

ỨNG DỤNG THUẬT TOÁN MÁY HỌC ĐÁNH GIÁ XU THẾ TĂNG GIẢM CỔ PHIẾU CỦA CÔNG TY NIÊM YẾT

GVHD: TS Vũ Quang Huy

Nhóm thực hiện:

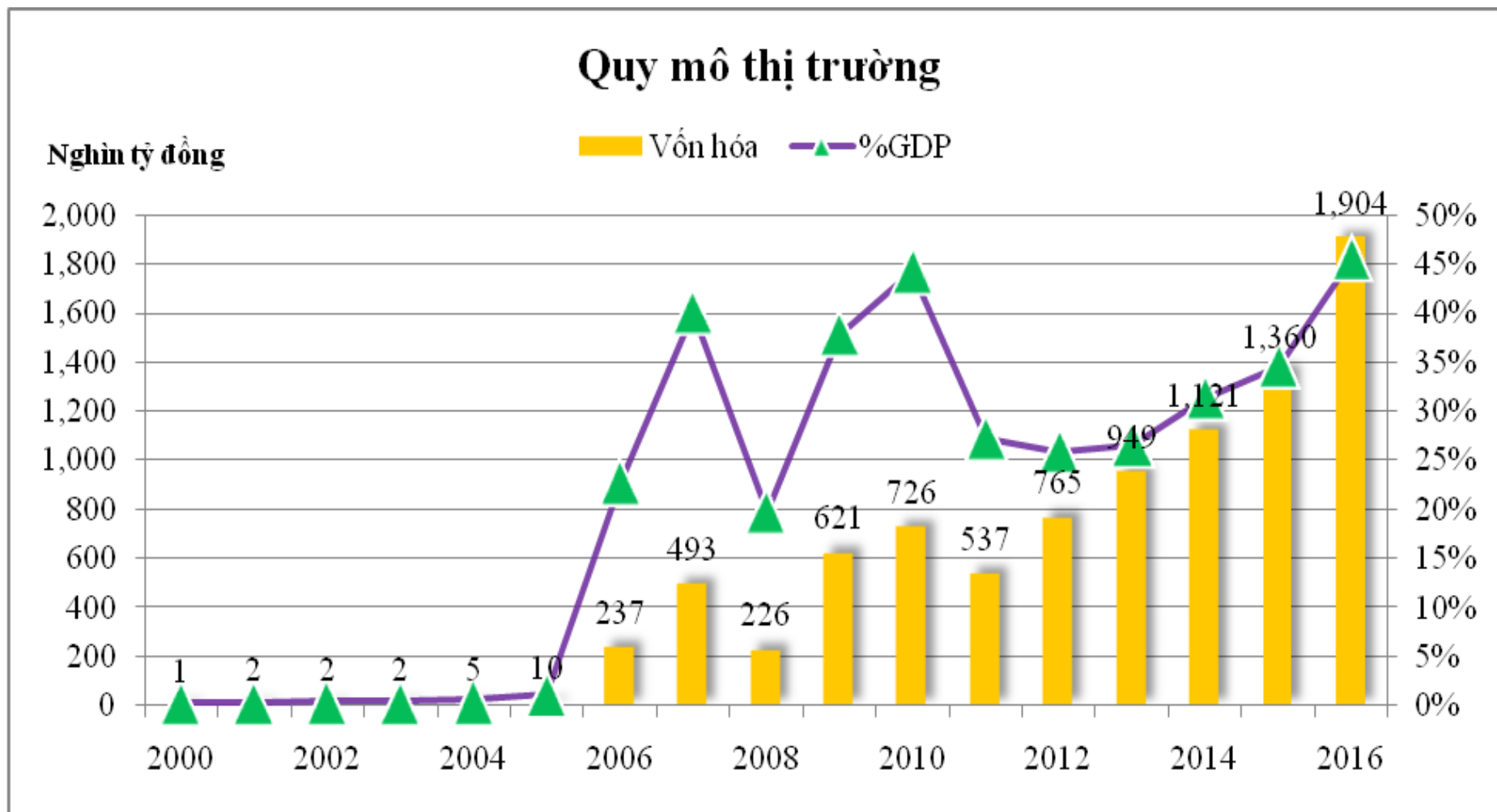
Nguyễn Ngọc Hoàng Phúc 15110099

Phạm Trường Giang 15110036

