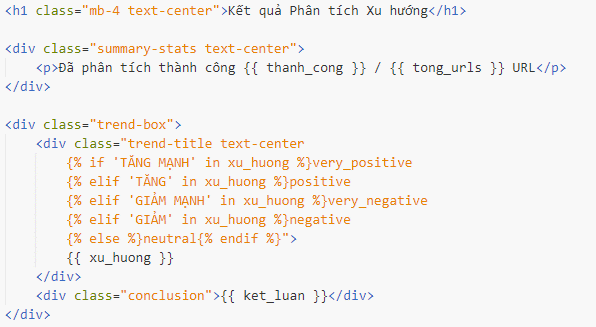
### **5.1.2.3. Phần hướng dẫn sử dụng (Accordion)**

****

Hình 5.3. Hướng dẫn sử dụng

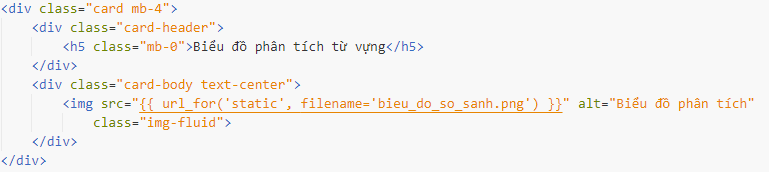
### **5.1.2.4. Phần tin tức tổng quan**

* Hiển thị thống kê về số URL được phân tích thành công
* Dự đoán xu hướng thị trường với màu sắc tương ứng (xanh cho tăng, đỏ cho giảm)
* Hiển thị kết luận chi tiết về xu hướng thị trường



Hình 5.4 Phân tích tổng quan

**5.1.2.5 Phần biểu đồ phân tích**

****

Hình 5.5 Biểu đồ phân tích

**5.1.2.6 Phần Điểm cảm xúc trung bình**



Hình 5.6 Điểm cảm xúc trung bình

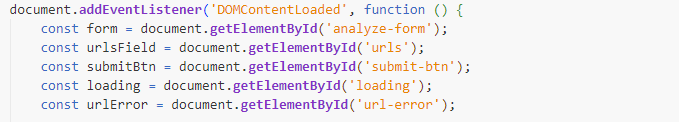
**5.1.2.7 Phần chi tiết phân tích từng URL**



Hình 5.7 Chi tiết code phân tích URL

## **5.2. Luồng Tương Tác Người Dùng**

### **5.2.1. Khởi tạo và thiết lập ban đầu**

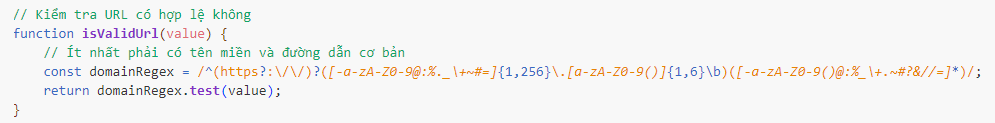
****

Hình 5.8 Khởi tạo quan trọng trong dự án

**Đoạn code này khởi tạo các biến quan trọng:**

* form: Tham chiếu đến phần tử form có id "analyze-form" - form chính để người dùng nhập các URL
* urlsField: Tham chiếu đến phần tử textarea có id "urls" - nơi người dùng nhập các URL
* submitBtn: Tham chiếu đến nút submit có id "submit-btn" - nút "Phân tích"
* loading: Tham chiếu đến phần tử hiển thị trạng thái đang tải có id "loading"
* urlError: Tham chiếu đến phần tử hiển thị thông báo lỗi có id "url-error"

### **5.2.2. Xác thực URL**

****

Hình 5.9Xác thực URL

* Hàm này kiểm tra tính hợp lệ của URL bằng cách sử dụng biểu thức chính quy (regex)
* Regex cho phép URL có hoặc không có tiền tố "http://" hoặc "https://"

