

UBND TỈNH BÌNH DƯƠNG
TRƯỜNG ĐẠI HỌC THỦ DẦU MỘT

NGUYỄN MINH LỢI

**DỰ ĐOÁN GIÁ CỒ PHIẾU BẰNG PHƯƠNG PHÁP HỌC
SÂU KHÔNG GIÁM SÁT
GENERATIVE ADVERSARIAL NETWORK (GAN)**

**CHUYÊN NGÀNH: HỆ THỐNG THÔNG TIN
MÃ SỐ: 8480104**

LUẬN VĂN THẠC SỸ

BÌNH DƯƠNG - 2020

**UBND TỈNH BÌNH DƯƠNG
TRƯỜNG ĐẠI HỌC THỦ DẦU MỘT**

NGUYỄN MINH LỢI

**DỰ ĐOÁN GIÁ CỐ PHIẾU BẰNG PHƯƠNG PHÁP
HỌC SÂU KHÔNG GIÁM SÁT
GENERATIVE ADVERSARIAL NETWORK (GAN)**

**CHUYÊN NGÀNH: HỆ THỐNG THÔNG TIN
MÃ SỐ: 8480104**

LUẬN VĂN THẠC SỸ

**NGƯỜI HƯỚNG DẪN KHOA HỌC:
TIẾN SĨ. BÙI THANH HÙNG**

BÌNH DƯƠNG – 2020

LỜI CAM ĐOAN

Tôi xin cam đoan rằng, luận văn “**Dự đoán giá cổ phiếu bằng phương pháp học không giám sát Generative Adversarial Network (GAN)**” là công trình nghiên cứu của tôi dưới sự hướng dẫn của TS. Bùi Thanh Hùng, xuất phát từ nhu cầu thực tiễn và nguyện vọng tìm hiểu của bản thân.

Ngoại trừ kết quả tham khảo từ các công trình khác đã ghi rõ trong luận văn, các nội dung trình bày trong luận văn này là kết quả nghiên cứu do chính tôi thực hiện và kết quả của luận văn chưa từng công bố trước đây dưới bất kỳ hình thức nào.

Bình Dương, thángnăm 20....

Tác giả

Nguyễn Minh Lợi

LỜI CẢM ƠN

Qua thời gian học tập và rèn luyện tại trường Đại học Thủ Dầu Một, được sự chỉ bảo và giảng dạy nhiệt tình của quý thầy cô, đặc biệt là các thầy của các trường đại học ở thành phố Hồ Chí Minh đã không ngại đường xá xa xôi để truyền đạt kiến thức cho tôi trong suốt thời gian học ở trường. Cùng với sự nỗ lực của bản thân, tôi đã hoàn thành luận văn của mình.

Từ những kết quả đạt được này, tôi xin chân thành cảm ơn quý thầy cô trường Đại học Thủ Dầu Một, đã truyền đạt cho tôi những kiến thức bổ ích trong thời gian qua. Đặc biệt, TS. Bùi Thanh Hùng đã tận tình hướng dẫn và giúp đỡ tôi hoàn thành báo cáo luận văn thạc sỹ này. Do kiến thức còn hạn hẹp nên không tránh khỏi những thiếu sót trong cách diễn đạt và trình bày. Tôi rất mong nhận được sự đóng góp ý kiến của quý thầy cô để báo cáo luận văn đạt được kết quả tốt nhất.

Tôi xin kính chúc quý thầy cô và các bạn thật nhiều sức khỏe, niềm vui, luôn thành công trong công việc và cuộc sống.

TÓM TẮT LUẬN VĂN

Thị trường cổ phiếu có một vai trò quan trọng trong sự phát triển của xã hội hiện đại. Chúng cho phép triển khai các nguồn lực kinh tế. Sự thay đổi giá cổ phiếu phản ánh những thay đổi trên thị trường. Với khả năng xử lý dữ liệu mạnh mẽ trên nhiều lĩnh vực, học sâu cũng được sử dụng một cách rộng rãi trong lĩnh vực tài chính như: dự đoán thị trường cổ phiếu, đầu tư tối ưu, xử lý thông tin tài chính và thực hiện các chiến lược giao dịch tài chính. Do đó, dự đoán thị trường cổ phiếu được xem là một trong những lĩnh vực khá phổ biến và quý giá nhất trong lĩnh vực tài chính.

Trong nghiên cứu này, chúng tôi đề xuất phương pháp học không giám sát Generative Adversarial Network (GAN) trong dự đoán giá cổ phiếu. Mô hình GAN gồm 2 lớp, một lớp bộ nhớ ngắn dài 2 chiều Bidirectional Long Short-Term Memory (Bi-LSTM) được dùng để phân biệt (Discriminator) và lớp Long Short-Term Memory (LSTM) được sử dụng để dự đoán giá cổ phiếu (Generator). LSTM dựa trên dữ liệu cổ phiếu đã giao dịch và tạo ra dữ liệu giả giống như dữ liệu đã phân phối, trong khi đó lớp phân biệt được thiết kế bởi giải thuật Bi-LSTM với mục đích phân biệt dữ liệu cổ phiếu thật và dữ liệu cổ phiếu giả được tạo ra.

Chúng tôi thực nghiệm trên cổ phiếu AMZN (Amazon) và một số cổ phiếu khác là cổ phiếu có chuỗi ngày giao dịch trong phạm vi rộng lớn và dùng chúng để thử dự đoán giá đóng hàng ngày. Kết quả thực nghiệm cho thấy phương pháp GAN đề xuất của chúng tôi có thể đạt được kết quả tốt trong việc dự đoán giá cổ phiếu so với nhiều mô hình dự đoán khác.

Luận văn cũng đề xuất xây dựng một ứng dụng web để trực quan hóa kết quả nghiên cứu và hỗ trợ người sử dụng dự đoán giá cổ phiếu thị trường từ các giao dịch cổ phiếu hiện tại đang hoạt động trên thị trường.

Kết quả thực nghiệm cho thấy mô hình đề xuất đạt được kết quả tốt trên các bộ dữ liệu được sử dụng để huấn luyện và đánh giá trên tất cả các độ đo: Độ chính xác (**Accuracy**) và các độ đo gồm lỗi bình quân tuyệt đối (**MAE**), lỗi hình vuông gốc trung bình (**RMSE**), phần trăm bình quân tuyệt đối (**MAPE**) và lợi nhuận trung bình (**AR**).

MỤC LỤC

LỜI CAM ĐOAN	i
LỜI CẢM ƠN.....	ii
TÓM TẮT LUẬN VĂN.....	iii
MỤC LỤC	v
DANH MỤC THUẬT NGỮ VÀ CÁC TỪ VIẾT TẮT.....	viii
DANH MỤC CÁC BẢNG	ix
DANH MỤC HÌNH VẼ, ĐỒ THỊ.....	x
CHƯƠNG 1	1
LỜI MỞ ĐẦU.....	1
1.1. Lý do chọn đề tài.....	1
1.3. Đối tượng, phạm vi nghiên cứu	3
1.4. Phương pháp nghiên cứu.....	3
1.5. Ý nghĩa khoa học và thực tiễn	4
1.6. Bố cục luận văn.....	4
CHƯƠNG 2	6
CƠ SỞ LÝ THUYẾT VÀ CÁC NGHIÊN CỨU LIÊN QUAN ...	6
2.1. Những vấn đề cơ bản về chứng khoán [22]	6
2.1.1. Khái niệm chứng khoán	6
2.1.2. Phân loại chứng khoán.....	6
2.2. Tổng quan về thị trường chứng khoán[24]	9
2.2.1. Khái niệm thị trường chứng khoán	9
2.2.2. Vai trò của thị trường chứng khoán:	10
2.2.3. Phân loại thị trường chứng khoán[23]	12
2.3. Học máy (Machine learning) và Thị trường chứng khoán:	14

2.4. Học sâu - Deep learning	17
2.4.1. Mạng nơ-ron hồi quy RNN (Recurrent Neural Network)	19
2.4.2. Mạng nơ-ron ngắn dài LSTM (Long-short term memory)	22
2.4.3. Mạng nơ-ron tái phát hai chiều BiLSTM (Bidirectional Long-short term memory)	25
2.4.4. Mạng sáng tạo đối nghịch GAN	28
2.5. Dự đoán giá cổ phiếu	32
2.5.1. Tổng quan	32
2.5.2. Các hướng tiếp cận	32
2.5.3 Hướng đề xuất nghiên cứu	34
CHƯƠNG 3	35
MÔ HÌNH ĐỀ XUẤT	35
3.1. Tổng quan mô hình đề xuất.....	35
3.2. Các đặc trưng của mô hình đề xuất.....	36
3.2.1. Lớp tạo - The Generator	36
3.2.2. Lớp phân biệt - The Discriminator	39
3.2.3. Cấu trúc GAN	40
3.2.4. Phương pháp đánh giá.....	41
CHƯƠNG 4	43
THỰC NGHIỆM	43
4.1. Dữ liệu.....	43
4.2. Xử lý dữ liệu	46
4.3. Huấn luyện	47
4.4. Đánh giá	51
4.5. Xây dựng ứng dụng trực quan hóa kết quả	52

CHƯƠNG 5	57
KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN.....	57
5.1. Kết quả đạt được	57
5.2. Hướng phát triển	58
CÔNG TRÌNH CÔNG BỐ.....	59
TÀI LIỆU THAM KHẢO	60

DANH MỤC THUẬT NGỮ VÀ CÁC TỪ VIẾT TẮT

Từ viết tắt	Từ đầy đủ
ANN	Mạng nơ-ron nhân tạo (Artificial Neural Network)
TTCK	Thị trường chứng khoán
CNN	Mạng nơ-ron tích chập (Convolutional Neural Network)
TTGDCK	Trung tâm giao dịch chứng khoán
CTNY	Công ty niêm yết
SGDCK	Sàn giao dịch chứng khoán
LSTM	Mạng nơ-ron dài ngắn
BiLSTM	Mạng nơ-ron dài ngắn 2 chiều (Bidirectional LSTM)
FC	Fully-connected
MLP	Multilayer Perceptron
NLP	Natural Language Processing
PL	Pooling Layer
ReLU	Rectified Linear Unit
RNN	Mạng nơ-ron hồi quy (Recurrent Neural Network)
GAN	Mạng sáng tạo đối nghịch (Generative Adversarial Network)

DANH MỤC CÁC BẢNG

Bảng 4. 1: Dữ liệu cổ phiếu của tập đoàn Amzon	44
Bảng 4. 2: Dữ liệu cổ phiếu của tập đoàn Apple	44
Bảng 4. 3: Dữ liệu cổ phiếu của tập đoàn Google	45
Bảng 4. 4: Dữ liệu cổ phiếu của tập đoàn Microsoft	45
Bảng 4. 5: Dữ liệu cổ phiếu của công ty Vietcombank	46
Bảng 4. 6: Kết quả đánh giá bằng mô hình LSTM	48
Bảng 4. 7: Kết quả đánh giá bằng mô hình SVM	48
Bảng 4. 8: Kết quả đánh giá bằng mô hình BAYES	49
Bảng 4. 9: Kết quả đánh giá bằng mô hình ANN	50
Bảng 4. 10: Kết quả độ đo của các mô hình	51

DANH MỤC HÌNH VẼ, ĐỒ THỊ

Hình 2. 1: Mô hình phân loại thị trường chứng khoán	13
Hình 2. 2: Mô hình về các loại thị trường cổ phiếu	14
Hình 2. 3: Lược sử học sâu Deep Learning	18
Hình 2. 4: Mô hình mạng hồi quy đơn giản.....	20
Hình 2. 5: Mô hình của một tế bào LSTM	22
Hình 2. 6: Cấu trúc của mạng nơ-ron hồi tiếp hai chiều.....	27
Hình 2. 7: Bidirectional LSTM = forward LSTM + backward LSTM ..	28
Hình 2. 8: Kiến trúc mô hình Gan	31
Hình 3. 1: Mô hình đề xuất tổng quát	35
Hình 3. 2: Mạng LSTM	37
Hình 3. 3: Lớp tạo được thiết kế với giải thuật LSTM.....	38
Hình 3. 4: Mô hình Bidirectional LSTM	39
Hình 3. 5: Cấu trúc Lớp phân biệt sử dụng Bi-LSTM.....	40
Hình 3. 6: Cấu trúc GAN đề xuất của chúng tôi.....	41
Hình 4. 1: Biểu đồ giá cổ phiếu sử dụng phương pháp LSTM.....	48
Hình 4. 2: Biểu đồ giá cổ phiếu sử dụng phương pháp SVM.....	49
Hình 4. 3: Biểu đồ giá cổ phiếu sử dụng phương pháp BAYES	50
Hình 4. 4: Biểu đồ giá cổ phiếu sử dụng phương pháp ANN.....	50
Hình 4. 5: Trang chủ giao diện Web.....	53
Hình 4. 6: Phân tích dữ liệu cổ phiếu	54

Hình 4. 7: Mô hình huấn luyện GAN (LSTM-BiLSTM)	55
Hình 4. 8: Đánh giá các mô hình	56

CHƯƠNG 1

LỜI MỞ ĐẦU

1.1. Lý do chọn đề tài

Ngày nay, tất cả các quốc gia phát triển và hầu hết các nước đang phát triển đều có thị trường chứng khoán, một thị trường không thể thiếu với mọi nền kinh tế muốn phát triển vững mạnh. Ở Việt Nam, dù đã trải qua 20 năm hình thành và phát triển kể từ khi Trung tâm Giao dịch Chứng khoán TP Hồ Chí Minh (sau này được đổi tên thành Sở Giao dịch Chứng khoán TP Hồ Chí Minh - HOSE) có phiên giao dịch đầu tiên ngày 28-7-2000, đến nay ngành Chứng khoán đã đạt được những thành tựu nhất định cùng với những chuyển mình ngày càng lớn mạnh của nền kinh tế đất nước.

Tính từ khi thành lập thị trường chứng khoán (TTCK) Việt Nam đến tháng 06/2020, thông qua TTCK, Chính phủ và các doanh nghiệp đã huy động được trên 2,4 triệu tỷ đồng để đưa vào sản xuất kinh doanh, riêng trong giai đoạn 2011 đến nay, quy mô huy động vốn qua TTCK đạt khoảng 1,7 triệu tỷ đồng, đóng góp bình quân 20% tổng vốn đầu tư toàn xã hội.

Quy mô vốn hóa TTCK có sự tăng trưởng vượt bậc, từ mốc sơ khai, vốn hóa thị trường chỉ đạt 986 tỷ đồng, chiếm 0,28% GDP vào năm 2000, đến nay (tính đến hết tháng 6/2020) giá trị vốn hóa thị trường cổ phiếu là 3.894 nghìn tỷ đồng, đạt 64,5% GDP (quy mô tăng 3.949 lần trong vòng 20 năm). Mức vốn hóa thị trường trái phiếu tăng trưởng tích cực, tương đương trên 30,3% GDP năm 2019, trong đó riêng thị trường

trái phiếu doanh nghiệp đạt gần 10,9% GDP. Tính chung, giá trị vốn hóa TTCK đạt 94,8% GDP đã góp phần định hình hệ thống tài chính hiện đại trên nền tảng hài hòa giữa TTCK và thị trường tiền tệ - tín dụng.

Sự phát triển của TTCK Việt Nam đã góp phần thúc đẩy quá trình cơ cấu lại nền kinh tế trên cả 3 trụ cột: Cải cách doanh nghiệp nhà nước thông qua cổ phần hóa, thoái vốn nhà nước bằng các cơ chế đấu giá minh bạch, hiện đại và gắn cổ phần hóa với đăng ký giao dịch, niêm yết trên TTCK; Tái cơ cấu đầu tư công thông qua việc trở thành kênh huy động vốn quan trọng cho ngân sách nhà nước; Hỗ trợ quá trình tái cơ cấu các tổ chức tín dụng, đặc biệt là các ngân hàng thương mại tham gia niêm yết trên TTCK. Qua đó, nhằm giúp các nhà quản lý doanh nghiệp, các nhà đầu tư và các cá nhân muốn tham gia thị trường chứng khoán có thông tin đầy đủ và rõ ràng hơn về giá cổ phiếu giúp mình đưa ra quyết định tham gia TTCK như thế nào để có lợi nhuận cao và bền vững, cụ thể hơn trong quá trình giao dịch cổ phiếu của mình. Từ đó nâng cao lòng tin, độ tin cậy cao vào thị trường cổ phiếu nhằm đem lại sự thỏa mãn tối đa cho các nhà đầu tư trên thị trường cũng là làm cho thị trường chứng khoán ngày một hiệu quả và hoạt động tốt hơn.

Chính vì lẽ đó dự đoán thị trường chứng khoán là một nhu cầu cấp thiết và có ý nghĩa thực tiễn. Chủ đề này đã được nhiều nhà nghiên cứu trong và ngoài nước quan tâm và đưa ra nhiều giải pháp. Mỗi giải pháp có những ưu nhược điểm khác nhau, tuy nhiên sử dụng học máy là giải pháp mang lại kết quả tốt. Vì các lý do trên tôi đã lựa chọn đề tài “**Dự đoán cổ phiếu bằng phương pháp học sâu không giám sát Generative Adversarial Network GAN**” là đề tài luận văn thạc sĩ của mình.

1.2. Mục tiêu nghiên cứu

Luận văn này tập trung nghiên cứu giải quyết bài toán dự đoán giá cổ phiếu trên thị trường cổ phiếu thế giới và Việt Nam với các cổ phiếu như **Google, Amazon, Facebook, Microsoft, Vietcombank...** Trên cơ sở dữ liệu thu thập được từ trang Web chuyên cung cấp, đánh giá, phân tích cổ phiếu tài chính trên thế giới Alphavantage chúng tôi tiến xử lý, trích xuất đặc trưng, áp dụng các phương pháp học máy và phương pháp học sâu không giám sát GAN cho bài toán dự đoán giá cổ phiếu từ đó đề xuất mô hình tối ưu nhất.

1.3. Đối tượng, phạm vi nghiên cứu

Đối tượng nghiên cứu của luận văn là các dữ liệu cổ phiếu từ trang web chuyên cung cấp, đánh giá, phân tích các đặc trưng cổ phiếu tài chính trên thế giới Alphavantage mà trong đó có liên quan đến các cổ phiếu của các tập đoàn, công ty lớn trên thế giới.

Phạm vi nghiên cứu: Các cổ phiếu có chuỗi ngày giao dịch trong phạm vi rộng lớn áp dụng các phương pháp học máy, học sâu cho bài toán dự đoán giá cổ phiếu.

