

**TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP. HỒ CHÍ MINH**

**KHOA ĐÀO TẠO CHẤT LƯỢNG CAO**

**NGÀNH CÔNG NGHỆ THÔNG TIN**

-----🙠🙟🕮🙝🙢------



**Giảng viên hướng dẫn: TS Vũ Quang Huy**

**Sinh viên thực hiện: Phạm Trường Giang - 15110036**

**Nguyễn Ngọc Hoàng Phúc - 15110099**

Tp. Hồ Chí Minh, tháng 12 năm 2018

**BÁO CÁO TIỂU LUẬN CHUYÊN NGÀNH**

**ỨNG DỤNG THUẬT TOÁN MÁY HỌC TRONG VIỆC ĐÁNH GIÁ XU THẾ TĂNG GIẢM CỔ PHIẾU CỦA CÔNG TY NIÊM YẾT TRÊN THỊ TRƯỜNG CHỨNG KHOÁN**

**Ý KIẾN NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN**

*…………Ngày …… tháng ….. năm 20..…*

Giáo viên hướng dẫn

**Ý KIẾN NHẬN XÉT CỦA GIẢNG PHẢN BIỆN**

*…………Ngày …… tháng ….. năm 20..…*

Giáo viên phản biện

# Lời cảm ơn

Để hoàn thành bài tiểu luận này, chúng em xin trân trọng cảm ơn các thầy cô Ban lãnh đạo, văn phòng Khoa Đào tạo Chất lượng. Xin cảm ơn sâu sắc thầy Vũ Quang Huy, giáo viên hướng dẫn tiểu luận và thầy Nguyễn Thiên Bảo, giáo viên dạy môn Học máy đã hướng dẫn, truyền đạt kiến thức và đưa các giải pháp, đóng góp quý báu cho chúng em hoàn thành bài tiểu luận.

Vì thời gian tìm hiểu cho tiểu luận còn hạn chế, cũng như mới làm quen với kiến thức lĩnh vực mới nên còn nhiều hạn chế và không thể tránh những thiếu sót, kính mong được sự nhận xét, phê bình, đóng góp từ các quý thầy cô ngành công nghệ thông tin trong trong việc chấm và phản biện tiểu luận để chúng em có thể học hỏi và cải hiện nâng cấp phần mềm ở thời gian sau này.

Cuối cùng xin kính chúc quý thầy cô lơi chúc sức khỏe, thành đạt và hạnh phúc trong cuộc sống và công việc.

Tp. Hồ Chí Minh, Ngày 14 tháng 12 năm 2018

Nhóm sinh viên

Phạm Trường Giang

Nguyễn Ngọc Hoàng Phúc

# Quá trình thực hiện tiểu luận

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Công việc/Tuần | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | Ghi chú |
| Nhận đề tài |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Tìm hiểu đề tài |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Nghiên cứu thuật toán trong Machine learning |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Ứng dụng vào bài toán |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Cài đặt |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Tổng hợp báo cáo |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

# **Mục lục**

[Lời cảm ơn 3](#_Toc532564678)

[Quá trình thực hiện tiểu luận 4](#_Toc532564679)

[Mục lục 5](#_Toc532564680)

[Danh mục các từ viết tắt 6](#_Toc532564681)

[Danh mục hình ảnh 7](#_Toc532564682)

[Lời mở đầu 8](#_Toc532564683)

[1. Giới thiệu tiểu luận 9](#_Toc532564684)

[1.1 Lý do lựa chọn đề tài 9](#_Toc532564685)

[1.2 Ý nghĩa đề tài 11](#_Toc532564686)

[1.3 Mục tiêu mong muốn 11](#_Toc532564687)

[2. Cơ sở lý thuyết 11](#_Toc532564688)

[2.1 Tổng quan về thị trường chứng khoán 11](#_Toc532564689)

[2.2 Cổ phiếu và định giá cổ phiếu 17](#_Toc532564690)

[3. Tổng quan về học máy 19](#_Toc532564691)

[3.1 Tổng quan 19](#_Toc532564692)

[3.2 Các ứng dụng của học máy 23](#_Toc532564693)

[4. Phương pháp thực hiện 25](#_Toc532564694)

[3.1 Đặt vấn đề 25](#_Toc532564695)

[3.2 Thuật toán Linear Regression 26](#_Toc532564696)

[3.3 Cách xác định đường hồi quy tuyến tính 27](#_Toc532564697)

[3.4 Hạn chế của Linear Regression 30](#_Toc532564698)

[5. Ứng dụng thuật toán Regression vào bài toán 33](#_Toc532564699)

[6. Kết quả và đánh giá 38](#_Toc532564700)

[7. Kết luận 40](#_Toc532564701)

[Tài liệu tham khảo 41](#_Toc532564702)

# Danh mục các từ viết tắt

1. AI – Artificial Intelligence (Trí Tuệ Nhân Tạo)
2. ML – Machine learning
3. TP.HCM – Thành phố Hồ Chí Minh.
4. TTCK – Thị trường chứng khoán.
5. CP – Cổ phiếu
6. CTCK – Công ty chứng khoán

# Danh mục hình ảnh

[Hình 1 Machine learning 9](file:///D:\DOCUMENTS\Nam-4\Hoc-ky-7\tieu-luan-chuyen-nganh\Bao-cao-tieu-luan-chuyen-nganh-v3.docx#_Toc532603648)

[Hình 2 Màn hình điện tử chứng khoán 10](file:///D:\DOCUMENTS\Nam-4\Hoc-ky-7\tieu-luan-chuyen-nganh\Bao-cao-tieu-luan-chuyen-nganh-v3.docx#_Toc532603649)

[Hình 3 Nhà đầu tư chăm chú theo dõi biến động cổ phiếu 11](file:///D:\DOCUMENTS\Nam-4\Hoc-ky-7\tieu-luan-chuyen-nganh\Bao-cao-tieu-luan-chuyen-nganh-v3.docx#_Toc532603650)

[Hình 4 Trung tâm giao dịch chứng khoán Hà Nội 13](file:///D:\DOCUMENTS\Nam-4\Hoc-ky-7\tieu-luan-chuyen-nganh\Bao-cao-tieu-luan-chuyen-nganh-v3.docx#_Toc532603651)

[Hình 5 Đoạn phố Wall trước Sở giao dịch chứng khoán New York 15](#_Toc532603652)

[Hình 6 Sở giao dịch chứng khoán Tokyo 16](#_Toc532603653)

[Hình 7 Sở Giao dịch Chứng khoán Thành phố Hồ Chí Minh (HOSE) 17](#_Toc532603654)

[Hình 8 Mối liên hê giữa AL, Machine learning, Deep Learning 20](file:///D:\DOCUMENTS\Nam-4\Hoc-ky-7\tieu-luan-chuyen-nganh\Bao-cao-tieu-luan-chuyen-nganh-v3.docx#_Toc532603655)

[Hình 9 AI Landscape 21](file:///D:\DOCUMENTS\Nam-4\Hoc-ky-7\tieu-luan-chuyen-nganh\Bao-cao-tieu-luan-chuyen-nganh-v3.docx#_Toc532603656)

[Hình 10 Phân loại Machine Learning 23](file:///D:\DOCUMENTS\Nam-4\Hoc-ky-7\tieu-luan-chuyen-nganh\Bao-cao-tieu-luan-chuyen-nganh-v3.docx#_Toc532603657)

[Hình 11 Machine Learning và một số ứng dụng 25](file:///D:\DOCUMENTS\Nam-4\Hoc-ky-7\tieu-luan-chuyen-nganh\Bao-cao-tieu-luan-chuyen-nganh-v3.docx#_Toc532603658)

[Hình 12 Quan hệ y=3+4x 27](file:///D:\DOCUMENTS\Nam-4\Hoc-ky-7\tieu-luan-chuyen-nganh\Bao-cao-tieu-luan-chuyen-nganh-v3.docx#_Toc532603659)

[Hình 13 Biểu diễn dữ liệu 28](#_Toc532603660)

[Hình 14 Đồ thị của đường hồi quy 31](file:///D:\DOCUMENTS\Nam-4\Hoc-ky-7\tieu-luan-chuyen-nganh\Bao-cao-tieu-luan-chuyen-nganh-v3.docx#_Toc532603661)

[Hình 15 Hạn chế của hồi quy 31](file:///D:\DOCUMENTS\Nam-4\Hoc-ky-7\tieu-luan-chuyen-nganh\Bao-cao-tieu-luan-chuyen-nganh-v3.docx#_Toc532603662)

