

TỐI ƯU HÓA DANH MỤC ĐẦU TƯ VỚI DỰ ĐOÁN LỢI NHUẬN SỬ DỤNG PHƯƠNG PHÁP HỌC MÁY

Phạm Quốc Khang¹, Nguyễn Minh Đức^{1*}

Ngày nhận bài: 28/10/2025

Ngày nhận bản sửa: 12/11/2025

Ngày duyệt đăng: 30/12/2025

Tóm tắt. Trong đầu tư tài chính hiện đại, việc tìm kiếm và xây dựng một công cụ mạnh mẽ và có khả năng hiệu chỉnh nhanh chóng trong tối ưu hóa danh mục đầu tư trở thành nhu cầu cấp thiết để các nhà đầu tư có thể ra quyết định chính xác hơn. Phương pháp học máy ngày càng trở thành ứng dụng phổ biến trong nghiên cứu lĩnh vực tài chính, từ phân tích dự báo đến phức tạp như trong quản lý và tối ưu hóa danh mục đầu tư. Bài viết sử dụng phương pháp Rừng ngẫu nhiên (Random Forest - RF) và Hồi quy Vector hỗ trợ (Support Vector Regression - SVR) để dự báo lợi nhuận cổ phiếu, đồng thời tiến hành so sánh hiệu quả của hai chiến lược tối ưu hóa danh mục đầu tư: MVP (Mean-Variance Portfolio) và MVF (Mean-Variance with Forecasting). Nghiên cứu sử dụng danh mục đầu tư gồm các cổ phiếu trong danh mục 30 cổ phiếu vốn hóa lớn nhất (VN30) của Sở giao dịch chứng khoán Thành phố Hồ Chí Minh. Sau khi sàng lọc các điều kiện, mẫu nghiên cứu còn 22 cổ phiếu. Kết quả nghiên cứu chỉ ra kết hợp của phương pháp SVR và MVF mang lại hiệu quả tốt nhất trong các chiến lược tối ưu hóa danh mục đầu tư.

Từ khóa: Tối ưu hóa danh mục đầu tư; Học máy; VN30; Rừng ngẫu nhiên; Hồi quy Vector hỗ trợ.

1. Giới thiệu

Tối ưu hóa danh mục đầu tư vẫn luôn là chủ đề được quan tâm ngày càng tăng của các nhà nghiên cứu và các nhà đầu tư. Ý tưởng chính của việc tối ưu hóa danh mục đầu tư là kết hợp các tài sản sao cho tạo ra một danh mục hiệu quả nhất đánh đổi giữa lợi nhuận và rủi ro của danh mục đầu tư, đó cũng là nguồn gốc cho lý thuyết của Markowitz (1952). Mô hình trung bình-phương sai (Mean - Variance) của Markowitz là điểm khởi đầu cho các lý thuyết danh mục đầu tư hiện đại về sau (Ma và cộng sự, 2021). Markowitz xây dựng một mô hình tối ưu hóa danh mục đầu tư bằng cách đồng thời tối đa hóa lợi nhuận kỳ vọng của danh mục đầu tư và giảm thiểu rủi ro đầu tư của danh mục đầu tư. Mô hình này hình thành một đường biên hiệu quả, biểu diễn danh mục đầu tư tối ưu giúp giảm tổng rủi ro theo lợi nhuận kỳ vọng được xác định trước (Bodie và cộng sự, 2024).

Các mô hình tối ưu hóa danh mục đầu tư truyền thống thường sử dụng mức trung bình của lợi nhuận cổ phiếu lịch sử làm lợi nhuận kỳ vọng, phù hợp với dự báo dài hạn hơn là ngắn hạn. Một số mô hình dự báo giá cổ phiếu và lợi nhuận truyền thống trong tài

¹Trường Đại học Kinh tế, Đại học Huế, *Tác giả liên hệ: nguyenminhduc@hueuni.edu.vn

chính thường sử dụng như là: mô hình trung bình động tích hợp hồi quy tự động (ARIMA), mô hình độ lệch tuyệt đối trung bình (MAD), mô hình rủi ro giảm (DSR), mô hình giá trị rủi ro tuyến tính (LVAR), mô hình giá trị rủi ro có điều kiện (CVaR). Tuy nhiên, trong ngắn hạn, giá cổ phiếu bị ảnh hưởng nhiều từ tâm lý nhà đầu tư, hành vi của thị trường chứng khoán, cho nên các ước tính về lợi nhuận ngắn hạn trong tương lai sẽ không chính xác (Ma và cộng sự, 2020). Vì thế việc sử dụng các mô hình dự báo lợi nhuận truyền thống không hợp lý khi sử dụng mức trung bình của lợi nhuận lịch sử làm lợi nhuận kỳ vọng ngắn hạn của từng cổ phiếu. Do đó, dự đoán lợi nhuận cổ phiếu nên được kết hợp với các mô hình tối ưu hóa danh mục đầu tư trong đầu tư tài chính.

Về vấn đề này, các nhà nghiên cứu đã cố gắng kết hợp nhiều kết quả dự đoán hơn trong việc hình thành các hàm mục tiêu trong các mô hình tối ưu hóa danh mục đầu tư và cải thiện hơn nữa hiệu suất của các mô hình tối ưu hóa danh mục đầu tư ban đầu (Ma và cộng sự, 2021). Kết quả nghiên cứu của các tác giả cung cấp bằng chứng thực nghiệm cho thấy các mô hình tối ưu hóa danh mục đầu tư tiên tiến này có thể cải thiện hiệu suất của các mô hình tối ưu hóa danh mục đầu tư ban đầu (Yu và cộng sự, 2020).

Các nghiên cứu gần đây chứng minh tính hiệu quả của các kỹ thuật khoa học máy tính trong việc tối ưu hóa danh mục đầu tư như là phương pháp học máy (machine learning), học sâu (deep learning). Các mô hình này có thể cho phép tối ưu hóa bằng việc dự báo lợi nhuận tương lai để làm lợi nhuận kỳ vọng trong việc xây dựng danh mục đầu tư tối ưu (Perrin và Roncalli, 2020). Các phương pháp tiếp cận này đã cho thấy hiệu suất vượt trội so với các phương pháp thông thường, đặc biệt là trong thời kỳ thị trường hỗn loạn, biến động mạnh. Phương pháp học máy cũng cho thấy triển vọng trong việc phân bổ danh mục đầu tư tối ưu, linh hoạt cho việc mở rộng danh mục đầu tư ngoài mẫu, mang lại lợi nhuận tương đương có độ chắc chắn cao (Perrin và Roncalli, 2020; Ma và cộng sự, 2021). Những phát hiện này làm nổi bật tiềm năng của phương pháp học máy trong việc phát triển các chiến lược danh mục đầu tư sáng tạo cho các nhà đầu tư.