1.4. Phương pháp nghiên cứu

- **Phương pháp nghiên cứu lý thuyết:** Tổng hợp, nghiên cứu các tài liệu về cổ phiếu chứng khoán; nghiên cứu các phương pháp, thuật toán sử dụng cho dự đoán giá cổ phiếu; nghiên cứu các phương pháp học sâu vào thị trường cổ phiếu. Tìm hiểu các kiến thức liên quan như thị trường chứng khoán, học máy, kỹ thuật lập trình trên máy tính.

- **Phương pháp nghiên cứu thực nghiệm:** Sau khi nghiên cứu lý thuyết, phát biểu bài toán, đề xuất mô hình; xây dựng và phát triển ứng dụng dựa trên mô hình đề xuất; cài đặt thử nghiệm chương trình, đánh giá các kết quả đạt được; công bố kết quả nghiên cứu trong các Hội thảo.

- **Phương pháp so sánh và đánh giá:** phân tích đánh giá mô hình đề xuất với các mô hình nghiên cứu trước bằng các độ đo khác nhau.

1.5. Ý nghĩa khoa học và thực tiễn

Ý nghĩa khoa học của luận văn. Luận văn đề xuất phương pháp giải quyết bài toán dự đoán giá cổ phiếu bằng cách trích xuất các đặc trưng, luận văn tiến hành thử nghiệm các hướng tiếp cận các mô hình khác nhau, từ đó đưa ra mô hình tối ưu nhất cho việc dự đoán giá. Để đánh giá mô hình đề xuất, luận văn đã so sánh hiệu suất giữa các mô hình **LSTM**, **BiLSTM** và **GAN** (LSTM-BiLSTM) trên tập dữ liệu bao gồm 10.000 mẫu của các cổ phiếu trên trang Web Alphavantage.

Ý nghĩa thực tiễn. Luận văn xây dựng trang ứng dụng trực quan các mô hình để đưa ra kết quả dự đoán giá cổ phiếu. Bên cạnh đó, ứng dụng cũng cung cấp số liệu thống kê hiệu suất các mô hình và phân tích dữ liệu.

1.6. Bố cục luận văn

Bố cục của luận văn được chia làm 5 chương chính, bao gồm những nội dung như sau:

- **Chương 1.** Giới thiệu tổng quan về lĩnh vực nghiên cứu của bài toán dự đoán giá cổ phiếu, phát biểu bài toán dự đoán và mục tiêu, đối tượng, phạm vi nghiên cứu cũng như phương pháp để đạt kết quả tối ưu nhất.

- **Chương 2.** Giới thiệu những kiến thức nền tảng sẽ được bao gồm trong luận văn. Cụ thể, chương này sẽ giới thiệu các kiến thức nền của học máy (mạng nơ-ron nhân tạo, kỹ thuật lan truyền ngược), các mô hình học sâu được huấn luyện (**LSTM**, **BiLSTM**, **GAN**). Tổng quan về

dự đoán thị trường cổ phiếu, cũng như hướng tiếp cận và đề xuất nghiên cứu.

- **Chương 3.** Trong chương này, tôi đi sâu vào mô hình đề xuất, đồng thời mô tả chi tiết các đặc trưng của mô hình GAN để dự đoán giá cổ phiếu. Sau đó là các phương pháp đánh giá hiệu quả mô hình đề xuất.

- **Chương 4.** Trình bày thực nghiệm, các bước tiền xử lý dữ liệu, huấn luyện mô hình, các phương pháp đánh giá mô hình và so sánh các kết quả được kiểm tra của từng mô hình, với từng hướng tổ chức tập dữ liệu khác nhau, từ đó rút ra được các nhận xét về các mô hình đã cài đặt và lựa chọn mô hình tối ưu trực quan hóa kết quả.

- **Chương 5.** Tóm tắt các ý chính của luận văn, phân tích các ưu và khuyết điểm của các mô hình được huấn luyện, từ đó đề xuất mô hình thích hợp cho bài toán dự đoán giá cổ phiếu và các hướng giải quyết khả thi có thể tiếp cận trong tương lai, ứng dụng tốt trong thực tiễn.

CHƯƠNG 2

CƠ SỞ LÝ THUYẾT VÀ CÁC NGHIÊN CỨU LIÊN QUAN

2.1. Những vấn đề cơ bản về chứng khoán

2.1.1. Khái niệm chứng khoán

Chứng khoán là các công cụ để huy động vốn trung và dài hạn, các giấy tờ có giá, có khả năng chuyển đổi, chuyển nhượng nhằm xác nhận quyền sở hữu, quan hệ vay nợ giữa người nắm giữ nó và chủ thể phát hành ra nó [22]. Đối với mỗi loại chứng khoán thường có các tính chất sau:

- Tính thanh khoản (tính lỏng) của một chứng khoán là khả năng chuyển đổi giữa chứng khoán đó sang tiền mặt. Tính lỏng của chứng khoán thể hiện qua việc chứng khoán đó được mua bán, trao đổi trên thị trường.

- Tính sinh lời: Thu nhập của nhà đầu tư được sinh ra từ việc tăng giá chứng khoán trên thị trường, hay các khoản tiền lãi được trả hàng năm.

- Tính rủi ro: Đây là đặc trưng cơ bản của chứng khoán. Trong quá trình trao đổi, mua đi bán lại, giá của chứng khoán bị giảm hoặc mất hoàn toàn ta gọi là rủi ro.

2.1.2. Phân loại chứng khoán

Cổ phiếu: Là loại chứng khoán xác nhận quyền sở hữu và lợi ích hợp pháp đối với thu nhập và tài sản của công ty cổ phần. Số vốn đóng góp để thành lập công ty được chia ra thành nhiều phần nhỏ bằng nhau gọi là cổ phần. Người mua cổ phần gọi là cổ đông. Cổ phiếu có thể được phát hành dưới dạng chứng chỉ vật chất hoặc bút toán ghi sổ, chỉ có công

ty cổ phần mới có cổ phiếu. Giá trị ban đầu ghi trên cổ phiếu là mệnh giá của cổ phiếu. Mệnh giá là giá trị danh nghĩa. Số tiền nhận được từ khoản góp vốn gọi là cổ tức. Giá cổ phiếu giao động qua các phiên giao dịch trên thị trường chứng khoán và tách rời so với mệnh giá. Cổ phiếu được chia thành 2 loại:

- Cổ phiếu thường: Là loại cổ phiếu không có kỳ hạn, tồn tại cùng với sự tồn tại của công ty phát hành ra nó, nó không có lãi suất cố định, số lãi được chia vào cuối mỗi niên độ quyết toán. Cổ đông nắm giữ cổ phiếu này có quyền bỏ phiếu, mua cổ phiếu mới, tham gia vào đại hội cổ đông.

- Cổ phiếu ưu đãi: Gồm 2 loại

- * Cổ phiếu ưu đãi biểu quyết: Là cổ phiếu dành cho các cổ đông sáng lập. Cổ đông nắm giữ loại cổ phiếu này phải nắm giữ trong một khoảng thời gian nhất định, không được chuyển nhượng, trao đổi.

- * Cổ phiếu ưu đãi tài chính: Tương tự như cổ phiếu thường nhưng có một số hạn chế: cổ đông nắm giữ cổ phiếu này không được tham gia bầu cử, ứng cử vào hội đồng quản trị, ban kiểm soát của công ty. Nhưng họ được hưởng ưu đãi về tài chính theo một mức cổ tức riêng biệt, có tính cố định hàng năm, được ưu tiên chia cổ tức và phân chia tài sản còn lại của công ty sau khi thanh lý, giải thể trước cổ phiếu thường.

Trái phiếu Là loại chứng khoán quy định nghĩa vụ của người phát hành phải trả cho người nắm giữ chứng khoán đó một khoản tiền xác định vào những thời hạn cụ thể và theo những điều kiện nhất định. Đây là những chứng khoán nợ, được phát hành dưới dạng chứng chỉ vật chất hoặc bút toán ghi sổ. Trái phiếu bao gồm các loại sau:

- Trái phiếu vô danh: Loại này không ghi tên trái chủ trên cả chứng chỉ và sổ sách của các tổ chức phát hành. Việc chuyển nhượng loại trái phiếu này rất dễ dàng nên nó thường được giao dịch trên thị trường chứng khoán.

- Trái phiếu ghi danh: Ghi tên, địa chỉ của trái chủ trên chứng chỉ và sổ sách của tổ chức phát hành. Loại này ít được đem trao đổi trên thị trường.

- Trái phiếu chính phủ: Là loại trái phiếu do chính phủ phát hành nhằm bù đắp cho thâm hụt ngân sách. Đây là loại chứng khoán mà các nhà đầu tư không ưa mạo hiểm rất ưa thích vì nó hầu như không có rủi ro thanh toán.

- Trái phiếu công trình: Là loại trái phiếu được phát hành để huy động vốn xây dựng các công trình cơ sở hạ tầng hay công trình phúc lợi công cộng.

- Trái phiếu công ty: Do công ty phát hành để vay vốn trung và dài hạn. Khi công ty bán trái phiếu thì công ty là người đi vay của người mua, người mua là chủ nợ. Công ty phải cam kết trả cả lãi và gốc cho trái chủ như đã nêu trong hợp đồng. Nó bao gồm các loại: Trái phiếu có đảm bảo, trái phiếu không có bảo đảm, trái phiếu có thể mua lại...

Chứng chỉ quỹ đầu tư Là chứng khoán được phát hành bởi công ty quản lý quỹ để huy động vốn từ các nhà đầu tư. Vốn được dùng để mua bán kinh doanh các loại chứng khoán khác để kiếm lời, sau đó chia tiền lời đó cho các nhà đầu tư. Có thể phân loại chứng chỉ quỹ đầu tư thành:

- Quỹ đầu tư chung: Đây là loại quỹ mà tất cả mọi người đều có thể tham gia.

- Quỹ đầu tư riêng: Chỉ giới hạn ở một nhóm người nhất định.

- Quỹ đầu tư dạng mở: Là quỹ đầu tư trong đó nhà đầu tư có quyền bán lại chứng chỉ quỹ đầu tư cho quỹ phát hành chứng khoán liên tục để huy động vốn và sẵn sàng mua lại số chứng khoán mà nó đã phát hành ra.

- Quỹ đầu tư cổ phiếu: Quỹ chuyên đầu tư vào một loại cổ phiếu nhất định.

- Quỹ đầu tư trái phiếu: Là quỹ để đầu tư vào một loại trái phiếu nào đó.

- Quỹ đầu tư hỗn hợp: Loại quỹ đầu tư vào bất kì loại chứng khoán nào nếu thấy có hiệu quả. Ngoài ra, cũng có một số loại chứng khoán khác (thường gọi là công cụ phái sinh) như: Giấy bảo đảm quyền mua cổ phiếu, các hợp đồng kì hạn, chứng quyền...

2.2. Tổng quan về thị trường chứng khoán

2.2.1. Khái niệm thị trường chứng khoán

Để thực hiện quá trình sản xuất, chúng ta cần phải có các yếu tố: lao động, vốn, đất đai, khoa học kỹ thuật. Trong đó vốn ngày càng đóng một vai trò quan trọng, nhưng để có một lượng vốn lớn thì không một cái nào có thể đảm nhiệm được. Trong khi đó những người có cơ hội đầu tư sinh lời thì thiếu vốn, còn những người có vốn thì lại ít có cơ hội đầu tư. Thực tế đó đòi hỏi cần phải có những tổ chức trung gian nối liền giữa người thừa vốn và những người thiếu vốn. Hệ thống ngân hàng ra đời đó đáp ứng được phần nào yêu cầu đó. Tuy nhiên, để thu hút được các nguồn tiền tệ nhàn rỗi của xã hội và một số nguồn vốn đầu tư trực tiếp vào sản xuất để kiếm lời tận gốc thì thị trường chứng khoán ra đời [24]. Thị trường chứng khoán là nơi giao dịch các loại chứng khoán theo các

mục đích khác nhau của các chủ thể tham gia vào thị trường chứng khoán.

2.2.2. Vai trò của thị trường chứng khoán:

Xu thế hội nhập kinh tế quốc tế và khu vực với sự ra đời của tổ chức thương mại thế giới (WTO), của liên minh Châu Âu, của khối thị trường chung đòi hỏi các quốc gia phải thúc đẩy phát triển kinh tế với tốc độ và hiệu quả cao. Thực tế phát triển kinh tế ở các quốc gia trên thế giới đã khẳng định vai trò của thị trường chứng khoán trong phát triển kinh tế.

Thứ nhất, thị trường chứng khoán với việc tạo ra các công cụ có tính thanh khoản cao, có thể tích tụ, tập trung và phân phối vốn, chuyển thời hạn của vốn phù hợp với yêu cầu phát triển kinh tế. Ở Hàn Quốc, thị trường chứng khoán đó đóng vai trò không nhỏ trong sự tăng trưởng kinh tế trong hơn ba mươi năm, cho tới thập kỷ 90 thị trường chứng khoán nước này đã lớn mạnh và đứng hàng thứ 13 trên thế giới góp phần tạo mức tăng trưởng kinh tế đạt bình quân 9% năm, thu nhập quốc dân bình quân đầu người năm 1995 là trên 10.000USD. Yếu tố thông tin là yếu tố cạnh tranh trên thị trường sẽ đảm bảo cho việc phân phối vốn một cách hiệu quả. Thị trường tài chính là nơi tiên phong áp dụng công nghệ mới và nhạy cảm với môi trường thường xuyên thay đổi. Trên thị trường chứng khoán, tất cả các thông tin được cập nhật và được chuyển tải tới tất cả các nhà đầu tư để họ có thể phân tích và định giá cho các chứng khoán. Thị trường chứng khoán đó tạo ra một sự cạnh tranh có hiệu quả trên thị trường tài chính, do đó các ngân hàng thương mại và các tổ chức tài chính phải quan tâm tới hoạt động tài chính của họ và làm giảm chi phí tài chính. Việc huy động vốn trên thị trường chứng khoán có thể làm tăng vốn tự có của các công ty và giúp họ tránh các khoản vay có chi phí

cao cũng như sự kiểm soát chặt chẽ của các ngân hàng thương mại. Thị trường chứng khoán khuyến khích tính cạnh tranh của các công ty trên thị trường. Sự tồn tại của các thị trường chứng khoán cũng là yếu tố quyết định để thu hút vốn đầu tư nước ngoài. Đây là yếu tố đảm bảo cho sự phân bổ có hiệu quả các nguồn lực trong một quốc gia cũng như trong phạm vi quốc tế.

Thứ hai, thị trường chứng khoán góp phần thực hiện tái phân phối công bằng hơn, giải toả sự tập trung quyền lực kinh tế của các tập đoàn, huy động tầng lớp trung lưu trong xã hội, tăng cường sự giám sát của xã hội đối với quá trình phân phối. Từ đó, tạo ra sự cạnh tranh công bằng hơn qua đó thúc đẩy phát triển kinh tế, tạo ra hiệu quả, tiến tới xây dựng một xã hội công bằng và dân chủ.

Thứ ba, thị trường chứng khoán tạo điều kiện cho việc tách biệt giữa sở hữu và quản lý doanh nghiệp. Khi quy mô của doanh nghiệp tăng lên thì nhu cầu về quản lý chuyên trách cũng tăng theo. Thị trường chứng khoán tạo điều kiện cho tiết kiệm vốn và chất xám thúc đẩy quá trình cổ phần hoá doanh nghiệp nhà nước, giảm tiêu cực trong quản lý, tạo điều kiện kết hợp hài hoà lợi ích của chủ sở hữu, nhà quản lý và những người làm công.

Thứ tư, hiệu quả của quốc tế hoá thị trường chứng khoán. Việc mở cửa thị trường chứng khoán làm tăng tính lỏng và cạnh tranh trên thị trường quốc tế, cho phép các công ty có thể huy động nguồn vốn rẻ hơn, tăng cường đầu tư từ bên ngoài, đồng thời tăng khả năng cạnh tranh quốc tế và mở rộng cơ hội kinh doanh cho các công ty trong nước. Hàn Quốc, Singapore, Thái Lan, Malaysia là những minh chứng điển hình về việc tận dụng các cơ hội do thị trường chứng khoán mang lại. Tuy nhiên,

chúng ta cũng cần phải xét các tác động tiêu cực có thể xảy ra như việc tăng cung tiền quá mức, áp lực của lạm phát, vấn đề chảy máu vốn.

Thứ năm, thị trường chứng khoán tạo điều kiện cho Chính phủ huy động các nguồn tài chính tiền tệ của chính phủ. Đồng thời, thị trường chứng khoán cũng cung cấp một dự báo tuyệt vời về các chu kỳ kinh doanh trong tương lai giúp Chính phủ và các công ty đánh giá kế hoạch đầu tư cũng như việc phân bổ các nguồn lực của họ. Thị trường chứng khoán cũng tạo điều kiện tái cấu trúc nền kinh tế. Ngoài những tác động tích cực trên, thị trường chứng khoán cũng có những tác động tiêu cực nhất định như hiện tượng đầu cơ, hiện tượng xung đột quyền lực, bong bóng giá,... làm giảm thiệt hại cho quyền lợi của các cổ đông thiểu số, việc mua bán nội gián, thao túng thị trường làm nản lòng các nhà đầu tư tác động tiêu cực tới tiết kiệm và đầu tư. Nhiệm vụ của các nhà quản lý thị trường là giảm thiểu các tiêu cực của thị trường nhằm bảo vệ quyền lợi của các nhà đầu tư và đảm bảo tính hiệu quả của thị trường.

Như vậy, vai trò của thị trường chứng khoán được thể hiện ở nhiều khía cạnh khác nhau. Song để vai trò đó thật sự phát huy được hiệu quả thì phụ thuộc đáng kể vào các chủ thể tham gia vào thị trường và sự quản lý của nhà nước.

2.2.3. Phân loại thị trường chứng khoán

Thị trường sơ cấp: Là nơi phát hành lần đầu chứng khoán ra công chúng, tạo vốn cho tổ chức phát hành thông qua việc bán chứng khoán cho các nhà đầu tư. Tại đây, có sự chuyển tài chính trực tiếp đưa các khoản tiền nhàn rỗi tạm thời trong dân chúng vào quá trình đầu tư, từ nguồn vốn ngắn hạn sang nguồn vốn dài hạn [23]. Nó tạo điều kiện cung cấp cho nền kinh tế nguồn vốn có thời gian sử dụng đáp ứng được thời gian cần thiết của các khoản đầu tư. Thị trường sơ cấp có chức năng là

huy động vốn cho đầu tư, làm tăng thêm vốn đầu tư cho nền kinh tế. Có hai phương pháp phát hành chứng khoán trên thị trường này là: phát hành riêng lẻ và phát hành công chúng.

