ĐẠI HỌC BÁCH KHOA HÀ NỘI

ĐỒ ÁN TỐT NGHIỆP

Hệ thống thu thập, lưu trữ, xử lý và dự báo giá chứng khoán Việt Nam

TRẦN NGUYỄN ANH TUẤN
tuan.tna200565@sis.hust.edu.vn

Ngành Khoa học máy tính

Giảng viên hướng dẫn: TS. Trần Việt Trung _____

Chữ kí GVHD

Khoa: Khoa học máy tính

Trường: Công nghệ Thông tin và Truyền thông

HÀ NỘI, 01/2025

LỜI CAM KẾT

Họ và tên sinh viên: . . . . . .
Điện thoại liên lạc: . . . . . .
Email: . . . . . .
Lớp: . . . . . .
Hệ đào tạo: . . . . . .

Tôi – *Trần Nguyễn Anh Tuấn* – cam kết Đồ án Tốt nghiệp (ĐATN) là công trình nghiên cứu của bản thân tôi dưới sự hướng dẫn của *TS. Trần Việt Trung*. Các kết quả nêu trong ĐATN là trung thực, là thành quả của riêng tôi, không sao chép theo bất kỳ công trình nào khác. Tất cả những tham khảo trong ĐATN – bao gồm hình ảnh, bảng biểu, số liệu, và các câu từ trích dẫn – đều được ghi rõ ràng và đầy đủ nguồn gốc trong danh mục tài liệu tham khảo. Tôi xin hoàn toàn chịu trách nhiệm với dù chỉ một sao chép vi phạm quy chế của nhà trường.

Hà Nội, ngày tháng năm

Tác giả ĐATN

Họ và tên sinh viên

LỜI CẢM ƠN

Quá trình học tập hơn 4 năm tại Đại học Bách Khoa Hà Nội là một quãng thời gian khó quên với em cùng bao thăng trầm trong học tập và cuộc sống. Trong quá trình đó, các thầy cô của các Khoa, Viện và Trường đã tạo tiền đề rèn luyện về mặt kiến thức và kỹ năng để em có khả năng thực hiện đồ án này.

Em xin gửi lời cảm ơn đến TS. Trần Việt Trung, người hướng dẫn em hoàn thành đồ án với những gợi ý, những lời nhận xét và đánh giá quá trình hoàn thiện. Từ đó, em nhận thức được và thúc đẩy bản thân hoàn thiện hết sức với khả năng của mình.

Em xin gửi lời cảm ơn đến anh, chị tại Tổng Công Ty Truyền Thông (VNPT MEDIA) đã tạo điều kiện cho em quan sát, khơi gợi một số ý tưởng nhỏ trong quá trình thực hiện đồ án.

Em xin gửi lời cảm ơn đến bố mẹ, dù bố mẹ không áp lực đến tiến độ nhưng là động lực to lớn để động viên những lúc em muốn từ bỏ.

Quá trình thực hiện đồ án đã giúp em nhận thức nhiều bài học về lập kế hoạch-đánh giá, về phân tích thiết kế, về lập trình, về trình bày-báo cáo để em nắm rõ hơn năng lực bản thân cho định hướng sau này.

Em xin chân thành cảm ơn!

TÓM TẮT NỘI DUNG ĐỒ ÁN

Thị trường tài chính hiện nay được biết là môi trường năng động và phức tạp, nơi việc dự báo giá cổ phiếu đóng vai trò quan trọng trong việc hỗ trợ các quyết định đầu tư hay quản lý rủi ro. Tuy nhiên, việc dự báo giá cổ phiếu gặp nhiều thách thức do các yếu tố như lượng dữ liệu khổng lồ sản sinh theo thời gian từ báo cáo tài chính, giá cổ phiếu lịch sử, và phức tạp hơn là tin tức thị trường. Như vậy, với xu hướng dữ liệu ngày càng lớn, đa dạng và phức tạp cần được khai thác và quản lý hiệu quả với hệ thống chuyên biệt.

Giải pháp đề xuất trong đồ án bao gồm các bước chính như sau. Thu thập dữ liệu được tự động hóa bằng cách sử dụng Selenium [1] và BeautifulSoup [2]. Dữ liệu được lưu trữ trên MongoDB [3] và MinIO [4]. Ngoài ra, việc thu thập dữ liệu mới được kiểm chứng với Evidently [5]. Quá trình xử lý dữ liệu bao gồm làm sạch, chuẩn hóa, và tích hợp dữ liệu; đồng thời phân tích cảm xúc được thực hiện bằng VADER Sentiment để định lượng tâm lý thị trường. Sau đó, mô hình dự báo XGBoost được áp dụng để phân tích và dự báo giá cổ phiếu. Mô hình sẽ được huấn luyện lại nếu phát hiện sai lệch phân phối từ dữ liệu mới. Toàn bộ hệ thống được container hóa với Docker và tự động hóa chức năng dự báo bằng Apache Airflow [6]. Kết quả các quá trình được hiển thị trên dashboard tương tác được xây dựng bằng Streamlit [7].

Đồ án mang lại các đóng góp quan trọng, bao gồm phát triển một luồng hoàn chỉnh từ thu thập, xử lý đến tự động dự báo giá cổ phiếu. Hệ thống tích hợp cảm xúc thị trường vào dữ liệu tài chính để nâng cao độ chính xác của dự báo. Ngoài ra, việc sử dụng các nền tảng lưu trữ và triển khai hiện đại như MinIO và Docker [8] giúp đảm bảo hiệu quả và khả năng mở rộng của hệ thống. Hệ thống cung cấp cách thức để kiểm tra chất lượng dữ liệu và chất lượng mô hình. Cuối cùng, dashboard được phát triển giúp cung cấp một giao diện thân thiện để trực quan hóa kết quả và hỗ trợ ra quyết định.

Sinh viên thực hiện

(Ký và ghi rõ họ tên)

MỤC LỤC

DANH MỤC HÌNH VẼ