Nghiên cứu tối ưu hóa danh mục đầu tư bằng phương pháp học máy là một lĩnh vực quan trọng và đầy triển vọng vì nó cung cấp các công cụ và phương pháp mạnh mẽ để tăng cường hiệu quả đầu tư. Trong nghiên cứu tối ưu hóa danh mục đầu tư, việc lựa chọn phương pháp phân tích và dự báo dữ liệu có vai trò quan trọng nhằm nâng cao độ chính xác và hiệu quả của mô hình. Hai phương pháp phổ biến thường được sử dụng là Rừng ngẫu nhiên (Random Forest - RF) và Hồi quy Vector hỗ trợ (Support Vector Regression - SVR) do những đặc điểm ưu việt của chúng trong việc xử lý dữ liệu tài chính phức tạp. Các thuật toán học máy có thể học từ dữ liệu lịch sử và điều chỉnh danh mục đầu tư dựa trên những dự đoán về xu hướng trong tương lai. Khả năng dự đoán tốt hơn dẫn đến quyết định đầu tư tốt hơn, tăng khả năng sinh lời cho nhà đầu tư. Nghiên cứu này sử dụng phương pháp Mean-Variance Portfolio (MVP), Mean-Variance Forecasting (MVF) nhằm xây dựng danh mục đầu tư tối ưu. MVP do Harry Markowitz đề xuất là nền tảng của lý

thuyết danh mục đầu tư hiện đại. MVP đặc biệt hữu ích vì nó cung cấp một cách tiếp cận định lượng để xây dựng danh mục đầu tư tối ưu dựa trên các nguyên tắc toán học và thống kê, giúp giảm thiểu rủi ro không hệ thống thông qua đa dạng hóa. Bên cạnh đó, phương pháp MVF đóng vai trò quan trọng trong việc dự báo lợi nhuận kỳ vọng và ma trận hiệp phương sai của các tài sản. Do thị trường tài chính luôn biến động, việc dự báo chính xác những tham số này giúp cải thiện hiệu suất danh mục và tránh các sai lệch do dữ liệu lịch sử không phản ánh đúng xu hướng tương lai. Việc sử dụng kết hợp MVP và MVF không chỉ giúp tối ưu hóa danh mục đầu tư theo nguyên tắc rủi ro - lợi nhuận mà còn tăng cường khả năng thích ứng với điều kiện thị trường thay đổi, giúp nhà đầu tư đưa ra quyết định hợp lý hơn trong dài hạn. Bài viết này tập trung nghiên cứu xây dựng danh mục đầu tư gồm các cổ phiếu niêm yết trên Sở giao dịch chứng khoán Thành phố Hồ Chí Minh.

Bài nghiên cứu gồm có 05 phần: phần 2 trình bày cơ sở lý thuyết và tổng quan các nghiên cứu liên quan tối ưu hóa danh mục đầu tư bằng phương pháp học máy, phần 3 trình bày về mô hình và phương pháp nghiên cứu, phần 4 trình bày kết quả thực nghiệm và thảo luận kết quả; phần 5 trình bày kết luận.

2. Tổng quan các nghiên cứu liên quan về tối ưu hóa danh mục đầu tư

Trong phần này, bài viết nêu ra cơ sở lý thuyết về tối ưu hóa danh mục đầu tư và tóm lược các phương pháp học máy trong tối ưu hóa danh mục đầu tư phổ biến đang được áp dụng trong các nghiên cứu gần đây. Các nghiên cứu cung cấp cơ sở lý thuyết và cơ sở thực nghiệm phong phú trong chủ đề này.

Theo Bodie và cộng sự (2024), danh mục đầu tư (portfolio) là tập hợp các tài sản đầu tư. Brealey và cộng sự (2022) nêu khái niệm danh mục đầu tư là sự kết hợp của các tài sản tài chính khác nhau do một nhà đầu tư nắm giữ. Brealey và cộng sự (2022) giải thích rằng danh mục đầu tư có thể bao gồm nhiều loại tài sản tài chính khác nhau, chẳng hạn như cổ phiếu, trái phiếu và các chứng khoán khác. Thành phần của danh mục đầu tư rất quan trọng để xác định các đặc điểm rủi ro và lợi nhuận. Họ cũng nhấn mạnh sự đa dạng hóa như một chiến lược chính trong quản lý danh mục đầu tư để giảm rủi ro trong khi tối đa hóa lợi nhuận tiềm năng.

Tối ưu hóa danh mục đầu tư là một lĩnh vực quan trọng trong tài chính nhằm tìm cách tối đa hóa lợi nhuận đồng thời giảm thiểu rủi ro. Theo Markowitz (1952), đa dạng hóa giúp giảm rủi ro trong danh mục đầu tư bằng cách phân bổ vốn vào nhiều loại tài sản khác nhau. Việc phân bổ tài sản có ảnh hưởng lớn đến kết quả của danh mục đầu tư. Rủi ro của danh mục đầu tư phụ thuộc vào tương quan giữa lợi nhuận của các tài sản trong danh mục, giảm nhiều nhất khi lợi nhuận của hai tài sản bù trừ lẫn nhau một cách đáng tin cậy. Đồng phương sai và hệ số tương quan là các giá trị thống kê giúp xác định xu hướng lợi nhuận của hai tài sản biến thiên cùng nhau hay đối lập nhau.

Những tiến bộ gần đây trong học máy, đặc biệt là phương pháp RF, cho thấy triển vọng đáng kể trong việc nâng cao các kỹ thuật tối ưu hóa danh mục đầu tư. RF là một

phương pháp tổng hợp phi tham số, tận dụng nhiều cây quyết định để cải thiện độ chính xác và độ mạnh mẽ của các dự đoán liên quan đến lợi nhuận và rủi ro của tài sản (Jiménez-Preciado và cộng sự, 2022). Trong bối cảnh tối ưu hóa danh mục đầu tư, các mô hình RF thường được tích hợp với các phương pháp khác để tổng hợp các kết quả tốt hơn. Chẳng hạn, Li và cộng sự (2022) đã phát triển một mô hình Black-Litterman dựa trên RF nhằm mục đích kết hợp quan điểm chủ quan của nhà đầu tư đồng thời nâng cao hiệu suất danh mục đầu tư. Khung phân tích này sử dụng các chiến lược tổng hợp để điều chỉnh lợi nhuận dự kiến được tạo ra thông qua các phương pháp tiếp cận phương sai trung bình truyền thống, cung cấp một chiến lược phân bổ tài sản năng động và phản ứng nhanh hơn. Ma và cộng sự (2021) kết hợp dự đoán lợi nhuận bằng học máy và học sâu vào việc hình thành danh mục đầu tư, từ đó nâng cao hiệu quả của mô hình tối ưu hóa danh mục đầu tư ban đầu. Ma và cộng sự (2021) xác nhận hiệu quả của các phương pháp tổng hợp, cho thấy danh mục đầu tư được thiết kế bằng cách kết hợp RF và các mô hình học máy khác mang lại lợi nhuận cao hơn so với danh mục đầu tư sử dụng các kỹ thuật tối ưu hóa cổ điển. Phương pháp tiếp cận này nhấn mạnh tầm quan trọng của việc huấn luyện dữ liệu tổng hợp trong tài chính, trong đó việc kết hợp các dự đoán từ nhiều mô hình giúp tăng cường tính ổn định và hiệu suất dự đoán. Nghiên cứu có một số hạn chế: nghiên cứu chỉ sử dụng lợi nhuận lịch sử đơn giản làm đặc trưng đầu vào. Tỷ lệ luân chuyển cao là một thách thức lớn đối với các mô hình.