Thị trường thứ cấp: Là nơi diễn ra các giao dịch chứng khoán đã được phát hành trên thị trường sơ cấp. Thị trường thứ cấp tạo điều kiện cho thị trường sơ cấp phát triển, tăng tiềm năng huy động vốn của các tổ chức phát hành chứng khoán. Trên thị trường chứng khoán vừa có giao dịch của thị trường sơ cấp, vừa có giao dịch của thị trường thứ cấp.

Thị trường tập trung: Là nơi mà việc giao dịch, trao đổi, mua bán chứng khoán được thực hiện thông qua sổ giao dịch chứng khoán(hay còn gọi là sàn giao dịch).

Thị trường phi tập trung: Hay còn gọi là thị trường OTC, thị trường này diễn ra ở bất kỳ nơi nào miễn là nơi đó diễn ra các hoạt động mua bán, trao đổi chứng khoán.



Thị trường giao dịch tập trung
(Sổ giao dịch chứng khoán)

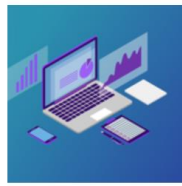


Thị trường phi tập trung
(Thị trường OTC)

Hình 2. 1: Mô hình phân loại thị trường chứng khoán

Thị trường trái phiếu (bond market): Là thị trường phát hành trái phiếu mới và mua đi bán lại trái phiếu cũ.

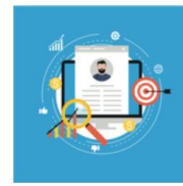
Thị trường cổ phiếu (the stock market): Là thị trường phát hành cổ phiếu mới và mua đi bán lại cổ phiếu cũ. Hoạt động của thị trường này gồm cả hai lĩnh vực đầu tư chứng khoán và kinh doanh chứng khoán.



Thị trường cổ phiếu



Thị trường trái phiếu



Thị trường các công cụ
chứng khoán phái sinh

Hình 2. 2: Mô hình về các loại thị trường cổ phiếu

Thị trường giao ngay: Là thị trường mà việc mua bán chứng khoán theo giá thoả thuận khi ký hợp đồng nhưng thanh toán và giao chứng khoán sau ngày ký hợp đồng một khoảng thời gian quy định. Chức năng của thị trường này là huy động và làm tăng nguồn vốn.

Thị trường giao dịch kỳ hạn: Thị trường mua và bán chứng khoán theo giá thoả thuận theo hợp đồng nhưng thanh toán và giao nhận chứng khoán sau ngày ký kết hợp đồng một khoảng thời gian nhất định thường là 30 hoặc 60 ngày.

Thị trường giao dịch tương lai: Giống như là thị trường giao dịch chứng khoán kỳ hạn nhưng khác ở chỗ, các hợp đồng mua bán đó được tiêu chuẩn hoá nội dung và điều kiện thực hiện hợp đồng và phải có tiền đặt cọc thông qua quy chế ký quỹ để đảm bảo hợp đồng.

2.3. Học máy (Machine learning) và Thị trường chứng khoán:

Cách mạng Công nghiệp 4.0 mang đến những thay đổi bước ngoặt, toàn diện trong mọi mặt của cuộc sống. “*Thế giới vận hành bằng công nghệ*” trở thành xu thế tất yếu, là mục tiêu tranh đua cho mọi nền kinh tế, mọi doanh nghiệp cùng hướng đến. Trong cuộc cách mạng Công nghiệp lần thứ IV, Trí tuệ nhân tạo (A.I) và Học máy học (Machine Learning) [25] là những thành phần quan trọng, nổi bật, đã và đang được nghiên cứu, ứng dụng rộng rãi trong nhiều lĩnh vực khác nhau trong đó có ngành ngân hàng nói chung và lĩnh vực công nghệ tài chính (FinTech)

nói riêng, tạo ra những bứt phá mạnh mẽ về tính hiệu quả, kịp thời, chính xác trong công tác giám sát, dự báo và hỗ trợ ra quyết định.

Học máy (tiếng Anh: *machine learning*) là một lĩnh vực của trí tuệ nhân tạo liên quan đến việc nghiên cứu và xây dựng các kỹ thuật cho phép các hệ thống "học" tự động từ dữ liệu để giải quyết những vấn đề cụ thể. Ví dụ như các máy có thể "học" cách phân loại thư điện tử xem có phải thư rác (spam) hay không và tự động xếp thư vào thư mục tương ứng. Học máy rất gần với suy diễn thống kê (statistical inference) tuy có khác nhau về thuật ngữ.

Học máy có liên quan lớn đến thống kê, vì cả hai lĩnh vực đều nghiên cứu việc phân tích dữ liệu, nhưng khác với thống kê, học máy tập trung vào sự phức tạp của các giải thuật trong việc thực thi tính toán. Nhiều bài toán suy luận được xếp vào loại bài toán NP-khó, vì thế một phần của học máy là nghiên cứu sự phát triển các giải thuật suy luận xấp xỉ mà có thể xử lý được.

Học máy hiện nay được áp dụng rộng rãi bao gồm máy truy tìm dữ liệu, chẩn đoán y khoa, phát hiện thẻ tín dụng giả, phân tích thị trường chứng khoán, phân loại các chuỗi DNA, nhận dạng tiếng nói và chữ viết, dịch tự động, chơi trò chơi và cử động rô-bốt (*robot locomotion*).

Trên thế giới, Machine Learning hiện được áp dụng rộng rãi trong nhiều lĩnh vực, bao gồm: Khai thác dữ liệu, chẩn đoán y khoa, phát hiện thẻ tín dụng giả, phân tích thị trường chứng khoán, phân loại các chuỗi DNA, nhận dạng tiếng nói và chữ viết, dịch tự động, chơi trò chơi và điều hướng rô-bốt (robot locomotion). Tất cả các ứng dụng trên có một điểm chung là sử dụng “bộ não logic” được cấu thành từ các thuật toán Machine Learning, tiếp nhận dữ liệu đầu vào đã được số hóa và thực

hiện xử lý, phân tích qua nhiều lớp, với độ phức tạp và mức độ “thông minh” ngày càng tăng (deep learning).

Thực tế đã chứng minh, nhờ áp dụng chính xác phương pháp Machine Learning trong vận hành các hệ thống trí tuệ nhân tạo, các tập đoàn và công ty lớn trên thế giới như Facebook, Amazon, Google... đều đạt được những bước phát triển thần kỳ, chỉ trong vòng dưới 10 năm.

Trong lĩnh vực tài chính, ngân hàng, TTCK, Machine Learning, khi được kết hợp với các mô hình phân tích định lượng, phát huy hiệu quả đặc biệt trong việc tìm kiếm các bộ mẫu dữ liệu, đưa ra những dự đoán, hỗ trợ hiệu quả ra quyết định giúp đảm bảo hoạt động kinh doanh liên tục và kiểm soát rủi ro. Trên thế giới, cuộc chạy đua trong ngành Ngân hàng cũng như các thị trường chứng khoán diễn ra đặc biệt sôi động. Từ các công ty công nghệ mới thành lập như Feedzai (trong mảng thanh toán), Shift Technology (trong mảng bảo hiểm), tới các tập đoàn công nghệ khổng lồ như IBM và nhóm dẫn đầu về công nghệ hiện tại như Google, Alibaba và các Fintech, đang dựa vào ưu thế công nghệ để cạnh tranh, lấn sân sang lĩnh vực ngân hàng, tài chính, cổ phiếu thị trường.

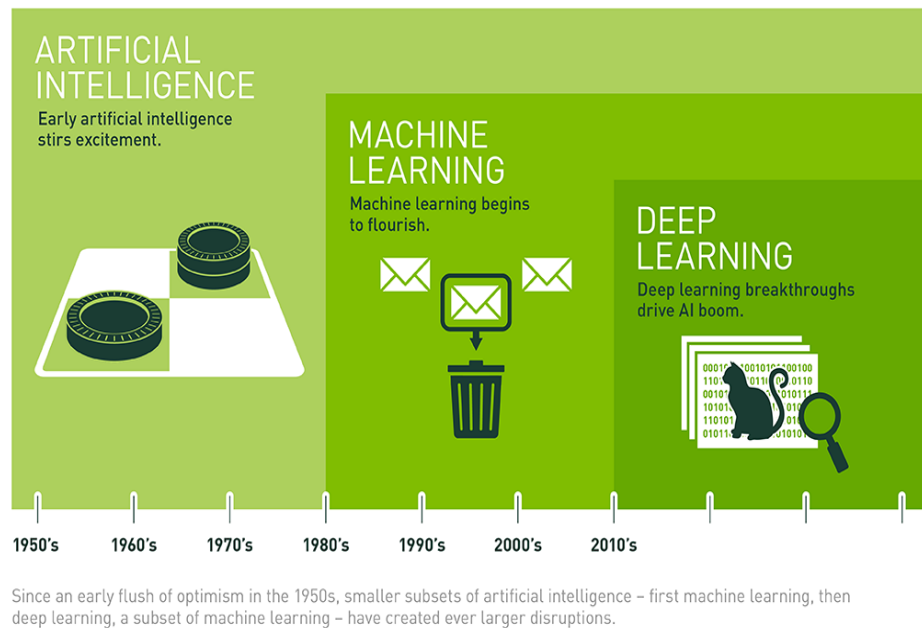
Rất nhiều ví dụ về ứng dụng hiệu quả Machine Learning trong lĩnh vực tài chính, ngân hàng, TTCK có thể kể đến như: Monzo - một ngân hàng khởi nghiệp tại Anh, đã xây dựng một mô hình phân tích, dự báo đủ nhanh, để kịp thời phát hiện và ngăn chặn những kẻ lừa đảo giả mạo trong quá trình hoàn tất giao dịch, giúp giảm tỷ lệ lừa đảo trên thẻ trả trước từ 0,85% vào tháng 6/2016 xuống dưới 0,1% vào tháng 1/2017 [26]. Các công ty công nghệ khác như Xcelerit hay Kinetica, cung cấp cho các ngân hàng và công ty đầu tư, một hệ thống có khả năng theo dõi

và phát hiện các rủi ro tiềm ẩn theo thời gian thực, cho phép ngân hàng giám sát chặt chẽ các yêu cầu về nguồn vốn. Trong năm 2017, JPMorgan Chase giới thiệu COiN, một nền tảng quản lý hợp đồng thông minh, sử dụng Machine Learning, có khả năng xem xét 12.000 hợp đồng tín dụng thương mại trong vài giây, tương đương khối lượng công việc trong 360.000 giờ làm việc của một nhân viên bình thường.

Trong TTCK, với khả năng xử lý dữ liệu mạnh mẽ trên nhiều lĩnh vực, học máy cũng được sử dụng một cách rộng rãi trong lĩnh vực tài chính như: dự đoán thị trường cổ phiếu, đầu tư tối ưu, xử lý thông tin tài chính và thực hiện các chiến lược giao dịch tài chính. Do đó, TTCK được xem là một trong những lĩnh vực khá phổ biến và quý giá nhất trong lĩnh vực tài chính trong quá trình nghiên cứu và phát triển của học máy.

2.4. Học sâu - Deep learning

Những năm gần đây, khi mà khả năng tính toán của máy tính được nâng lên một tầm cao mới cùng với lượng dữ liệu khổng lồ được thu thập, Machine Learning đã tiến thêm một bước dài và Deep Learning [27] (DL) một lĩnh vực mới được ra đời. Deep Learning được lấy cảm hứng từ mạng nơ-ron sinh học và bao gồm nhiều lớp trong mạng nơ-ron nhân tạo được tạo thành từ phần cứng và GPU. Deep Learning sử dụng một tầng các lớp đơn vị xử lý phi tuyến để trích xuất hoặc chuyển đổi các tính năng (hoặc biểu diễn) của dữ liệu. Đầu ra của một lớp phục vụ như là đầu vào của lớp kế tiếp. Deep learning tập trung giải quyết các vấn đề liên quan đến mạng thần kinh nhân tạo nhằm nâng cấp các công nghệ như nhận diện giọng nói, dịch tự động (machine translation), xử lý ngôn ngữ tự nhiên...



Hình 2. 3: Lược sử học sâu Deep Learning¹

Trong số các thuật toán học máy hiện đang được sử dụng và phát triển, học sâu thu hút được nhiều dữ liệu nhất và có thể đánh bại con người trong một số nhiệm vụ nhận thức. Do những thuộc tính này, học tập sâu đã trở thành phương pháp tiếp cận có tiềm năng đáng kể trong lĩnh vực trí tuệ nhân tạo.

Học sâu cũng như Học máy có thể chia thành 4 nhóm chính:

- Học sâu không giám sát
- Học sâu có giám sát
- Học sâu bán giám sát
- Học sâu tăng cường

¹<https://nordiccoder.com/blog/su-khac-nhau-giua-ai-machine-learning-va-deep-learning/>

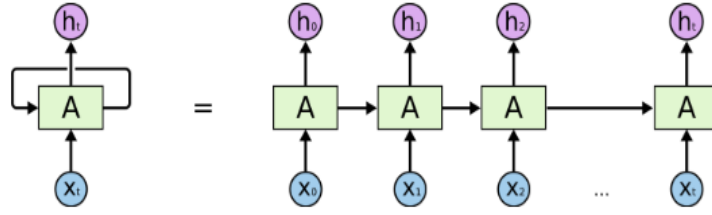
Trong phần tiếp theo chúng tôi sẽ trình bày các mô hình học sâu được nghiên cứu trong luận văn: Học sâu có giám sát (LSTM, Bi-LSTM) và Học sâu không giám sát GAN.

2.4.1. Mạng nơ-ron hồi quy RNN (Recurrent Neural Network)

Con người thường không ai suy nghĩ mọi thứ lại từ đầu hết. Ví dụ khi đọc một bài luận văn như thế này, ta sẽ đọc và hiểu từng chữ dựa theo những chữ đã biết ở trước, chứ không phải đọc từng chữ riêng lẻ và bỏ hết tất cả ý nghĩa các câu đã hiểu được. Nói cách khác, một mạng nơ-ron muốn giải được suy nghĩ của con người, thì cũng phải “nhớ” được các thông tin mà nó đã và đang đọc, từ đó suy ra mối quan hệ của những thứ sau này dựa trên trí nhớ ấy.

Ví dụ như ta cần phân biệt các sự kiện trong một đoạn phim, dựa trên sự kiện trước đó đã chiếu của đoạn phim ấy. Lúc này tín hiệu đầu vào có thể là toàn bộ chi tiết trên bộ phim (một vector tín hiệu khổng lồ), và cho vào mạng nơ-ron nhân tạo cơ bản ANN (mạng nơ-ron nhân tạo) để thử giải quyết. Tuy nhiên mạng hồi quy sẽ không làm vậy, mà thay vào đó chia nhỏ bộ phim ra từng thời điểm khác nhau (ví như là từng khung hình). Lúc đó, tín hiệu đầu vào của mạng hồi quy sẽ là một loạt các vector \mathbf{x}_t , ứng với tín hiệu đầu vào của mỗi khung hình t từ đầu tới cuối phim. Khi ấy, mạng hồi quy sẽ sử dụng thông tin đã dự đoán ở các khung hình trước, để dự đoán tiếp các khung hình sau.

Về bản chất, nó là mạng nơ-ron nhân tạo như bình thường, nhưng dữ liệu đầu ra hiện tại ở một thời điểm bất kỳ của nó được giữ lại bằng cách truyền vòng ngược lại, giúp cho thông tin được giữ từ điểm thời gian này tới điểm thời gian khác của chung một điểm dữ liệu [16].



Hình 2. 4: Mô hình mạng hồi quy đơn giản

Trong mô hình trên mũi tên quay ngược chính là vòng lặp. Bên trái là hình rút gọn của mạng (thể hiện vòng lặp). Bên phải là mạng sau khi được bung ra. Điểm đáng lưu ý như đã nói ở trên, đó là mạng sẽ nhận vào không còn đơn giản là một vector tín hiệu x , mà là một loạt các vector tín hiệu x_1, x_2, \dots, x_t tương ứng với tín hiệu đầu vào của điểm dữ liệu tại thời điểm $t = 1 \rightarrow t$ ở thời điểm hiện tại. Hơn nữa, mạng cũng sẽ cho ra các kết quả ẩn h_t , tương đương với vector tín hiệu đầu ra tại thời điểm t ấy.

Mạng hồi quy A ở trên hình sẽ nhận vào tín hiệu đầu vào x_t kết hợp với kết quả ẩn h_{t-1} . Hai tín hiệu khác nhau này sẽ được gộp lại tùy theo ma trận tham số tương ứng của chúng, để tính được kết quả ẩn h_t . Cứ tiếp tục như vậy cho đến khi t kết thúc (tức là đến khi điểm dữ liệu này đã đọc hết một cách tuần tự từ đầu tới cuối).

Công thức cơ bản của mạng hồi quy để tính các vector tín hiệu ẩn h_t là:

$$h_t = \varphi (WX_t + Uh_{t-1}) \quad (2.1)$$

Ma trận W sẽ tương đương ma trận trọng số như mạng nơ-ron nhân tạo cơ bản. Ma trận U thể hiện trọng số để tham chiếu các tín hiệu kết quả cũ. Cuối cùng hàm φ sẽ dùng để “ép” mỗi giá trị của vector h_t về một không gian giá trị log (như là từ $-1 \rightarrow 1$, và có thể tính đạo hàm

được). Nó tương đương với các hàm kích hoạt như nói ở phần giải thuật lan truyền ngược. Thông thường ta hay dùng hàm tanh để giải quyết.

Vấn đề về sự phụ thuộc tới thông tin quá lâu trong quá khứ

Với ý tưởng và công thức trên, mạng hồi quy sẽ gặp một số vấn đề khó giải quyết. Điển hình là vấn đề “mức độ phụ thuộc bị mất” (vanishing gradient problem) [14].

Thử xét phần tính lan truyền ngược dựa trên ý tưởng ở phần trên bao gồm một chuỗi các tích các mức độ phụ thuộc. Điều này có nghĩa nếu các mức độ phụ thuộc có giá trị tuyệt đối nhỏ hơn 1, thì ta có:

$$\lim_{t \rightarrow \infty} \left(\frac{\partial E_t}{\partial W} \right) = 0 \quad (2.2)$$

Trong đó:

E là sai số bình phương.

t là thời điểm đầu ra cho huấn luyện mẫu.

w: ma trận trọng số.