[Hình 16 Ví dụ thực tế của hồi quy 33](#_Toc532603663)

[Hình 17 So sánh y và y(predict) 34](file:///D:\DOCUMENTS\Nam-4\Hoc-ky-7\tieu-luan-chuyen-nganh\Bao-cao-tieu-luan-chuyen-nganh-v3.docx#_Toc532603664)

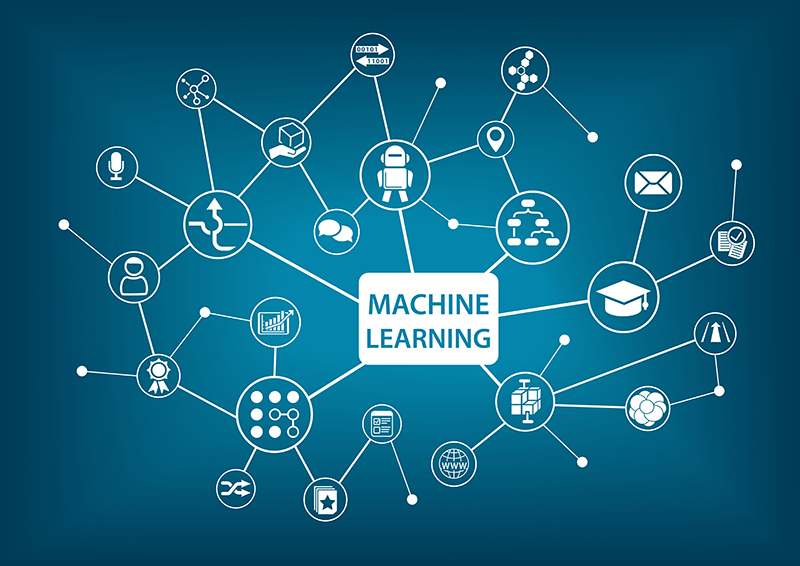
[Hình 18 Dữ liệu cổ phiếu 35](file:///D:\DOCUMENTS\Nam-4\Hoc-ky-7\tieu-luan-chuyen-nganh\Bao-cao-tieu-luan-chuyen-nganh-v3.docx#_Toc532603665)

[Hình 19 Đồ thị thể hiện sự biến động và dự đoán của cổ phiếu 39](file:///D:\DOCUMENTS\Nam-4\Hoc-ky-7\tieu-luan-chuyen-nganh\Bao-cao-tieu-luan-chuyen-nganh-v3.docx#_Toc532603666)

[Hình 20 So sánh thực tế và dự đoán được 40](file:///D:\DOCUMENTS\Nam-4\Hoc-ky-7\tieu-luan-chuyen-nganh\Bao-cao-tieu-luan-chuyen-nganh-v3.docx#_Toc532603667)

# Lời mở đầu

Vài năm gần đây, cụm từ "cách mạng công nghiệp lần thứ 4" hay "cách mạng công nghiệp 4.0", tập trung chủ yếu vào sản xuất thông minh dựa trên sự phát triển đột phá của cách ngành công nghệ thông tin, công nghệ sinh học, công nghệ nano, ... Mỗi một cuộc cách mạng công nghệ đều sẽ mang đến một bước ngoặt lớn với cách thức chúng ta sản xuất, lao động, hãy nhìn lại thế giới xung quanh bạn đang thay đổi từng ngày như thế nào: chúng ta có các sản phẩm trí tuệ nhân tạo mô phỏng được các hoạt động y hệt con người, thậm chí là giỏi hơn khi AlphaGo của google đã đánh bại Lee Sedol, kì thủ cờ vây hàng đầu thế giới, rồi chụp x-quang 3 chiều giúp phát hiện sớm ung thư, công nghệ nano giúp chữa trị ung thư cho con người, công nghệ thực tế ảo trong pokemon go từng gây sốt cho toàn thế giới, ... Thế giới thay đổi từng ngày, góp một phần không nhỏ trong đó chính là công nghệ thông tin, và cụ thể hơn, một trong các công nghệ góp phần vào bước phát triển của công nghệ thông tin, chính là Machine learning.

Machine Learning (ML) là một trong các cụm từ được nhắc đến khá nhiều gần đây, mà ví dụ tiêu biểu là AlphaGo của google. Không chỉ trong lĩnh vực trí tuệ nhân tạo, mà rất, rất nhiều lĩnh vực khác cũng đang có sự tham gia của ML và kiến thức về ML thì rất nhiều và khó khăn khi tiếp cận.

Hình 1 Machine learning

# Giới thiệu tiểu luận

***Đề tài:*** Ứng dụng thuật toán máy học trong việc đánh giá xu thế tăng giảm cổ phiếu của công ty niêm yết trên thị trường chứng khoán.

***Mô tả:***

Tìm hiểu về cổ phiếu và thị trường chứng khoán sau đó áp dụng thuật toán máy học để dự đoán giá trị tăng hay giảm cổ phiếu 1 công ty.

Kiến thức đã học về machine learning, kiến thức một số thuật toán, kiến thức toán học về các thuật toán đã áp dụng để phục vụ bài toán dự đoán cổ phiếu.

Chương trình dự đoán viết bằng ngôn ngữ Python và sử dụng data cổ phiếu của Google có mã là GOOGL lấy từ database WIKI của thư viện Quandl.

Hình 2 Màn hình điện tử chứng khoán

## Lý do lựa chọn đề tài

* **Hoàn cảnh:**

**Cổ phiếu:** Cổ phiếu là một phần của thị trường chứng khoán. Từ lâu thị trường chứng khoán luôn là vấn đề nóng đặc biệt ở thị trường mới nổi như Việt Nam. Để tránh những rủi ro đáng tiếc, cũng như dự đoán, dự báo thì nhu cầu cần có một hệ thống dự đoán dự báo tin cậy, chính xác cao, nhanh chóng rất được quan tâm.

**Công nghệ:** Machine Learning nói riêng và AI nói chung đang nổi lên và trong tương lai sẽ đang phát triển rất nhanh chóng và hỗ trợ, phục vụ con người trong rất nhiều lĩnh vực như: kinh tế, công nghiệp, đời sống, giáo dục,… và lĩnh vực mà nhóm đang thực hiện là liên quan quan đến dự đoán cổ phiếu.

* ***Lý do chọn nội dung liên quan đến cổ phiếu:***

Thị trường chứng khoán sẽ là kênh lựa chọn hấp dẫn nhà đầu tư trong những năm tới. Việc đầu tư vào thị trường giai đoạn hiện nay ở Việt Nam đã chuyên nghiệp hơn rất nhiều so với trước đây.

Là nhà đầu tư chứng khoán ai cũng muốn mình giao dịch có lãi nhưng việc xác định thời điểm tốt nhất để tham gia thị trường thì không hề đơn giản chút nào. Nếu nhìn thị trường bằng mắt thường và thao tác thủ công chỉ nhìn thấy được góc độ nào đó, chưa thể nhìn được tổng quan thị trường tại một thời điểm.

Ứng dụng công nghệ giúp nhà đầu tư phân tích nhanh chóng, tăng tính hiệu quả.

Hình 3 Nhà đầu tư chăm chú theo dõi biến động cổ phiếu

## Ý nghĩa đề tài

Mục đích hỗ trợ nhà đầu tư giao dịch hiệu quả mà không mất nhiều thời gian theo dõi thị trường. Mọi thứ đã được hệ thống máy tính xử lý giúp nhà đầu tư có chiến lược giao dịch đơn giản, kiểm soát tâm lý trong các thời điểm quan trọng.

Học hỏi kiến thức mới về lĩnh vực đang rất nóng hổi, và áp dụng sâu hơn trong các hệ thống phức tạp trong tương lai.

## Mục tiêu mong muốn

* Đáp ứng được yêu cầu đề ra của giáo viên hướng dẫn.
* Học hỏi thêm những kiến thức và công nghệ mới.
* Kết quả đầu ra có độ chính xác cao.
* Áp dụng nhiều hơn nữa trong các dự án sau này.