Kaczmarek và Perez (2022) cũng đã khám phá các kỹ thuật học máy, bao gồm RF, để tối ưu hóa các thành phần danh mục đầu tư. Họ đã chứng minh rằng các danh mục đầu tư được xây dựng thông qua các dự đoán học máy có thể hoạt động tốt hơn các mô hình truyền thống, do RF có khả năng nắm bắt các mối quan hệ phi tuyến tính phức tạp giữa các tài sản (Kaczmarek và Perez, 2021). Nghiên cứu của họ đã so sánh các danh mục đầu tư được tối ưu hóa bằng học máy với các danh mục đầu tư có trọng số bằng nhau và phương sai trung bình, làm nổi bật tiềm năng của RF trong việc nắm bắt các sắc thái thị trường mà các mô hình đơn giản hơn có thể bỏ qua. Chen và cộng sự (2021) chỉ ra RF mặc dù không phải lúc nào cũng vượt trội hơn các phương pháp như LSTM về khả năng dự đoán, nhưng nó lại vượt trội về khả năng diễn giải, điều này rất quan trọng đối với việc ra quyết định chiến lược. Hơn nữa, các nghiên cứu đã chỉ ra khả năng thích ứng của các mô hình RF trong nhiều bối cảnh tài chính khác nhau, từ lựa chọn cổ phiếu đến đánh giá rủi ro. Zhang (2024) nghiên cứu tích hợp các dự báo lợi nhuận cổ phiếu hàng ngày từ mô hình học máy; kết quả thực nghiệm cho thấy phương pháp RF mang lại độ chính xác cao hơn Phương pháp tăng cường độ dốc nhẹ (Light Gradient Boosting Machine). Nghiên cứu này nhấn mạnh tính hữu ích thực tế của các kỹ thuật học máy và tiềm năng đáng kể của việc kết hợp các dự báo dựa trên học máy với các kỹ thuật tối ưu hóa danh mục đầu tư. Những phát hiện như vậy cho thấy việc tích hợp các mô hình RF vào quá trình tối ưu hóa danh mục đầu tư mang lại một con đường đầy hứa hẹn để nâng cao các quy trình ra

quyết định tài chính. Bằng cách nắm bắt các mối quan hệ phức tạp và cung cấp các dự đoán có thể diễn giải được, RF có thể cải thiện đáng kể các phương pháp truyền thống, phù hợp với nhu cầu hiện đại về các chiến lược đầu tư năng động và mạnh mẽ hơn.

Mô hình Hồi quy Vector hỗ trợ (SVR) đã nổi lên như một phương pháp mạnh mẽ trong tối ưu hóa danh mục đầu tư, tích hợp các kỹ thuật từ máy học để nâng cao độ chính xác dự đoán và giải quyết các tình huống tài chính phức tạp. Một trong những lợi thế chính của việc sử dụng SVR trong tối ưu hóa danh mục đầu tư là khả năng xử lý hiệu quả các mối quan hệ phi tuyến tính và dữ liệu đa chiều. Nghiên cứu của Henrique và cộng sự (2018) cho thấy việc cập nhật định kỳ mô hình SVR làm giảm lỗi bình phương trung bình so với việc sử dụng mô hình cứng nhắc mà không cập nhật định kỳ. Mô hình đạt được hiệu suất tốt hơn khi sử dụng cửa sổ trượt trên dữ liệu đào tạo. Hsu và cộng sự (2016) khẳng định rằng SVR cũng đóng vai trò quan trọng trong các mô hình dự đoán hiện đại được sử dụng để tối ưu hóa các khoản đầu tư bằng cách cung cấp một khuôn khổ ít có khả năng bị quá mức so với các phương pháp hồi quy truyền thống. Ma và cộng sự (2021) đã so sánh SVR với các kỹ thuật học sâu trong các mô hình tối ưu hóa danh mục đầu tư dựa trên dự đoán của họ, chứng minh rằng SVR có thể nắm bắt hiệu quả các cơ chế cơ bản của hành vi tài sản. Khả năng thích ứng này trong việc tạo ra các mô hình dự đoán khiến SVR trở nên đặc biệt có giá trị trong bối cảnh thị trường tài chính biến động.

Hơn nữa, việc khám phá SVR trong các khuôn khổ đã thiết lập của lý thuyết danh mục đầu tư, chẳng hạn như tối ưu hóa phương sai trung bình của Markowitz, đã chỉ ra sự chuyển dịch sang các phương pháp động và dựa trên dữ liệu. Các phương pháp mới hơn được bổ sung bằng các kỹ thuật học máy như SVR đang chứng minh khả năng giảm thiểu các vấn đề liên quan đến độ không chính xác trong ước tính tham số, thường dẫn đến phân bổ trọng số cực đoan trong danh mục đầu tư (Chen và cộng sự, 2021). Silva và cộng sự (2024) nghiên cứu sử dụng mô hình SVR để hỗ trợ tiên chọn danh mục đầu tư bằng cách sử dụng mô hình Markowitz, tối đa hóa tỷ lệ Sharpe từ danh mục 89 cổ phiếu công ty niêm yết của Brazil. Kết quả thực nghiệm cho thấy mô hình SVR mang lại độ chính xác 61% khi sàng lọc cổ phiếu tiềm năng thông qua chỉ số tài chính cơ bản của các công ty. Danh mục đầu tư được lựa chọn bằng SVR giúp cải thiện hiệu suất đầu tư cao hơn so với các phương pháp truyền thống, đồng thời có khả năng hiệu chỉnh phục hồi nhanh hơn trong giai đoạn thị trường suy giảm. Tóm lại, SVR thể hiện sự cải tiến đáng kể cho các chiến lược tối ưu hóa danh mục đầu tư bằng cách thích ứng với sự phức tạp của dữ liệu tài chính và cung cấp sức mạnh dự đoán mạnh mẽ. Việc tích hợp SVR với các phương pháp phân tích khác không chỉ hợp lý hóa việc lựa chọn và quản lý danh mục đầu tư mà còn cung cấp một phương pháp toàn diện để tối đa hóa lợi nhuận trong khi kiểm soát rủi ro. Điều này phù hợp với sự phát triển hiện đại của các lý thuyết tài chính, ngày càng ưu tiên các kỹ thuật tính toán để bổ sung cho các mô hình truyền thống.