Nói cách khác, với thời gian cần nhớ thông tin đủ dài, thì đến một mức độ nào đó, đột nhiên mạng sẽ không thể học được nữa. Điều này giống với việc mạng nơ-ron học quá nhiều và bị bão hòa.

Có thể xét ví dụ đơn giản ngoài thực tế như khi nói rằng “Chim bay trên ...”, thì thông tin từ 3 từ ấy có thể giúp ta đoán từ tiếp theo như là “trời”. Do khoảng cách giữa thông tin quá khứ cần biết “chim” nó rất gần với thông tin tương lai, ta có thể đoán “trời”, cho nên mạng hồi quy có thể học được. Tuy nhiên với ví dụ thứ hai là “Tôi sống ở Việt Nam từ nhỏ, cho nên tôi có thể nói tiếng Việt”. Ta thấy khoảng cách đang rất là xa nhau, cho nên để có thể dự đoán chính xác, thì mạng phải nhớ được những thông tin rất xa như vậy.

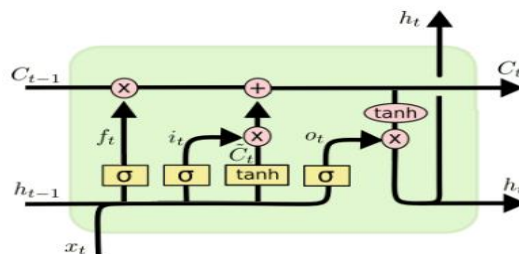
Hiển nhiên theo lý thuyết, ta vẫn có thể giải quyết chúng, nếu ta mở rộng số lượng thông tin cần phải nhớ, và mở rộng hệ số tốc độ học η cho khớp với bài toán. Tuy nhiên trên thực nghiệm, mạng hồi quy vẫn sẽ không học được thông tin quá xa, và cái lý do chủ yếu là do cái vấn đề “mức độ phụ thuộc bị mất” như trên.

Khó khăn của mạng RNN là với thông tin quá khứ “quan trọng” cần nhớ ở x_0, x_1 thì nếu khoảng cách t đủ lâu, mạng sẽ khó sử dụng chúng để dự đoán h_{t+1} , do từ $1 \rightarrow t$ có thể đã quá xa, khiến cho mức độ phụ thuộc tính được gần thành 0, dẫn tới toàn bộ thời điểm $t + 1$ trở đi, mạng sẽ không còn học được gì nữa.

2.4.2. Mạng nơ-ron ngắn dài LSTM (Long-short term memory)

Là phương án để giải quyết vấn đề phụ thuộc thông tin quá khứ xa, vì lý do mạng hồi quy không học được thông tin quá xa, nên mạng nơ-ron trí nhớ ngắn dài (gọi tắt LSTM) được tạo ra. Nó được giới thiệu bởi Hochreiter và Schmidhuber (1997) [15], và được nhiều người khác cải tiến tiếp.

Thay vì một hệ thống lặp lại chỉ chứa một tầng \tanh đơn giản như của mạng hồi quy bình thường, mạng LSTM chứa tới 4 tầng khác nhau trong hệ thống lặp lại của nó, mô tả đơn giản như hình sau:



Hình 2. 5: Mô hình của một tế bào LSTM

Một tế bào LSTM gồm 4 tầng khác nhau. Xét tại thời điểm t , h_t thể hiện kết quả X_t ẩn, thể hiện tín hiệu vào của dữ liệu, C_t là đầu ra của mạng LSTM. Với sự kết hợp khác nhau của các tầng (ở đây là các cổng) như trên, mạng LSTM khi lan truyền tiến (để dự đoán kết quả) sẽ phức tạp hơn một chút so với mạng nơ-ron nhân tạo cơ bản.

Ý tưởng chính của mạng LSTM đó là: Với mỗi thời điểm t , ta sẽ có một trạng thái của ô LSTM (1 cell) là tương ứng. Thể hiện trên hình đó là đường thẳng chạy ngang từ C_{t-1} tới C_t , ứng với việc ta sẽ truyền kết quả từ trạng thái trước đến trạng thái sau. Tuy nhiên điều đó không có nghĩa là toàn bộ thông tin đều đi mà không bị gì cả. Tương tác với các giá trị C_{t-1} , ta sẽ có các cổng (như hình có các hàm kích hoạt sigmoid với kí hiệu σ , và hàm tanh) và các phép toán trên ma trận ($\times, +$). Danh sách các cổng như sau

Cổng quên (forget gate - f_t):

Dùng để quyết định xem có bao nhiêu thông tin cũ của C_{t-1} được giữ lại. Cổng này sẽ sử dụng giá trị kết quả ẩn trước đó (h_{t-1}), và tín hiệu đầu vào hiện tại x_t và gom lại với nhau, kết quả đó sẽ được “kích hoạt lại” thông qua hàm sigmoid của cổng. Giá trị mới này sẽ từ $0 \rightarrow 1$, thể hiện mức độ từ “quên hết” tới “nhớ hết”:

$$f_t = \sigma(W_{fx_t} + U_f h_{t-1} + b_f) \quad (2.3)$$

Cổng đầu vào (input gate - i_t):

Dùng để quyết định có bao nhiêu thông tin mới nhập vào \widetilde{C}_t . Thông tin mới ấy được tạo ra dựa trên kết quả ẩn trước đó (h_{t-1}), và tín hiệu đầu vào hiện tại x_t , và được kích hoạt qua hàm \tanh , tạo ra các giá trị từ $-1 \rightarrow 1$. Giá trị thực tế của cổng đầu vào cũng sẽ được kích hoạt

tương tự cổng quên ở trên, thông qua hàm sigmoid, thể hiện mức độ “bỏ hết” tới “lấy hết” của thông tin mới:

$$f_t = \sigma(W_x x_t + W_h h_{t-1} + W_c c_{t-1} + b_f) \quad (2.4)$$

$$i_t = \sigma(W_x x_t + U_h h_{t-1} + b_i) \quad (2.5)$$

$$\widetilde{C}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (2.6)$$

Tới đây ta cần kết hợp cổng quên và cổng đầu vào, để tạo ra trạng thái C_t thực sự:

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \quad (2.7)$$

Như đã nói, cổng quên sẽ quyết định thông tin cũ có được giữ hay không, và cổng đầu vào sẽ quyết định thông tin mới có được sử dụng hay không.

Cổng kết quả (output gate - o_t):

Trạng thái C_t mới chắc chắn sẽ được truyền để xét tiếp tại thời điểm $t + 1$ tiếp theo. Tuy nhiên, việc ta có cho phép kết quả ở bước hiện tại h_t hay không thì lại phải phụ thuộc tiếp cổng ra. Tương tự hai cổng trước, ta cũng dùng hàm sigmoid để ép giá trị kết hợp của h_{t-1} và x_t để quyết định điều này. Giá trị của trạng thái hiện tại C_t cũng được ép trong khoảng $-1 \rightarrow 1$ bằng \tanh trước khi nhân với lại cổng ra:

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (2.8)$$

$$h_t = o_t * \tanh(C_t) \quad (2.9)$$

Mạng LSTM sau khi truyền tiến như trên là đã có thể dự đoán kết quả của một điểm dữ liệu nào đó. Ta có thể áp dụng kỹ thuật lan truyền ngược để cập nhật trọng số cho phù hợp tương tự như các phần trên, bằng cách tính đạo hàm từng phần ngược lại từ đầu ra tới đầu vào.

2.4.3. Mạng nơ-ron tái phát hai chiều BiLSTM (Bidirectional Long-short term memory)

Một mạng lưới nơ-ron tái phát (RNN) là một lớp của mạng nơ-ron nhân tạo nơi kết nối giữa các nút, tạo thành một đồ thị có hướng dọc theo một trình tự thời gian. Điều này cho phép nó thể hiện hành động tạm thời. Bắt nguồn từ mạng nơ-ron truyền thẳng, RNN có thể sử dụng trạng thái bên trong (bộ nhớ) của chúng để xử lý chuỗi đầu vào có độ dài thay đổi. Điều này làm cho chúng có thể áp dụng cho các tác vụ như nhận dạng chữ viết tay được kết nối, không phân đoạn hoặc nhận dạng giọng nói.

Với Mạng nơ-ron tái phát hai chiều (BiLSTM) là sự kết nối hai lớp ẩn có hướng ngược lại với cùng một đầu ra. Với hình thức học sâu chung này, lớp đầu ra có thể lấy thông tin từ trạng thái quá khứ (ngược) và tương lai (chuyển tiếp) đồng thời. Được phát minh vào năm 1997 bởi Schuster và Paliwal, BiLSTM đã được giới thiệu để tăng lượng thông tin đầu vào có sẵn cho mạng. Ví dụ: perceptron nhiều lớp (MLP) và mạng nơ-ron trễ thời gian (TDNN) có những hạn chế về tính linh hoạt của dữ liệu đầu vào, vì chúng yêu cầu dữ liệu đầu vào của chúng phải được sửa. Mạng nơ-ron tái phát tiêu chuẩn (RNN) cũng có những hạn chế vì không thể truy cập thông tin đầu vào trong tương lai từ trạng thái hiện tại. Ngược lại, BiLSTM không yêu cầu cố định dữ liệu đầu vào của chúng. Hơn nữa, thông tin đầu vào trong tương lai của họ có thể truy cập được từ trạng thái hiện tại.

Cho một bước thời gian t , đầu vào minibatch là $X_t \in R^{n \times d}$ (n là số lượng mẫu, d là số lượng đầu vào) và hàm kích hoạt của tầng ẩn là ϕ . Trong kiến trúc hai chiều, ta giả định rằng trạng thái ẩn xuôi và ngược của bước thời gian này lần lượt $\vec{H}_t \in R^{n \times h}$ và $\overleftarrow{H}_t \in R^{n \times h}$. h ở đây chỉ số

lượng nút ẩn. Chúng ta tính toán việc cập nhật xuôi và ngược của trạng thái ẩn như sau:

$$\vec{H}_t = \phi(X_t W_{xh}^{(f)} + \vec{H}_{t-1} W_{hh}^{(f)} + b_h^{(f)}), \quad (2.10)$$

$$\vec{H}_t = \phi(X_t W_{xh}^{(b)} + \vec{H}_{t+1} W_{hh}^{(b)} + b_h^{(b)}), \quad (2.11)$$

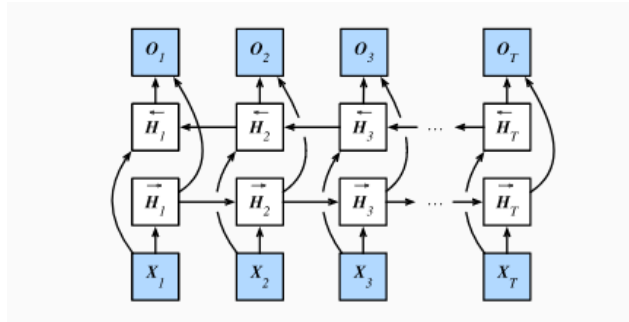
Ở đây, các trọng số $W_{xh}^{(f)} \in R^{dxh}$, $W_{hh}^{(f)} \in R^{h \times h}$, $W_{xh}^{(b)} \in R^{dxh}$ và $W_{hh}^{(b)} \in R^{h \times h}$ và các độ chệch $b_h^{(f)} \in R^{1 \times h}$ và $b_h^{(b)} \in R^{1 \times h}$ đều là tham số mô hình.

Sau đó, chúng ta nối các trạng thái ẩn xuôi và ngược (\vec{H}_t, \vec{H}_t) để thu được trạng thái ẩn $H_t \in R^{n \times 2h}$ và truyền nó đến tầng đầu ra. Trong các mạng nơ-ron hồi tiếp hai chiều sâu, thông tin được truyền đi như là đầu vào cho tầng hai chiều (bidirectional layer) tiếp theo. Cuối cùng, tầng đầu ra sẽ tính toán đầu ra $H_t \in R^{n \times 2h}$ (q là số lượng đầu ra) như sau:

$$O_t = H_t W_{hq} + b_q \quad (2.12)$$

Ở đây, trọng số $W_{hq} \in R^{2h \times q}$ và độ chệch $b_q \in R^{1 \times q}$ là các tham số mô hình của tầng đầu ra. Hai chiều ngược và xuôi có thể có số nút ẩn khác nhau.

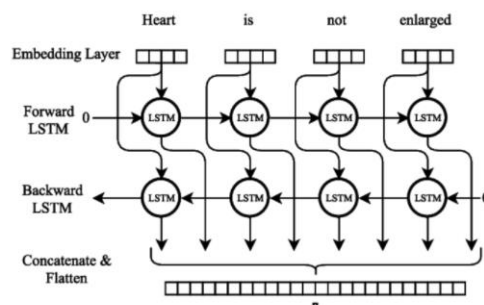
Mạng nơ-ron hồi tiếp hai chiều (Bidirectional recurrent neural network) sẽ thêm một tầng ẩn cho phép xử lý dữ liệu theo chiều ngược lại một cách linh hoạt hơn so với RNN truyền thống. Hình 2.11 mô tả cấu trúc của mạng nơ-ron hồi tiếp hai chiều với một tầng ẩn.



Hình 2. 6: Cấu trúc của mạng nơ-ron hồi tiếp hai chiều

Trên thực tế, điều này không quá khác biệt với phép đệ quy xuôi và ngược mà ta đã đề cập ở phần trước. Điểm khác biệt chính là trước đây các phương trình này có một ý nghĩa thống kê nhất định. Còn bây giờ thì chúng không còn mang một ý nghĩa dễ hiểu nào nhất định, thay vào đó ta sẽ chỉ xét chúng như những hàm tổng quát. Quá trình chuyển đổi này là điển hình cho nhiều nguyên tắc thiết kế các mạng học sâu hiện đại: đầu tiên, sử dụng các dạng quan hệ phụ thuộc hàm của các mô hình thống kê cổ điển, sau đó sử dụng các mô hình này dưới dạng tổng quát.

Nguyên tắc của BiLSTM là chia các nơ-ron của một RNN thông thường thành hai hướng, một hướng thời gian dương (trạng thái thuận) và hướng khác cho hướng thời gian âm (trạng thái lùi). Đầu ra của hai trạng thái đó không được kết nối với đầu vào của các trạng thái có hướng ngược lại. Cấu trúc chung của RNN và BiLSTM có thể được mô tả trong sơ đồ bên phải. Bằng cách sử dụng hai hướng thời gian, thông tin đầu vào từ quá khứ và tương lai của khung thời gian hiện tại có thể được sử dụng không giống như RNN tiêu chuẩn đòi hỏi sự chậm trễ để bao gồm thông tin trong tương lai.



Hình 2. 7: Bidirectional LSTM = forward LSTM + backward LSTM

(Nguồn: <https://paperswithcode.com>)

BiLSTM có thể được huấn luyện bằng cách sử dụng các thuật toán tương tự như RNN, vì hai nơ-ron định hướng không có bất kỳ tương tác nào. Tuy nhiên, khi áp dụng truyền ngược, các quy trình bổ sung là cần thiết vì không thể thực hiện cập nhật các lớp đầu vào và đầu ra cùng một lúc. Các thủ tục chung để huấn luyện như sau: Đối với chuyển tiếp, trạng thái chuyển tiếp và trạng thái lùi được chuyển trước, sau đó chuyển tiếp các nơ-ron đầu ra. Đối với truyền ngược, các nơ-ron đầu ra được chuyển trước, sau đó các trạng thái tiến và trạng thái lùi được chuyển tiếp theo. Sau khi thực hiện các đường chuyển tiến và lùi, trọng số được cập nhật.

2.4.4. Mạng sáng tạo đối nghịch GAN

Mạng sáng tạo đối nghịch (Generative Adversarial Network) với tên ngắn phổ biến GAN là một lớp các mô hình sinh mẫu sử dụng các kiến trúc học sâu. Mô hình này được đề xuất bởi Ian Goodfellow vào năm 2014 trong bài báo: Goodfellow, Ian; Pouget-Abadie, Jean; Mirza, Mehdi; Xu, Bing; Warde-Farley, David; Ozair, Sherjil; Courville, Aaron; Bengio, Yoshua (2014). Generative Adversarial Networks. Proceedings

of the International Conference on Neural Information Processing Systems (NIPS 2014). pp. 2672–2680.[12]

Mô hình sinh là một bài toán trong học máy liên quan đến việc tự động phát hiện và học các mẫu phổ biến trong dữ liệu để từ đó, mô hình có thể sáng tạo ra các ví dụ mới phù hợp với dữ liệu gốc.

GAN là một cách tiếp cận thông minh để huấn luyện mô hình sinh bằng cách nhìn nhận bài toán sinh như một bài toán học có giám sát với hai thành phần: mô hình sinh được huấn luyện để sáng tạo ví dụ mới, và một mô hình đánh giá để phân biệt các ví dụ là một thực thể thật hay một thực thể giả (được sáng tạo ra). Hai mô hình được huấn luyện song song, đối nghịch, cho đến khi mô hình phân biệt bị đánh lừa trên hơn 50% số mẫu, tức là mô hình sinh thành công trong việc tạo ra dữ liệu giả chất lượng cao.

GAN đã làm thay đổi bức tranh công nghệ một cách nhanh chóng và đầy cảm hứng. Nó cho phép các mô hình tạo ra các ví dụ thực tế trong nhiều vấn đề khác nhau.

- Học có giám sát và không giám sát: Bài toán học máy điển hình thường liên quan đến việc sử dụng một mô hình để đưa ra dự đoán. Do đó thành phần không thể thiếu là tập dữ liệu huấn luyện, bao gồm nhiều ví dụ, với biến đầu vào (X) và kết quả đầu ra (y). Hiểu đơn giản nhất, một mô hình được huấn luyện bằng cách dựa vào các biến đầu vào, dự đoán kết quả đầu ra và được sửa lại với mỗi lần đoán sai. Sau nhiều lần huấn luyện, khả năng dự đoán của mô hình dần khớp với dữ liệu huấn luyện. Phương pháp này gọi là học có giám sát.

Một biến thể khác là học không giám sát (còn được gọi là học mô tả), trong các bài toán này, dữ liệu chỉ có các biến đầu vào mà không có

kết quả đầu ra. Mô hình được xây dựng bằng cách trích xuất và tổng hợp các mẫu tồn tại trong dữ liệu. Do đó không có mô hình chính xác vì mô hình trong các bài toán này không có chức năng dự đoán. Cùng với các bài toán khác như phân cụm, **GAN** thuộc lớp các bài toán học không giám sát.