# 2. Cơ sở lý thuyết

## 2.1 Tổng quan về thị trường chứng khoán

**a) Khái quát về thị trường chứng khoán**

**Quá trình hình thành và phát triển thị trường chứng khoán**

Hình thức sơ khai của thị trường chứng khoán đã xuất hiện cách đây hàng trăm năm. Quá trình phát triển của thị trường chứng khoán trải qua nhiều bước thăng trầm (1929, 1987). Chính phủ các nước cải cách hệ thống thuế nhằm kích thích tiết kiệm và đầu tư. Một loạt các công cụ tài chính mới được tạo ra. Các nước tiến hành thành lập các thị trường không chính thức. Thị trường chứng khoán có sự đổi mới mạnh mẽ về kỹ thuật. Thị trường vay nợ của Chính phủ được hiện đại hóa.

**Khái niệm về thị trường chứng khoán**

Thị trường chứng khoán là đặc trưng của nền kinh tế hiện đại. Thị trường chứng khoán được hiểu đơn giản là nơi trao đổi, mua bán các chứng khoán hay các giấy tờ có giá. Việc trao đổi mua bán này được thực hiện theo những quy tắc ấn định trước. Chứng khoán (security) là giấy tờ có giá và có khả năng chuyển nhượng, xác nhận quyền và lợi ích hợp pháp của người sở hữu đối với tài sản và phần vốn góp của tổ chức phát hành.

Hình 4 Trung tâm giao dịch chứng khoán Hà Nội

**Các chủ thể tham gia thị trường chứng khoán**

* Chủ thể phát hành
* Nhà đầu tư:
  + Các nhà đầu tư cá nhân.
  + Các nhà đầu tư có tổ chức.
* Cơ quan quản lý và giám sát hoạt động thị trường chứng khoán:
  + Cơ quan quản lý Nhà nước về thị trường chứng khoán.
  + Các tổ chức tự quản.
* Các tổ chức cung ứng dịch vụ trên thị trường chứng khoán.

**Cấu trúc của thị trường chứng khoán**

* Phân loại theo hình thức tổ chức của thị trường:
  + Sở giao dịch (Stock Exchange).
  + Thị trường giao dịch qua quầy hay thị trường chứng khoán phi tập trung (OTC: Over The Counter Market).
* Phân loại theo quá trình luân chuyển vốn:
  + Thị trường sơ cấp hay thị trường cấp 1 (Primary Market).
  + Thị trường thứ cấp hay thị trường cấp 2 (Secondary Market).
* Phân loại theo hàng hoá:
  + Thị trường trái phiếu (Bond Markets).
  + Thị trường cổ phiếu (Stock Markets).
  + Thị trường các công cụ dẫn suất (Derivative Markets).

*Thị trường sơ cấp*

* Thị trường phát hành chứng khoán.
* Giá phát hành được ấn định, hoặc qua đấu giá.
* Việc mua bán chứng khoán làm tăng giá trị vốn cho tổ chức phát hành.

*Thị trường thứ cấp*

* Mua bán lại chứng khoán đã được phát hành.
* Thị giá, phụ thuộc giá trị và quan hệ cung cầu.
* Việc mua bán chứng khoán không làm thay đổi lượng vốn đã huy động của nhà phát hành.

**b) Vai trò của thị trường chứng khoán**

***Đối với nhà phát hành***

* Tạo ra một kênh huy động vốn mới với chi phí vốn hợp lý.
* Thúc đẩy quá trình cổ phần hóa doanh nghiệp.
* Tạo ra môi trường cạnh tranh lành mạnh nhằm nâng cao hiệu quả sử dụng vốn, kích thích áp dụng công nghệ mới, cải tiến sản phẩm;
* Hiệu quả của quốc tế hóa thị trường chứng khoán.
* Tạo điều kiện cho việc tách biệt giữa sở hữu và quản lý.

***Đối với nhà đầu tư***

* Phân tích và sàng lọc một danh mục đầu tư hấp dẫn và đầu tư vốn theo danh mục đó;
* Hình thành văn hóa đầu tư;
* Thị trường chứng khoán cung cấp môi trường đầu tư lành mạnh với các cơ hội đầu tư phong phú.

***Đối với Chính phủ và nền kinh tế***

* Huy động các nguồn lực tài chính và tạo ra các công cụ cho việc thực hiện chính sách tài chính tiền tệ.
* Thu hút vốn đầu tư trực tiếp và gián tiếp từ nước ngoài.
* Tạo ra kênh huy động vốn có hiệu quả cho nền kinh tế.

***Một số tiêu cực của thị trường chứng khoán***

* Thông tin không hoàn hảo.
* Hiện tượng đầu cơ.
* Mua bán nội gián.
* Thao túng thị trường.
* Xung đột quyền lợi.

**Một số thị trường chứng khoán thế giới**

*Thị trường chứng khoán mỹ*

**

Hình 5 Đoạn phố Wall trước Sở giao dịch chứng khoán New York

* Thị trường chứng khoán Mỹ ra đời năm 1972 là thị trường phát triển mạnh và lâu đời bậc nhất.
* Cơ cấu:
  + Sở giao dịch chứng khoán New York (NYSE);
  + Sở giao dịch chứng khoán Mỹ (AMEX) (1908);
  + Thị trường OTC Nasdaq (1971).
* Các chỉ số chứng khoán: DowJones, S&P 500, NASDAQ
* Ngày 4/4/2007, NYSE và Euronext ký thỏa thuận sáp nhập để trở thành Sở giao dịch lớn nhất thế giới NYSE Euronext.
  + Trụ sở: Số 11 Phố Wall, Manhattan, New York, USA.
  + Mức vốn hóa thị trường: 16,613 tỉ USD (5/2013).
  + Giá trị giao dịch hàng ngày: 153 tỉ USD(2008).
  + 2308 công ty niêm yết.
* Tháng 10/2008, NYSE Euronext mua lại AMEX với giá 260 triệu USD bằng cổ phiếu.
* NASDQD hiện vẫn là sàn lớn thứ 2 về mức vốn hóa thị trường sau NYSE Euronext.

*Thị trường chứng khoán nhật bản*

**

Hình 6 Sở giao dịch chứng khoán Tokyo

* Lớn nhất là Sở giao dịch chứng khoán Tokyo, lớn thứ 3 thế giới về mức vốn hóa thị trường.
* Thành lập năm 1878
* Thành viên: 2292
* Vốn hóa thị trường: 3.3 ngàn tỉ USD (12/2011)
* Hai chỉ số chứng khoán: Nikkei 225 và TOPIX
* Ngoài ra, ở Nhật còn có thị trường chứng khoán phi tập trung Jasdaq (do Sở Osak quản lý) và 7 sở giáo dịch chứng khoán khác.

*Thị trường chứng khoán việt nam*

**

Hình 7 Sở Giao dịch Chứng khoán Thành phố Hồ Chí Minh (HOSE)

* Ngày 11-7-1998 Chính phủ đã ký Nghị định số 48/CP ban hành về chứng khoán và TTCK chính thức khai sinh cho Thị trường chứng khoán Việt Nam ra đời.
* Cùng ngày, Chính phủ cũng ký quyết định thành lập Trung tâm Giao dịch Chứng khoán đặt tại TP.HCM và Hà Nội. Trung tâm Giao dịch Chứng khoán TP.HCM (TTGDCK TP.HCM) được thành lập theo Quyết định số 127/1998/QĐ-TTg ngày 11-7-1998 và chính thức đi vào hoạt động thực hiện phiên giao dịch đầu tiên vào ngày 28-7-2000. Trung tâm Giao dịch chứng khoán (TTGDCK) Hà Nội đã chính thức chào đời vào ngày 8/3/2005.
* Khác với TTGDCK TP.HCM (vốn là nơi niêm yết và giao dịch chứng khoán của các công ty lớn), TTGDCK Hà Nội sẽ là “sân chơi” cho các DN nhỏ và vừa (với vốn điều lệ từ 5 đến 30 tỷ đồng). Từ đó đến nay, TTCK VN đã không ngừng biến động và ngày càng chứng tỏ rằng nó là 1 trong những thị trường được quan tâm nhất hiện nay.
* Các chỉ số chứng khoán: Phiên giao dịch đầu tiên với 2 cổ phiếu là REE và SAM. Hiện nay thị trường chứng khoán đang giao dịch với các chỉ số chính: VN-INDEX, VN30-INDEX, HNX-INDEX, HNX30-INDEX, UPCOM-INDEX.
* Trải qua gần 20 năm vận hành, thị trường chứng khoán Việt Nam đã phát triển tới mức: giá trị bình quân giao dịch đạt trên 3.000 tỷ đồng/phiên; hơn 700 công ty và quỹ đầu tư niêm yết trên HOSE và HNX; thị trường đã huy động được khoảng 80 tỷ USD cho nền kinh tế.