Để tối ưu hóa danh mục đầu tư, hai phương pháp được sử dụng phổ biến trong các

nghiên cứu là MVP và MVF, chẳng hạn, Ma và cộng sự (2021), Ashrafzadeh và cộng sự (2023). Nhiều nghiên cứu đã tập trung vào việc áp dụng học máy và học sâu nhằm cải thiện hiệu quả của danh mục đầu tư. Tuy nhiên, hầu hết các nghiên cứu chỉ sử dụng lợi nhuận lịch sử đơn giản làm đặc trưng đầu vào, bỏ qua khả năng phát triển các mô hình dự đoán lợi nhuận phức tạp hơn. Cụ thể, phương pháp MVP tối ưu hóa danh mục đầu tư mà không tính đến sai số của dự báo, trong khi MVF tích hợp sai số dự báo vào quá trình tối ưu hóa. Nghiên cứu này nhằm kiểm tra độ tin cậy của dự báo lợi nhuận trong tối ưu danh mục đầu tư thông qua việc so sánh hiệu quả của MVF với MVP. Đồng thời, nghiên cứu cũng gợi ý cần có thêm các nghiên cứu bổ sung để đánh giá hiệu quả của các mô hình này trên các thị trường chứng khoán đang phát triển, với Việt Nam là một trường hợp điển hình.

3. Mẫu nghiên cứu và phương pháp nghiên cứu

Nghiên cứu lựa chọn xây dựng danh mục đầu tư gồm các cổ phiếu trong danh mục VN30 của Sở giao dịch chứng khoán Thành phố Hồ Chí Minh. VN30 bao gồm 30 cổ phiếu có vốn hóa lớn, thanh khoản tốt và đại diện cho các ngành kinh tế quan trọng của Việt Nam. Việc lựa chọn cổ phiếu trong danh sách VN30 để làm danh mục đầu tư là một chiến lược phổ biến do tính thanh khoản cao và mức độ ổn định của các doanh nghiệp trong rổ chỉ số này. Khi chọn cổ phiếu trong VN30, nhà đầu tư thường xem xét các yếu tố như tăng trưởng doanh thu, lợi nhuận, khả năng chi trả cổ tức, cũng như xu hướng thị trường và chính sách vĩ mô. Ngoài ra, việc phân bổ danh mục theo ngành nghề và theo dõi biến động giá cả cũng giúp tối ưu hóa lợi nhuận và kiểm soát rủi ro. Bằng cách tập trung vào các cổ phiếu đầu ngành và có nền tảng tài chính vững mạnh, nhà đầu tư có thể xây dựng danh mục hiệu quả, phù hợp với chiến lược dài hạn.

Phần này trình bày phương pháp sử dụng hai mô hình học máy, Hồi quy Vector hỗ trợ SVR và Rừng ngẫu nhiên (RF), để dự đoán lợi nhuận chứng khoán và tích hợp vào mô hình Danh mục đầu tư trung bình - phương sai có dự báo (MVFB). Kết quả này được so sánh với Danh mục đầu tư trung bình - phương sai (MVP). Quy trình thực nghiệm bao gồm thu thập dữ liệu, huấn luyện mô hình, xây dựng danh mục và đánh giá hiệu suất dựa trên các chỉ số tài chính, nhằm tối đa hóa lợi nhuận và giảm thiểu rủi ro đầu tư.

3.1. Mô hình dự đoán lợi nhuận

3.1.1. Hồi quy Vector hỗ trợ SVR

Hồi quy Vector hỗ trợ là một mô hình học máy phổ biến, được áp dụng rộng rãi trong lĩnh vực dự đoán lợi nhuận thị trường chứng khoán. Mô hình này dựa trên nguyên lý tối thiểu hóa rủi ro cấu trúc (Structural Risk Minimization - SRM) của Vapnik, giúp giải quyết các bài toán hồi quy bằng cách cân bằng giữa độ phức tạp của mô hình và sai số thực nghiệm. SVR xuất phát từ lý thuyết học thống kê và có khả năng điều chỉnh mức độ khái quát hóa của mô hình. SVR có khả năng tổng quát hóa tốt với lượng dữ liệu hạn chế, đồng thời có thể xử lý dữ liệu có nhiễu bằng cách sử dụng hàm kernel để chuyển đổi

không gian đặc trưng. Điều này đặc biệt hữu ích trong bối cảnh thị trường tài chính, nơi dữ liệu thường có tính biến động cao và không tuyến tính.

Trong nghiên cứu này, hàm nhân cơ sở xuyên tâm (Radial Basis Function - RBF) được chọn làm hàm nhân chính cho SVR, với công thức:

$$K(x_i, x_j) = \exp\left(-\gamma\|x_i - x_j\|^2\right) \tag{1}$$

Trong đó, γ là tham số của hàm nhân RBF, quyết định mức độ ảnh hưởng của từng điểm dữ liệu trong không gian đặc trưng.

Mô hình SVR được điều chỉnh thông qua các tham số chính, bao gồm hệ số điều chỉnh C và tham số γ . Các giá trị tối ưu của những tham số này được xác định bằng phương pháp tìm kiếm theo lưới nhằm đạt được hiệu suất dự đoán tốt nhất trong quá trình huấn luyện. Bảng 1 trình bày các giá trị được sử dụng cho việc tìm kiếm giá trị tối ưu cho các tham số.

Bảng 1. Các tham số của mô hình SVR

Tham số	Giá trị tìm kiếm
C	1, 2, 4, 8, 16, 25
γ	$2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^0$

Nguồn: Tác giả tổng hợp

3.1.2. Rừng ngẫu nhiên RF

Rừng ngẫu nhiên là một mô hình học máy mạnh mẽ, được ứng dụng rộng rãi trong dự đoán lợi nhuận chứng khoán nhờ khả năng xử lý dữ liệu phi tuyến tính và giảm thiểu hiện tượng quá khớp. RF là một phương pháp học tổ hợp, trong đó nhiều cây quyết định được huấn luyện trên các tập dữ liệu con được lấy mẫu ngẫu nhiên và đưa ra dự đoán tổng hợp bằng cách trung bình hóa kết quả. Nhằm đảm bảo tính ổn định của mô hình và hiệu suất dự đoán tối ưu, nghiên cứu này sử dụng số lượng cây quyết định là 500 theo đề xuất của Breiman (2001). Dự đoán lợi nhuận được tính dựa trên công thức trung bình của các cây quyết định như sau:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N f_i(x) \tag{2}$$

Trong đó, \hat{y} là giá trị dự đoán cuối cùng, N là số lượng cây quyết định (500), $f_i(x)$ là kết quả dự đoán từ cây quyết định thứ i . Các tham số quan trọng được tối ưu trong nghiên cứu này được thể hiện ở Bảng 2. Phương pháp tìm kiếm lưới cũng đã được sử dụng để xác định các giá trị tối ưu cho các tham số trong Bảng 2 nhằm đảm bảo hiệu suất dự đoán cao nhất.

