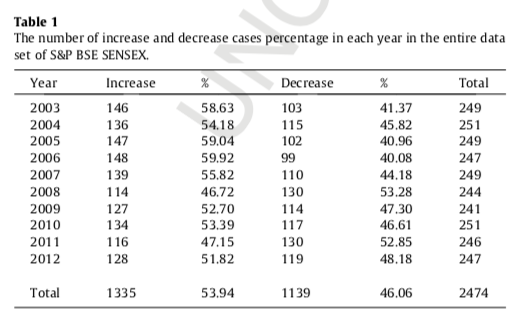
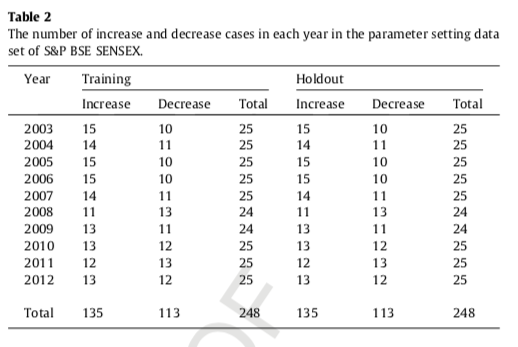
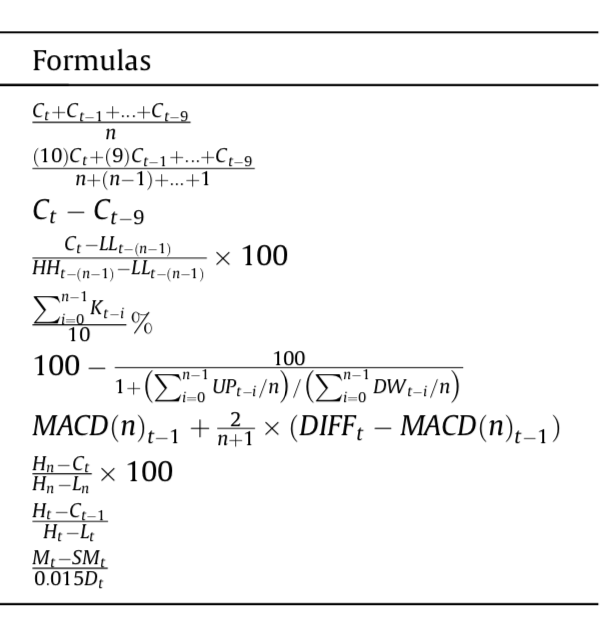
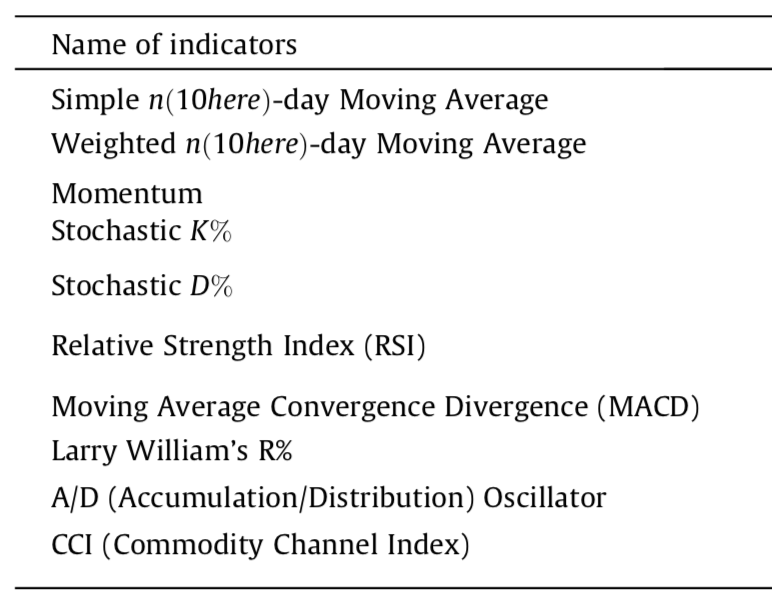
Dự đoán xu hướng của chỉ số chứng khoán và giá cổ phiếu cho thị trường chứng khoán Ấn Độ.

* Abstract
* Ở Indian.
* Các mô hình được sử dụng là:
  + ANN
  + SVM
  + Random forest
  + Naïve – Bayes
* Có 2 cách tiếp cận đầu vào các mô hình trên
  + Dữ liệu đầu vào liên quan đến việc tính toán mười thông số kỹ thuật sử dụng dữ liệu giao dịch chứng khoán (giá mở, cao, thấp và đóng).
  + Trong khi cách tiếp cận thứ hai tập trung vào việc biểu diễn các tham số kỹ thuật này dưới dạng dữ liệu xác định xu hướng.
* Việc đánh giá được thực hiện trên 10 năm dữ liệu lịch sử từ 2003 đến 2012 của
  + Hai cổ phiếu là Reliance Industries và Infosys Ltd.
  + Hai chỉ số giá cổ phiếu CNX Nifty và S & P Bombay (BSE) Sensex.
* Các kết quả thử nghiệm cho thấy đối với cách tiếp cận đầu tiên của dữ liệu đầu vào trong đó mười tham số kỹ thuật được biểu diễn dưới dạng các giá trị liên tục, giải thuật Random forest đã có hiệu suất tổng thể cao hơn 3 giải thuật còn lại.
* Kết quả thử nghiệm cũng cho thấy hiệu suất của tất cả các mô hình dự đoán được cải thiện khi các tham số kỹ thuật này được biểu thị dưới dạng dữ liệu xác định xu hướng.
* Introduction
* Dự đoán chỉ số chứng khoán và giá cổ phiếu là khó khăn.
* Có hai loại phân tích mà các nhà đầu tư thực hiện trước khi đầu tư vào một cổ phiếu.
  + Đầu tiên là phân tích cơ bản.
    - Trong đó, các nhà đầu tư nhìn vào:
      * giá trị thực sự của cổ phiếu
      * hiệu suất của ngành công nghiệp và nền kinh tế
      * môi trường chính trị
      * vv…
    - Kết luận có đầu tư hay không
  + Thứ hai là phân tích kỹ thuật là đánh giá cổ phiếu bằng phương pháp nghiên cứu thống kê:
    - Được tạo ra bởi hoạt động thị trường, chẳng hạn như:
      * giá cả và khối lượng trong quá khứ.
    - Các nhà phân tích kỹ thuật không cố gắng để đo lường giá trị an toàn thực tế nhưng thay vì sử dụng các biểu đồ chứng khoán để xác định các mẫu và xu hướng có thể đề xuất cách một cổ phiếu sẽ cư xử trong tương lai.
    - Giả thuyết thị trường hiệu quả của Malkiel và Fama (1970) nói rằng giá cổ phiếu có hiệu quả về mặt thông tin, điều đó có nghĩa là có thể dự đoán giá cổ phiếu dựa trên dữ liệu giao dịch.