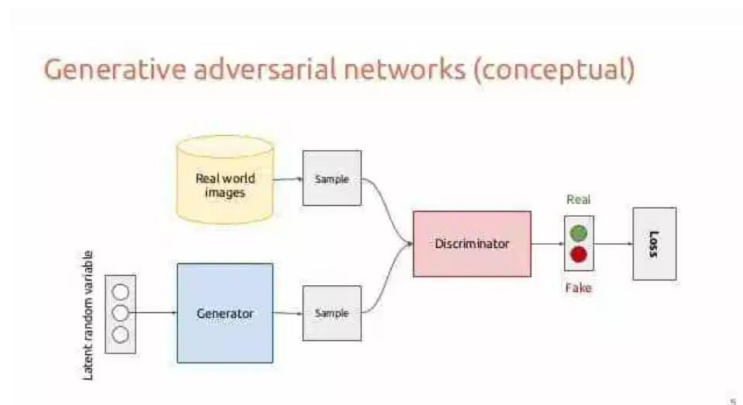
Kiến trúc của **GAN** bao gồm hai thành phần nhỏ là một mô hình sinh để sinh ra ví dụ mới và một mô hình phân biệt để xác định các thực thể là giả (được tạo ra bởi mô hình sinh) hay là một thực thể thật.

- **Mô hình sinh.** Mô hình sinh ra ví dụ mới từ dữ liệu đầu vào.

- **Mô hình phân biệt.** Mô hình xác định thật giả.

Mô hình sinh nhận vào một véc tơ ngẫu nhiên có độ dài cố định và sinh ra một thực thể giả trong miền dữ liệu. Véc tơ này được sinh ra ngẫu nhiên từ phân bố Gauss (là một phân phối xác suất cực kì quan trọng trong nhiều lĩnh vực.... Nó là họ phân phối có dạng tổng quát giống nhau, chỉ khác tham số vị trí (giá trị trung bình μ) và tỉ lệ (phương sai σ^2) và được sử dụng để khởi tạo quá trình sinh. Sau khi được huấn luyện, các điểm trong không gian véc tơ đa chiều này sẽ phù hợp với các điểm từ dữ liệu thật với phân bố dữ liệu tự nhiên.

Mô hình phân biệt nhận đầu vào là các ví dụ (thật và giả) và thực hiện bài toán phân lớp nhị phân để xác định xem các ví dụ này là thật hay giả. Các dữ liệu thật được lấy từ tập dữ liệu huấn luyện, các dữ liệu giả được lấy trực tiếp từ đầu ra của mô hình sinh. Mô hình phân biệt là một mô hình phân lớp dễ hiểu.



Hình 2. 8: Kiến trúc mô hình Gan

(Nguồn: trituenhantao.io)

Dù là một bài toán học không giám sát nhưng với kiến trúc được thiết kế thông minh, **GAN** được huấn luyện như một bài toán học có giám sát. Mô hình sinh và mô hình phân biệt được huấn luyện đồng thời. Mô hình phân biệt sẽ dần thông minh hơn trong việc phát hiện các ví dụ thật / giả và do đó mô hình sinh cũng tinh tế hơn trong việc sáng tạo của mình.

Nói với ngôn ngữ của Lý thuyết trò chơi, trong bối cảnh này, hai mô hình thi đấu với nhau và đối nghịch trong một game có tổng bằng 0.

Biến thể quan trọng của **GAN** là các mô hình sinh tạo ra các ví dụ thỏa mãn các điều kiện cho trước. Các điều kiện này được đưa vào mạng nơ-ron dưới dạng các đặc trưng. Các giá trị này có thể là giá trị biểu thị các lớp, giá trị số hay một ma trận (ví dụ bức ảnh). Đến lượt mình, mô hình phân biệt cũng nhận được những thông tin này và “ra yêu cầu” cho mô hình sinh thông qua những quyết định phân lớp của mình.

Cách tiếp cận này là khởi nguồn của hàng loạt các ứng dụng rất ấn tượng của **GAN** như thay đổi nội dung của ảnh, sáng tạo phong cách

nghệ thuật, tô tranh hay chuyển một bức ảnh mùa hè thành mùa đông, ban ngày thành ban đêm, hay dự đoán giá cổ phiếu, v.v..

2.5. Dự đoán giá cổ phiếu

2.5.1. Tổng quan

Trên thế giới có rất nhiều công trình nghiên cứu về giá cổ phiếu. Phần lớn các nhà nghiên cứu đều xoay quanh các thuật toán học máy, với các kỹ thuật phân tích lý thuyết đầu tư chính là phương pháp để dự đoán trực tiếp giá cổ phiếu thông qua việc nghiên cứu dữ liệu thị trường cổ phiếu trong quá khứ. Dự đoán thị trường cổ phiếu là một trong những vấn đề quan trọng và nhiều thách thức. Dự đoán thị trường cổ phiếu được xem như là vấn đề dự báo về chuỗi thời gian và một trong những thuật toán giải quyết bài toán này là thuật toán **Autoregressive Integrated Moving Average** (gọi tắt là **ARIMA**) [2]. Thuật toán **ARIMA** thực hiện tốt trên chuỗi thời gian tuyến tính và không thay đổi nhưng không thực hiện tốt trên dữ liệu phi tuyến tính và biến đổi như trong thị trường cổ phiếu. Với các đặc điểm này của thuật toán **ARIMA**, cùng với sự biến đổi phi tuyến của giá cổ phiếu trên thị trường nên các nhà nghiên cứu càng về sau hướng nghiên cứu của mình theo sự phát triển kết hợp các thuật toán để đưa ra phương pháp nghiên cứu tối ưu cho bài toán của mình về dự đoán giá cổ phiếu hiện nay.

2.5.2. Các hướng tiếp cận

Có nhiều hướng tiếp cận khác nhau được sử dụng để dự đoán giá cổ phiếu như:

Hướng tiếp cận kết hợp giữa **ARIMA** và **SVM** [3]. Vì trong thực tiễn dự báo có thể được cấu thành bởi một thành phần tuyến tính và phi tuyến, vì thế có thể dự đoán một phần tuyến tính với **ARIMA** và một phần phi tuyến tính với **SVM**.

Sự kết hợp **Wavelet** với **SVM** [4], trong đó Wavelet thực hiện phân tích dữ liệu cổ phiếu và SVM dùng để dự báo.

Các nghiên cứu về sau sử dụng mạng nơ ron nhân tạo Artificial Neural Network (**ANN**) kết hợp với thuật toán **ARIMA** để dự đoán một phần dữ liệu giá cổ phiếu phi tuyến tính [5].

Hỗn hợp biến đổi **Wavelet** và **ANN** cũng được đề xuất là giải pháp hiệu quả trong việc giải quyết vấn đề này [6].

Thuật toán Convolutional Neural Network (**CNN**) cũng được dùng trong việc dự đoán giá cổ phiếu [7].

Một số nghiên cứu đề xuất mô hình mạng nơ ron hồi quy - **RNN**, học tăng cường Reinforcement learning [8-11].

Nghiên cứu sử dụng phương pháp học không giám sát **GAN** dự đoán giá cổ phiếu trong đó sử dụng 2 lớp **LSTM** dùng để dự đoán và **CNN** dùng để phân biệt dữ liệu cổ phiếu dự đoán và cổ phiếu thật [18].

Hướng kết hợp bằng phương pháp dự đoán học không giám sát **Generative Adversarial Network** gọi tắt là **GAN**, trong đó có 2 lớp phân biệt. Một là lớp **MLP** để dự đoán giá cổ phiếu và lớp còn lại là lớp **LSTM** dùng làm lớp phân biệt dữ liệu cổ phiếu dự đoán với dữ liệu cổ phiếu thật [17].

Trong đề tài này, chúng tôi xin giới thiệu hướng tiếp cận khác đó là Dự đoán giá cổ phiếu bằng phương pháp **GAN** mà trong đó sử dụng hai lớp. Một lớp là **LSTM** dùng làm lớp tạo, lớp còn lại là lớp **BiLSTM** dùng làm lớp phân biệt dữ liệu cổ phiếu do lớp tạo dự đoán ra. Để thực hiện dự đoán giá của cổ phiếu từ việc nghiên cứu dữ liệu thị trường cổ phiếu trong quá khứ.

2.5.3 Hướng đề xuất nghiên cứu

Nếu **LSTM** có khả năng trích xuất thông tin từ việc nghiên cứu dữ liệu cổ phiếu trong quá khứ với việc giải quyết tốt về khoảng cách dài ngắn của dữ liệu cổ phiếu trước đó, thì **BiLSTM** có thể giải quyết tốt hơn vấn đề bộ nhớ dài ngắn thông qua việc chạy 2 vòng **LSTM (Forward - Backward)** này để xác nhận lại các đặc trưng của dữ liệu để kiểm chứng dữ liệu tối ưu.

Trong nghiên cứu này, chúng tôi đề xuất mô hình **GAN** kết hợp 2 mô hình (**LSTM** và **BI-LSTM**) cho bài toán dự đoán giá cổ phiếu vì những nguyên nhân và đóng góp đáng kể của phương pháp so với các phương pháp trước kia như sau:

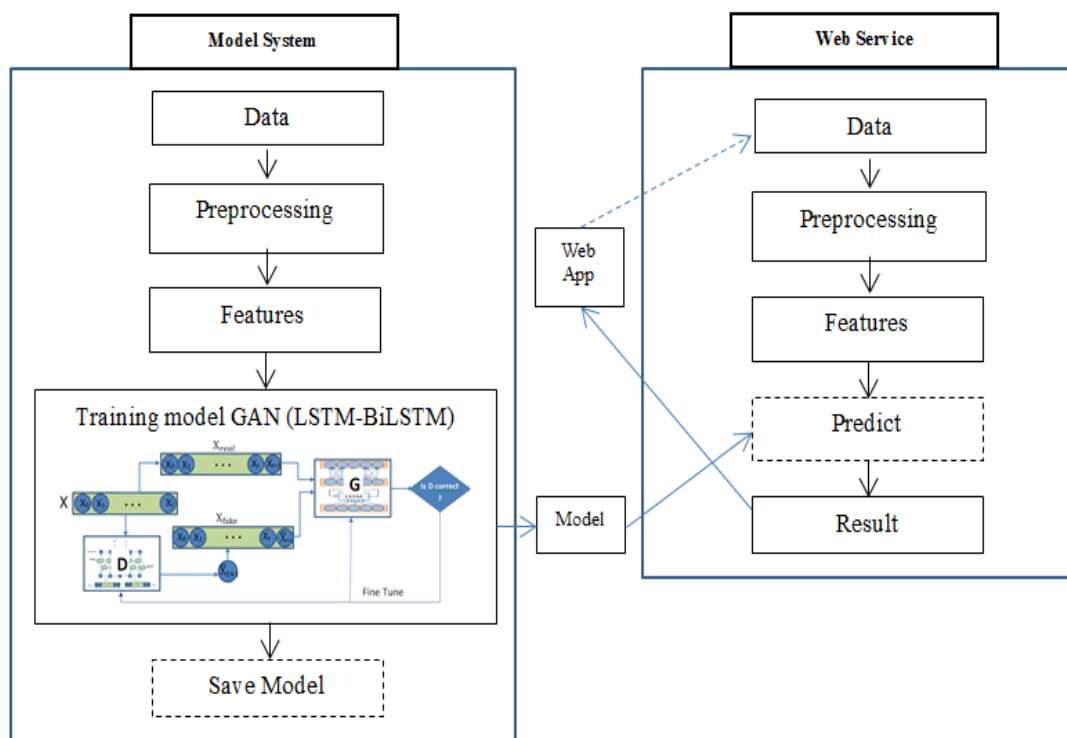
- Giải pháp GAN cũ khi thực nghiệm cho kết quả chưa tốt, chúng tôi đã thực nghiệm trên các mô hình khác là cải tiến của MLP và LSTM để cho kết quả tốt hơn. Qua quá trình thử nghiệm chúng tôi nhận thấy LSTM cho kết quả tốt hơn MLP và Bi-LSTM cho kết quả tốt hơn LSTM, đó là lý do chúng tôi đề xuất mô hình GAN (**LSTM** và **BI-LSTM**)
- Các tham số của mô hình như số neuron, số tầng, số lớp, các tham số khác,...chúng tôi cũng đã thử nghiệm và các tham số mà chúng tôi trình bày trong luận văn là các tham số cho kết quả tốt nhất, đó là lý do chúng tôi lựa chọn các tham số này trong mô hình cuối cùng để xây dựng ứng dụng trực quan hóa kết quả.

CHƯƠNG 3

MÔ HÌNH ĐỀ XUẤT

3.1. Tổng quan mô hình đề xuất

Mô hình tổng quát được trình bày trong hình 3.1.



Hình 3. 1: Mô hình đề xuất tổng quát

Trong mô hình đề xuất tổng quát gồm hai phần lớn: Huấn luyện mô hình và ứng dụng web.

- Huấn luyện mô hình: Từ dữ liệu thô, là một tập các dữ liệu cổ phiếu, ta tiền xử lý nó. Sau đó tiến hành xử lý lấy đặc trưng từ dữ liệu đã có để huấn luyện ra model, cuối cùng huấn luyện sử dụng mô hình học sâu **GAN** với 2 thành phần **LSTM-BiLSTM**.

- Ứng dụng web: Cho loại cổ phiếu trong số 5 loại cổ phiếu: **AMZN, MSFT, GOOGL, AAPL** và **VCB**, tiền xử lý dữ liệu. Sau đó, sử dụng 1 trong 5 mô hình có sẵn: **ANN, LSTM, BAYES, SVM** và **GAN** để dự đoán và in ra kết quả ra giao diện web.

Chi tiết về mô hình đề xuất được trình bày ở các phần tiếp theo.

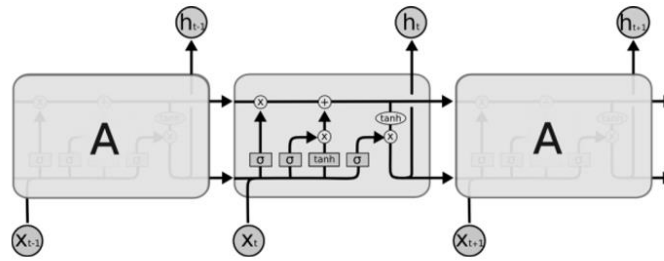
3.2. Các đặc trưng của mô hình đề xuất

GAN là mạng đối nghịch tạo sinh chứa hai mạng thần kinh (mạng phân biệt và mạng tạo sinh) cạnh tranh lẫn nhau trong một trò chơi (ở dạng một trò chơi có tổng bằng 0, trong đó bên thắng sẽ chiếm phần của bên thua, hay nói cách khác là bên thua sẽ bị mất phần cho bên thắng). [12]. Trong quá trình truyền nghịch, Một lớp được dùng để tạo ra dữ liệu giả giống hệt như dữ liệu thật, trong khi lớp kiểm tra đóng vai trò thẩm phán để phân biệt dữ liệu thật và dữ liệu giả được tạo ra. Chúng có thể đạt tới điểm lý tưởng mà lớp phân biệt không thể phân biệt hai loại dữ liệu. Tại điểm này lớp tạo sẽ chụp lại dữ liệu phân phối tại trò chơi này. Dựa trên nguyên tắc này, chúng tôi đề xuất cấu trúc **GAN** để dự đoán giá cổ phiếu.

3.2.1. Lớp tạo - The Generator

Mô hình lớp tạo của chúng tôi được thiết kế bởi Mạng bộ nhớ dài-ngắn (Long Short Term Memory networks - **LSTM**), **LSTM** có khả năng xử lý mạnh mẽ đối với dữ liệu chuỗi thời gian, đây là một dạng đặc biệt của **RNN**, nó có khả năng học được các phụ thuộc xa. **LSTM** được giới thiệu bởi Hochreiter & Schmidhuber (1997) [13], và sau đó đã được cải tiến và phổ biến bởi rất nhiều người trong ngành. Chúng hoạt động cực kì hiệu quả trên nhiều bài toán khác nhau nên dần đã trở nên phổ biến như hiện nay.

LSTM được thiết kế để tránh được vấn đề phụ thuộc xa (long-term dependency). Việc nhớ thông tin trong suốt thời gian dài là đặc tính mặc định của chúng, chứ ta không cần phải huấn luyện nó để có thể nhớ được. Tức là ngay nội tại của nó đã có thể ghi nhớ được mà không cần bất kì can thiệp nào. **LSTM** cũng có kiến trúc dạng chuỗi như **RNN**, nhưng các mô-đun trong nó có cấu trúc khác với mạng **RNN** chuẩn. Thay vì chỉ có một tầng mạng nơ-ron, chúng có 4 tầng tương tác với nhau một cách đặc biệt.