Bảng 2. Các tham số của mô hình Rừng ngẫu nhiên

Tham số	Giá trị
Độ sâu tối đa của mỗi cây	10, 15, 20, 25, 30
Số lượng đặc trưng chọn tại mỗi lần chia	5, 10, 15, 20, 25

Số lượng mẫu tối thiểu để chia nút	5, 10, 15, 20, 25
Số lượng mẫu tối thiểu trong nút lá	10, 20, 30, 40, 50

Nguồn: Tác giả tổng hợp

3.2. Mô hình tối ưu danh mục đầu tư

3.2.1. Danh mục đầu tư trung bình - phương sai MVP

Markowitz (1959) đã giới thiệu mô hình trung bình - phương sai trong xây dựng danh mục đầu tư như một giải pháp để kết hợp các tài sản vào danh mục đầu tư. Theo Markowitz (1959), lợi nhuận và rủi ro của danh mục đầu tư được đo lường bằng lợi nhuận kỳ vọng và phương sai. Mô hình MVP giúp xác định cơ cấu danh mục đầu tư đạt hiệu quả tốt nhất trong việc tối đa hóa lợi nhuận và tối thiểu hóa chi phí với điều kiện ràng buộc về ngân sách nhất định, không có bán khống. Trong nghiên cứu này, mô hình MVP được xây dựng dựa trên dự báo lợi nhuận từ các mô hình học máy đã giới thiệu. Các phương trình sau đây mô tả mô hình MVP:

$$\min \sum_{i,j=1}^n x_i x_j \sigma_{ij} \quad (3)$$

$$\max \sum_{i=1}^n x_i \hat{r}_i \quad (4)$$

Với các ràng buộc: $\sum_{i=1}^n x_i = 1, 0 \leq x_i \leq 1, i = 1, 2, \dots, n$

Trong đó, x_i là tỷ trọng của tài sản i trong danh mục, n là số lượng tài sản trong danh mục, σ_{ij} là hiệp phương sai giữa tài sản i và j , \hat{r}_i là lợi nhuận dự báo của tài sản i . Với phương pháp trọng số bằng nhau, bài toán tối ưu đa mục tiêu có thể được chuyển thành một bài toán tối ưu đơn mục tiêu như sau:

$$\min \left(\sum_{i,j=1}^n x_i x_j \sigma_{ij} - \sum_{i=1}^n x_i \hat{r}_i \right) \quad (5)$$

Với các ràng buộc: $\sum_{i=1}^n x_i = 1, 0 \leq x_i \leq 1, i = 1, 2, \dots, n$

Để giải bài toán tối ưu trên, phương pháp lập trình bậc hai (Quadratic Programming) được sử dụng nhằm tìm ra phân bổ tối ưu của danh mục đầu tư. Sau khi tối ưu hóa, lợi nhuận danh mục đầu tư được tính toán và so sánh với danh mục tham chiếu để đánh giá hiệu suất.

3.2.2. Danh mục đầu tư trung bình - phương sai có dự báo MVF

Dựa trên khung mô hình MVP, nghiên cứu này kết hợp các dự báo lợi nhuận để xây dựng mô hình tối ưu hóa danh mục MVF. MVF là một bài toán tối ưu hóa đa mục tiêu, trong đó không chỉ xem xét lợi nhuận kỳ vọng mà còn xét đến phần lợi nhuận vượt dự báo, được đo bằng giá trị trung bình của chênh lệch giữa lợi nhuận thực tế và lợi nhuận dự báo. Việc bổ sung đại lượng này giúp mô hình ưu tiên những tài sản có mức sinh lợi thực tế cao hơn so với mức dự báo, từ đó cải thiện hiệu quả phân bổ tài sản. Các phương trình sau đây mô tả mô hình MVF:

$$\min \sum_{i,j=1}^n x_i x_j \sigma_{ij} \quad (6)$$

$$\max \sum_{i=1}^n x_i \hat{r}_i \tag{7}$$

$$\max \sum_{i=1}^n x_i \bar{\epsilon}_i \tag{8}$$

Với các ràng buộc: $\sum_{i=1}^n x_i = 1, 0 \leq x_i \leq 1, i = 1, 2, \dots, n$

Trong đó, x_i là tỷ trọng của tài sản i trong danh mục, n là số lượng tài sản trong danh mục, σ_{ij} là hiệp phương sai giữa tài sản i và j , \hat{r}_i là lợi nhuận dự báo của tài sản i , $\bar{\epsilon}_i$ là giá trị trung bình của chênh lệch giữa lợi nhuận thực tế và lợi nhuận dự báo của tài sản i trong khoảng thời gian mẫu.

Bằng cách sử dụng phương pháp trọng số bằng nhau, bài toán tối ưu đa mục tiêu có thể được chuyển thành một bài toán tối ưu đơn mục tiêu như sau:

$$\min \left(\sum_{i,j=1}^n x_i x_j \sigma_{ij} - \sum_{i=1}^n x_i \hat{r}_i - \sum_{i=1}^n x_i \bar{\epsilon}_i \right) \tag{9}$$

Với các ràng buộc: $\sum_{i=1}^n x_i = 1, 0 \leq x_i \leq 1, i = 1, 2, \dots, n$

Mô hình MVF cho phép tối ưu hóa danh mục đầu tư không chỉ dựa trên rủi ro và lợi nhuận mà còn xem xét mức độ không chắc chắn trong dự báo. Phương pháp này giúp cải thiện hiệu suất danh mục và giảm thiểu tác động tiêu cực của sai số dự báo.

Trong nghiên cứu này, chúng tôi đã áp dụng phương pháp trọng số bằng nhau cho các thành phần trong hàm mục tiêu của mô hình MVP và MVF. Cách tiếp cận này có một số ưu điểm. Thứ nhất, phương pháp này giúp giảm thiểu sự chủ quan trong việc lựa chọn trọng số, vì việc gán các hệ số khác nhau có thể làm thay đổi đáng kể kết quả tối ưu và làm tăng tính thiên lệch. Thứ hai, việc sử dụng trọng số bằng nhau đảm bảo rằng cả ba mục tiêu đều được coi là quan trọng ngang nhau, phản ánh sự cân bằng giữa các yếu tố trong bài toán. Thứ ba, đây là một phương pháp tham chiếu phổ biến trong nghiên cứu tối ưu danh mục, thường được sử dụng khi chưa có bằng chứng mạnh mẽ để ưu tiên một mục tiêu hơn hẳn mục tiêu khác. Cuối cùng, cách tiếp cận này giúp đơn giản hóa bài toán, tránh việc phải hiệu chỉnh thêm siêu tham số và hạn chế nguy cơ quá khớp “overfitting” trong bối cảnh dữ liệu tài chính nhiều biến động.

