

THÔNG TIN CHUNG CỦA BÁO CÁO


- Link YouTube video của báo cáo:

<https://www.youtube.com/watch?v=omCzoGzIefg>

- Link slides (dạng .pdf đặt trên Github):

<https://github.com/tranphuongduy20/CS2205.MAR2024/blob/main/Duy%20Tr%E1%BA%A7n%20Ph%C6%B0%C6%A1ng%20-%20xCS2205.DeCuong.FinalReport.Template.Slide.pdf>

- Mỗi thành viên của nhóm điền thông tin vào một dòng theo mẫu bên dưới
- Sau đó điền vào Đề cương nghiên cứu (tối đa 5 trang), rồi chọn Turn in

<ul style="list-style-type: none">• Họ và Tên: Trần Phương Duy• MSHV: 230201007 	<ul style="list-style-type: none">• Lớp: CS2205.MAR2024• Tự đánh giá (điểm tổng kết môn): 9/10• Số buổi vắng: 0• Link Github: https://github.com/tranphuongduy20/CS2205.MAR2024
--	--

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI (IN HOA)

NGHIÊN CỨU VÀ TRIỂN KHAI HỆ THỐNG DỰ ĐOÁN BIẾN ĐỘNG THỊ TRƯỜNG CHỨNG KHOÁN DỰA TRÊN PHÂN TÍCH CHỈ SỐ CẢM XÚC NHÀ ĐẦU TƯ KẾT HỢP VỚI MÔ HÌNH LSTM

TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

RESEARCH AND IMPLEMENTATION OF A STOCK MARKET VOLATILITY FORECASTING SYSTEM BASED ON ANALYSIS OF INVESTOR EMOTIONAL INDEX COMBINED WITH LSTM MODEL

TÓM TẮT

Đề tài nghiên cứu tập trung vào việc sử dụng phân tích cảm xúc nhà đầu tư và mô hình Long Short-Term Memory (LSTM) để dự đoán biến động thị trường chứng khoán. Thị trường chứng khoán là một môi trường phức tạp và biến động, và cảm xúc của nhà đầu tư có thể ảnh hưởng đáng kể đến các quyết định giao dịch và biến động giá cả. Điều này đã thúc đẩy nhu cầu nghiên cứu về việc sử dụng các chỉ số cảm xúc nhà đầu tư để dự đoán biến động thị trường chứng khoán. Về cảm xúc sử dụng xây dựng mô hình phân tích tâm lý với mô hình BERT còn về dự đoán thị trường sử dụng mô hình LSTM là một mô hình học sâu (deep learning) thuộc lớp mô hình RNN, được sử dụng rộng rãi trong việc xử lý dữ liệu chuỗi thời gian. Đối tượng nghiên cứu của đề tài là phân tích dữ liệu cảm xúc nhà đầu tư từ các nguồn thông tin, chẳng hạn như mạng xã hội, trang tin tức và diễn đàn chứng khoán. Sau đó, dữ liệu cảm xúc sẽ được kết hợp với mô hình LSTM để xây dựng một hệ thống dự đoán biến động thị trường chứng khoán. Quá trình nghiên cứu bao gồm các giai đoạn như thu thập dữ liệu cảm xúc, xử lý và phân tích dữ liệu, xây dựng và huấn luyện mô hình LSTM, và đánh giá hiệu suất của hệ thống dự đoán. Đánh giá hiệu suất sẽ được thực hiện bằng cách so sánh các dự đoán của hệ thống với thực tế biến động thị trường chứng khoán. Nghiên cứu này sẽ cung cấp thông tin quan trọng về khả năng dự đoán biến động thị

trường chứng khoán dựa trên cảm xúc nhà đầu tư và mô hình LSTM. Nếu kết quả nghiên cứu cho thấy tính khả thi và hiệu quả của hệ thống dự đoán, nó có thể được triển khai trong thực tế để hỗ trợ quyết định giao dịch và đầu tư trong lĩnh vực chứng khoán.

GIỚI THIỆU

Thị trường chứng khoán là một môi trường phức tạp và biến động, nơi các nhà đầu tư phải đưa ra quyết định giao dịch trong bối cảnh thông tin liên tục thay đổi và tác động của cảm xúc con người. Việc dự đoán biến động thị trường chứng khoán đã trở thành một nhiệm vụ quan trọng cho các nhà đầu tư và các chuyên gia tài chính.