## 2.2 Cổ phiếu và định giá cổ phiếu

**a) Cổ phiếu**

*Khái niệm*

Là giấy chứng nhận số tiền nhà đầu tư đóng góp vào công ty phát hành. Cổ phiếu là chứng chỉ do công ty cổ phần phát hành hoặc bút toán ghi sổ xác nhận quyền sở hữu một hoặc một số cổ đông của công ty đó. Nguời nắm giữ cổ phiếu trở thành cổ đông và đồng thời là chủ sở hữu của công ty phát hành.

*Phân loại cổ phiếu*

Có nhiều cách phân loại cổ phiếu.

* Căn cứ vào việc lưu hành trên thị trường
  + Cổ phiếu lưu hành: số cổ phiếu đang nằm trong tay cổ đông của công ty.
  + Cổ phiếu ngân quỹ: là cổ phiếu công ty đã phát hành và mua lại từ cổ đông trên thị trường thứ cấp. Cổ phiếu ngân quỹ không có quyền bầu cử và không được nhận cổ tức. Công ty có thể giữ cổ phiếu vô hạn, huỷ hay bán ra.
* Căn cứ vào việc phát hành vốn điều lệ
  + Cổ phiếu sơ cấp là cổ phiếu được phát hành lần đầu tiên và tiền thuộc về công ty phát hành.
  + Cổ phiếu thứ cấp là cổ phiếu lưu hành sau khi đã phát hành cổ phiếu sơ cấp.
* Căn cứ vào quyền của cổ đông (cách phổ biến nhất)
  + Cổ phiếu phổ thông: là loại cổ phiếu cơ bản của công ty, cho phép cổ đông được hưởng các quyền lợi thông thường trong công ty, được phát hành ra đầu tiên và thu hồi về cuối cùng, cổ đông có quyền biểu quyết và tham gia quyết định tất cả các vấn đề quan trọng ảnh hưởng đến hoạt động của công ty, được hưởng cổ tức theo kết quả kinh doanh.
  + Cổ phiếu ưu đãi: là loại cổ phiếu phát hành sau cổ phiếu thường, cho phép người nắm giữ được hưởng một số quyền lợi ưu đãi hơn so với cổ phiếu thường.

**b) Công ty niêm yết trên thị trường chứng khoán là gì?**

Công ty niêm yết (listed company) là công ty công cộng có cổ phiếu được phép mua bán trên các thị trường chứng khoán.

Điều 6, Luật chứng khoán 2006 có đưa ra định nghĩa về niêm yết chứng khoán, theo đó, niêm yết chứng khoán là việc đưa các chứng khoán có đủ điều kiện vào giao dịch tại Sở giao dịch chứng khoán hoặc Trung tâm giao dịch chứng khoán.

Điều kiện trở thành công ty niêm yết

Điều 12 - Luật chứng khoán năm 2006 có quy định về điều kiện chào bán cổ phiếu ra công chúng gồm 3 điều kiện.

Cụ thể, doanh nghiệp có mức vốn điều lệ đã góp tại thời điểm đăng ký chào bán từ 10 tỷ đồng Việt Nam trở lên tính theo giá trị ghi trên sổ kế toán; Hoạt động kinh doanh của năm liền trước năm đăng ký chào bán phải có lãi, đồng thời không có lỗ luỹ kế tính đến năm đăng ký chào bán; Có phương án phát hành và phương án sử dụng vốn thu được từ đợt chào bán được Đại hội đồng cổ đông thông qua.

# 3. Tổng quan về học máy

## 3.1 Tổng quan

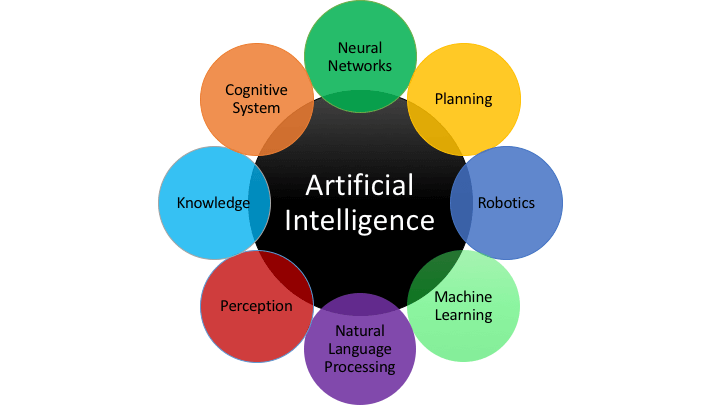
Những năm gần đây, AI - Artificial Intelligence (Trí Tuệ Nhân Tạo), và cụ thể hơn là Machine Learning (Học Máy hoặc Máy Học) nổi lên như một bằng chứng của cuộc cách mạng công nghiệp lần thứ tư.

AI là các kỹ thuật giúp cho máy tính thực hiện được những công việc của con người chúng ta. Ví dụ như một chương trình cờ vua tự động có thể coi là một chương trình có sử dụng AI hay viết tắt là một chương trình AI.

Trong lĩnh vực AI có một nhánh nghiên cứu về khả năng tự học của máy tính được gọi là học máy (machine learning). Hiện nay không có 1 định nghĩa chính thức nào về học máy cả nhưng có thể hiểu rằng nó là các kỹ thuật giúp cho máy tính có thể tự học mà không cần phải cài đặt các luật quyết định. Thường một chương trình máy tính cần các quy tắc, luật lệ để có thể thực thi được một tác vụ nào đó như dán nhãn cho các email là thư rác nếu nội dung email có chứ từ khoá “quảng cáo”. Nhưng với học máy, các máy tính có thể tự động phân lại các thư rác thành mà không cần chỉ trước bất kỳ quy tắc nào cả. Nói hơi khó thoát ý, nhưng có thể hiểu đơn giản là nó giúp cho máy tính có được cảm quan và suy nghĩ được như con người. Còn nếu nói nôm na kỹ thuật một chút thì học máy là phương pháp vẽ các đường thể hiện mối quan hệ của tập dữ liệu. Ví dụ như đường ngăn cách 2 loại dữ liệu cho nhãn khác nhau, đường thể hiện xu hướng của giá nhà phụ thuộc vào diện tích và trí hay các đường phân cụm dữ liệu.

Hình 8 Mối liên hê giữa AL, Machine learning, Deep Learning

Một nhánh nhỏ trong học máy gần đây rất được ưu chuộng là học sâu (deep learning). Học sâu là kỹ thuật sử dụng các mạng nơ-ron tương tự như các nơ-ron của não người để xây dựng hệ thống học máy. Đây là một sự kết hợp tuyệt vời giữa toán học và khoa học thần kinh. Kết quả của nó mang lại cực kỳ to lớn, có thể coi là khởi nguyên của ngành công nghiệp mới. Tại thời điểm này, hầu hết các anh lớn cả trong ngành công nghệ lẫn các ngành khác như ôto, điện tử đều đang tập trung phát triển và ứng dụng kỹ thuật học sâu cho bài toán của mình. Ví dụ như AlphaGo của Google đã chiến thắng nhà vô địch cờ vây Lee Sedol vào tháng 3 năm 2016. Tính năng nhận diện khuôn mặt khá chính xác của Facebook được triển khai vào năm 2016. Trợ lý ảo Siri của Apple được giới thiệu từ năm 2006. Xe tự lái của Google được thử nghiệm chính thức trên đường phố vào năm 2015,…