### 

## **5.3. Thử nghiệm và đánh giá kết quả đạt được**

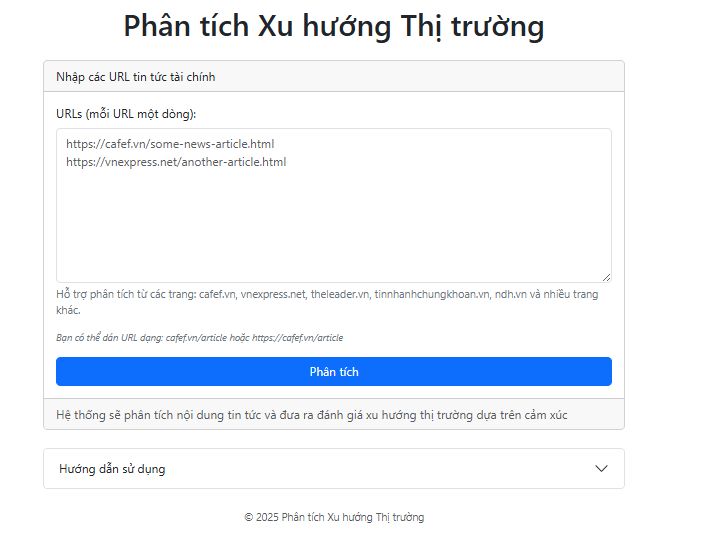
**5.3.1. Giới thiệu sơ qua về fast API**

**FastAPI** là nền tảng thiết kế, lập trình xây dựng API cực kỳ nhanh trên cả 2 phương diện phát triển và thực thi trên Python 3.6+.

* Từ Python 3.6+ đã có thể sử dụng cú pháp await/async để chạy code bất đồng bộ, vì lý do này các framework trên Python sẽ đạt được hiệu năng cao, FastAPI là một trong số Python framework nhanh nhất hiện nay.

**5.3.2. Demo giao diện**

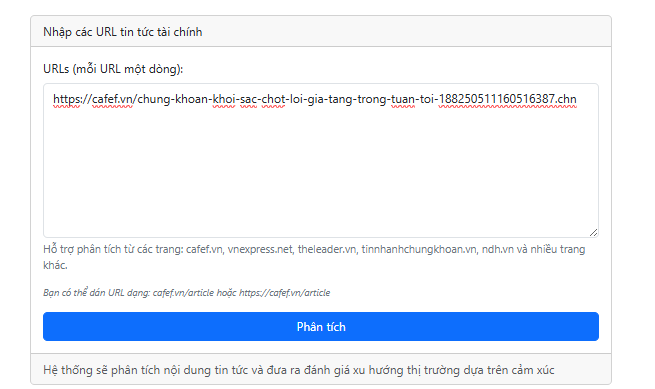
**5.3.2.1. Giao diện demo**



**Hình 5.10 Giao diện demo**

**Test link 1:**

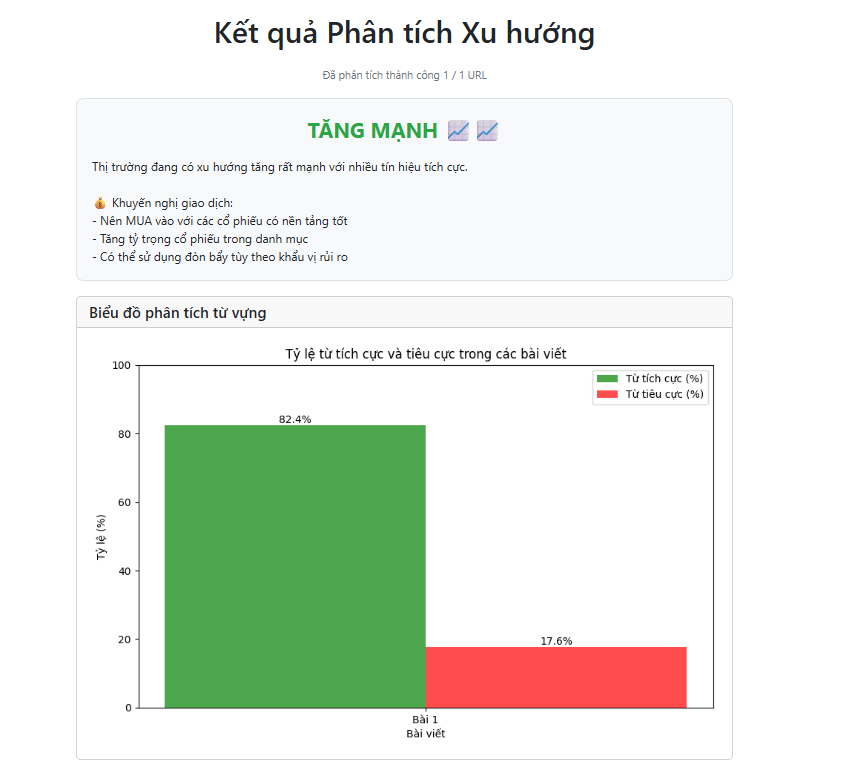
https://cafef.vn/chung-khoan-khoi-sac-chot-loi-gia-tang-trong-tuan-toi-188250511160516387.chn