Đề tài hướng đến việc phát triển một hệ thống tiên tiến có khả năng dự đoán chính xác biến động thị trường chứng khoán. Hệ thống này sử dụng mô hình BERT (Bidirectional Encoder Representations from Transformers) để phân tích và hiểu ngữ cảnh của các văn bản chứa thông tin về cảm xúc nhà đầu tư. BERT, với khả năng xử lý ngôn ngữ tự nhiên mạnh mẽ, cho phép trích xuất các đặc trưng ngữ nghĩa từ dữ liệu văn bản một cách hiệu quả.

Sau khi dữ liệu cảm xúc được phân tích và biểu diễn bằng BERT, mô hình LSTM (Long Short-Term Memory) sẽ được áp dụng để xử lý chuỗi thời gian và dự đoán xu hướng biến động của thị trường chứng khoán. Mô hình LSTM, với khả năng ghi nhớ thông tin trong khoảng thời gian dài và quản lý các mối quan hệ phức tạp trong dữ liệu, sẽ giúp nâng cao độ chính xác của dự đoán.

Sự kết hợp giữa BERT và LSTM trong việc phân tích và dự đoán giúp tăng cường khả năng nhận diện xu hướng và biến động của thị trường.

Input của đề tài nghiên cứu là:

- Dữ liệu cảm xúc nhà đầu tư: Dữ liệu này được thu thập từ các nguồn thông tin như mạng xã hội, trang tin tức và diễn đàn chứng khoán. Nó có thể chứa các thông tin về tâm trạng, ý kiến và cảm nhận của nhà đầu tư đối với thị trường chứng khoán.
- Dữ liệu thị trường chứng khoán: Đây là dữ liệu biến động giá cả, chỉ số và

thông tin thị trường chứng khoán. Nó có thể bao gồm các tham số kỹ thuật và cơ bản liên quan đến các công ty và thị trường chứng khoán.

Output của đề tài nghiên cứu là:

- Hệ thống dự đoán biến động thị trường chứng khoán



Hình 1: Kết quả dự đoán xu hướng thị trường VN-Index

MỤC TIÊU

- Xây dựng hệ thống dự đoán biến động thị trường chứng khoán dựa trên phân tích chỉ số cảm xúc nhà đầu tư bằng mô hình BERT kết hợp với LSTM.
- Đánh giá tính khả thi và hiệu quả của hệ thống dự đoán.
- Triển khai hệ thống để hỗ trợ quyết định giao dịch và đầu tư trên thị trường chứng khoán.

NỘI DUNG

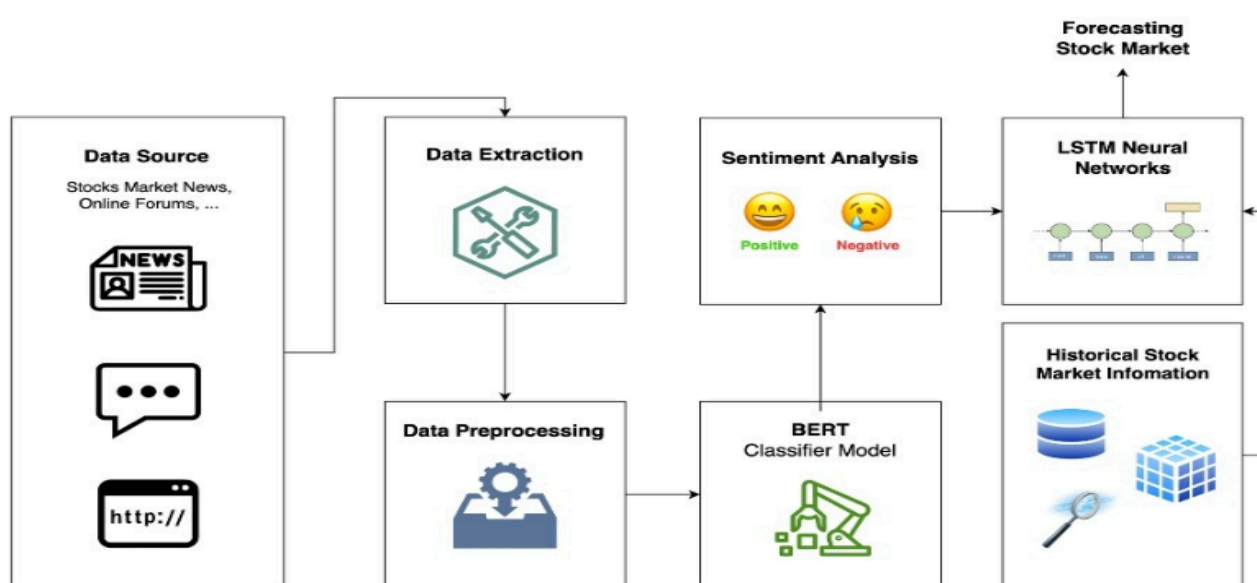
- Nghiên cứu tập trung vào phát triển hệ thống dự đoán biến động thị trường chứng khoán bằng phân tích chỉ số cảm xúc của nhà đầu tư và kết hợp mô hình BERT và LSTM.
- Sử dụng mô hình BERT để phân tích dữ liệu văn bản và trích xuất thông tin về thị trường chứng khoán và cảm xúc nhà đầu tư.
- Sử dụng mô hình LSTM để dự đoán biến động thị trường dựa trên kết quả từ