Hình 9 AI Landscape

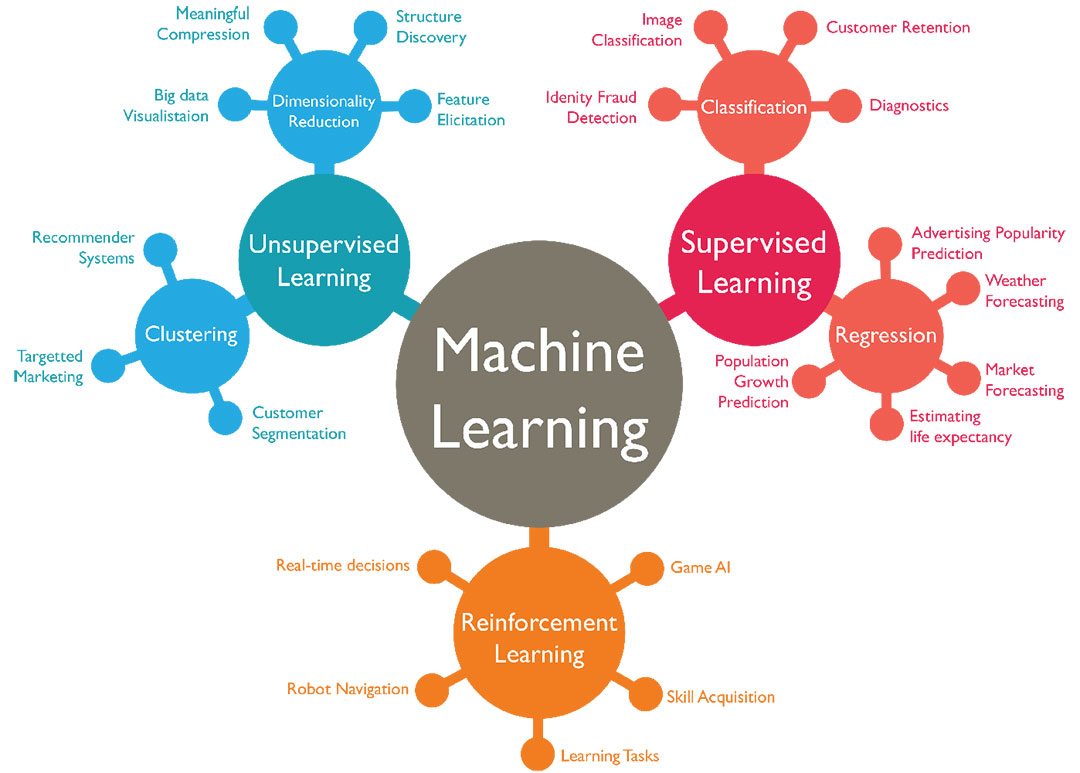
Học máy (Machine Learning) là một ngành khoa học nghiên cứu các thuật toán cho phép máy tính có thể học được các khái niệm (concept).

Phân loại: Có hai loại phương pháp học máy chính

* **Phương pháp quy nạp:** Máy học/phân biệt các khái niệm dựa trên dữ liệu đã thu thập được trước đó. Phương pháp này cho phép tận dụng được nguồn dữ liệu rất nhiều và sẵn có.
* **Phương pháp suy diễn:** Máy học/phân biệt các khái niệm dựa vào các luật. Phương pháp này cho phép tận dụng được các kiến thức chuyên ngành để hỗ trợ máy tính.

Hiện nay, các thuật toán đều cố gắng tận dụng được ưu điểm của hai phương pháp này.

Các ngành khoa học liên quan:

* Lý thuyết thống kê: các kết quả trong xác suất thống kê là tiền đề cho rất nhiều phương pháp học máy. Đặc biệt, lý thuyết thống kê cho phép ước lượng sai số của các phương pháp học máy.
* Các phương pháp tính: các thuật toán học máy thường sử dụng các tính toán số thực/số nguyên trên dữ liệu rất lớn. Trong đó, các bài toán như: tối ưu có/không ràng buộc, giải phương trình tuyến tính, v.v… được sử dụng rất phổ biến.
* Khoa học máy tính: là cơ sở để thiết kế các thuật toán, đồng thời đánh giá thời gian chạy, bộ nhớ của các thuật toán học máy.

Hình 10 Phân loại Machine Learning

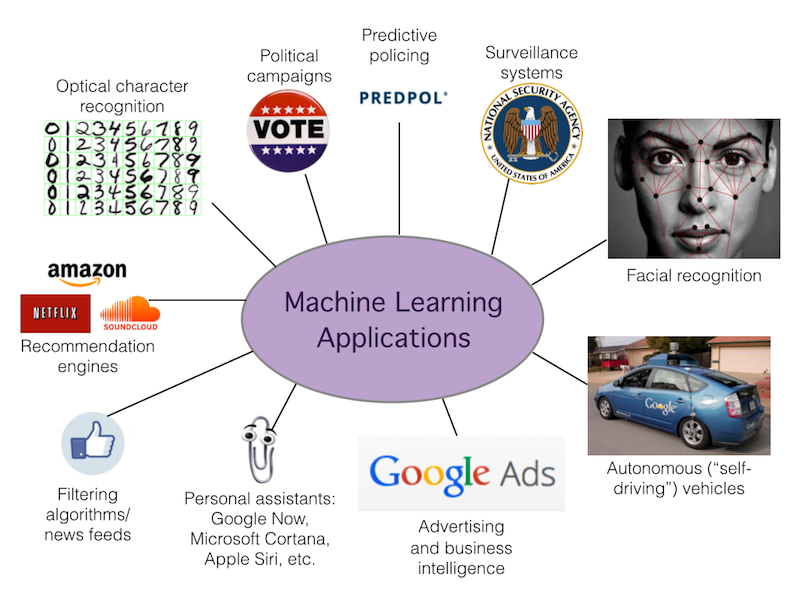
Các nhóm giải thuật học máy:

* **Học có giám sát (Supervised Learning):** Máy tính được sử dụng mẫu gồm đầu vào (input) và đầu ra (output) tương ứng trước. Sau khi học xong các mẫu này, máy tính dự đoán một đầu vào mới và cho ra kết quả.
* **Học không giám sát (Unsupervised Learning):** Máy tính chỉ được xem các mẫu không có đầu ra, sau đó máy tính phải tự tìm cách phân loại các mẫu này và các mẫu mới.
* **Học nửa giám sát (Semi-Supervised Learning):** Một dạng lai giữa hai nhóm giải thuật trên.
* **Học tăng cường (Reinforcement Learning):** Máy tính đưa ra quyết định hành động (action) và nhận kết quả phản hồi (response/reward) từ môi trường (environment). Sau đó máy tính tìm cách chỉnh sửa cách ra quyết định hành động của mình.

## 3.2 Các ứng dụng của học máy

Học máy có ứng dụng rộng khắp trong các ngành khoa học/sản xuất, đặc biệt những ngành cần phân tích khối lượng dữ liệu khổng lồ. Một số ứng dụng thường thấy:

* Xử lý ngôn ngữ tự nhiên (Natural Language Processing): xử lý văn bản, giao tiếp người – máy,…
* Nhận dạng (Pattern Recognition): nhận dạng tiếng nói, chữ viết tay, vân tay, thị giác máy (Computer Vision).
* Tìm kiếm (Search Engine).
* Chẩn đoán trong y tế: phân tích ảnh X-quang, các hệ chuyên gia chẩn đoán tự động.
* Tin sinh học: phân loại chuỗi gene, quá trình hình thành gene/protein.
* Vật lý: phân tích ảnh thiên văn, tác động giữa các hạt.
* Phát hiện gian lận tài chính (financial fraud): gian lận thẻ tỉn dụng.
* Phân tích thị trường chứng khoán (stock market analysis).
* Chơi trò chơi: tự động chơi cờ, hành động của các nhân vật ảo.



Hình 11 Machine Learning và một số ứng dụng

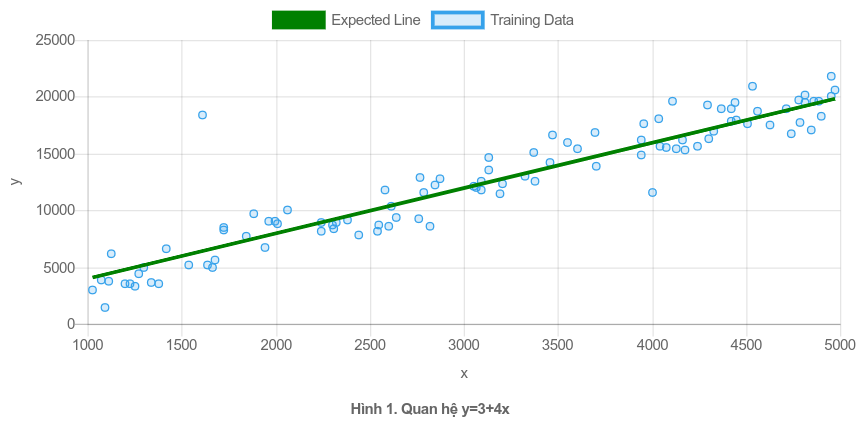
# 4. Phương pháp thực hiện

## 3.1 Đặt vấn đề

Theo nghiên cứu, thì trong machine learning có rất nhiều thuật toán dùng để dự báo (dự đoán) kết quả dựa trên số liệu thống kê chúng ta có trong tập training data. Ở bài toán này, nhóm quyết định sử dụng một thuật toán có thể xem là cơ bản nhất trong machine learning đó là thuật toán Linear Regression (Hồi quy truyến tính). Thuật toán này hiểu một cách đơn giản nhất là chỉ ra quan hệ tuyến tính giữa một biến phụ thuộc vào một hay nhiều biến độc lập. Mục tiêu của thuật toán hồi quy truyến tính là dự đoán giá trị của một hoặc nhiều biến liên tục y dựa trên một vec-tơ đầu vào x.