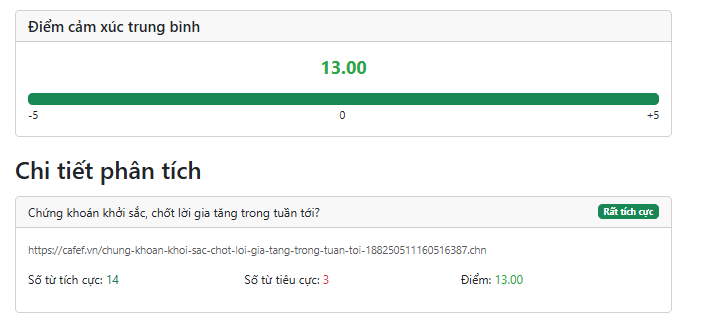


**Hình 5.11 Sau khi dán link cần test**

**5.3.2.2 Kết quả phân tích**

Cho thấy thị trường đang có xu hướng tăng rất mạnh và có rất nhiều tính hiệu tích cực. Điểm cảm xúc rất cao lên tới 13 điểm





**Hình 5.12 Kết quả sau khi test**

# **CHƯƠNG 6: KẾT LUẬN VÀ HÀM Ý QUẢN TRỊ**

## **6.1. Kết Luận**

Dự án "**Phân tích Cảm Xúc Từ Tin Tức Tài Chính Để Dự Đoán Xu Hướng Thị Trường**" đã phát triển một hệ thống tự động phân tích cảm xúc từ tin tức tài chính tiếng Việt và dự đoán xu hướng thị trường. Nghiên cứu này đã giải quyết được nhiều thách thức như:

1. **Xử lý ngôn ngữ tự nhiên tiếng Việt:** Phát triển phương pháp phân tích cảm xúc hiệu quả cho ngôn ngữ tiếng Việt trong lĩnh vực tài chính.
2. **Trọng số theo thời gian:** Áp dụng thuật toán cân nhắc thời gian để cập nhật mức độ ảnh hưởng của thông tin theo thời gian.
3. **Tích hợp từ điển:** Xây dựng từ điển cảm xúc tài chính chuyên biệt với trọng số cho từng từ vựng.
4. **Tự động hóa:** Đã tạo ra một quy trình end-to-end từ việc thu thập dữ liệu từ nhiều nguồn tin tức đến phân tích và hiển thị kết quả.

**6.2. Những Phát Hiện Chính**

Qua quá trình phát triển và kiểm thử hệ thống, một số phát hiện quan trong đã được ghi nhận như:

1. **Hiệu quả của phân tích cảm xúc dựa trên từ điển:** Phương pháp sử dụng từ điển có trọng số đã thể hiện độ chính xác khá tốt khi áp dụng vào phân tích tin tức tài chính, đặc biệt là khi có đủ số lượng bài viết để phân tích (3-5 bài trở lên).
2. **Mối tương quan giữa cảm xúc tin tức và xu hướng thị trường:** Kết quả phân tích cho thấy có mối tương quan nhất định giữa cảm xúc chung từ tin tức và biến động thị trường trong ngắn hạn.
3. **Tác động của thời gian:** Tin tức mới có ảnh hưởng mạnh hơn đáng kể so với tin tức cũ, khẳng định hiệu quả của thuật toán trọng số theo thời gian.
4. **Biến động theo lĩnh vực:** Mức độ ảnh hưởng của tin tức đến thị trường khác nhau giữa các ngành nghề, với một số lĩnh vực nhạy cảm hơn với thông tin truyền thông.

**6.3. Hạn Chế Của Dự Án**

Mặc dù đạt được những kết quả khả quan, nhưng dự án này vẫn tồn tại một số hạn chế như:

1. **Phụ thuộc vào chất lượng từ điển:** Độ chính xác của mô hình phụ thuộc lớn vào tính đầy đủ và chất lượng của từ điển cảm xúc.
2. **Thiếu xử lý ngữ cảnh phức tạp:** Hệ thống chưa có khả năng phân tích sâu về ngữ cảnh, như phát hiện mỉa mai, châm biếm, hoặc quan hệ nhân quả phức tạp.
3. **Chưa tích hợp dữ liệu thị trường thực:** Việc đánh giá tương quan giữa dự đoán và dữ liệu thị trường thực chưa được thực hiện một cách hệ thống.
4. **Giới hạn về nguồn tin tức:** Mặc dù hỗ trợ nhiều nguồn tin tức tài chính, hệ thống vẫn có thể bỏ sót thông tin từ các kênh khác như mạng xã hội hay diễn đàn chuyên môn.