Hình 3. 2: Mạng LSTM

LSTM có 4 thành phần cơ bản là: Tế bào trạng thái (cell state), Cổng (gates), Sigmoid và Tanh. Cho một chuỗi các vector (x_1, x_2, \dots, x_n), σ là hàm sigmoid logistic, trạng thái ẩn h_t của **LSTM** tại thời điểm t được tính như sau:

$$h_t = o_t * \tanh(c_t) \quad (3.1)$$

$$o_t = \tanh(Wx_0x_t + Wh_0t_t + Wc_0c_t + b_0) \quad (3.2)$$

$$o_t = f_t * c_{t-1} + i_t * \tanh(Wx_0x_t + Wh_c t_{t-1} + b_c) \quad (3.3)$$

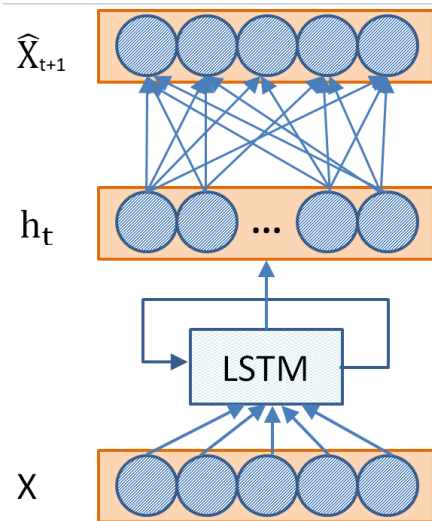
$$f_t = \sigma(Wx_f x_t + Wh_f t_{t-1} + Wc_f c_{t-1} + b_f) \quad (3.4)$$

$$i_t = \sigma(Wx_i x_t + Wh_i t_{t-1} + Wc_i c_{t-1} + b_i) \quad (3.5)$$

Trong nghiên cứu này, chúng tôi xây dựng mạng **LSTM** xử lý dữ liệu đầu vào với 5 yếu tố tài chính để dự đoán giá đóng trong tương lai. Năm yếu tố dữ liệu cổ phiếu trong một ngày là **High Price, Low Price, Open Price, Close Price, Volume**. 5 yếu tố này là những đặc trưng quan trọng và không có trong dự đoán giá của các phương pháp phân tích **Mean Reversion**, hoặc **MAR**. Vì thế những yếu tố này có thể được dùng như là 5 đặc trưng của dữ liệu cho dự đoán giá. Giả sử chúng tôi nhập vào $\mathbf{X} = \{\mathbf{X}_1, \dots, \mathbf{X}_t\}$, bao gồm dữ liệu cổ phiếu hàng ngày của t ngày. Mỗi \mathbf{X}_k trong \mathbf{X} là một véc tơ được tạo ra từ 5 đặc trưng như sau:

$$[\mathbf{X}_{k,i}]_i^k = [X_{k,\text{high}}, X_{k,\text{low}}, X_{k,\text{open}}, X_{k,\text{close}}, X_{k,\text{volume}}] \quad (3.6)$$

Cấu trúc của lớp tạo được chỉ ra như trong Hình 1. Để đơn giản chúng tôi bỏ qua những chi tiết của thuật toán **LSTM**. Với lớp tạo chúng tôi trích xuất ra \mathbf{h}_t của thuật toán **LSTM** và đặt nó vào trong 5 lớp kết nối đầy đủ với 5 tế bào thần kinh để tạo ra $\hat{\mathbf{X}}_{t+1}$ để đạt được gần đúng \mathbf{X}_{t+1} và chúng tôi lấy $\hat{\mathbf{X}}_{t+1}$ để dự đoán giá đóng trong $t+1$ ngày.



Hình 3. 3: Lớp tạo được thiết kế với giải thuật LSTM

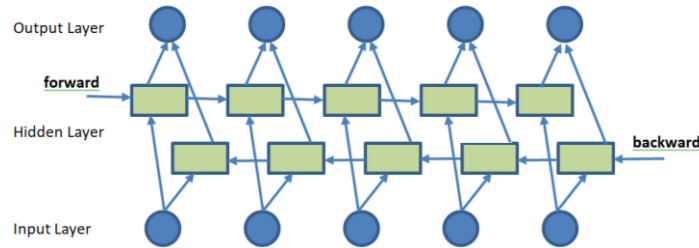
Đầu ra của lớp tạo $G(X)$ được định nghĩa như sau: $h_t = g(X)$,

$$G(\mathbf{X}) = \hat{\mathbf{X}}_{t+1} = \delta(\mathbf{W}_h^T \mathbf{h}_t + \mathbf{b}_h) \quad (3.7)$$

$g(.)$ biểu thị đầu ra của LSTM và \mathbf{h}_t là đầu ra của LSTM với $\mathbf{X} = \{\mathbf{X}_1, \dots, \mathbf{X}_t\}$ được nhập vào. δ là nơi để hàm Leaky Rectified Linear Unit (ReLU) kích hoạt. \mathbf{W}_h và \mathbf{b}_h biểu thị cho trọng số và bias trong lớp kết nối đầy đủ. Chúng tôi dùng hàm **Dropout** như là phương pháp chính để tránh bị quá tải. Hơn thế nữa chúng tôi có thể tiếp tục dự đoán $\hat{\mathbf{X}}_{t+2}$ với $\hat{\mathbf{X}}_{t+1}$ và \mathbf{X} .

3.2.2. Lớp phân biệt - The Discriminator

Mục đích của lớp phân biệt là tạo ra hàm phân biệt D để phân loại dữ liệu vào. Lớp phân biệt mong đợi Output là 0 khi nhập vào là dữ liệu giả. Output là 1 khi nhập vào là dữ liệu thật. Ở đây, chúng tôi chọn lớp **Bi-LSTM** là lớp phân biệt của chúng tôi với 3 lớp input, lớp output và lớp ẩn. Trong đó, lớp ẩn gồm 2 lớp mạng LSTM đơn được sử dụng đồng thời và độc lập để mô hình hoá chuỗi đầu vào theo 2 hướng: từ trái sang phải (forward LSTM) và từ phải sang trái (backward LSTM). Hàm Leaky ReLU được dùng như hàm kích hoạt giữa các lớp với lớp ẩn và hàm sigmoid được dùng như một lớp ra. Trong đó, Entropy chéo được chọn như hàm loss để tối ưu **Bi-LSTM**.



Hình 3. 4: Mô hình Bidirectional LSTM

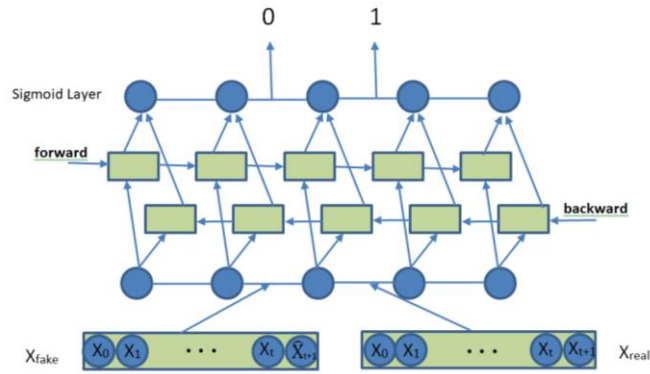
Cụ thể, chúng tôi so sánh tập $\mathbf{X} = \{\mathbf{X}_1, \dots, \mathbf{X}_t\}$ và $\hat{\mathbf{X}}_{t+1}$ để lấy $\{\mathbf{X}_1, \dots, \mathbf{X}_t, \hat{\mathbf{X}}_{t+1}\}$ như là dữ liệu giả \mathbf{X}_{fake} và so sánh $\mathbf{X} = \{\mathbf{X}_1, \dots, \mathbf{X}_t\}$ và

\mathbf{X}_{t+1} để lấy $\{\mathbf{X}_1, \dots, \mathbf{X}_t, \mathbf{X}_{t+1}\}$ như là dữ liệu thật \mathbf{X}_{real} . Out put của lớp phân biệt được định nghĩa như sau:

$$D(\mathbf{X}_{\text{fake}}) = \sigma(d(\mathbf{X}_{\text{fake}})), \quad (3.8)$$

$$D(\mathbf{X}_{\text{real}}) = \sigma(d(\mathbf{X}_{\text{real}})) \quad (3.9)$$

$d(\cdot)$ biểu thị cho Output của **Bi-LSTM** và biểu thị cho hàm kích hoạt Sigmoid. Cả hai \mathbf{X}_{fake} và \mathbf{X}_{real} có Output là một giá trị vô hướng. Hình 3 chỉ ra cấu trúc của lớp phân biệt.



Hình 3. 5: Cấu trúc Lớp phân biệt sử dụng Bi-LSTM

3.2.3. Cấu trúc GAN

Với hai mô hình được đề cập ở trên chúng tôi đề xuất cấu trúc **GAN** của chúng tôi. Theo đó trong 2 người chơi trò chơi minimax, cả **G** và **D** đều cố gắng tối ưu giá trị của hàm. Tương tự, chúng tôi định nghĩa sự tối ưu giá trị của $\mathbf{V}(\mathbf{G}, \mathbf{D})$ như sau:

$$\min_G \max_D \mathbf{V}(\mathbf{G}, \mathbf{D}) = E[\log D(\mathbf{X}_{\text{real}})] + E[\log(1 - D(\mathbf{X}_{\text{fake}}))] \quad (3.10)$$

Chúng tôi định nghĩa hàm loss của lớp tạo là G_{loss} và hàm loss của lớp phân biệt là D_{loss} để tối ưu giá trị hàm. Cụ thể, chúng tôi kết hợp hàm Mean Square Error (**MSE**) với hàm loss của lớp tạo **GAN** để cấu thành G_{loss} của cấu trúc của chúng tôi. G_{loss} và D_{loss} được định nghĩa như sau:

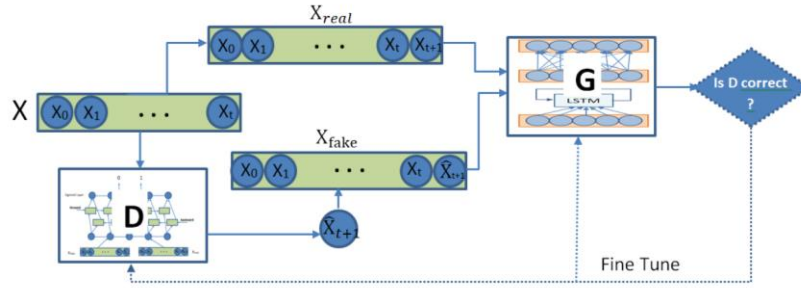
$$D_{\text{loss}} = -\frac{1}{m} \sum_{i=1}^m \log D(\mathbf{X}_{\text{real}}^i) - \frac{1}{m} \sum_{i=1}^m \log(1 - D(\mathbf{X}_{\text{fake}}^i)) \quad (3.11)$$

$$g_{\text{MSE}} = \frac{1}{m} \sum_{i=1}^m (\hat{X}_{t+1}^i - X_{t+1}^i)^2 \quad (3.12)$$

$$g_{\text{loss}} = \frac{1}{m} \sum_{i=1}^m \log(1 - D(X_{\text{fake}}^i)) \quad (3.13)$$

$$G_{\text{loss}} = \lambda_1 g_{\text{MSE}} + \lambda_2 g_{\text{loss}} \quad (3.14)$$

Hàm mất mát G_{loss} được tạo bởi g_{MSE} và g_{loss} với λ_1 và λ_2 , tương ứng λ_1 và λ_2 là những tham số mà chúng tôi đặt tự nhiên. Hình 4 chỉ ra cấu trúc GAN của chúng tôi. Lý do tại sao chúng tôi chọn X_{fake} và X_{real} hơn là \hat{X}_{t+1} và X_{t+1} trong Máy phân biệt là vì chúng tôi mong đợi Máy phân biệt chụp được sự tương quan và thông tin chuỗi thời gian giữa X_{t+1} và X .



Hình 3. 6: Cấu trúc GAN đề xuất của chúng tôi

3.2.4. Phương pháp đánh giá

Chúng tôi đánh giá hiệu quả dự báo của mô hình của chúng tôi bằng các tiêu chí thống kê như: Lỗi bình quân tuyệt đối (**MAE**), lỗi hình vuông gốc trung bình (**RMSE**), phần trăm bình quân tuyệt đối (**MAPE**) và lợi nhuận trung bình (**AR**) được tính theo các công thức trình bày sau đây:

Lỗi bình quân tuyệt đối (MAE):

$$\text{MAE} = \frac{1}{N} \sum_{k=1}^N |\hat{y}_k - y_k| \quad (3.15)$$

Lỗi hình vuông gốc trung bình (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{k=1}^N (\hat{y}_k - y_k)^2} \quad (3.16)$$

Phần trăm bình quân tuyệt đối (MAPE):

$$\text{MAPE} = \frac{1}{N} \sum_{k=1}^N \frac{|\hat{y}_k - y_k|}{y_k} \quad (3.17)$$

Lợi nhuận trung bình (AR):

$$\text{AR} = \frac{1}{N-1} \sum_{k=1}^{N-1} (y_{k+1} - y_k), \text{ if } \hat{y}_{k+1} > \hat{y}_k \quad (3.18)$$

Trong đó: y_k là dữ liệu cổ phiếu thật.

\hat{y}_k là dữ liệu cổ phiếu dự đoán.

N là tổng số ngày giao dịch của cổ phiếu

Các độ đo **MAE**, **RMSE** và **MAPE** có giá trị thấp có nghĩa là dự đoán giá đóng gần đúng với dữ liệu thật.

CHƯƠNG 4

THỰC NGHIỆM

4.1. Dữ liệu

Chúng tôi đánh giá mô hình của chúng tôi trên dữ liệu cổ phiếu **AMZN** from Amazon.com Inc. Tất cả dữ liệu cổ phiếu được lấy từ trang web <https://www.alphavantage.co> của công ty tài chính Alphavantage. Chúng tôi chọn thời gian giao dịch trong khoảng 10 năm (bao gồm 5000 mảnh dữ liệu). Ví dụ, một vài cổ phiếu được chỉ ra trong Bảng 4.1. Ngày giao dịch thì không liên tục do giới hạn giao dịch vào cuối tuần và ngày nghỉ.

Cổ phiếu của tập đoàn AMAZON, với giá khởi đầu, cao nhất, thấp nhất và giá kết thúc giao dịch, số lượng 10000 mẫu.

Tên	Thời gian	Open	High	Low	Close	Volume
AMZN	10/30/2019	210.45	214.67	207.99	208.25	33600749
	10/31/2019	210.07	213.26	208.69	212.83	33602226
	11/1/2019	216.33	216.38	211.47	214.32	33122280
	11/4/2019	213.62	214.08	211.08	213.67	26178663
	11/5/2019	214.48	215.8	206.5	207.07	38135816
	11/6/2019	206.13	208.85	202.03	208.35	37598121
	11/7/2019	209.56	211.33	205.03	208.04	32248893
	11/8/2019	205.4	205.7	202.31	203.92	29940718
	11/11/2019	204.47	205.04	201.39	202.88	31636070
	11/12/2019	205	212.3	203.01	211.6	36386080

Bảng 4. 1: Dữ liệu cổ phiếu của tập đoàn Amzon

Cổ phiếu của tập đoàn APPLE, với với giá khởi đầu, cao nhất, thấp nhất và giá kết thúc giao dịch, số lượng 10000 mẫu.

Tên	Thời gian	Open	High	Low	Close	Volume
AAPL	10/30/2019	209.2	212.3	208.39	211.75	33600749
	10/31/2019	207.19	210.92	202.15	202.54	33602226
	11/1/2019	200.42	202.86	197.51	201.3	33122280
	11/4/2019	201.47	203.97	200.86	203.85	26178663
	11/5/2019	203.61	204.7	201.74	202.02	38135816
	11/6/2019	202.5	204.65	202.01	204.06	37598121
	11/7/2019	201	204.46	199.57	203.9	32248893
	11/8/2019	204.4	205.1	199.01	205.01	29940718
	11/11/2019	211.52	217.64	210.44	216.54	31636070
	11/12/2019	214.17	214.77	210.31	213.29	36386080

Bảng 4. 2: Dữ liệu cổ phiếu của tập đoàn Apple

Cổ phiếu của tập đoàn GOOGLE, với với giá khởi đầu, cao nhất, thấp nhất và giá kết thúc giao dịch, số lượng 10000 mẫu.

Tên	Thời gian	Open	High	Low	Close	Volume
GOOGL	10/30/2019	211.67	211.88	206.35	208.25	36716462
	10/31/2019	207.155	207.65	203.14	203.38	36446459
	11/1/2019	205.29	210.28	204.75	209.19	28041364
	11/4/2019	209.44	211.35	208.15	208.7	22588870
	11/5/2019	208.76	209.59	207.51	208.9	17958936

	11/6/2019	209.6	211.1874	208.915	210.28	20063496
	11/7/2019	210.53	212.36	209.21	211.49	21336167
	11/8/2019	211.49	212.1	209.25	209.7	27627561
	11/11/2019	209.54	215	208.91	214.58	26981478
	11/12/2019	213.86	216.25	212.85	213.02	36249319

Bảng 4. 3: Dữ liệu cổ phiếu của tập đoàn Google

Cổ phiếu của tập đoàn MICROSOFT, với với giá khởi đầu, cao nhất, thấp nhất và giá kết thúc giao dịch, số lượng 10000 mẫu.

Tên	Thời gian	Open	High	Low	Close	Volume
MSFT	10/30/2019	209.2	212.3	208.39	211.75	49605692
	10/31/2019	207.19	210.92	202.15	202.54	67457035
	11/1/2019	200.42	202.86	197.51	201.3	39826989
	11/4/2019	201.47	203.97	200.86	203.85	30160867
	11/5/2019	203.61	204.7	201.74	202.02	23251388
	11/6/2019	202.5	204.65	202.01	204.06	19632602
	11/7/2019	201	204.46	199.57	203.9	25079596
	11/8/2019	204.4	205.1	199.01	205.01	51247969
	11/11/2019	211.52	217.64	210.44	216.54	78983009
	11/12/2019	214.17	214.77	210.31	213.29	49280056

Bảng 4. 4: Dữ liệu cổ phiếu của tập đoàn Microsoft

Cổ phiếu của Ngân hàng Vietcombank, với với giá khởi đầu, cao nhất, thấp nhất và giá kết thúc giao dịch, số lượng 10000 mẫu.

Tên	Thời gian	Open	High	Low	Close	Volume
VCB	10/30/2019	81.1	82.5	80.5	80.6	809820
	10/31/2019	81	81.7	80.5	81.5	527230
	11/1/2019	81.5	82.8	81	81	2794080
	11/4/2019	81.6	81.9	79	79	1187100
	11/5/2019	79.5	80.8	78.9	80.6	791470
	11/6/2019	80	80.9	79.4	80.7	729340
	11/7/2019	80.7	81.3	80.3	81	1040190
	11/8/2019	81.3	83.1	80.8	82.9	1388490
	11/11/2019	82.9	83.3	81.9	82.1	775770
	11/12/2019	82.1	83.6	81.8	83.4	795100

Bảng 4. 5: Dữ liệu cổ phiếu của công ty Vietcombank

4.2. Xử lý dữ liệu

Tiền xử lý dữ liệu

Chúng tôi chuẩn hóa dữ liệu với giả định **MAR** được đề cập ở trên như sau:

$$\mathbf{X}_i = (\mathbf{X}_i - \mu^t) / \tau^t \quad (4.1)$$

μ^t và τ^t : là giá trị trung bình và độ lệch chuẩn của **X**. Chúng tôi chọn $t=5$ theo kinh nghiệm bởi chúng tôi cố gắng dự đoán dữ liệu trong ngày tới bằng dữ liệu một tuần trước đó (giao dịch được giới hạn cuối tuần). Ví dụ, chúng tôi tính toán mô tả độ sai lệch chuẩn của dữ liệu trong 5 ngày để chuẩn hóa dữ liệu. Sau đó, dữ liệu chuẩn hóa được dùng để dự đoán dữ liệu cho ngày thứ 6. Dữ liệu trong quá trình **Training** và **Testing** được xử lý cùng lúc như nhau.