3.3. Quy trình thực nghiệm

Để đánh giá các phương pháp được đề xuất, nghiên cứu này sử dụng lịch sử giao dịch của các cổ phiếu thuộc rổ VN30, bao gồm các cổ phiếu vốn hóa lớn trên thị trường chứng khoán Việt Nam. Tập dữ liệu thực nghiệm là giá đóng cửa của các cổ phiếu hàng ngày trong khoảng thời gian 5 năm, từ năm 2019 đến 2023. Sau khi loại bỏ các cổ phiếu có thời gian niêm yết không đủ 5 năm hoặc có dữ liệu giá bị thiếu hụt, số lượng cổ phiếu còn lại 22, được sử dụng trong các thí nghiệm của nghiên cứu này.

Đối với từng cổ phiếu, lợi nhuận hàng ngày của 60 ngày gần nhất được sử dụng làm đặc trưng đầu vào để dự đoán lợi nhuận của ngày tiếp theo. Dữ liệu đầu vào được chuẩn hóa để thống nhất phạm vi dao động trước khi huấn luyện mô hình, sử dụng phương trình:

$$\frac{d_i - \mu}{\sigma} \quad \hat{d}_i = \quad (10)$$

Với μ và σ lần lượt là giá trị trung bình và độ lệch chuẩn của chuỗi dữ liệu $\{d_i\}$.

Quá trình thí nghiệm được thực hiện theo phương pháp cửa sổ trượt, trong đó 3 năm đầu (2019-2021) được sử dụng làm tập huấn luyện, 1 năm (2022) được sử dụng làm tập xác thực, và 1 năm còn lại (2023) được sử dụng để kiểm tra hiệu suất của các mô hình. Điều này giúp đánh giá khả năng tổng quát hóa của mô hình trên dữ liệu chưa từng thấy trước đó. Trong thí nghiệm này, các mô hình SVR và RF được triển khai bằng thư viện Scikit-learn của Python.

4. Kết quả thực nghiệm

4.1. Dự đoán lợi nhuận cổ phiếu

Chúng tôi đánh giá hiệu suất của các mô hình học máy (SVR và RF) được sử dụng để dự đoán lợi nhuận cổ phiếu. Hiệu suất của các mô hình được đánh giá thông qua các chỉ số sai số phổ biến, cụ thể là Sai số tuyệt đối trung bình (Mean Absolute Error - MAE) và Sai số trung bình bình phương (Mean Squared Error - MSE). Cụ thể, MAE đo lường độ chênh lệch trung bình giữa giá trị thực tế và giá trị dự báo, phản ánh mức độ chính xác của mô hình mà không bị ảnh hưởng bởi hướng dự đoán. Công thức được tính như sau:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (11)$$

Trong đó, y_i là giá trị thực tế của lợi nhuận cổ phiếu, \hat{y}_i là giá trị dự báo của mô hình, n là tổng số mẫu thử nghiệm.

Bên cạnh đó, MSE đo lường độ lệch bình phương trung bình giữa giá trị thực tế và giá trị dự báo, nhấn mạnh vào các sai số lớn và giúp đánh giá hiệu suất tổng thể của mô hình. Công thức tính MSE như sau:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (12)$$

Chỉ số MAE phản ánh sai số trung bình theo đơn vị gốc của dữ liệu, trong khi MSE phản ánh sai số bình phương, giúp làm nổi bật những dự đoán có sai số lớn hơn. Bảng 3 trình bày kết quả dự đoán lợi nhuận của các mô hình SVR và RF được đo lường bằng MAE và MSE trên tập dữ liệu kiểm tra (năm 2023):

Bảng 3. Hiệu suất dự đoán lợi nhuận của các mô hình

Mô hình		MAE	MSE
SVR	Trung bình	$1,40 \cdot 10^{-1}$	$3,93 \cdot 10^{-2}$
	Độ lệch chuẩn	$2,85 \cdot 10^{-2}$	$1,54 \cdot 10^{-2}$
RF	Trung bình	$1,35 \cdot 10^{-1}$	$3,82 \cdot 10^{-2}$
	Độ lệch chuẩn	$3,35 \cdot 10^{-2}$	$1,66 \cdot 10^{-2}$

Nguồn: Tính toán của tác giả

Kết quả cho thấy mô hình RF có giá trị MAE trung bình thấp hơn ($1,35 \cdot 10^{-1}$) so với SVR ($1,40 \cdot 10^{-1}$), điều này cho thấy mô hình RF có độ chính xác cao hơn trong việc dự đoán giá trị trung bình của lợi nhuận cổ phiếu. Tương tự, giá trị MSE của RF ($3,82 \cdot 10^{-2}$) cũng thấp hơn so với SVR ($3,93 \cdot 10^{-2}$), chứng tỏ RF có khả năng dự đoán hiệu quả hơn khi xét đến sai số bình phương, giúp hạn chế các sai số lớn hơn.

Đối với độ lệch chuẩn của sai số, mô hình SVR có giá trị độ lệch chuẩn MAE thấp hơn ($2,85 \cdot 10^{-2}$) so với RF ($3,35 \cdot 10^{-2}$), cho thấy dự đoán của SVR ổn định hơn với ít biến động hơn so với RF. Tương tự, độ lệch chuẩn của MSE trong mô hình SVR ($1,54 \cdot 10^{-2}$) cũng nhỏ hơn so với RF ($1,66 \cdot 10^{-2}$), cho thấy mức độ biến động trong sai số của mô hình SVR thấp hơn.

Kết quả này cho thấy mặc dù mô hình RF đạt được hiệu suất dự đoán tốt hơn về mặt trung bình, nhưng SVR lại có lợi thế về tính ổn định trong dự báo. Do đó, việc lựa chọn mô hình phù hợp phụ thuộc vào mục tiêu cụ thể của nhà đầu tư: nếu ưu tiên độ chính xác, mô hình RF là lựa chọn tối ưu; trong khi nếu cần sự ổn định và ít biến động, SVR là sự lựa chọn phù hợp hơn.