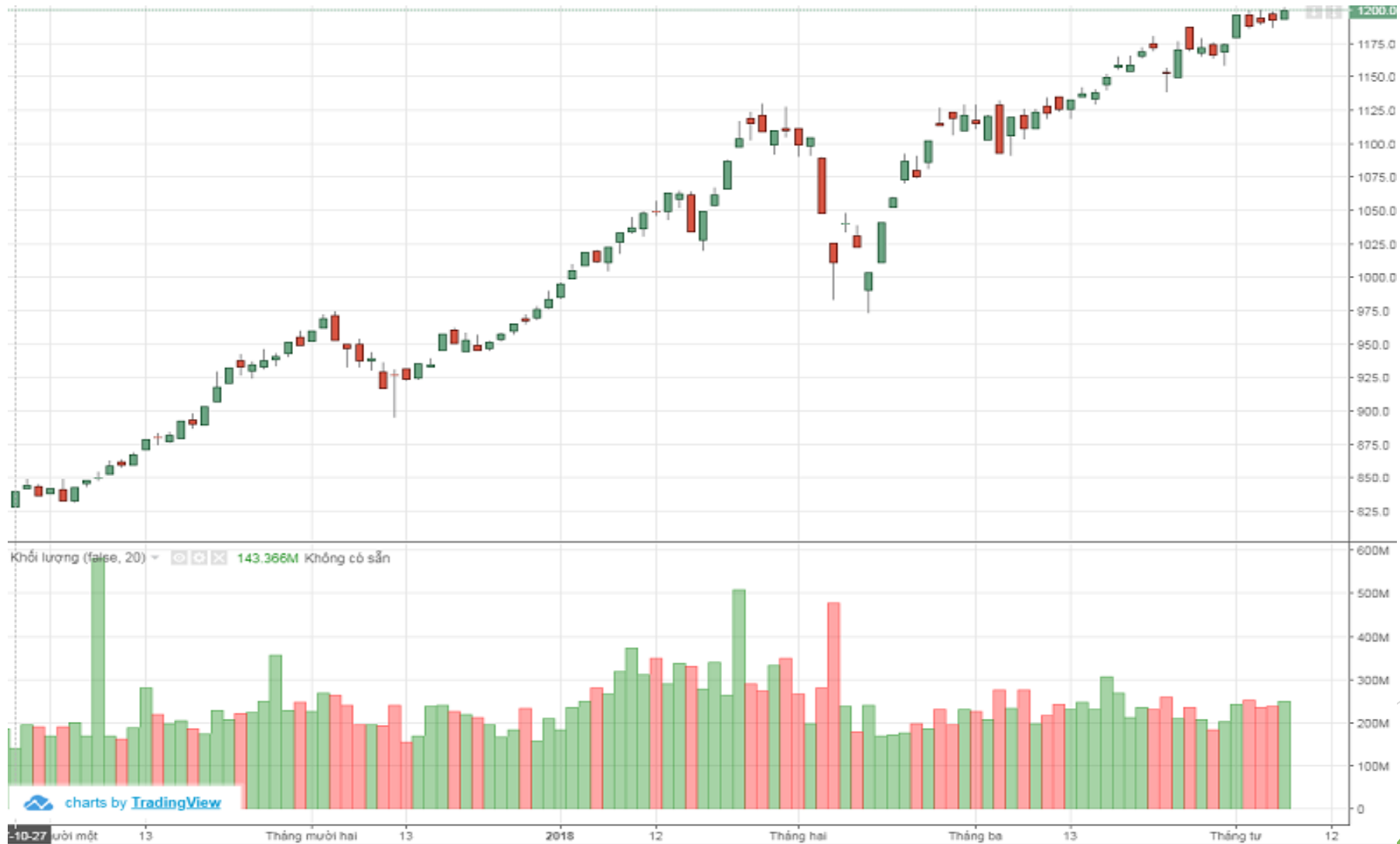
Mục lục.

- ▶ Nhu cầu dự đoán giá trị cổ phiếu.
- ▶ Thuật toán máy học trong dự đoán giá trị cổ phiếu.
- ▶ Ứng dụng.
- ▶ Kết quả, đánh giá.
- ▶ Phương hướng phát triển.