Phân chia tập dữ liệu

Mục đích của chúng tôi là từ 5 yếu tố này để dự đoán giá đóng của ngày tới từ dữ liệu t ngày trước đó. Lý do sử dụng 5 yếu tố trong ngày tới là vì lớp tạo sẽ hỗ trợ tìm ra sự phân phối của dữ liệu thật và chúng tôi có thể lấy giá đóng từ dữ liệu đã được tạo ra. Dữ liệu được chia làm hai phần cho **Training** và **Testing**. Chúng tôi chọn đầu tiên từ 90% - 95% dữ liệu cổ phiếu cho **Training** và sử dụng 5%-10% (khoảng 500-1000 mẫu dữ liệu) cho **Testing**.

Hàm loss trong quá trình **Training** do không có tiến trình xử lý đối nghịch giữa lớp phân biệt và lớp tạo suốt trong quá trình huấn luyện. Cả hai lớp phân biệt và lớp tạo được tối ưu hóa suốt tiến trình đối nghịch.

4.3. Huấn luyện

Chúng tôi tiến hành chọn 5 cổ phiếu của các tập đoàn lớn có khối lượng giao dịch lớn và phạm vi hoạt động rộng khắp trên thế giới và Việt Nam gồm Cổ phiếu của tập đoàn Amazon (**AMZN**), Google (**GOOGL**), Apple (**AAPL**), Microsoft (**MSFT**) và một cổ phiếu ở Việt Nam là **VCB** (Ngân hàng Vietcombank) với khoảng thời gian giao dịch là 10 năm từ năm 2009 đến 2019, với khoảng 10000 mẫu (9000-9500 mẫu cho **Training**, 500-1000 cho **Testing**).

Môi trường thực nghiệm trên máy PC sử dụng hệ điều hành Windows 10 Pro bản 64 bit với RAM 8GB, Intel(R) Core(TM) i5-4200M CPU 2.50GHz. Trong quá trình huấn luyện, tôi sử dụng Python kết hợp với thư viện mã nguồn mở Keras do Google phát triển.

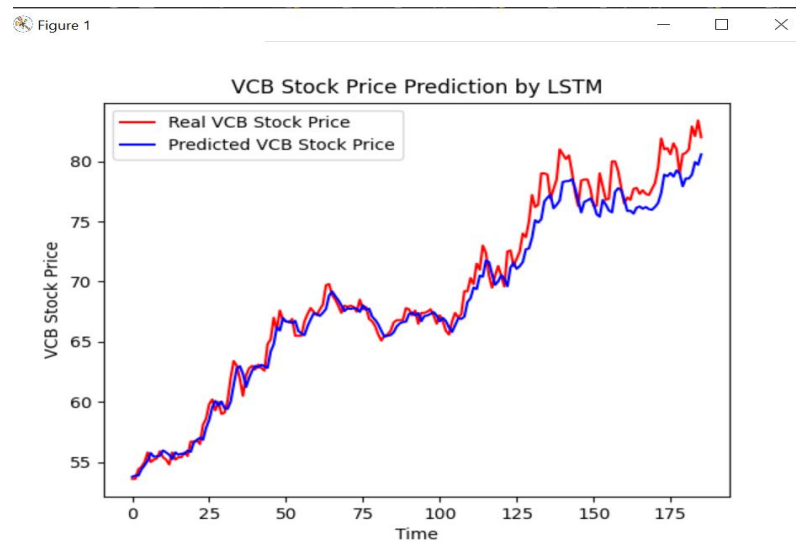
Thực hiện thực nghiệm từng loại cổ phiếu trên với các mô hình khác nhau gồm các mô hình **LSTM**, **SVM**, **BAYES**, **ANN** và cuối cùng là phương pháp **GAN** với mỗi loại mô hình chúng tôi thực hiện khoảng 100 epochs.

Qua quá trình thực nghiệm các phương pháp với cùng loại dữ liệu, cùng độ đo và số đặc trưng như nhau. Kết quả chúng tôi thu được như sau:

- Mô hình LSTM

Mô hình	RMSE	AR	MAE	MAPE
LSTM	5.3526	0.6725	4.0268	0.0168

Bảng 4. 6: Kết quả đánh giá bằng mô hình LSTM

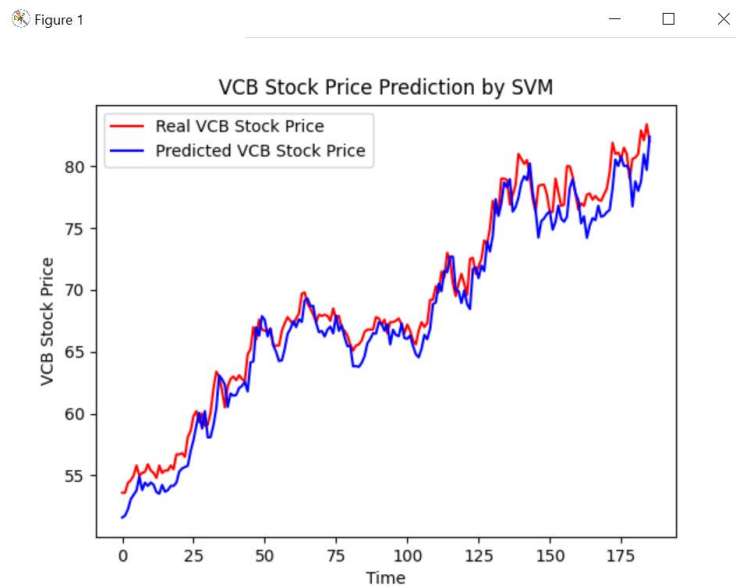


Hình 4. 1: Biểu đồ giá cổ phiếu sử dụng phương pháp LSTM

- Mô hình SVM

Mô hình	RMSE	AR	MAE	MAPE
SVR	8.1623	0.7134	4.7236	0.0412

Bảng 4. 7: Kết quả đánh giá bằng mô hình SVM

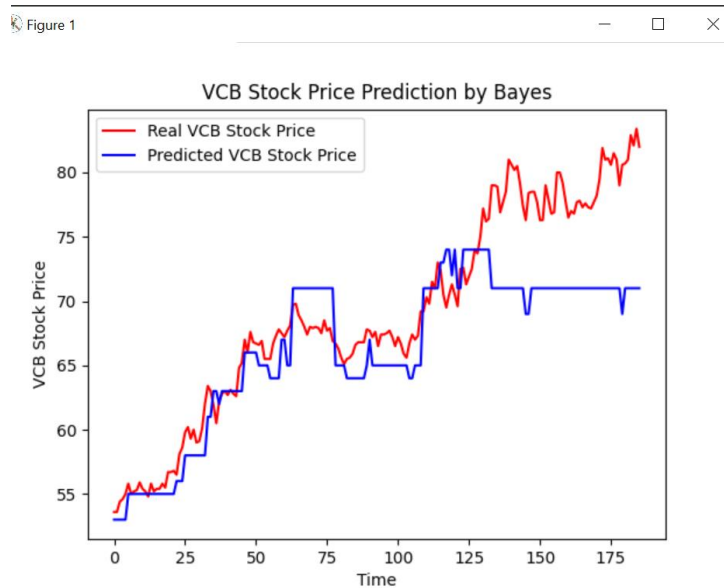


Hình 4. 2: Biểu đồ giá cổ phiếu sử dụng phương pháp SVM

- Mô hình BAYES

Mô hình	RMSE	AR	MAE	MAPE
BAYES	7.4632	0.6326	5.2134	0.0571

Bảng 4. 8: Kết quả đánh giá bằng mô hình BAYES

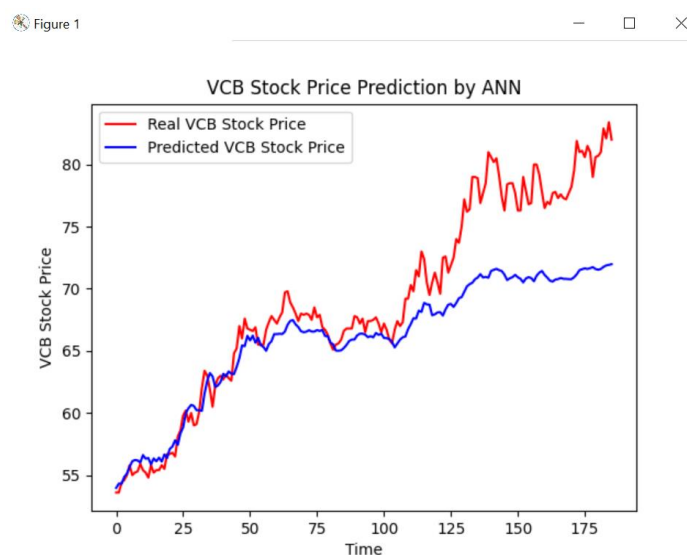


Hình 4. 3: Biểu đồ giá cổ phiếu sử dụng phương pháp BAYES

- **Mô hình ANN**

Mô hình	RMSE	AR	MAE	MAPE
ANN	9.0264	0.5068	7.2035	0.0708

Bảng 4. 9: Kết quả đánh giá bằng mô hình ANN



Hình 4. 4: Biểu đồ giá cổ phiếu sử dụng phương pháp ANN

So sánh các mô hình trên với mô hình GAN, chúng tôi được bảng kết quả như sau:

Mô hình	RMSE	AR	MAE	MAPE
ANN	9.0264	0.5068	7.2035	0.0708
Bayes	7.4632	0.6326	5.2134	0.0571
SVM	8.1623	0.7134	4.7236	0.0412
LSTM	5.3526	0.6725	4.0268	0.0168
Mô hình đề xuất GAN	4.0346	0.7478	3.4201	0.0145

Bảng 4. 10: Kết quả độ đo của các mô hình

Từ các biểu đồ thể hiện, các độ đo của các mô hình. Chúng tôi nhận thấy mô hình **Gan** của chúng tôi đã đạt được những kết quả quý giá và đáng khích lệ. Qua đó khẳng định mô hình của chúng tôi đã đạt kết quả tốt trong quá trình huấn luyện mô hình.

4.4. Đánh giá

Chúng tôi tính **RMSE** trên 5 loại dữ liệu bằng giá trị trung bình. **MAE** và **AR** cũng được tính toán bằng cách này. Véc tơ hỗ trợ hồi quy (**SVM**), **ANN**, **BAYES** và **LSTM** là phương pháp phân lớp cho dự đoán thị trường cổ phiếu và chúng tôi chọn chúng như là những phương pháp cơ bản để so sánh với mô hình đề xuất của chúng tôi. Kết quả dự đoán được chỉ ra trong Bảng trên với phần tô đậm là kết quả tốt nhất. Các độ đo **MAE**, **RMSE** và **MAPE** có giá trị thấp có nghĩa là dự đoán giá đóng gần đúng với dữ liệu thật và **AR** chỉ ra giá trị trung bình của những cổ phiếu này dựa trên 5 phương pháp dự đoán. Có thể thấy phương pháp dự đoán của chúng tôi đã đạt được kết quả tốt so với những phương pháp khác.

Kết quả hàng **in đậm** là kết quả tốt nhất của hướng tiếp cận dữ liệu mà mô hình đã đề xuất (theo cột).

Nhận xét và đề xuất:

- Từ kết quả trên cho thấy, mô hình **GAN** kết hợp LSTM - BiLSTM cho kết quả tốt nhất so với kết quả của các mô hình khác theo các tiêu chí **RMSE**, **AR**, **MAE**, **MAPE**, kèm theo đó là tốc độ huấn luyện nhanh hơn nhiều và dự đoán chính xác hơn, cho thấy mạng này khá thích hợp trong việc huấn luyện để thử nghiệm và học tập.
- Các tiêu chí của mô hình **Gan** cho kết quả thấp nhất. Điều này có nghĩa việc dự đoán kết quả tương đối chính xác và gần với giá cổ phiếu thật.

4.5. Xây dựng ứng dụng trực quan hóa kết quả

Sau khi xây dựng ra các mô hình học sâu dự đoán giá cổ phiếu, tôi xây dựng một giao diện web để người dùng có thể dự đoán xem kết quả dự đoán có chính xác hay không qua dữ liệu người dùng nhập vào.

Cấu trúc trang web gồm 3 trang con: Trang chủ, Minh họa, Dự đoán kết quả. Trang web được xây dựng trên thư viện Flask kết nối với Python server, sử dụng các công cụ thiết kế giao diện và ngôn ngữ lập trình: html, css, javascript và bootstrap. Ngoài ra, trang web còn sử dụng thư viện biểu diễn dữ liệu bokeh để xây dựng các biểu đồ trong trang web.

Người dùng chọn một mã cổ phiếu trong số 5 cổ phiếu: **AAPL**, **AMZN**, **MSFT**, **GOOGL** và **VCB**. Sau đó người dùng chọn phương pháp học sâu, trang Web sẽ trả về kết quả dự đoán giá dựa vào phương pháp học sâu đã chọn. Ngoài ra, giao diện Web còn bao gồm các thống

kê về dữ liệu dùng để huấn luyện và phân tích kết quả thu được từ các mô hình khác như: **SVM**, **BAYES**, **ANN**, **LSTM** sau khi huấn luyện.



Hình 4. 5: Trang chủ giao diện Web



Tiền xử lý dữ liệu

127.0.0.1:5001/PhanTichDuLieu

Cổ phiếu AMZN

Tên	Thời gian	Open	High	Low	Close	Volumn
AMZN	10/30/2019	143.52	145	142.79	144.61	18496591
	10/31/2019	149.49	144.93	142.99	143.37	24605135
	11/1/2019	144.26	144.92	142.965	143.72	33128366
	11/4/2019	144.83	145	144.16	145.55	16911999
	11/5/2019	144.97	145.02	143.905	144.46	18250172
	11/6/2019	144.37	144.52	143.2	144.06	16575798
	11/7/2019	143.84	144.88	143.77	144.26	17786715
	11/8/2019	143.98	145.99	143.76	145.96	16752939
	11/11/2019	145.34	146.42	144.73	146.11	14370178
11/12/2019	146.28	147.57	146.06	147.07	18648712	

Cổ phiếu GOOGL

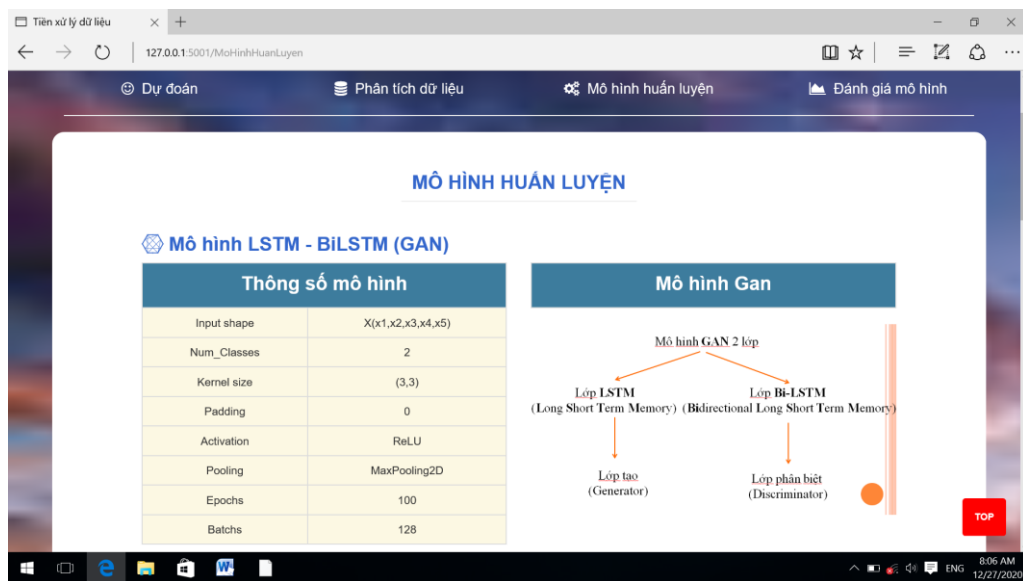
Tên	Thời gian	Open	High	Low	Close	Volumn
	10/30/2019	143.52	145	142.79	144.61	18496591

TOP

8:05 AM

12/27/2020

Hình 4. 6: Phân tích dữ liệu cổ phiếu





Hình 4. 8: Đánh giá các mô hình

CHƯƠNG 5

KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

5.1. Kết quả đạt được

Trong bài nghiên cứu này, luận văn đã thực nghiệm 5 mô hình học máy và học sâu. Kết quả cho thấy mô hình **GAN** 2 lớp LSTM-BiLSTM cho kết quả tốt nhất trong bộ dữ liệu.

Kết quả các mô hình và mô hình học không giám sát **GAN** được phân tích, đánh giá trên các phương pháp khác nhau đã làm nổi bật tính ưu việt của mô hình đề xuất. Các phần kết quả, thống kê dữ liệu và giao diện tương tác với người dùng cũng được xây dựng khá tốt. Nhìn chung, luận văn đã giải quyết phần nào được bài toán dự đoán giá cổ phiếu với hướng tiếp cận học sâu không giám sát **GAN**.

Tuy nhiên, luận văn vẫn còn gặp một số khó khăn cần giải quyết trong thời gian tới như:

Trước hết là về mặt dữ liệu. Do đây là dữ liệu được lấy từ nguồn dữ liệu từ các công ty tài chính Alphavantage, cho nên phần thu nhập dữ liệu không thể quản lý được và chưa phong phú. Dữ liệu cũng chứa những cổ phiếu bị lỗi khi load về, và còn hạn chế chuỗi thời gian liên tục của dữ liệu.

Điều này dẫn tới sự thiếu đa dạng của dữ liệu sẽ khó lấy ra kết quả chính xác.

Các cách xử lý và lập trình cũng tập trung vào sử dụng những đặc trưng đã có dựa trên 4 tiêu chí **RMSE**, **AR**, **MAE**, **MAPE**, chưa khai thác các đặc trưng khác và chưa đánh giá toàn diện các phương pháp thử nghiệm.

5.2. Hướng phát triển

Trong nghiên cứu này chúng ta đã đề xuất phương pháp học không giám sát **GAN** để dự đoán giá cổ phiếu. Chúng tôi đã tiến hành thực nghiệm trên bộ dữ liệu thực và đánh giá kết quả với các phương pháp khác. Kết quả cho thấy phương pháp **GAN** chúng tôi đề xuất đạt kết quả tốt nhất.