4.2. Hiệu suất của mô hình tối ưu hóa danh mục đầu tư

Ở phần này, chúng tôi trình bày kết quả đánh giá hiệu suất của bốn mô hình tối ưu hóa danh mục đầu tư, bao gồm SVR-MVP, RF-MVP, SVR-MVF, và RF-MVF, dựa trên các chỉ số hiệu suất quan trọng. Các chỉ số được sử dụng để so sánh hiệu suất giữa các mô hình bao gồm: trung bình của lợi nhuận hằng ngày (Mean Daily Return – MDR), độ lệch chuẩn của lợi nhuận hằng ngày (Standard Deviation of Daily Return - SDDR), và lợi nhuận tích lũy (Cumulative Return – CR). Cụ thể, MDR phản ánh mức sinh lời trung bình của danh mục đầu tư mà không sử dụng danh mục tham chiếu, giúp đo lường khả năng tạo ra lợi nhuận theo ngày. SDDR đánh giá mức độ biến động của lợi nhuận so với giá trị trung bình, cho biết mức độ ổn định và rủi ro liên quan đến danh mục đầu tư. Cuối cùng, CR phản ánh tổng số lợi nhuận tích lũy trong suốt giai đoạn đầu tư, cung cấp cái nhìn toàn diện về hiệu suất danh mục theo thời gian. Bảng 4 trình bày hiệu suất của các mô hình tối ưu danh mục đầu tư dựa trên ba chỉ số đánh giá vừa nêu trên.

Bảng 4. Hiệu suất của mô hình tối ưu danh mục đầu tư

Mô hình	MDR (%)	SDDR	CR (%)
SVR-MVP	0,086	0,021	16,161
RF-MVP	0,058	0,020	10,832
SVR-MVF	0,103	0,020	19,353
RF-MVF	0,054	0,021	10,154

Nguồn: Tính toán của tác giả

Về lợi nhuận trung bình hàng ngày (MDR), mô hình SVR-MVF đạt giá trị cao nhất (0,103%), cho thấy khả năng sinh lời tốt nhất trong ngắn hạn so với các mô hình khác. Trong khi đó, mô hình RF-MVF ghi nhận lợi nhuận trung bình thấp nhất (0,054%). Các

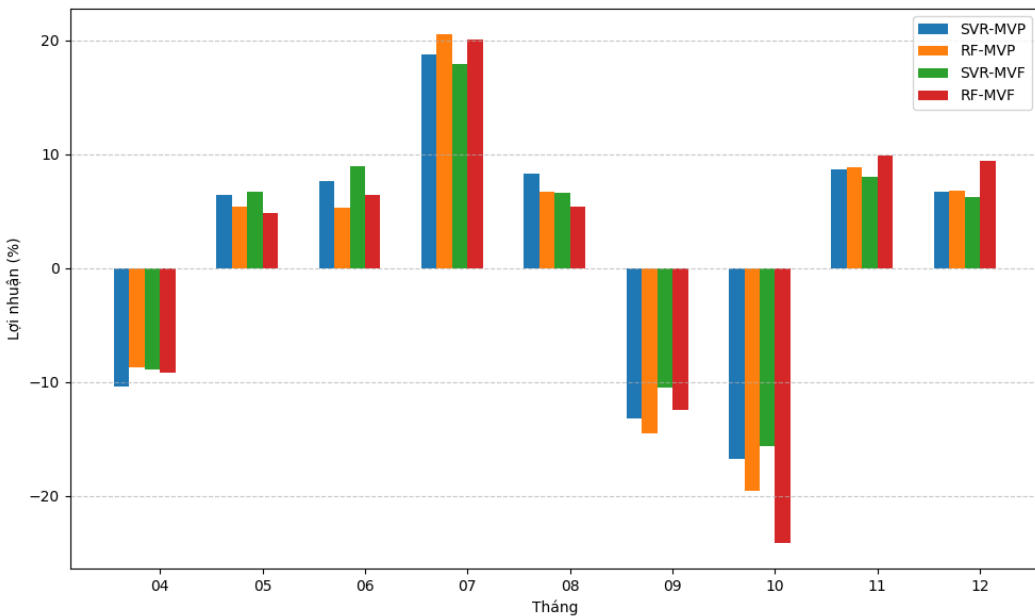
mô hình sử dụng SVR trong cả hai chiến lược (MVP và MVF) đều có lợi nhuận cao hơn các mô hình sử dụng RF, cho thấy ưu thế của SVR trong việc dự báo và tối ưu danh mục đầu tư.

Độ lệch chuẩn của lợi nhuận (SDDR) dao động từ 0,020 đến 0,021 giữa các mô hình, phản ánh mức độ biến động tương đối đồng đều. Các mô hình SVR-MVP và RF-MVF có độ lệch chuẩn cao nhất (0,021), cho thấy mức độ biến động lớn hơn một chút so với các mô hình còn lại. Tuy nhiên, mức chênh lệch này không đáng kể, điều này chỉ ra rằng rủi ro giữa các mô hình là tương đồng và có thể chấp nhận được.

Tổng lợi nhuận tích lũy (CR) tiếp tục khẳng định hiệu suất vượt trội của mô hình SVR-MVF, với giá trị lên tới 19,353%, vượt xa các mô hình còn lại. Trong khi đó, mô hình RF-MVF chỉ đạt 10,154%, là mức thấp nhất trong các mô hình được so sánh. Mô hình SVR-MVP cũng đạt kết quả tích cực với tổng lợi nhuận tích lũy 16,161%, trong khi RF-MVP đạt 10,832%.

Nhìn chung, các mô hình dựa trên SVR cho thấy hiệu suất tốt hơn về cả lợi nhuận trung bình và tổng lợi nhuận tích lũy, đặc biệt là khi sử dụng chiến lược MVF. Mô hình SVR-MVF nổi bật nhất khi cân bằng giữa lợi nhuận và rủi ro, trong khi các mô hình RF có mức lợi nhuận thấp hơn dù không chênh lệch nhiều về mức độ rủi ro. Do đó, mô hình SVR-MVF có thể được xem là lựa chọn ưu tiên cho các nhà đầu tư mong muốn đạt được lợi nhuận cao với rủi ro chấp nhận được.

Ngoài ra, Hình 1 so sánh lợi nhuận theo tháng của bốn mô hình tối ưu danh mục đầu tư (SVR-MVP, RF-MVP, SVR-MVF, RF-MVF) trên tập dữ liệu kiểm tra, trong giai đoạn từ tháng 04/2023 đến 12/2023. Các tháng 01, 02, 03 đã được sử dụng cho quá trình dự đoán nên không có trong biểu đồ.



Hình 1. So sánh lợi nhuận theo tháng giữa các mô hình tối ưu danh mục đầu tư.

Nguồn: Tác giả vẽ dựa trên kết quả tính toán

Cụ thể, trong các tháng mà thị trường sụt giảm mạnh như tháng 9 và tháng 10, mô hình SVR-MVF có mức sụt giảm ít nhất, thể hiện khả năng kiểm soát rủi ro tốt hơn. Ngược lại, các mô hình MVP (SVR-MVP và RF-MVP) chịu mức giảm sâu trong tháng 9, còn RF-MVF sụt giảm mạnh nhất vào tháng 10. Trong tháng tích cực như tháng 7, hai mô hình sử dụng RF dẫn đầu về lợi nhuận, cho thấy sự vượt trội trong điều kiện thị trường thuận lợi. Nhìn chung, SVR-MVF thể hiện khả năng duy trì sự ổn định tốt giữa hiệu suất và rủi ro qua các thời kỳ.