    - Điều này khá logic khi nhiều yếu tố không chắc chắn như kịch bản chính trị của đất nước, hình ảnh công khai của công ty sẽ bắt đầu phản ánh trong giá cổ phiếu.
    - Vì vậy, nếu thông tin thu được từ giá cổ phiếu được xử lý trước một cách hiệu quả và áp dụng các thuật toán thích hợp thì xu hướng của chỉ số giá cổ phiếu hoặc giá cổ phiếu có thể được dự đoán.
* Từ nhiều năm, nhiều kỹ thuật đã được phát triển để dự đoán xu hướng chứng khoán.
  + Ban đầu phương pháp hồi quy cổ điển được sử dụng để dự đoán xu hướng chứng khoán.
  + Vì dữ liệu chứng khoán có thể được phân loại thành dữ liệu chuỗi thời gian không cố định, các kỹ thuật học máy phi tuyến tính cũng đã được sử dụng.
  + Mạng nơ ron nhân tạo (ANN) và Máy vectơ hỗ trợ (SVM) là hai thuật toán học máy được sử dụng rộng rãi nhất để dự đoán biến động chỉ số chứng khoán và giá cổ phiếu.
  + Mỗi thuật toán có cách riêng để tìm hiểu các mẫu. ANN mô phỏng chức năng của bộ não của chúng ta để học bằng cách tạo ra mạng lưới các nơ-ron. Hassan, Nath, và Kirley (2007) đề xuất và thực hiện một mô hình phản ứng tổng hợp bằng cách kết hợp các Hidden Markov Model (HMM), Artificial Neural Networks (ANN) và thuật toán di truyền (GA) để dự báo hành vi thị trường tài chính.
  + Sử dụng ANN, giá cổ phiếu hàng ngày được chuyển đổi thành các bộ giá trị độc lập trở thành đầu vào cho HMM. Wang và Leu (1996) đã phát triển một hệ thống dự đoán hữu ích trong việc dự báo xu hướng giá giữa kỳ tại thị trường chứng khoán Đài Loan.
  + Hệ thống của họ dựa trên một mạng lưới thần kinh tái phát được đào tạo bằng cách sử dụng các tính năng được trích xuất từ ​​các phân tích ARIMA. Kết quả thực nghiệm cho thấy các mạng được đào tạo sử dụng dữ liệu hàng tuần trong vòng 4 năm có khả năng dự đoán xu hướng thị trường lên đến 6 tuần với độ chính xác chấp nhận được.
  + Các kỹ thuật điện toán mềm lai hóa để dự báo thị trường chứng khoán tự động và phân tích xu hướng đã được giới thiệu bởi Abraham, Nath và Mahanti (2001). Họ đã sử dụng chỉ số Nas-daq của thị trường chứng khoán Nasdaq với mạng lưới thần kinh (Neural network) cho một ngày trước dự báo chứng khoán và một hệ thống thần kinh (neuro-fuzzty )để phân tích xu hướng của các giá trị chứng khoán dự đoán. Các kết quả dự báo và dự đoán xu hướng sử dụng hệ thống lai được đề xuất là đầy hứa hẹn.
  + Chen, Leung và Daouk (2003) đã điều tra mạng lưới thần kinh song phương (PNN) để dự báo hướng của chỉ số sau khi nó được đào tạo bởi dữ liệu lịch sử. Kết quả thực nghiệm cho thấy các chiến lược đầu tư dựa trên PNN thu được lợi nhuận cao hơn các chiến lược đầu tư khác đã được nghiên cứu như chiến lược mua và nắm giữ cũng như các chiến lược đầu tư được hướng dẫn bởi các dự báo được ước tính bởi mô hình đi bộ ngẫu nhiên (random walk) và mô hình GMM tham số.
  + Một thuật toán SVM rất nổi tiếng được phát triển bởi Vapnik (1999) tìm kiếm một siêu mặt phẳng ở không gian nhiều chiều để phân lớp. (SVM) là một loại thuật toán học tập rất đặc trưng được đặc, việc sử dụng các kernel function và scarcity solution. Huang, Nakamori và Wang (2005) đã điều tra tính khả thi của hướng di chuyển tài chính với SVM bằng cách dự báo hướng di chuyển hàng tuần của chỉ số NIKKEI 225.
  + Họ đã so sánh SVM với Linear Discriminant Analysis, Quadratic Discriminant Analysis and Elman Backpropagation Neural Networks. Kết quả thí nghiệm cho thấy SVM vượt trội so với các phương pháp phân loại khác. SVM đã được Kim (2003) sử dụng để dự đoán hướng thay đổi giá cổ phiếu hàng ngày trong chỉ số giá cổ phiếu tổng hợp Hàn Quốc (KOSPI). Mười hai chỉ số kỹ thuật đã được chọn để tạo nên các thuộc tính ban đầu. Nghiên cứu này đã so sánh SVM với Elman Backpropagation Neural Networks (BPN) và lý luận dựa trên trường hợp (CBR). Rõ ràng từ các kết quả thử nghiệm cho thấy SVM vượt trội so với BPN và CBR.
  + Random forest tạo ra n lớp phân loại bằng cách sử dụng mẫu với sự thay thế và dự đoán lớp dựa trên phần lớn các cây dự đoán, một lớp đã được huấn luyện nên sẽ là đại diện cho chính lớp đó. Tuy nhiên, giả thuyết này không nhất thiết được chứa trong không gian giả thuyết của các mô hình mà nó được xây dựng.
  + Do đó, các lớp có thể được chứng minh là có tính linh hoạt hơn trong các chức năng mà chúng có thể đại diện. Về mặt lý thuyết, tính linh hoạt này có thể cho phép họ phù hợp với dữ liệu đào tạo hơn một mô hình đơn lẻ, nhưng trên thực tế, một số kỹ thuật tập hợp (đặc biệt là bagging) có xu hướng giảm các vấn đề liên quan đến việc khớp quá mức dữ liệu đào tạo. Tsai, Lin, Yen và Chen (2011) đã điều tra hiệu suất dự đoán sử dụng phương pháp phân loại tập hợp để phân tích lợi nhuận chứng khoán. Các phương pháp lai của bỏ phiếu đa số(voting) và bagging đã được xem xét.
  + Hơn nữa, hiệu suất sử dụng hai loại phân loại phân loại được so sánh với các loại sử dụng phân loại đường cơ sở duy nhất (nghĩa là mạng thần kinh, cây quyết định và hồi quy logistic). Kết quả chỉ ra rằng nhiều phân loại vượt trội hơn các phân loại đơn về độ chính xác dự đoán và lợi tức đầu tư. Sun và Li (2012) đã đề xuất phương pháp dự tình trạng khó khăn tài chính (FDP) mới dựa trên tập hợp SVM. Thuật toán để chọn các bộ phân loại cơ sở của SVM từ các ứng cử viên được thiết kế bằng cách xem xét cả phân tích hiệu suất và phân tích riêng lẻ. Kết quả thí nghiệm chỉ ra rằng nhóm SVM vượt trội hơn hẳn so với phân loại SVM riêng lẻ. Ou và Wang (2009) đã sử dụng tổng số mười kỹ thuật khai thác dữ liệu để dự đoán biến động giá của chỉ số Hang Seng của thị trường chứng khoán Hồng Kông. Các phương pháp tiếp cận bao gồm Linear discriminant analysis (LDA), Quadratic discriminant analysis (QDA), K-nearest neighbor classification, Naive Bayes based on kernel estimation, Logit model, Tree based classification, neural network, Bayesian classification với thực hiện Gaussian, ( SVM) và (LS-SVM). Kết quả thử nghiệm cho thấy SVM và LS-SVM tạo ra hiệu suất dự đoán vượt trội trong số các mô hình khác.