BERT và dữ liệu thời gian trước đó.

- Nghiên cứu thu thập dữ liệu, tiền xử lý và trích xuất đặc trưng cảm xúc từ BERT, sau đó sử dụng LSTM để dự đoán biến động thị trường.
- Mục tiêu là cung cấp công cụ hỗ trợ quyết định đầu tư thông minh cho nhà đầu tư và chuyên gia tài chính.

PHƯƠNG PHÁP

Quy trình dự báo thị trường chứng khoán (TTCK) được thực hiện theo luồng quy trình sau đây, như được minh họa trong Hình 2:



Hình 2: Phương pháp xây dựng mô hình dự báo TTCK

- Xác định nguồn dữ liệu: Đầu tiên, xác định các nguồn dữ liệu cần thiết để dự báo TTCK. Các nguồn này có thể bao gồm dữ liệu về thị trường chứng khoán, thông tin tài chính, tin tức và bình luận từ nhà đầu tư.
- Quy trình trích xuất dữ liệu: Tiếp theo, thực hiện quy trình trích xuất dữ liệu từ các nguồn đã xác định. Quá trình này có thể bao gồm việc thu thập dữ liệu từ các nguồn trực tuyến, sử dụng công cụ tự động hoặc kết hợp các phương pháp khác nhau để lấy được thông tin cần thiết.
- Tiền xử lý dữ liệu: Sau khi thu thập được dữ liệu, thực hiện bước tiền xử lý để chuẩn bị dữ liệu cho việc phân tích. Quá trình này bao gồm làm sạch dữ liệu,

loại bỏ nhiễu, xử lý văn bản và chuẩn hóa dữ liệu để đảm bảo tính nhất quán và chính xác.

- Xây dựng mô hình phân tích tâm lý với mô hình BERT: Tiếp theo, sử dụng mô hình BERT (Bidirectional Encoder Representations from Transformers) để phân tích tâm lý từ dữ liệu đã được tiền xử lý. Mô hình BERT có khả năng hiểu và trích xuất thông tin từ dữ liệu văn bản liên quan đến TTCK và chỉ số cảm xúc của nhà đầu tư.
- Tích hợp các yếu tố đầu vào: Sau khi có kết quả từ mô hình BERT, tích hợp các yếu tố đầu vào khác như dữ liệu thời gian trước đó, chỉ số TTCK và các biến động khác để tạo ra mô hình dự đoán TTCK hoàn chỉnh.

KẾT QUẢ MONG ĐỢI

- Xây dựng một hệ thống với một phương pháp tiên tiến để dự đoán biến động thị trường chứng khoán dựa trên chỉ số cảm xúc nhà đầu tư và kết hợp mô hình BERT và LSTM.
- Cung cấp công cụ hỗ trợ quyết định đầu tư thông minh cho nhà đầu tư và chuyên gia tài chính.
- Cải thiện khả năng dự đoán và hiệu suất so với các phương pháp truyền thống.
- Đóng góp vào lĩnh vực nghiên cứu và ứng dụng trong việc dự đoán thị trường chứng khoán.

TÀI LIỆU THAM KHẢO

- [1]. Boyacioglu, M. A., and Avcı, D. (2010), An Adaptive Network-Based Fuzzy Inference System (ANFIS) for the prediction of stock market return: The case of the Istanbul Stock Exchange, *Expert Systems with Applications*, 37(12), 7908-7912.
- [1]. Yeh, C.-Y., Huang, C.-W., and Lee, S.-J. (2011), A multiple-kernel support vector regression approach for stock market price forecasting, *Expert Systems with Applications*, 38(3), 2177-2186.
- [1]. Brown, G. W., and Cliff, M. T. (2004), Investor sentiment and the near-term stock market. *Journal of Empirical Finance*, 11(1), 1–27.