Ví dụ: Dự đoán giá nhà ở Thành phố Hồ Chí Minh dựa vào diện tích, vị trí, năm xây dung. Thì ở đây y là giá của ngôi nhà và x = () với là diện tích của ngôi nhà, là vị trí ngôi nhà, là năm xây dựng của ngôi nhà.

Đây chính là phương pháp phân tích hồi quy của xác xuất thống kê. Mọi lý thuyết cơ bản của phương pháp này vẫn được giữ nguyên tuy nhiên khi áp dụng cho máy tính thì mặt cài đặt có thay đổi đôi chút. Về cơ bản thì ta sẽ có một tập huấn luyện chứa các cặp tương ứng và nhiệm của ta là phải tìm giá trị y ứng với một đầu vào x mới. Để làm điều này, ta cần tìm được quan hệ giữa x và y để từ đó đưa ra dự đoán. Hay nói cách trừu tượng hơn là ta cần vẽ được một đường quan hệ thể hiện mối quan hệ trong tập dữ liệu.



Hình 12 Quan hệ y=3+4x

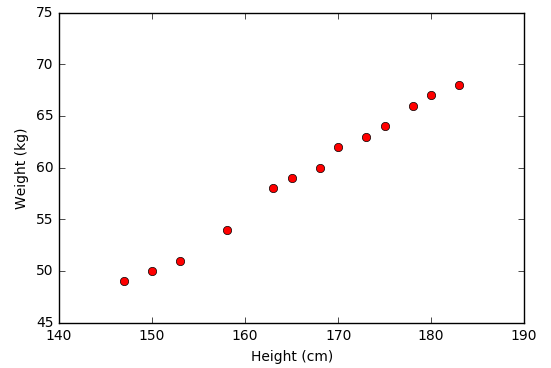
Như hình trên minh họa một đường xanh lá cây thể hiện quan hệ giữa x và y dựa trên các điểm dữ liệu huấn luyện đã biết. Thuật toán hồi quy tuyến tính sẽ giúp ta tự động tìm được đường màu xanh đó để ta có thể dự đoán được y cho một x chưa từng xuất hiện bao giờ.

## 3.2 Thuật toán Linear Regression

Ví dụ: Ở đây ta có một bảng dữ liệu về chiều cao và cân nặng của 13 người:

|  |  |
| --- | --- |
| Chiều cao (cm) | Cân nặng (kg) |
| 147 | 49 |
| 150 | 50 |
| 153 | 51 |
| 158 | 54 |
| 163 | 58 |
| 165 | 59 |
| 168 | 60 |
| 170 | 62 |
| 173 | 63 |
| 175 | 64 |
| 178 | 66 |
| 180 | 67 |
| 183 | 68 |

Ở đây, việc dự đoán cân nặng dự vào chiều cao của một người trên thực tế là không khả thi vì nó còn phụ thuộc vào nhiều yếu tố khác. Tuy nhiên để có thể đưa ra ví dụ về thuật toán thì ta giả sử việc này là khả thi.



Hình 13 Biểu diễn dữ liệu

Ở hình này thì các cặp điểm giá trị (x,y) được sắp xếp gần như là một đường thẳng nên khả năng cao mô hình thuật toán Linear Regression sẽ cho kết quả tốt.

## 3.3 Cách xác định đường hồi quy tuyến tính

Đường hồi quy tuyến tính là đường thẳng có dạng phương trình:

Với x,y là cặp giá trị tương ứng với chiều cao và cân nặng (y là biến phụ thuộc và x là biến độc lập). và là 2 tham số cần tìm ( là hệ số mối quan hệ giữa x và y, là hằng số).

Ta xác định 2 giá trị và như sau:

= (Tương quan x và y) \* (Độ lệch chuẩn y / Độ lệch chuẩn x)

= Trung bình (y) - \* Trung bình (x)

**Bước 1**: Ta tìm giá trị trung bình của x và y

**Bước 2**: Tìm độ lệch chuẩn của x và y

Giá trị độ lệch chuẩn là căn của phương sai và phương sai có công thức

Vậy độ lệch chuẩn của x và y lần lượt là:

**Bước 3**: Tính độ tương quan giữa x và y:

Độ tương quan = (hiệp phương sai x và y) / (độ lệch chuẩn x \* độ lệch chuẩn y)

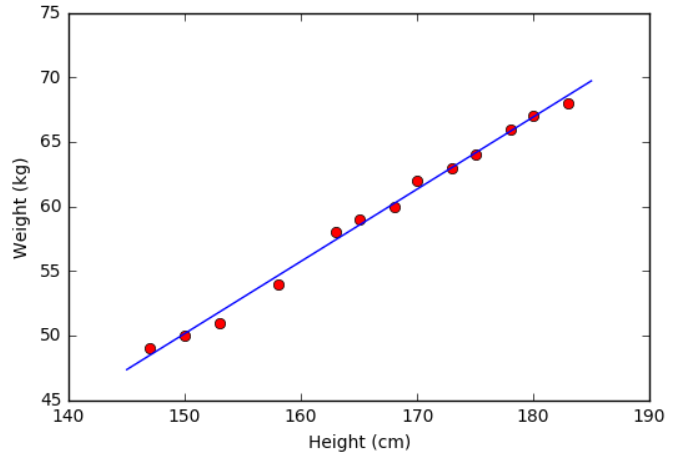
Hiệp phương sai có công thức:

Vậy

**Bước 4**: Như vậy ta sẽ tính được và từ những thông số đã tính được ở trên.

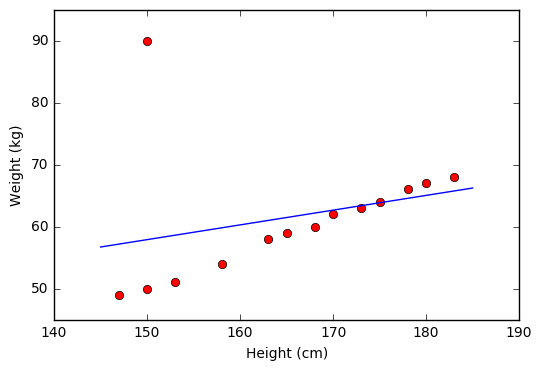
Vậy đường hồi quy của bảng dữ liệu trên sẽ có phương trình



Kết quả khi chạy thuật toán trên Python so với kết quả khi chạy tay cũng khá tương đồng. Và đồ thị sẽ được biểu diễn như sau:

Hình 14 Đồ thị của đường hồi quy

## 3.4 Hạn chế của Linear Regression

Hạn chế đầu tiền của Linear Regression là nó rất nhạy cảm với nhiễu. Chỉ cần một tập nhiễu thì kết quả sẽ khác đi rất nhiều như sau:

Hình 15 Hạn chế của hồi quy

Vì vậy trước khi thực hiện thuật toán Linear Regression thì các nhiễu cần phải được loại bỏ và bước làm này được gọi là tiền xử lý.

Hạn chế thứ hai của Linear Regression là nó không biểu diễn được các mô hình phức tạp. Mặc dù trong phần trên, ta thấy rằng phương pháp này có thể áp dụng nếu quan hệ giữa outcome và input không nhất thiết phải tuyến tính, nhưng mối quan hệ này vẫn đơn giản hơn nhiều so với mô hình thực tế.