  + Rõ ràng từ các cuộc thảo luận ở trên rằng mỗi thuật toán theo cách riêng của nó có thể giải quyết vấn đề này. Cũng cần lưu ý rằng mỗi thuật toán có những hạn chế riêng.
  + Kết quả dự đoán cuối cùng không chỉ phụ thuộc vào thuật toán dự đoán được sử dụng mà còn bị ảnh hưởng bởi sự dữ liệu của đầu vào. Xác định các độ quan trọng của feature và chỉ sử dụng chúng làm đầu vào thay vì tất cả các tính năng có thể cải thiện độ chính xác dự đoán của các mô hình dự đoán.
  + Một kiến ​​trúc hai giai đoạn được phát triển bởi Hsu, Hsieh, Chih và Hsu (2009). Họ tích hợp bản đồ tự tổ chức và hỗ trợ hồi quy véc tơ để dự đoán giá cổ phiếu. Họ đã kiểm tra bảy chỉ số thị trường chứng khoán lớn. Cụ thể, bản đồ tự tổ chức (SOM) lần đầu tiên được sử dụng để phân tách toàn bộ không gian đầu vào thành các khu vực nơi các điểm dữ liệu có phân phối thống kê tương tự được nhóm lại với nhau, để chứa và nắm bắt thuộc tính không cố định của chuỗi tài chính. Sau khi phân tách các điểm dữ liệu không đồng nhất thành một số vùng đồng nhất, hồi quy vectơ hỗ trợ (SVR) đã được áp dụng để dự báo các chỉ số tài chính. Kết quả cho thấy kiến ​​trúc hai giai đoạn cung cấp một sự thay thế đầy hứa hẹn cho dự đoán giá cổ phiếu. Lập trình di truyền (GP) và các biến thể của nó đã được áp dụng rộng rãi để mô hình hóa thị trường chứng khoán để cải thiện khả năng khái quát hóa của mô hình, GP đã được lai với các biến thể của riêng nó (lập trình biểu hiện gen (GEP), lập trình đa biểu thức (MEP)) hoặc với các phương pháp khác như mạng thần kinh và tăng cường. Khả năng khái quát hóa của mô hình GP cũng có thể được cải thiện bằng một lựa chọn phù hợp về tiêu chí lựa chọn mô hình.
  + Garg, Sriram và Tai (2013) đã làm việc để phân tích ảnh hưởng của ba tiêu chí lựa chọn mô hình qua hai biến đổi dữ liệu đối với hiệu suất của GP trong khi mô hình hóa cổ phiếu được lập chỉ mục tại Sở giao dịch chứng khoán New York (NYSE). Người ta thấy rằng các tiêu chí FPE đã cho thấy sự phù hợp tốt hơn cho mô hình GP trên cả hai phép biến đổi dữ liệu so với các tiêu chí lựa chọn mô hình khác.
  + Nair và cộng sự. (2011) dự đoán ngày hôm sau giá trị đóng cửa của năm chỉ số chứng khoán quốc tế sử dụng hệ thống dựa trên Mạng thần kinh nhân tạo thích ứng (ANN). Hệ thống tự thích ứng với sự thay đổi của thị trường với sự trợ giúp của thuật toán di truyền điều chỉnh các tham số của mạng lưới thần kinh vào cuối mỗi phiên giao dịch.
  + Nghiên cứu của Ahmed (2008) đã nghiên cứu bản chất của mối quan hệ nhân quả giữa giá cổ phiếu và các biến số kinh tế vĩ mô quan trọng đại diện cho lĩnh vực tài chính và thực tế của nền kinh tế Ấn Độ trong giai đoạn tháng 3 năm 1995 sử dụng dữ liệu. Nghiên cứu cho thấy sự chuyển động của giá cổ phiếu không chỉ là kết quả của hành vi của các biến số kinh tế vĩ mô quan trọng mà nó còn là một trong những nguyên nhân của sự dịch chuyển trong khía cạnh vĩ mô khác trong nền kinh tế. Mantri, Gahan và Nayak (2010) đã tính toán các biến động của thị trường chứng khoán Ấn Độ bằng cách sử dụng các mô hình GARCH, EGARCH, GJR-GARCH, IGARCH & ANN. Nghiên cứu này đã sử dụng mười bốn năm dữ liệu của BSE Sen-sex & NSE Nifty để tính toán các biến động.
  + Nó đã được kết luận rằng không có sự khác biệt trong các biến động của Sensex, & Nifty được ước tính theo các mô hình GARCH, EGARCH, GJR GARCH, IGARCH & ANN. Mishra, Sehgal và Bhanumurthy (2011) đã kiểm tra sự hiện diện của sự phụ thuộc phi tuyến tính và sự hỗn loạn xác định trong tỷ lệ lợi nhuận của sáu chỉ số thị trường chứng khoán Ấn Độ. Kết quả phân tích cho thấy chuỗi trả về không tuân theo quy trình walking random.
  + Thay vào đó, có vẻ như sự gia tăng hàng ngày trong lợi nhuận chứng khoán có mối tương quan chặt chẽ và số mũ Hurst ước tính là dấu hiệu cho thấy sự tồn tại cận biên trong lợi nhuận vốn chủ sở hữu. Liu và Wang (2012) đã điều tra và dự báo biến động giá của một mạng lưới thần kinh Legendre được cải thiện bằng cách giả định rằng các nhà đầu tư quyết định vị trí đầu tư của họ bằng cách phân tích dữ liệu lịch sử trên thị trường chứng khoán.
  + Họ cũng giới thiệu một hàm cường độ thời gian ngẫu nhiên trong mô hình dự báo. Phương pháp Dự báo tuyến tính xếp hạng hình thái (EMRLF) được đề xuất bởi Araújo và Ferreira (2013). Một phân tích thử nghiệm đã được thực hiện và kết quả được so sánh với các mạng đa lớp Perceptron (MLP) và phương pháp Dự báo tiến hóa được thêm thời gian trễ (TAEF).
  + => Nghiên cứu này tập trung vào việc so sánh hiệu suất dự đoán của các thuật toán ANN, SVM, random forest và thuật toán Naive-Bayes cho nhiệm vụ dự đoán biến động cổ phiếu và giá cổ phiếu. Mười thông số kỹ thuật được sử dụng làm đầu vào cho các mô hình này.