1.Nhu cầu dự đoán giá trị cổ phiếu.



Biểu đồ VNINDEX



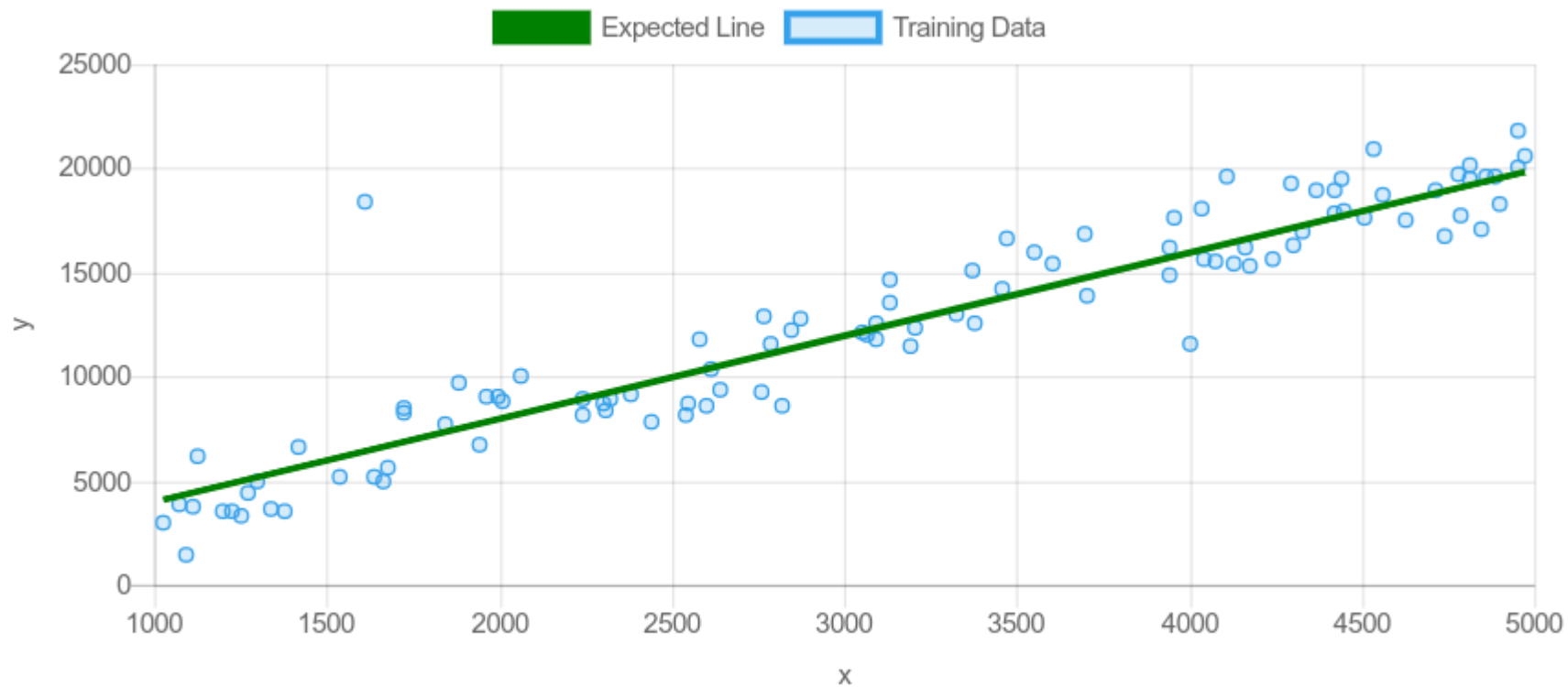
2. Thuật toán máy học trong dự đoán giá trị cổ phiếu.