Ở đây ta có một demo nhỏ về thuật toán này và bạn có thể lấy source code tại đường link:

<https://github.com/littlepea77/ML_PredictingStockPrices>

Tuy nhiên để sử dụng thì các bạn cần cài một số thư viện Python như sau:

**Scikit-learn**: thư viện viết sẵn cách giải các thuật toán, trong đó có Linear Regression.

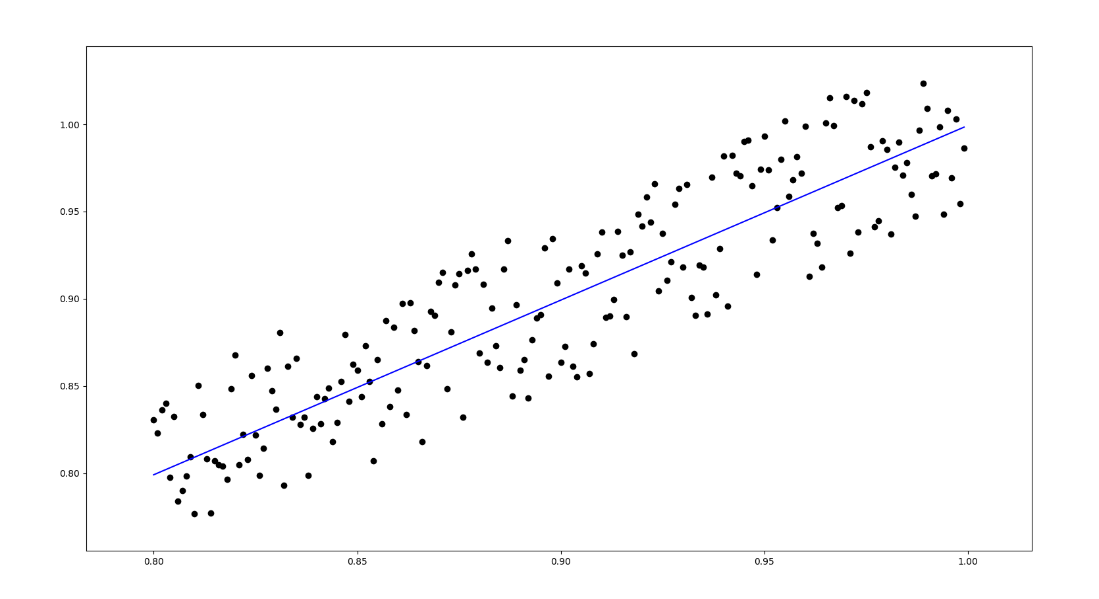
**Scipy**: thư viện xử lý toán học phức tạp, bắt buộc cài để có thể dùng scikit-learning.

**Numpy**: thư viện tính toán các dạng véc-tơ, ma trận. Ta dùng để tính toán sai số của đường hồi quy.

**Matplotlib**: thư viện mô phỏng dữ liệu bằng hình ảnh để ta có thể quan sát một cách trực quan, ta dùng để vẽ đường hồi quy.

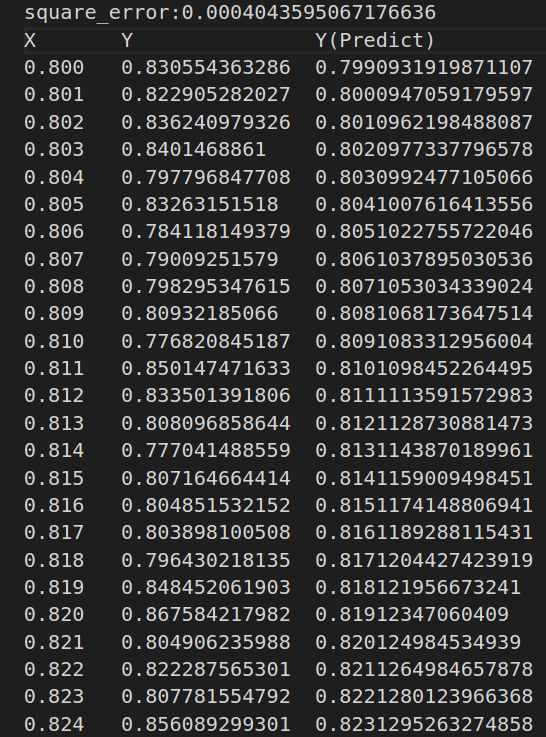
**Pandas**: thư viện tương tác hệ điều hành, ta dùng để đọc file input.

Sau khi chạy thì chương trình sẽ giúp ta nhìn thấy được đường hồi quy thể hiện mối tương quan một cách rõ ràng.



Hình 16 Ví dụ thực tế của hồi quy

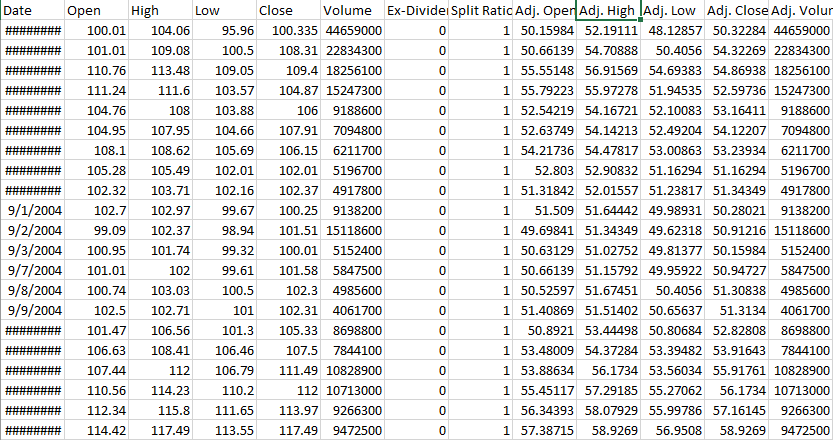
Ngoài ra chương trình giúp ta thấy được giá trị y được dự đoán (Y Predict) là bao nhiêu và sai số (Square\_error) so với thực tế như thế nào.



Hình 17 So sánh y và y(predict)

# 5. Ứng dụng thuật toán Regression vào bài toán

Trước tiên, ta có một file dữ liệu trong đó gồm các cột lần lượt chứa các thông tin như: Date, Open, High, Low, Close, Volume,…



Hình 18 Dữ liệu cổ phiếu

Tuy nhiên ở đây ta chỉ chú ý đến các cột Adj.High, Adj.Low, Adj.Open, Adj.Close, Adj.Volume và ta sẽ tiến hành get data ở các cột đó.



Ở đây ta dự đoán dựa trên cột “Adj.Close”. Và dự đoán đầu ra là 1% của bộ dữ liệu hay nói cách khác là ta dùng 99% của bộ dữ liệu để training và dùng 1% còn lại để kiểm tra đường hồi quy tạo ra giá trị dự đoán y có sát với thực tế hay không. Hàm preprocessing.scale() sẽ hỗ trợ ta làm việc đó. Phương thức Scale chia tỉ lệ dựa trên tất cả dữ liệu đã biết được đưa vào nó.

forecast\_col = 'Adj. Close'

df.fillna(value=-99999, inplace=True)

forecast\_out = int(math.ceil(0.01 \* len(df)))

df['label'] = df[forecast\_col].shift(-forecast\_out)

X = np.array(df.drop(['label'], 1))

X = preprocessing.scale(X)

X\_lately = X[-forecast\_out:]

X = X[:-forecast\_out]

Bộ dữ liệu training của ta ở đây cũng không quá lớn và thời gian thực hiện cũng không quá dài nên ta có thể chia tỉ lệ trên tập training, tập testing và thực hiện dự báo/dự đoán dữ liệu cùng một lúc.

Giá cổ phiếu thì tăng/giảm mỗi ngày nên ta sẽ thực hiện việc dự báo một cách đơn giản với ý tưởng là vẽ từng dự báo như thể nó chỉ đơn giản là hết 1 ngày. Trước tiên ta cần import một vài thư viện Python sau:

import datetime

import matplotlib.pyplot as plt

from matplotlib import style

Datetime: thư viện giúp ta làm việc với thời gian thực.

Matplotlib: thư viện giúp ta mô phỏng dữ liệu bằng hình ảnh.

Trước tiên ta cần lấy ngày cuối cùng trong khung dữ liệu và bắt đầu gán từng dự báo cho một ngày mới.

last\_date = df.iloc[-1].name

last\_unix = last\_date.timestamp()

one\_day = 86400

next\_unix = last\_unix + one\_day

for i in forecast\_set:

next\_date = datetime.datetime.fromtimestamp(next\_unix)

next\_unix += 86400

df.loc[next\_date] = [np.nan for \_ in range(len(df.columns)-1)]+[i]

Ở đây, việc ta làm là lặp đi lặp lại thông qua forecast\_set, lấy từng dự báo và ngày của dự báo đó, sau đó thiết lập một giá trị trong bộ dữ liệu (tạo một “features” NaN). Hiểu đơn giản là lấy tất cả giá trị ở cột đầu tiên, thiết lập chúng thành NaN và cột cuối cùng là dự báo trong bài toán này.