  + Bài viết này cung cấp đầu vào chuẩn bị dữ liệu xác định xu hướng chuyển đổi các đầu vào có giá trị liên tục thành các đầu vào riêng biệt.
  + Mỗi tham số đầu vào ở dạng riêng biệt cho thấy xu hướng tăng hoặc giảm có thể được xác định dựa trên đặc tính vốn có của nó. Trọng tâm cũng là để so sánh hiệu suất của các mô hình dự đoán này khi các yếu tố đầu vào được biểu diễn dưới dạng giá trị thực và dữ liệu xác định xu hướng.
  + Tất cả các thí nghiệm được thực hiện bằng 10 năm dữ liệu lịch sử của hai cổ phiếu Reliance Industries và Infosys Ltd. và hai chỉ số S & P BSE Sensex và CNX Nifty. Cả hai cổ phiếu và chỉ số đều rất mạnh và được giao dịch mạnh mẽ và vì vậy chúng phản ánh toàn bộ nền kinh tế Ấn Độ.
  + Phần còn lại của bài viết này được tổ chức thành các phần sau. Phần 2 mô tả dữ liệu nghiên cứu, xử lý trước dữ liệu và tính toán các thông số tài chính đóng vai trò đầu vào. Nó cũng thảo luận về việc chuẩn bị dữ liệu xác định xu hướng. Bốn mô hình dự đoán được sử dụng trong nghiên cứu này sẽ được thảo luận trong Phần 3. Phần 4 cho thấy kết quả thử nghiệm. Thảo luận về kết quả đạt được được báo cáo trong Phần 5. Mục 6 kết luận nghiên cứu.
* Dữ liệu nghiên cứu
  + Mười năm dữ liệu của tổng hai chỉ số giá cổ phiếu (CNX Nifty, S & P BSE Sensex) và hai cổ phiếu (Reliance Industries, Infosys Ltd.) từ tháng 1 năm 2003 đến tháng 12 năm 2012 được sử dụng trong nghiên cứu này. Tất cả dữ liệu được lấy từ các trang web <http://www.nseindia.com/> và <http: //www.bseindi- a.com/>. Những dữ liệu này tạo thành toàn bộ tập dữ liệu của chúng tôi. Tỷ lệ phần trăm tăng và giảm khôn ngoan của mỗi năm trong toàn bộ tập dữ liệu được thể hiện trong Bảng 1.
  + 
  + Nghiên cứu này sử dụng 20% toàn bộ dữ liệu làm dữ liệu lựa chọn tham số. Dữ liệu này được sử dụng để xác định các tham số thiết kế của các mô hình dự đoán. Tập dữ liệu lựa chọn tham số được xây dựng bằng cách lấy tỷ lệ dữ liệu bằng nhau từ mỗi mười năm. Tỷ lệ phần trăm tăng giảm trong mỗi năm cũng được duy trì. Phương pháp lấy mẫu này cho phép tập dữ liệu cài đặt tham số trở thành đại diện tốt hơn cho toàn bộ tập dữ liệu. Dữ liệu lựa chọn tham số này được chia thành tập huấn luyện và holdout. Mỗi bộ bao gồm 10% toàn bộ dữ liệu. Bảng 2 mô tả số lượng các trường hợp tăng và giảm cho tập dữ liệu lựa chọn tham số. Những thống kê này dành cho S & P BSE Sensex. Phân tích dữ liệu tương tự được thực hiện cho CNX Nifty, Reliance Industries và Infosys Ltd.
  + 
  + Các tham số tối ưu cho các mô hình dự đoán thu được bằng các thí nghiệm trên dữ liệu lựa chọn tham số. Sau đó, để so sánh ANN, SVM, Random forest và Naive-Bayes, bộ dữ liệu so sánh được đưa ra. Bộ dữ liệu này bao gồm toàn bộ mười năm dữ liệu. Nó cũng được chia trong đào tạo (50% toàn bộ dữ liệu) và giữ (50% toàn bộ dữ liệu) như bảng. Chi tiết về bộ dữ liệu này của S & P BSE SENSEX được hiển thị trong Bảng 3.
  + Các cài đặt thử nghiệm này giống như trong Kara, Acar Boyacioglu và Baykan (2011).
  + Có một số chỉ số kỹ thuật thông qua đó người ta có thể dự đoán sự chuyển động của cổ phiếu trong tương lai. Ở đây trong nghiên cứu này, tổng cộng mười chỉ số kỹ thuật được sử dụng trong Kara et al. (2011) được sử dụng. Các chỉ số này được thể hiện trong Bảng 4. Bảng 5 cho thấy số liệu thống kê tóm tắt cho các chỉ số được chọn của hai chỉ số và hai cổ phiếu. Hai cách tiếp cận để biểu diễn dữ liệu đầu vào được sử dụng trong nghiên cứu này. Cách tiếp cận đầu tiên sử dụng biểu diễn giá trị liên tục, tức là chuỗi thời gian thực tế trong khi cách thứ hai sử dụng biểu diễn xác định xu hướng (có tính chất rời rạc) cho các đầu vào. Cả hai đại diện được thảo luận ở đây.
  + 2.1. Continuous representation – the actual time series
    - Mười chỉ số kỹ thuật được tính toán dựa trên công thức như được thảo luận trong Bảng 4 được đưa ra làm đầu vào cho các mô hình dự đoán. Rõ ràng là mỗi chỉ số kỹ thuật được tính toán dựa trên công thức đã đề cập ở trên có giá trị liên tục. Các giá trị của tất cả các chỉ báo kỹ thuật được chuẩn hóa trong phạm vi giữa [-1, +1], để giá trị lớn hơn của một input params không áp đảo chỉ báo có giá trị nhỏ hơn. Hiệu suất của tất cả các mô hình đang nghiên cứu được đánh giá cho đại diện đầu vào này.



* + 2.2 Discrete representation – trend prediction data
    - Một lớp quyết định mới được sử dụng để chuyển đổi các tham số kỹ thuật có giá trị liên tục thành giá trị rời rạc, đại diện cho xu hướng. Chúng tôi gọi lớp này là ‘‘Trend Deterministic Data Preparation movement” Công việc của lớp mới này là chuyển đổi các giá trị liên tục này thành +1 hoặc -1 bằng cách xem xét thuộc tính này trong quá trình rời rạc. Bằng cách này, dữ liệu đầu vào của từng mô hình dự đoán được chuyển đổi thành +1 hoặc -1, trong đó +1 chuyển động đi lên và -1 hiển thị chuyển động xuống. Chi tiết về cách lấy ý kiến của từng chỉ số kỹ thuật được nêu ra dưới đây.