Tôi cũng sẽ thử nghiệm trên các mô hình học sâu khác nhau để từ đó so sánh, đánh giá, tìm ra mô hình tối ưu cho bài toán dự báo giá cổ phiếu.

Một hướng đi trong tương lai khá hứa hẹn là áp dụng các phương pháp học sâu cho việc tiếp tục tìm hiểu cách trích xuất những đặc trưng quan trọng và tối ưu mô hình của chúng tôi để có thể học dữ liệu phân phối một cách chính xác hơn, từ đó chúng tôi có thể đạt kết quả dự đoán tốt hơn bởi phương pháp chúng tôi đề xuất. Đồng thời, với tốc độ mạng internet phát triển ngày càng nhanh, có thể luôn cập nhật mô hình, xử lý dữ liệu theo thời gian thực giúp cho dự đoán giá cổ phiếu chính xác hơn trong tương lai,.

CÔNG TRÌNH CÔNG BỐ

Bùi Thanh Hùng, Nguyễn Minh Lợi. **“Dự đoán giá cổ phiếu bằng phương pháp học sâu không giám sát Generative Adversarial Network GAN”**. Hội nghị nghiên cứu khoa học Quốc gia về nghiên cứu cơ bản và ứng dụng Công nghệ thông tin (**FAIR** – Nha Trang năm 2020).

TÀI LIỆU THAM KHẢO

- [1] Li, B., Hoi, S.C.H., 2012. On-line portfolio selection with moving average reversion, in: Proceedings of the 29th International Conference on Machine Learning, ICML 2012, Edinburgh, Scotland, UK, June 26 - July 1, 2012.
- [2] Box, G.E.P., Jenkins, G.M., 1976. Time series analysis: Forecasting and control. *Journal of Time* 31, 238–242.
- [3] Pai, P.F., Lin, C.S., 2005. A hybrid arima and support vector machines model in stock price forecasting. *Omega* 33, 497–505.
- [4] Huang, S., Wang, H., 2006. Combining time-scale feature extractions with svms for stock index forecasting, in: Neural Information Processing, 13th International Conference, ICONIP 2006, Hong Kong, China, October 3-6, 2006, Proceedings, Part III, pp. 390–399.
- [5] Areekul, P., Senjyu, T., Toyama, H., Yona, A., 2010. A hybrid arima and neural network model for short-term price forecasting in deregulated market. *IEEE Transactions on Power Systems Pwrs*
- [6] Chandar, S.K., Sumathi, M., Sivanandam, S.N., 2016. Prediction of stock market price using hybrid of wavelet transform and artificial neural network. *Indian Journal of Science & Technology* 9.
- [7] Tsantekidis, A., Passalis, N., Tefas, A., Kannianen, J., Gabbouj, M., Iosifidis, A., 2017. Forecasting stock prices from the limit order book using convolutional neural networks, in: 19th IEEE Conference on Business Informatics, CBI 2017, Thessaloniki, Greece, July 24-27, 2017, Volume 1: Conference Papers, pp. 7–12.
- [8] Rather, A.M., Agarwal, A., Sastry, V.N., 2015. Recurrent neural network and a hybrid model for prediction of stock returns. *Expert Syst. Appl.* 42, 3234–3241.

- [9] Saad, E.W., Prokhorov, D.V., II, D.C.W., 1998. Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks. *IEEE Trans. Neural Networks* 9, 1456–1470.
- [10] Ding, X., Zhang, Y., Liu, T., Duan, J., 2015. Deep learning for event-driven stock prediction, in: *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015*, Buenos Aires, Argentina, July 25-31, 2015, pp. 2327–2333.
- [11] Nevmyvaka, Y., Feng, Y., Kearns, M.J., 2006. Reinforcement learning for optimized trade execution, in: *Machine Learning, Proceedings of the Twenty-Third International Conference (ICML 2006)*, Pittsburgh, Pennsylvania, USA, June 25-29, 2006, pp. 673–680.
- [12] Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A.C., Bengio, Y., 2014. Generative adversarial nets, in: *Advances in Neural Information Processing Systems: Annual Conference on Neural Information Processing Systems 2014*, December 8-13 2014, Montreal, Quebec, Canada, pp. 2672–2680.
- [13] S. Hochreiter and J. Schmidhuber, Long Short-Term Memory, *Neural Computation*, vol. 9, pp. 1735–1780, 1997
- [14] Zhiheng Huang, Wei Xu, and Kai Yu. Bidirectional LSTM-CRF models for sequence tagging. *arXiv preprint arXiv:1508.01991*. 2015
- [14] S. Hochreiter, “The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions”, 1997, World Scientific
- [15] S. Hochreiter, J. Schmidhuber, ”Long short-term memory”, 1997.
- [16] Denny Britz, “Recurrent Neural Networks Tutorial”, 2015.
- [17] Kang Zhang, Guoqiang Zhong, Junyu Dong, Shengke Wang, Yong Wang, “Stock Market Prediction Based on Generative Adversarial Network”, 2018.

- [18] Xingyu Zhou, Zhisong Pan, Guyu Hu, Siqi Tang, Cheng Zhao, “Stock Market Prediction on High-Frequency Data Using Generative Adversarial Nets”, 2018.
- [19] Dương Ngọc Mai Phương, Vũ Thị Phương Anh, Đỗ Thị Trúc Đào & Nguyễn Hữu Tuấn, “Tác động của chính sách tiền tệ đến thị trường chứng khoán: Bằng chứng tại Việt Nam”, năm 2015.
- [20] Đinh Bảo Ngọc, Nguyễn Chí Cường, “Các nhân tố ảnh hưởng đến chính sách cổ tức của các doanh nghiệp niêm yết trên thị trường chứng khoán Việt Nam”, năm 2014.
- [21] Đỗ Văn Thành, Nguyễn Minh Hải, “Phân tích và dự báo chỉ số thị trường chứng khoán bằng sử dụng chỉ số báo trước“, năm 2016.
- [22] <https://text.123doc.net/document/230610-nhung-van-de-co-ban-cua-thi-truong-chung-khoan.htm>
- [23] <https://www.apsi.vn/kien-thuc-can-ban-ve-chung-khoan-va-thi-truong-chung-khoan.html>
- [24] <https://dautucophieu.net/tong-quan-ve-ttck-viet-nam-15-nam-hinh-thanh-va-phat-trien/>
- [25] Thân Quang Khoát, “Học Máy (Machine Learning)”, năm 2016.
- [26] http://vnba.org.vn/index.php?option=com_k2&view=item&id=10720:vietinbank-voi-ung-dung-machine-learning-trong-hoat-dong-ngan-hang&Itemid=272&lang=en&cid=117:fin-tech-block-chain.
- [27] <https://machinelearningcoban.com/2018/06/22/deeplearning>.

Bình Dương, ngày 11 tháng 3 năm 2021

**BIÊN BẢN CUỘC HỌP HỘI ĐỒNG
CHẤM LUẬN VĂN THẠC SĨ**

Của học viên: **Nguyễn Minh Lợi**

Về đề tài: **Dự đoán giá cổ phiếu bằng phương pháp học sâu không giám sát Generative adversarial network (GAN)**

Tại Trường Đại học Thủ Dầu Một, ngày 11 tháng 3 năm 2021

Hội đồng chấm luận văn được thành lập theo Quyết định số, ngày cấp: **114/QĐ-ĐHTDM, ngày 15/01/2021** của Hiệu trưởng Trường Đại học Thủ Dầu Một đã nghe học viên: **Nguyễn Minh Lợi** trình bày tóm tắt nội dung luận văn thạc sĩ.

Nhận xét, đặt câu hỏi của thành viên Hội đồng chấm luận văn:

.....luận văn về ý nghĩa khoa học và thực tiễn, phù hợp
.....với chương trình đào tạo
.....Các câu hỏi của thành viên Hội đồng:
.....1. Vì sao phải generation chỉ cần dùng một mạng
.....LSTM không phải discriminator cần một kiến trúc
.....ĐL-LSTM?
.....2. Mô tả những ưu khuyết điểm của mô hình đề xuất về
.....độ bao phủ phiếu?
.....3. Hãy mô tả cách thu thập dữ liệu mô hình đề xuất
.....4. Trong mô hình đề xuất cần lấy data như thế nào?
.....Preprocessing như thế nào? các features nào được
.....sử dụng trong mô hình

Sau khi nghe tác giả trả lời và người hướng dẫn khoa học nhận xét luận văn. Hội đồng họp riêng và thống qua kết luận. Khi thông qua kết luận, Hội đồng bầu Ban kiểm phiếu gồm:

- 1) TS. Hoàng Minh Hà Trưởng ban
 2) PGS.TS. Quan Thanh Thọ Ủy viên
 3) TS. Văn Thế Thành Ủy viên

Căn cứ vào kết quả bỏ phiếu của Hội đồng chấm luận văn là:

- Số thành viên Hội đồng tham gia bỏ phiếu: 5...
- Tổng số điểm chấm luận văn của Hội đồng: 40...
- Điểm bình quân: 8.0...

Hội đồng kết luận:

- 1) Bản luận văn của học viên Nguyễn Minh Lợi đáp ứng yêu cầu của luận văn Thạc sĩ. Cụ thể là:

..... luận văn có ý nghĩa khoa học và thực tiễn
 Phương pháp nghiên cứu phù hợp
 Đề tài phù hợp với chuyên ngành đào tạo
 luận văn trình bày, phân tích rõ ràng và hợp lý
 Tổng hợp luận văn phù hợp với đề tài luận
 văn thạc sĩ

2) Các nội dung đề nghị chỉnh sửa (nếu có)

- Cần bổch dẫn tài liệu cho phân lý thuyết có số và các văn đề tiếp cần tổng quan
- Nên khảo sát và phân tích thêm các công trình liên quan về di báo quá có phiên
- Cần phân tích nguyên nhân, sự đóng góp về việc đưa ra mô hình để xuất hiện luận văn
- Học viên cần biên luận thêm cho việc lựa chọn các biến tự many cho các thành phần GAN
- cần minh, đặc biệt là discriminator
- Các mô tả kỹ thuật chưa nên được những lý do sai dụng, mô tả được quan tại sao

dùng các lớp input, output, ẩn, ...
 cần chú ý hệ ly kết cấu các chương
 lại hệ ly bản
 cần dẫn nguồn các thông tin mang tính
 lượng
 - Xem lại danh sách công thức


3) Đề nghị công nhận học vị Thạc sĩ cho học viên

Nguyễn Minh Sơn khi hoàn
 thành việc làm của mình.

THƯ KÝ HỘI ĐỒNG


 Huỳnh Hiền Nghĩa

CHỦ TỊCH HỘI ĐỒNG


 Hoàng Mạnh Hòa

NHẬN XÉT LUẬN VĂN THẠC SĨ

Tên đề tài: Dự đoán giá cổ phiếu bằng phương pháp học không giám sát Generative Adversarial Network (GAN)

Họ và tên học viên: Nguyễn Minh Lợi.

Người viết nhận xét: TS. Văn Thế Thành

Cơ quan công tác: Trường Đại học Công nghiệp Thực phẩm Tp.HCM

NỘI DUNG NHẬN XÉT

1. Tính cấp thiết, ý nghĩa khoa học và thực tiễn của luận văn

Thị trường cổ phiếu có một ý nghĩa quan trọng và mang nhiều tính bức thiết trong lĩnh vực kinh doanh. Luận văn tiếp cận việc dự đoán thị trường cổ phiếu bằng phương pháp học không giám sát Generative Adversarial Network (GAN). Hệ thống dự đoán dựa trên dữ liệu cổ phiếu đã giao dịch. Luận văn được thực nghiệm trên cổ phiếu AMZN (Amazon) và một số cổ phiếu khác. Vì vậy, đề tài luận văn có tính cấp thiết trong xã hội và có ý nghĩa khoa học.

2. Sự phù hợp của đề tài với chuyên ngành đào tạo

Đề tài tiếp cận việc xây dựng một hệ thống dự báo dựa trên dữ liệu thị trường chứng khoán dựa trên phương pháp học máy bằng cách sử dụng mạng nơ-ron dài ngắn LSTM và mạng nơ-ron hai chiều BiLSTM. Từ đó, tác giả đưa ra mô hình dự báo giá cổ phiếu và thực nghiệm trên tập dữ liệu cổ phiếu AMAZON có 10000 mẫu. Đề tài phù hợp với chuyên ngành đào tạo về hệ thống thông tin.

3. Về phương pháp nghiên cứu đã sử dụng

Luận văn sử dụng ba phương pháp nghiên cứu gồm phương pháp nghiên cứu lý thuyết, phương pháp nghiên cứu thực nghiệm, phương pháp so sánh và đánh giá. Tác giả đã lần lượt nghiên cứu lý thuyết về học máy, cổ phiếu chứng khoán, từ đó tác giả đề xuất mô hình dự báo và xây dựng, phát triển ứng dụng dự báo giá cổ phiếu. Kết quả thực nghiệm được đánh giá về sai số giữa các mô hình dự báo khác nhau.

4. Về cấu trúc, hình thức của luận văn

Luận văn luận văn bao gồm phần tóm tắt, 5 chương, công trình công bố và tài liệu tham khảo, trong đó phần tóm tắt nêu lên được phương pháp và ý nghĩa của vấn đề; nội dung các chương tiếp cận chủ yếu như sau:

Chương 1: Tiếp cận bài toán dự đoán giá cổ phiếu, mục tiêu, đối tượng, phương pháp và phạm vi nghiên cứu.

Chương 2: Tiếp cận lý thuyết về học máy (mạng nơ-ron nhân tạo, kỹ thuật lan truyền ngược), các mô hình học sâu (LSTM, BiLSTM, GAN), tổng quan về dự đoán thị trường cổ phiếu và hướng tiếp cận và đề xuất nghiên cứu.

Chương 3: Đề xuất mô hình dự đoán giá cổ phiếu theo mô hình GAN và đưa ra các phương pháp đánh giá hiệu quả của mô hình đã đề xuất.

Chương 4: Trình bày thực nghiệm, các bước tiền xử lý dữ liệu, huấn luyện mô hình, các phương pháp đánh giá mô hình và so sánh các kết quả của từng mô hình.

Chương 5: Tóm tắt và phân tích các ưu khuyết điểm của các mô hình, từ đó đề xuất mô hình thích hợp cho bài toán dự đoán giá cổ phiếu và ứng dụng tốt trong thực tiễn.

5. Về kết quả nghiên cứu, những đóng góp của luận văn

Luận văn đã tiếp cận bài toán dự báo giá cổ phiếu dựa trên cơ sở phương pháp học máy, cụ thể là sử dụng mạng LSTM và BiLSTM để xây dựng mô hình GAN 2 lớp theo tiếp cận học sâu không giám sát. Luận văn đã thực nghiệm quá trình huấn luyện mạng để có kết quả mô hình thực thi dự báo trên ứng dụng Web. Dữ liệu thực nghiệm được sử dụng trên dữ liệu cổ phiếu Amazon của công ty tài chính Alphavantage và một số dữ liệu khác như Apple, Google, Microsoft, Vietcombank. Kết quả thực nghiệm được đánh giá, so sánh sai số giữa các mô hình dự báo khác nhau và đưa ra kết luận về mô hình GAN 2 lớp.

6. Góp ý các thiếu sót cần chỉnh sửa, bổ sung

- Cần trích dẫn tài liệu cho những lý thuyết cơ sở và các vấn đề tiếp cận tổng quan.
- Nên khảo sát và phân tích thêm các công trình liên quan về dự báo giá cổ phiếu.
- Cần phân tích nguyên nhân, sự đóng góp về việc đưa ra mô hình đề xuất trong luận văn.

7. Kết luận chung

Luận văn đáp ứng yêu cầu đối với một luận văn thạc sĩ. Luận văn có thể đưa ra bảo vệ để nhận học vị Thạc sĩ

Câu hỏi:

1. Mô tả nhưng ưu khuyết điểm của mô hình đề xuất về dự báo cổ phiếu?
2. Hãy minh chứng tính đúng đắn của mô hình đề xuất?

Xác nhận của cơ quan công tác
HIỆU TRƯỞNG



Nguyễn Xuân Hoàn

....., ngày 11 tháng 3 năm 2021

Người nhận xét

(ký tên)

Trần Thị Thanh
Trần Thị Thanh



NHẬN XÉT LUẬN VĂN THẠC SĨ

Tên đề tài: Dự đoán giá cổ phiếu bằng phương pháp học sâu không giám sát
Generative adversarial network (GAN)

Họ và tên học viên: Nguyễn Minh Lợi

Người viết nhận xét: TS. Huỳnh Hữu Nghĩa

Cơ quan công tác: Viện Đào Tạo Sau Đại Học, Đại học Thủ Dầu Một

NỘI DUNG NHẬN XÉT

1. Tính cấp thiết, ý nghĩa khoa học và thực tiễn của luận văn

Luận văn có ý nghĩa khoa học và thực tiễn.

2. Sự phù hợp của đề tài với chuyên ngành đào tạo

Đề tài phù hợp với chuyên ngành đào tạo.

3. Về phương pháp nghiên cứu đã sử dụng

Luận văn đã sử dụng phương pháp nghiên cứu phân tích, khảo sát và lập trình.

4. Về cấu trúc, hình thức của luận văn

Cấu trúc và hình thức luận văn có 5 chương:

Chương 1: Tổng quan về lĩnh vực nghiên cứu

Chương 2: Cơ sở lý thuyết và các nghiên cứu liên quan

Chương 3: Mô hình đề xuất

Chương 4: Thực nghiệm

Chương 5: Kết luận và hướng phát triển

5. Về kết quả nghiên cứu, những đóng góp của luận văn

Tác giả luận văn đã nghiên cứu về thị trường chứng khoán, một số phương pháp máy học, học sâu. Sau đó đề xuất mô hình cho bài toán dự đoán giá cổ phiếu và đã thực nghiệm, đánh giá kết quả. Đóng góp của luận văn là đề xuất mô hình dự đoán giá cổ phiếu.

6. Góp ý các thiếu sót cần chỉnh sửa, bổ sung

Kết cấu chương chưa hợp lý, có thể kết cấu lại 3 chương chính (chương 2, 3 và 4), chương 1 thay bằng lời mở đầu và chương 5 thay bằng kết luận và hướng phát triển.

7. Kết luận chung

Đáp ứng yêu cầu luận văn thạc sĩ.

Câu hỏi:

1. Trong mô hình đề xuất cách lấy data như thế nào? Preprocessing như thế nào? Các features nào được sử dụng trong mô hình?

2.

....., ngày 11 tháng 3 năm 2021

Người nhận xét

(ký tên)