Tên thuật toán	Sự miêu tả	Kiểu
Hồi quy tuyến tính (Linear regression)	Tìm cách tương quan từng tính năng với đầu ra để giúp dự đoán các giá trị trong tương lai.	hồi quy
Hồi quy logistic (Logistic regression)	Mở rộng hồi quy tuyến tính được sử dụng cho các nhiệm vụ phân loại. Biến đầu ra là nhị phân (ví dụ: chỉ đen hoặc trắng) chứ không phải liên tục (ví dụ: danh sách vô hạn các màu tiềm năng)	Phân loại
Cây quyết định (Decision tree)	Mô hình phân loại hoặc hồi quy có thể hiểu sâu, phân chia các giá trị tính năng dữ liệu thành các nhánh tại các nút quyết định (ví dụ: nếu một tính năng là một màu, mỗi màu có thể trở thành một nhánh mới) cho đến khi đưa ra quyết định cuối cùng	Phân loại hồi quy
Naive Bayes	Phương pháp Bayes là một phương pháp phân loại sử dụng định lý Bayes. Định lý cập nhật kiến thức trước về một sự kiện với xác suất độc lập của từng tính năng có thể ảnh hưởng đến sự kiện.	Phân loại hồi quy
Support vector machine	Support Vector Machine, hoặc SVM, thường được sử dụng cho nhiệm vụ phân loại. Thuật toán SVM tìm thấy một siêu phẳng chia tối ưu các lớp. Nó được sử dụng tốt nhất với một bộ giải phi tuyến tính.	Phân loại hồi quy (không phổ biến)
Rừng ngẫu nhiên (Random forest)	Thuật toán được xây dựng dựa trên cây quyết định để cải thiện độ chính xác một cách quyết liệt. Rừng ngẫu nhiên tạo ra nhiều cây quyết định đơn giản và sử dụng phương pháp 'bỏ phiếu đa số' để quyết định nhãn nào sẽ trả lại. Đối với nhiệm vụ phân loại, dự đoán cuối cùng sẽ là người có nhiều phiếu bầu nhất; trong khi đối với nhiệm vụ hồi quy, dự đoán trung bình của tất cả các cây là dự đoán cuối cùng.	Phân loại hồi quy

Thuật toán hồi quy tuyến tính.

- ▶ Mô hình chỉ ra một quan hệ tuyến tính giữa một biến phụ thuộc vào một hay nhiều biến độc lập.
- ▶ Đây chính là phương pháp phân tích hồi quy của xác suất thống kê.
- ▶ Mục tiêu của thuật toán hồi quy tuyến tính là dự đoán giá trị của một hoặc nhiều biến liên tục y dựa trên một vec-tơ đầu vào x .
- ▶ (x_i, y_i)

Thuật toán hồi quy tuyến tính.



Hình 1. Quan hệ $y=3+4x$

Thuật toán hồi quy tuyến tính.

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

$$y = b_1x + b_0$$

- ▶ b_1 và b_0 có công thức tính lần lượt là:

$$b_1 = (\text{Tương quan } x \text{ và } y) * \left(\frac{\sigma_y}{\sigma_x}\right)$$

$$b_0 = \bar{y} - b_1\bar{x}$$

Thuật toán hồi quy tuyến tính.

- Công thức Phương sai:

$$\sigma^2 = \frac{\sum_i^n (a_i - \bar{a})^2}{n-1}$$

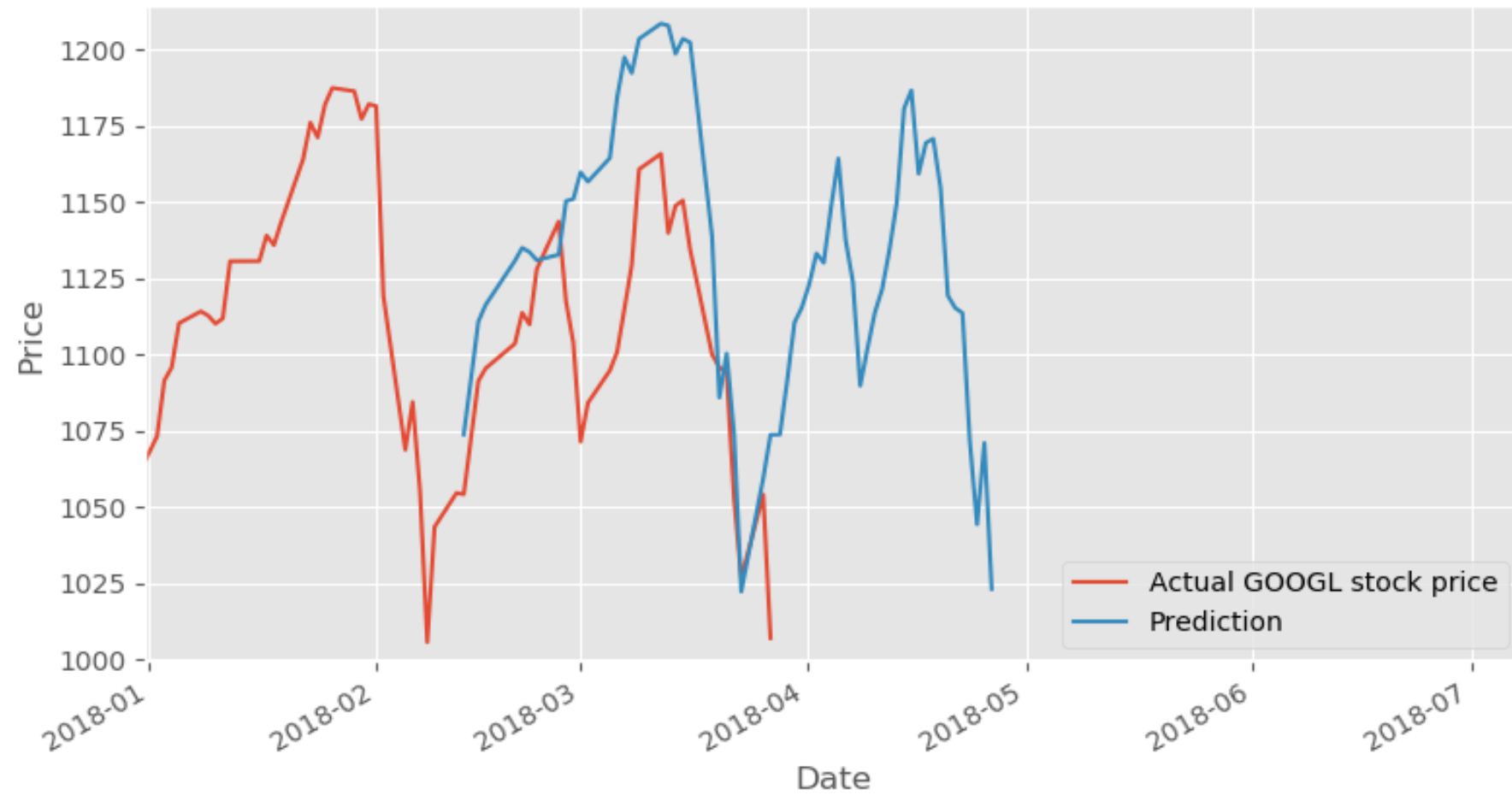
- Độ tương quan = (hiệp phương sai x và y) / (độ lệch chuẩn x * độ lệch chuẩn y).
- Hiệp phương sai:

$$\text{Cov}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

3.Ứng dụng.

Date	Adj. Open	Adj. High	Adj. Low	Adj. Close	Adj. Volume
#####	50.15984	52.19111	48.12857	50.32284	44659000
#####	50.66139	54.70888	50.4056	54.32269	22834300
#####	55.55148	56.91569	54.69383	54.86938	18256100
#####	55.79223	55.97278	51.94535	52.59736	15247300
#####	52.54219	54.16721	52.10083	53.16411	9188600
#####	52.63749	54.14213	52.49204	54.12207	7094800
#####	54.21736	54.47817	53.00863	53.23934	6211700
#####	52.803	52.90832	51.16294	51.16294	5196700
#####	51.31842	52.01557	51.23817	51.34349	4917800
9/1/2004	51.509	51.64442	49.98931	50.28021	9138200
9/2/2004	49.69841	51.34349	49.62318	50.91216	15118600
9/3/2004	50.63129	51.02752	49.81377	50.15984	5152400
9/7/2004	50.66139	51.15792	49.95922	50.94727	5847500
9/8/2004	50.52597	51.67451	50.4056	51.30838	4985600
9/9/2004	51.40869	51.51402	50.65637	51.3134	4061700

4. Kết quả, đánh giá.



5. Phương hướng phát triển.

THANKS FOR WATCHING