    - Hai chỉ số kỹ thuật đầu tiên đang di chuyển trung bình. Đường trung bình động (MA) là công cụ phân tích kỹ thuật đơn giản giúp làm mịn dữ liệu giá bằng cách tạo ra giá trung bình được cập nhật liên tục. Trong bài báo này, 10 ngày trung bình di chuyển đơn giản (SMA) và trung bình di chuyển có trọng số (WMA) được sử dụng vì chúng tôi dự đoán tương lai ngắn hạn. Theo nguyên tắc chung, nếu giá cao hơn mức trung bình động thì xu hướng tăng. Nếu giá dưới mức trung bình động, xu hướng giảm <http://www.investopedia.com>, <http: // www.stockcharts.com>. Vì vậy, theo những điều này, chúng tôi đã rút ra ý kiến của cả hai chỉ số SMA và WMA cho mỗi ngày từ giá trị của SMA và WMA so với giá hiện tại. Nếu giá hiện tại cao hơn các giá trị trung bình di động thì xu hướng sẽ tăng lên và được biểu thị là +1, và nếu giá hiện tại nằm dưới các giá trị trung bình di chuyển thì xu hướng đó sẽ giảm xuống và được biểu thị dưới dạng -1.
    - STCK%, STCD% và Williams R% là các dao động ngẫu nhiên. Những dao động là chỉ số xu hướng rõ ràng cho bất kỳ cổ phiếu. Khi các dao động ngẫu nhiên đang tăng lên, giá cổ phiếu có thể sẽ tăng lên và tiếp tục. Điều này ngụ ý rằng nếu giá trị của các bộ dao động ngẫu nhiên tại thời điểm đó lớn hơn giá trị tại thời điểm t thì ý kiến về xu hướng là lên và được biểu thị dưới dạng +1.
    - MACD đi theo xu hướng của cổ phiếu, tức là nếu MACD tăng thì giá cổ phiếu cũng tăng và ngược lại. Vì vậy, nếu giá trị của MACD tại thời điểm 't' lớn hơn giá trị tại thời điểm 't-1', ý kiến về xu hướng là 'lên' và được biểu thị là '+1' và nếu giá trị của MACD tại thời điểm 't' nhỏ hơn giá trị tại thời điểm 't-1', ý kiến về xu hướng là 'xuống' và được gửi lại là -1.
    - RSI thường được sử dụng để xác định các điểm mua quá mức và bán quá mức <http://www.stockcharts.com>. Nó nằm trong khoảng từ 0 đến 100. Nếu giá trị của RSI vượt quá 70 mức, điều đó có nghĩa là cổ phiếu bị mua quá mức, do đó, nó có thể giảm trong tương lai gần (biểu thị '-1') và nếu giá trị của RSI giảm xuống dưới 30 mức độ, điều đó có nghĩa là cổ phiếu bị bán quá mức, vì vậy, nó có thể sẽ tăng trong tương lai gần (biểu thị ý kiến '+1'). Đối với các giá trị nằm trong khoảng (30, 70), nếu RSI tại thời điểm đó lớn hơn RSI tại thời điểm ’t-1, thì ý kiến về xu hướng được biểu thị là +1 và ngược lại.
    - CCI đo lường sự khác biệt giữa thay đổi giá cổ phiếu và thay đổi giá trung bình. Chỉ số tích cực cao cho thấy giá cao hơn mức trung bình của họ, đó là một sự thể hiện sức mạnh. Chỉ số tiêu cực thấp cho thấy giá thấp hơn mức trung bình của họ, đây là một biểu hiện của sự yếu kém. CCI cũng được sử dụng để xác định mức mua quá mức và bán quá mức. Trong bài báo này, chúng tôi đã đặt 200 là mức mua quá mức và -200 là mức bán quá mức vì 200 là đại diện cho một thái cực thực sự <http://www.stockcharts.com>. Điều này có nghĩa là nếu giá trị CCI vượt quá 200 cấp, ý kiến cho xu hướng là -1 và nếu dưới mức 200 thì ý kiến cho xu hướng là ‘+1. Đối với các giá trị trong khoảng (-200, 200), nếu CCI tại thời điểm đó lớn hơn CCI tại thời điểm t-1, thì ý kiến về xu hướng là ‘+1 và ngược lại.
    - Bộ tạo dao động A / D cũng tuân theo xu hướng chứng khoán có nghĩa là nếu giá trị của nó tại thời điểm đó lớn hơn giá trị tại thời điểm đó, thì ý kiến về xu hướng là +1 và ngược lại.
    - Động lượng đo tốc độ tăng giảm của giá cổ phiếu. Giá trị dương của động lượng biểu thị xu hướng tăng và được biểu thị bằng +1 trong khi giá trị âm biểu thị xu hướng giảm và được gửi lại dưới dạng -1.
    - Tóm lại, dữ liệu xác định xu hướng được chuẩn bị bằng cách khai thác thực tế rằng mỗi chỉ số kỹ thuật có ý kiến cố hữu riêng về biến động giá cổ phiếu. Khi chúng tôi cung cấp các dữ liệu này làm đầu vào cho mô hình trái ngược với giá trị liên tục thực tế của chúng, chúng tôi đã nhập thông tin xu hướng theo cảm nhận của từng chỉ số kỹ thuật riêng lẻ. Đây là một bước tiến theo nghĩa mà bây giờ các mô hình dự đoán phải xác định mối tương quan giữa xu hướng đầu vào và xu hướng đầu ra.
    - Sử dụng các giá trị chỉ báo này, bộ đầu vào xác định xu hướng được chuẩn bị và đưa ra cho các mô hình dự đoán. Hiệu suất của tất cả các mô hình đang nghiên cứu cũng được đánh giá cho đại diện đầu vào này.
* Các mô hình dự đoán
  + Mô hình ANN (Artificial Neuron Network):
    - Là một mạng lưới dày đặc các nơ-ron liên kết với nhau được kích hoạt dựa trên các đầu vào. Một mạng nơ-ron 3 lớp feed-forward được sử dụng trong nghiên cứu. Đầu vào cho mạng là 10 chỉ số kỹ thuật được biểu thị bằng 10 nơ-ron trong lớp đầu vào. Lớp đầu ra có một nơ-ron duy nhất ứng với log sigmoid như là hàm chuyển. Điều này dẫn đến đầu ra giá trị liên tục trong khoảng từ 0 - 1. Ngưỡng 0,5 được sử dụng để xác định dự đoán chuyển động lên hoặc xuống. Lớn hơn 0,5 dự đoán là chuyển động tăng, ngược lại là giảm. Mỗi tầng nơ-ron ẩn sử dụng tan sigmoid như là hàm chuyển.
    - 
    - Kiến trúc 3 lớp feed-forward ANN được minh hoạ ở hình trên. Độ dốc giảm với động lượng được sử dụng để điều chỉnh các trọng số, trong đó tại mỗi thời kỳ, trọng số được điều chỉnh sao cho có thể đạt mức tối thiểu toàn cục.
    - Các thử nghiệm thiết lập tham số toàn diện để xác định tham số cho từng cổ phiếu và chỉ số chứng khoán. Các tham số cho mô hình ANN là số lớn nơ-ron ẩn (n) giá trị tỉ lệ học (lr), hằng số động lượng (mc) và số thời kỳ (ep). Để xác định chúng một cách hiệu quả, mười cấp (n), chín cấp (mc) và mười cấp (ep) được kiểm tra trong các thử nghiệm cài đặt tham số. Ban đầu, giá trị (lr) được cố định là 0,1. Các tham số này và mức độ đã được kiểm tra được tóm tắt trong bảng dưới.