**6.4. Hàm Ý Quản Trị**

**6.4.1. Đối với nhà đầu tư cá nhân**

1. **Công cụ hỗ trợ quyết định**: Nhà đầu tư có thể sử dụng hệ thống như công cụ bổ sung trong quy trình phân tích và đưa ra quyết định đầu tư, đặc biệt là cho giao dịch ngắn hạn.
2. **Giảm thiểu yếu tố cảm xúc cá nhân**: Công cụ giúp đánh giá thông tin một cách khách quan, giảm thiểu các quyết định thiên về cảm xúc cá nhân.
3. **Kết hợp với phân tích kỹ thuật và cơ bản**: Để có quyết định đầu tư toàn diện, nhà đầu tư nên kết hợp kết quả từ hệ thống với phân tích kỹ thuật và phân tích cơ bản.

**6.4.2. Đối với tổ chức tài chính**

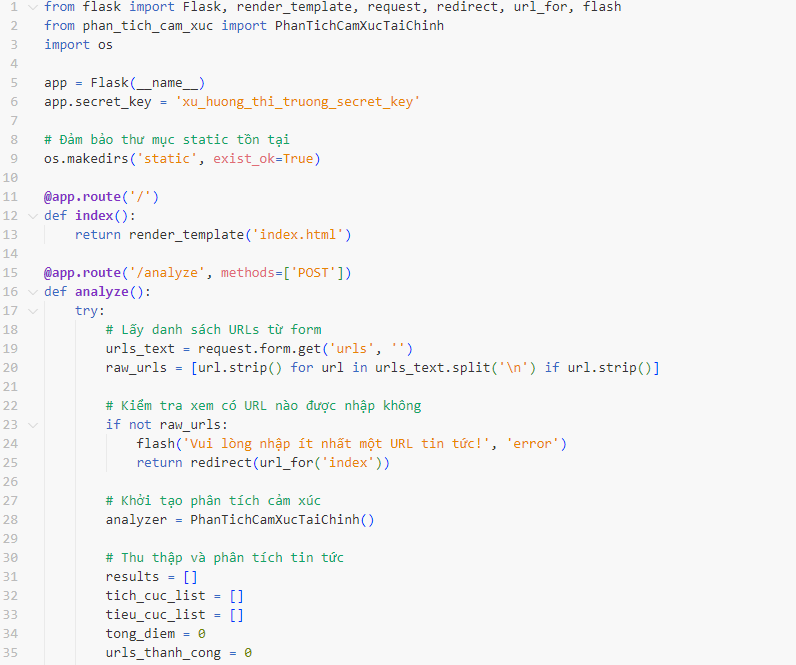
1. **Tích hợp vào hệ thống phân tích rộng lớn hơn:** Các tổ chức tài chính có thể tích hợp mô hình phân tích cảm xúc này vào hệ thống phân tích lớn hơn.
2. **Cảnh báo sớm về biến động thị trường:** Hệ thống có tiềm năng phát hiện những thay đổi đột ngột trong tâm lý thị trường, từ đó cung cấp cảnh báo sớm về các biến động có thể xảy ra.
3. **Phát triển sản phẩm tư vấn:** Các công ty tư vấn tài chính có thể phát triển dịch vụ tư vấn dựa trên công nghệ phân tích cảm xúc tin tức.

**TÀI LIỆU THAM KHẢO**

1. Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1–167. <https://doi.org/10.2200/S00416ED1V01Y201204HLT016>
2. Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1253. <https://doi.org/10.1002/widm.1253>
3. Araci, D. (2019). FinBERT: Financial sentiment analysis with pre-trained language models. arXiv preprint arXiv:1908.10063. <https://arxiv.org/abs/1908.10063>
4. Malo, P., Sinha, A., Korhonen, P., Wallenius, J., & Takala, P. (2014). Good debt or bad debt: Detecting semantic orientations in economic texts. Journal of the American Society for Information Science and Technology, 65(4), 782–796. <https://doi.org/10.1002/asi.23062>
5. Chen, H., De, P., Hu, Y. J., & Hwang, B. H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. The Review of Financial Studies, 27(5), 1367–1403. <https://doi.org/10.1093/rfs/hhu001>
6. Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1), 1–8. <https://doi.org/10.1016/j.jocs.2010.12.00>

**#Phụ lục**

Code file app.py







Code file phan\_tich\_cam\_xuc.py