5. Kết luận

Trong nghiên cứu này, chúng tôi đã tích hợp dự báo lợi nhuận từ các mô hình học máy (cụ thể là Hồi quy Vector Hỗ trợ SVR và Rừng Ngẫu nhiên RF) vào quá trình tối ưu hóa danh mục đầu tư qua hai chiến lược MVP và MVF. Thông qua các thí nghiệm trên dữ liệu cổ phiếu VN30 của Sở Giao dịch Chứng khoán Thành phố Hồ Chí Minh, kết quả cho thấy mô hình dự báo của SVR không chỉ cho độ ổn định cao mà còn khi tích hợp sai số dự báo theo chiến lược MVF thì danh mục đầu tư đạt được lợi nhuận trung bình hàng ngày và tổng lợi nhuận tích lũy vượt trội so với các mô hình khác. Điều này khẳng định vai trò quan trọng của việc xem xét sai số dự báo trong việc cải thiện hiệu quả tối ưu hóa danh mục đầu tư, đặc biệt trong bối cảnh thị trường luôn biến động. Tuy nhiên, nghiên cứu vẫn tồn tại hạn chế như việc chỉ sử dụng dữ liệu lợi nhuận lịch sử đơn giản và chưa tích hợp đầy đủ các yếu tố vĩ mô cũng như chi phí giao dịch, từ đó mở ra hướng nghiên cứu tiếp theo nhằm hoàn thiện mô hình dự báo và kiểm chứng hiệu quả trên nhiều thị trường khác.

Lời cảm ơn: Nghiên cứu này được tài trợ bởi Trường Đại học Kinh tế, Đại học Huế, Việt Nam, theo đề tài mã số DHKT2025-02-09. Các tác giả xin chân thành cảm ơn sự hỗ trợ tài chính của Trường đã giúp công trình này được công bố.

TÀI LIỆU THAM KHẢO

- Ashrafzadeh, M., Taheri, H. M., Gharehgozlou, M., & Zolfani, S. H. (2023), 'Clustering-based return prediction model for stock pre-selection in portfolio optimization using PSO-CNN+ MVF', *Journal of King Saud University-Computer and Information Sciences*, 35(9), 101737.
- Bodie, Z., Kane, A., & Marcus, A. J. (2024), *Essentials of investments*, McGraw-Hill.
- Breiman, L. (2001), 'Random Forests', *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Brealey, R. A., Myers, C. S., Allen, F., & Edmus, A. (2022), *Principles of Corporate Finance (14 Ausg.)*, New York: McGraw-Hill Education

- Chen, W., Zhang, H., Mehlawat, M. K., & Jia, L. (2021), 'Mean–variance portfolio optimization using machine learning-based stock price prediction', *Applied Soft Computing*, 100, 106943.
- Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2018), 'Stock price prediction using support vector regression on daily and up to the minute prices', *The Journal of finance and data science*, 4(3), 183-201.
- Hsu, M. W., Lessmann, S., Sung, M. C., Ma, T., & Johnson, J. E. (2016), 'Bridging the divide in financial market forecasting: machine learners vs. financial economists', *Expert systems with Applications*, 61, 215-234.
- Jiménez-Preciado, A. L., Venegas-Martínez, F., & Ramírez-García, A. (2022), 'Stock portfolio optimization with competitive advantages (moat): A machine learning approach', *Mathematics*, 10(23), 4449.
- Kaczmarek, T., & Perez, K. (2021), 'Building portfolios based on machine learning predictions', *Economic research-Ekonomska istraživanja*, 35(1), 19-37.
- Li, C., Chen, Y., Yang, X., Wang, Z., Lu, Z., & Chi, X. (2022), 'Intelligent black–litterman portfolio optimization using a decomposition-based multi-objective direct algorithm', *Applied Sciences*, 12(14), 7089. <https://doi.org/10.3390/app12147089>
- Ma, Y., Han, R., & Wang, W. (2020), 'Prediction-based portfolio optimization models using deep neural networks', *Ieee Access*, 8, 115393-115405. <https://ieeexplore.ieee.org/abstract/document/9121212>
- Ma, Y., Han, R., & Wang, W. (2021), 'Portfolio optimization with return prediction using deep learning and machine learning', *Expert Systems with Applications*, 165, 113973.
- Markowitz, H. (1952), 'Portfolio Selection', *The Journal of Finance*, 7: 77-91. <https://doi.org/10.1111/j.1540-6261.1952.tb01525.x>
- Markowitz, H. M. (1959), 'Efficient Diversification of Investments', *Yale University Press*. <http://www.jstor.org/stable/j.ctt1bh4c8h>
- Perrin, S., & Roncalli, T. (2020), 'Machine learning optimization algorithms & portfolio allocation', *Machine learning for asset management: New developments and financial applications*, 261-328.
- Silva, N. F., de Andrade, L. P., da Silva, W. S., de Melo, M. K., & Tonelli, A. O. (2024), 'Portfolio optimization based on the pre-selection of stocks by the Support Vector Machine model', *Finance Research Letters*, 61, 105014.
- Yu, J. R., Chiou, W. J. P., Lee, W. Y., & Lin, S. J. (2020), 'Portfolio models with return forecasting and transaction costs', *International Review of Economics & Finance*, 66, 118-130.
- Zhang, Y. (2024), 'Integrating Forecasting and Mean-Variance Portfolio Optimization: A Machine Learning Approach', *Advances in Economics, Management and Political Sciences*, 90, 157-168.

PORTFOLIO OPTIMIZATION WITH RETURN PREDICTION USING MACHINE LEARNING

Pham Quoc Khang, Nguyen Minh Duc

Abstract. In modern financial investment, finding and building a powerful and fast-correcting tool in portfolio optimization has become an urgent need for investors to make more accurate decisions. Machine learning methods are becoming increasingly popular in financial research, from predictive analysis to complex ones such as in portfolio management and optimization. This paper uses Random Forest (RF) and Support Vector Regression (SVR) methods to forecast stock returns, and conducted a comparative test of two portfolio optimization strategies: MVP (Mean-Variance Portfolio) and MVF (Mean-Variance with Forecasting). The study used a portfolio of stocks in the 30 largest capitalization stocks (VN30) of the Ho Chi Minh City Stock Exchange. Following the application of the screening conditions, the research sample comprises 22 stocks. The research results showed that the combination of the SVR and MVF methods brought the best efficiency in portfolio optimization strategies.

Keywords: Portfolio optimization; Machine learning; VN30; Random Forest; Support Vector Regression.