    - 
    - Các thiết lập tham số này mang lại tổng cộng 10x10x9 = 900 phương pháp ANN cho một cổ phiếu. Xét 2 cổ phiếu và 2 chỉ số chứng khoán, tổng số 3600 phương pháp cho ANN được thực hiện. Ba kết hợp tham số hàng đầu dẫn tới kết quả huấn luyện và kiểm tra trung bình tốt nhất được chọn là 3 mô hình ANN hàng đầu cho các thử nghiệm so sánh trên tập dữ liệu. Đối với các mô hình hiệu suất này, tỷ lệ học tập (lr) được thay đổi trong khoảng [0,1, 0,9].
  + Mô hình SVM (Support Vector Machine):
    - Lần đầu tiên được giới thiệu bởi Vapnik (1999). Có 2 loại chính là: Support Vector Classification (SVC) và Support Vector Regression (SVR). SVM là hệ thống học sử dụng không gian đặc trưng có chiều lớn. Khemframani và Chandra (2009) tuyên bố rằng trong SVM, các điểm được phân loại bằng cách gán chúng vào một trong hai nửa không gian khác nhau, trong không gian mẫu hoặc trong không gian đặc trưng có số chiều cao hơn.
    - Mục tiêu của SVM là xác định siêu phẳng có biên tối đa. Ý tưởng là khoảng cách từ mặt phẳng tới điểm mẫu positive và negative là tôi đa và bằng nhau (Xu, Zhou, & Wang, 2009). Nó tìm siêu phẳng có biên tối đa là ranh giới quyết định cuối cùng.
    - Giả sử rằng tạo thành tập các vector đầu vào với nhãn tương ứng .
    - SVM có thể ánh xạ các vector đầu vào vào một không gian đặc trưng có số chiều lớn . Một hàm kernel thực hiện ánh xạ.
    - Ranh giới quyết định kết quả được xác định trong biểu thức:
    - Vấn đề về phương trình bậc 2 được thể hiện trong các phương trình dưới đây có thể lấy giá trị của
    - )
    - Sự đánh đổi giữa khoảng cách và lỗi phân loại được kiểm soát bởi tham số chính quy (c). Các hàm nhân đa thức và cơ sở xuyên tâm được sử dụng và hiển thị trong các biểu thức dưới đây

    - Trong đó ( d ) là độ của hàm đa thức và ( ) là hằng số của hàm cơ sở xuyên tâm.
    - Lựa chọn hàm kernel, mức độ của hàm kernel ( d ) trong trường hợp kernel đa thức, gamma trong hàm kernel () trong trường hợp kernel cở sở xuyên tâm và hằng số chính quy ( c ) là các tham số SVM.
    - Để xác định chúng một cách hiệu quả, bốn cấp ( d ), mười cấp () và 4 đến 5 cấp ( c ) được thử nghiệm trong các thử nghiệm thiết lập tham số. Các tham số này và mức độ của chúng được kiểm tra và tóm tắt trong bảng dưới.
    - 
    - Với một cổ phiếu, các thiết lập tham số này mang lại tổng cộng 20 và 40 phương pháp cho SVM sử dụng các hàm nhân cơ sở đa thức và hướng tâm tương ứng. Xem xét hai chỉ số chứng khoán và hai cổ phiếu, tổng số 240 phương pháp cho SVM được thực hiện. Một kết hợp tham số cho mỗi SVM kernel đa thức và SVM kernel cơ sở xuyên tâm dẫn đến hiệu suất đào tạo và tổ chức trung bình tốt nhất được chọn làm hai mô hình SVM hàng đầu cho các thử nghiệm so sánh.
  + Mô hình Random forest:
    - Cây quyết định là một trong những kỹ thuật phổ biến nhất để phân loại. Độ chính xác tương đương với các phương pháp phân loại khác, và nó rất hiệu quả. Mô hình học được thông qua kỹ thuật này được biểu diễn dưới dạng cây và được gọi là cây quyết định.
    - Quinlan (1986), C4.5 Quinlan (1993) và Cart Breiman, Friedman, Stone và Olshen (1984) là các thuật toán học cây quyết định. Chi tiết có thể được tìm thấy trong Han, Kamber và Pei (2006).
    - Random forest thuộc về loại thuật toán học. Nó sử dụng cây quyết định làm cơ sở cho việc học. Ý tưởng của việc học là một bộ phân loại duy nhất không đủ để xác định lớp dữ liệu kiểm tra. Lý do là, dựa trên dữ liệu mẫu, bộ phân loại không thể phân biệt giữa nhiễu và mẫu. Vì vậy, nó thực hiện lấy mẫu với sự thay thế sao cho ( n ) cây được học dựa trên các mẫu tập dữ liệu này. Cũng trong các thí nghiệm, mỗi cây được học bằng 3 đặc trưng được chọn ngẫu nhiên. Điều này cũng tránh được vấn đề về quá khớp dữ liệu. Việc thực hiện thuật toán Random forest được tóm tắt như dưới đây.
    - 
    - Số lượng cây trong toàn bộ n cây được coi là tham số của rừng ngẫu nhiên.
    - Để xác định nó một cách hiệu quả, nó được thay đổi từ 10 đến 200 với mức tăng 10 mỗi lần trong các thử nghiệm thiết lập tham số. Đối với một cổ phiếu, các cài đặt tham số này mang lại tổng cộng 20 phương pháp. Xem xét 2 chỉ số chứng khoán và 2 cổ phiếu, tổng số 80 phương pháp điều trị được thực hiện. Ba giá trị tham số hàng đầu dẫn đến kết quả trung bình tốt nhất về hiệu suất đào tạo và tổ chức được chọn là ba mô hình rừng ngẫu nhiên hàng đầu cho các thử nghiệm so sánh.
  + Phân lớp Naive-Bayes:
    - Phân lớp Naive-Bayes giả định lớp độc lập có điều kiện. Phân lớp Bayes dự đoán xác suất dữ liệu thuộc về một lớp cụ thể. Để dự đoán xác suất, nó sử dụng khái niệm định lý Bayes. Định lý Bayes, hữu ích ở chỗ nó cung cấp cách tính xác suất, cụ thể là:
    - Trong đó là xác suất giả thuyết C là đúng với sự kiện X đã xảy ra.