Ta cũng dễ dàng thấy được dựa vào kết quả của bộ dữ liệu đã được training (2004-2018) ta sẽ dự báo được giá trị của cổ phiếu ở thời điểm tiếp theo. Và dưới đây là toàn bộ phần source code mà mình đã chỉnh sửa để dự đoán kết quả ***từ 1 ngày thành 60 ngày****:*

import pandas as pd

import numpy as np

import math, datetime, time

from sklearn import preprocessing, svm

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LinearRegression

import matplotlib.pyplot as plt

from matplotlib import style

style.use('ggplot')

# run download\_dataset.py first!

df = pd.read\_csv('wiki-googl.csv', index\_col='Date', parse\_dates=True)

# compute new stats to use as features

df['HL\_PCT'] = (df['Adj. High'] - df['Adj. Low']) / df['Adj. Low'] \* 100.0

df['PCT\_change'] = (

df['Adj. Close'] - df['Adj. Open']) / df['Adj. Open'] \* 100.0

# we are only going to use those columns

df = df[['Adj. Close', 'HL\_PCT', 'PCT\_change', 'Adj. Volume']]

forecast\_col = 'Adj. Close'

#forecast\_col = 'PCT\_change'

df.fillna(-99999, inplace=True)

forecast\_out = 30

X = np.array(df)

X = preprocessing.scale(X) # normalization

y = np.array(df[forecast\_col])

# training

X\_train = X[:-2 \* forecast\_out]

y\_train = y[forecast\_out:-forecast\_out]

X\_train, X\_test, y\_train, y\_test = train\_test\_split(

X\_train, y\_train, test\_size=0.2)

clf = LinearRegression()

clf.fit(X\_train, y\_train)

accuracy = clf.score(X\_test, y\_test)

print('Accuracy:', accuracy)

# predict last 30 days

X\_eval30 = X[-2 \* forecast\_out:-forecast\_out]

y\_eval\_forecast = clf.predict(X\_eval30)

# y\_eval30 = X[-forecast\_out:] # expected result

df['Prediction'] = np.nan

df['Prediction'][-forecast\_out:] = y\_eval\_forecast

# predict 30 days into future

X\_pred30 = X[-forecast\_out:]

y\_pred\_forecast = clf.predict(X\_pred30)

last\_date = df.iloc[-1].name

dt = datetime.datetime.fromtimestamp(last\_date.timestamp())

for i in y\_pred\_forecast:

dt += datetime.timedelta(days=1)

df.loc[dt] = [np.nan for \_ in range(len(df.columns) - 1)] + [i]

df[forecast\_col].plot()

df['Prediction'].plot()

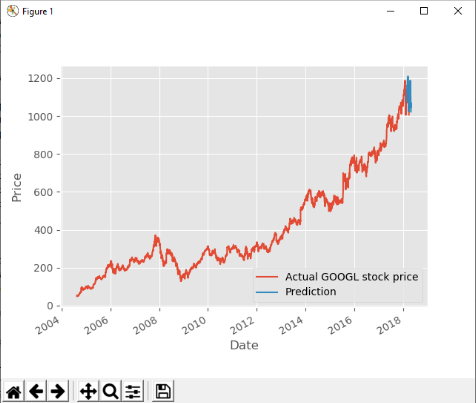
plt.xlabel('Date')

plt.ylabel('Price')

plt.legend(['Actual GOOGL stock price', 'Prediction'], loc=4)

plt.show()

# 6. Kết quả và đánh giá

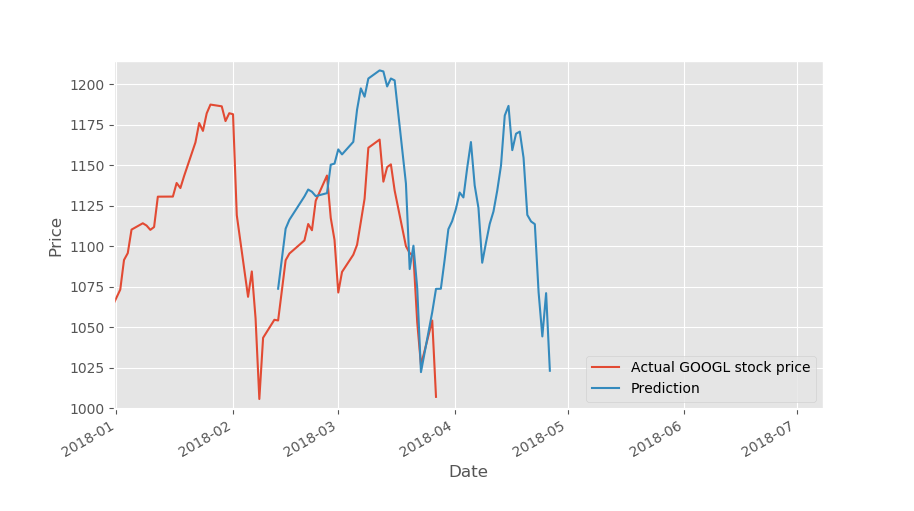
Cuối cùng là đồ thị để ta có thể nhìn một cách trực quan hơn.

Hình 19 Đồ thị thể hiện sự biến động và dự đoán của cổ phiếu

***Đánh giá***

Nhóm đã sử dụng thuật toán Linear Regression là thuật toán tuyến tính, dự đoán sự tăng/giảm trong 1 ngày sau đó ghép lại thành một đường đồ thị biểu thị việc dự đoán. Do đó, việc này gặp nhiều sai sót hạn chế khi so với thuật toán phi tuyến tính khác. Khi so sánh kết quả dự đoán và kết quả thực tế, ta thấy có sự sai lệch đáng kể.

Như hình dưới, ta thấy rằng đường dự đoán màu xanh chênh lệch tương đối so với đường màu đỏ, nhưng nhìn chung có nhiều đoạn có xu hướng tăng giảm giống với thực tế đã xảy ra. Đây là điểm bất lợi mà ta đã nói ở trên.



Hình 20 So sánh thực tế và dự đoán được

# 7. Kết luận

Việc thực hiện đồ án tiểu luận đã giúp nhóm tiếp cận với các khía cạnh của ML hiểu được tầm quan trọng của nó. Đồ án còn giúp ta phân biệt được các thuật toán Regression. Tìm hiểu về chứng khoán và thị trường chứng khoán khác biệt như thể nào trong thời đại Cách mạng công nghiệp 4.0 so với thời kì sơ khai và hoàng kim của nó. Dù cho kết quả cuối cùng chưa thực sự tốt, việc dự báo còn ở mức cơ bản, chưa đi sâu chi tiết do còn hạn chế về mặt kiến thức nhưng trong tương lai nhóm sẽ cố gắng hoàn thiện sản phẩm hơn.

Tổng kết lại, ta có thể thấy con người luôn cố gắng vượt qua giới hạn bản thân (mà ở đây là sử dụng máy tính và các thiết bị điện tử) để cải thiện, gia tăng hiệu suất công việc của mình. Machine learning là một trong những công cụ giúp cho chúng ta có thể giải quyết những bài toán mà phương pháp lập trình truyền thống chỉ dựa trên tiến trình không thể nào nhanh và hiệu quả được. Bằng cách vay mượn một lượng lớn dữ liệu có sẵn, ta có thể “dạy” cho máy tính “học” để chúng có khả năng tự động thực hiện những tác vụ hữu ích cho con người và điển hình nhất là trong kinh tế - tài chính, mà ở trong đề tài là dự đoán xu hướng tăng giảm của cổ phiếu.

# Tài liệu tham khảo

1. GS.TS Nguyễn Văn Nam & PGS.TS Vương Trọng Nghĩa, Giáo trình thị trường chứng khoán, NXB Tài chính, 2002.
2. Trần Đăng Khâm, Thị trường chứng khoán – Phân tích cơ bản, NXB ĐH Kinh tế Quốc dân, 2009.
3. Tom M. Mitchell, Machine Learning, McGraw Hill, 1997.
4. Giáo trình thị trường chứng khoán, Bộ môn Thị trường chứng khoán, Trường Đại học Kinh tế Quốc dân.
5. Vũ Hữu Tiệp , blog Machine Learning cơ bản, <https://machinelearningcoban.com/>
6. Andrew Ng, course Machine Learning, Stanford university, <https://www.coursera.org/learn/machine-learning>
7. Harrison Kinsley, course Machine Learning with Python, sentdex Youtube channel, <https://www.youtube.com/user/sentdex>