    - Trong trường hợp giả thuyết C là xác suất thuộc về lớp Up = Down và sự kiện X là dữ liệu thử nghiệm. là một xác suất có điều kiện xảy ra sự kiện X với giả thuyết C là đúng. Nó có thể được ước tính từ dữ liệu đào tạo. Công việc của Bayes đơn giản được tóm tắt như sau:
    - Giả sử có m lớp và sự kiện xảy ra của tập dữ liệu kiểm tra X được đưa ra. Bayes phân loại dữ liệu thành lớp với xác suất cao nhât. Theo định lý Bayes:
    - Tập dữ liệu đã cho với nhiều thuộc tính ( ), nó sẽ cực kì tốn kém khi tính toán . Để giảm tính toán, phương pháp này giả định lớp độc lập có điều kiện được tạo. Điều này giả định rằng các giá trị của các thuộc tính độc lập có điều kiện với nhau, được gắn nhãn của bộ dữ liệu (nghĩa là không có mối quan hệ phụ thuộc giữa các thuộc tính). Do đó:
    - Ở đây thể hiện giá trị của thuộc tính cho bộ X. Tính toán phụ thuộc vào việc phân loại hay liên tục. Nếu là phân loại, thì là số lượng quan sát của lớp trong tập huấn luyện có giá trị cho , chia cho số lượng quan sát của lớp trong tập huấn luyện. Nếu có giá trị liên tục, thì phân phối Gaussian được gắn vào dữ liệu và giá trị của được tính toán dựa trên biểu thức:
    - Vậy nên
    - Ở đây và lần lượt là giá trị trung bình và tiêu chuẩn của các giá trị thuộc tính cho bộ huấn luyện của lớp . Hai đại lượng này sau đó được gắn vào biểu thức ở trên cùng với . Để ước lượng , được ước tính cho mỗi lớp để dự đoán nhãn của lớp X. Nhãn quan sát của X được dự đoán là lớp khi và chỉ khi:
    - Các bộ phân lớp Bayes cũng đóng vai trò là lý thuyết cho các phân lớp khác không sử dụng rõ ràng định lý Bayes. Ví dụ, theo các giả định cụ thể, có thể chứng minh rằng nhiều mạng thần kinh và thuật toán khớp đường cong đưa ra giả thuyết posteriori tối đa, giống như phân lớp Naive-Bayes.
* Kết quả thực nghiệm:
  + - Độ chính xác và độ đo f được sử dụng để đánh giá hiệu suất của các mô hình được đề xuất. Việc tính toán các biện pháp đánh giá này đòi hỏi phải ước tính Độ rõ ràng (Precision) và Thu hồi (Recall) được đánh giá từ True Positive (TP), False Positive (FP), True Negative (TN) và False Negative (FN). Các tham số này được xác định trong các phương trình.
    - Độ rõ ràng (Precision) là trung bình có trọng số của độ chính xác positive và negative trong khi Thu hồi (Recall) là trung bình có trọng số của thu hồi positive và negative. Độ chính xác và độ đo f được tính bằng các phương trình:
    - Giai đoạn thử nghiệm đầu tiên coi đầu vào là dữ liệu có giá trị liên tục. Các kết hợp tham số tốt nhất được xác định bằng các thử nghiệm trên tập dữ liệu cài đặt tham số cho từng mô hình dự đoán. Các kết hợp tham số này với Dộ chính xác và Dộ đo f đáp ứng trong các thử nghiệm cài đặt tham số được báo cáo trong bảng dưới:
    - 
    - Cần lưu ý rằng không có tham số nào được điều chỉnh cho phân lớp Naive-Bayes.
    - Mục đích của các thử nghiệm trên tập dữ liệu so sánh là để hoàn thành hiệu suất dự đoán của các mô hình này cho các kết hợp tương đương tốt nhất được báo cáo trong các thử nghiệm cài đặt tham số. Trong thí nghiệm so sánh này, mỗi mô hình dự đoán được học dựa trên các tham số tốt nhất được báo cáo bởi các thử nghiệm cài đặt tham số.
    - 
    - Bảng trên báo cáo độ chính xác trung bình và số đo f của từng mô hình trong quá trình thử nghiệm so sánh. Độ chính xác trung bình và số đo f được báo cáo trung bình trên các mô hình hoạt động hàng đầu. Có thể thấy rằng Naive-Bayes với quy trình Gaussian là ít chính xác nhất trong khi Random forest là chính xác nhất với độ chính xác trung bình gần 84%. Hình dưới mô tả quá trình dự đoán khi dữ liệu được định giá liên tục.
    - 
    - Giai đoạn thử nghiệm thứ hai giống hệt với giai đoạn đầu tiên ngoại trừ đầu vào của các mô hình là dữ liệu xác định xu hướng. Ý tưởng được mô tả trong hình dưới đây
    - 
    - Bảng 12 cho thấy kết quả của các kết hợp hình thành tốt nhất cho ANN, SVM và Random forest tương ứng trong các thí nghiệm cài đặt tham số.
    - 
    - Cần lưu ý rằng khi dữ liệu được biểu diễn dưới dạng dữ liệu xác định xu hướng, phân lớp Naive-Bayes được học bằng cách phù hợp với phân phối đa biến Bernoulli cho dữ liệu. Kết quả về tập dữ liệu so sánh cho tất cả các mô hình được đề xuất được báo cáo trong Bảng 15.
    - 
    - So sánh cuối cùng cho thấy tất cả các mô hình hoạt động tốt với dữ liệu đầu vào riêng biệt nhưng SVM, Random forest và Naive-Bayes hoạt động tốt hơn ANN. Độ chính xác của 3 mô hình trên là gần 90%.
* Thảo luận:
  + - Dữ liệu thị trường chứng khoán là một ví dụ về dữ liệu không cố định. Tại thời điểm cụ thể có thể có xu hướng (trends) , chu kỳ (cycles), “random walk” hoặc kết hợp cả ba. Người ta mong muốn rằng nếu một năm cụ thể là một phần của chu kỳ nói rằng một năm tăng thì mô hình của chúng ta nên theo mô hình này để dự đoán xu hướng. Tương tự có thể được xem xét cho một năm xu hướng. Tuy nhiên, thông thường giá trị cổ phiếu của một năm cụ thể không bị cô lập và có những ngày “random walk”. Giá trị cổ phiếu cũng bị ảnh hưởng bởi các yếu tố bên ngoài tạo ra xu hướng và tình trạng của nền kinh tế của đất nước. Chính trị cũng là yếu tố ảnh hưởng có thể dẫn đến chu kỳ.
    - Có thể thấy từ kết quả rằng tất cả các mô hình hoạt động tốt khi chúng được học từ các đầu vào có giá trị liên tục nhưng hiệu suất của từng mô hình được cải thiện hơn nữa khi chúng được học bằng cách sử dụng dữ liệu xác định xu hướng. Dữ liệu xác định xu hướng được chuẩn bị bằng cách phân biệt dữ liệu có giá trị liên tục. Ý tưởng này dựa trên thực tế là mỗi tham số có giá trị liên tục khi so sánh với giá trị ngày trước của nó cho thấy xu hướng tăng hoặc giảm trong tương lai. Các dữ liệu được rời rạc dựa trên các heuristic này. Khi dữ liệu này được cung cấp làm đầu vào cho mô hình, chúng tôi đã nhập xu hướng dựa trên từng tham số đầu vào. Đây thực sự là tình huống mà mỗi tham số đầu vào biểu thị về xu hướng tương lai có thể xảy ra và chúng ta có xu hướng tương lai thực tế để xác định sự chuyển đổi từ xu hướng có thể xảy ra sang xu hướng chính xác. Đây là một bước tiến từ việc chuyển đổi tập dữ liệu từ bộ dữ liệu không cố định sang tập dữ liệu xác định xu hướng. Bây giờ các mô hình của chúng tôi phải xác định mối quan hệ đồng giữa xu hướng đầu vào và xu hướng đầu ra. Mặc dù nó không tuyến tính, nhưng việc tạo ra một mô hình có thể chuyển đổi xu hướng đầu vào thành xu hướng đầu ra sẽ dễ dàng hơn.
    - Khi đưa ra các chỉ số kỹ thuật có giá trị liên tục làm đầu vào cho các mô hình, chúng sẽ bị tước đi, đây là xu hướng mà mỗi chỉ dẫn kỹ thuật đưa ra. Điều này khiến các mô hình dự đoán phân loại dựa trên giá trị của các chỉ số kỹ thuật này nhưng thông tin từ quá trình chuyển đổi giá trị của cổ phiếu bị mất và không được sử dụng bởi các mô hình dự đoán. Vì mục tiêu là dự đoán hướng di chuyển hoặc xu hướng, nên dữ liệu xác định xu hướng phù hợp hơn.
    - Ngoài ra, đối với bất kỳ cổ phiếu hoặc chỉ số nào cũng có kịch bản khi họ giao dịch ở một số giá trị là 200, sau đó do một số yếu tố bên ngoài, họ có thể bắt đầu giao dịch ở mức giá cao hơn 400 và sau đó ổn định ở giá trị cao hơn đó.
    - Nếu mô hình được cung cấp đầu vào có giá trị liên tục trực tiếp, thì có thể nó cố gắng thiết lập quan hệ giữa các giá trị trong 200 và trong 400 không yêu cầu xa hơn giá trị tuyệt đối của thay đổi. Do đó, dữ liệu xác định xu hướng về bản chất là rời rạc, là dấu hiệu thống kê về việc cổ phiếu được mua quá mức hay bán quá mức và độc lập với giá trị. Do đó, các tham số đầu vào này, khi được biểu diễn dưới dạng các xu hướng trong tương lai có thể đóng vai trò là thước đo tốt hơn về tình trạng cổ phiếu thay vì kịch bản khi chúng được gửi lại dưới dạng giá trị liên tục.
* Kết luận:
  + - Nhiệm vụ tập trung trong bài viết này là dự đoán hướng di chuyển cho các chỉ số chứng khoán và giá cổ phiếu. Hiệu suất dự đoán của bốn mô hình là ANN, SVM, Random forest và Naive-Bayes được so sánh dựa trên mười năm (2003, 2012) của dữ liệu lịch sử của CNX Nifty, S & P BSE Sensex, Infosys Ltd. và Reliance Industries từ thị trường chứng khoán Ấn Độ.
    - Mười thông số kỹ thuật phản ánh tình trạng của chỉ số chứng khoán và giá cổ phiếu được sử dụng để tìm hiểu từng mô hình này. Lớp dữ liệu xác định xu hướng được sử dụng để chuyển đổi từng báo cáo kĩ thuật liên tục thành +1 hoặc -1 cho thấy khả năng tăng hoặc giảm tương lai có thể xảy ra trong tương lai.
    - Các thử nghiệm với dữ liệu có giá trị liên tục cho thấy mô hình Naive (quy trình Gaussian) thể hiện hiệu suất thấp nhất với độ chính xác 73,3% và Random forest với hiệu suất cao nhất với độ chính xác 83,56%. Hiệu suất của tất cả các mô hình này được cải thiện đáng kể khi chúng được học thông qua dữ liệu xác định xu hướng. ANN kém chính xác hơn một chút về độ chính xác dự đoán so với ba mô hình khác thực hiện gần như giống hệt nhau. Độ chính xác của 86,69%, 89,33%, 89,98% và 90,19% đạt được bằng ANN, SVM, Random forest và Naive-Bayes (Quy trình đa biến Bernoulli).
    - Lớp dữ liệu xác định xu hướng khai thác từng chỉ số kĩ thuật về biến động giá cổ phiếu. Nó giống những chuyên gia về thị trường chứng khoán. Trong các thí nghiệm trước đó, các chỉ số kỹ thuật được sử dụng trực tiếp để dự đoán trong khi nghiên cứu này lần đầu tiên trích xuất thông tin liên quan đến xu hướng từ mỗi chỉ số kỹ thuật và sau đó sử dụng tương tự để dự đoán, giúp cải thiện độ chính xác đáng kể.
    - Nhờ có sự cải thiện đáng chú ý trong khả năng dự đoán, hệ thống được đề xuất có thể được triển khai trong thời gian thực để dự đoán xu hướng của cổ phiếu, giúp đầu tư có lợi hơn và an toàn hơn. Cải thiện độ chính xác với sự trợ giúp của phương pháp này dựa trên các phương pháp phổ biến của nhà đầu tư để đầu tư chứng khoán, cũng cung cấp ý tưởng xử lý trước dữ liệu dựa trên miền sử dụng thuật toán học máy. Ý tưởng này không những có thể được mở rộng không chỉ trong lĩnh vực chứng khoán bằng cách kết hợp các phương pháp đầu tư khác của con người mà còn trong các lĩnh vực khác, nơi các hệ thống chuyên gia và kỹ thuật học máy được sử dụng.
    - Mười chỉ số kỹ thuật được sử dụng trong bài viết này để xây dựng nền tảng kiến thức, tuy nhiên, các biến động kinh tế vĩ mô khác như tỷ giá hối đoái, lạm phát, chính sách của chính phủ, lãi suất, v.v. cũng là một yếu tố có thể hữu ích trong việc quyết định xu hướng.
    - Trong bài báo này, các chỉ số kĩ thuật về sự biến động giá cổ phiếu được phân loại thành “tăng” hoặc “giảm”. Đi cùng nhiều thể loại như “rất có thể tăng”, “rất có thể giảm” hoặc trung lập. Điều này có thể cung cấp đầu vào chính xác hơn cho các thuật toán dự đoán trong bài viết này.
    - Ngoài ra, trọng tâm của bài viết này là dự đoán ngắn hạn. Dự đoán dài hạn cũng có thể được coi là một trong những định hướng trong tương lai có thể liên quan đến phân tích hiệu suất hàng quý của cổ phiếu, doanh thu, lợi nhuận, sự ổn định của tổ chức công ty, v.v. (ví dụ SMA, WMA, v.v.).
    - Trên hết, thành công của phương pháp đề xuất dựa trên phương pháp đầu tư của con người, khuyến khích mô phỏng phương pháp ra quyết định của con người trong khi phát triển hệ thống chuyên gia và sử dụng thuật toán học máy cho các vấn đề trong các lĩnh vực khác nhau.