**TRƯỜNG ĐẠI HỌC THỦY LỢI**

**KHOA CÔNG NGHỆ THÔNG TIN**

****

**BÁO CÁO MÔN HỌC MÁY**

**Ứng dụng phương pháp hồi quy tuyến tính dự đoán giá cổ phiếu của CTCP Tập đoàn hoá chất Đức Giang**

**GIẢNG VIÊN MÔN HỌC :** NGUYỄN THỊ KIM NGÂN

**NHÓM SINH VIÊN THỰC HIỆN (NHÓM 13):**

Phạm Minh Tiến - 1851061636

Trần Đức Huy - 1851060586

Phạm Hoài Nam - 1851061318

*Hà Nội, 09/2021*

### Mô tả bài toán

1. Tên đề tài: Sử dụng phương pháp hồi quy tuyến tính dự đoán Close(giá đóng cửa) cổ phiếu DGC
2. Bài toán được xây dựng dựa trên thuật toán  Linear Regression ( hồi quy tuyến tính) thuộc nhóm Supervised learning ( học có giám sát ). Hồi quy tuyến tính (Linear Regression) là một phương pháp thống kê để thiết lập mối quan hệ giữa một biến phụ thuộc và một nhóm tập hợp các biến độc lập rất đơn giản nhưng đã được chứng minh được tính hữu ích cho một số lượng lớn các tình huống.
3. Tóm tắt công việc:

* Thu thập dữ liệu của 105ngày của cổ phiếu DGC tính từ ngày 15/3/2021 - 10/8/2021(Không bao gồm ngày lễ)
* Sử dụng phương pháp hồi quy tuyến tính dự đoán close(giá đóng cửa) của cổ phiếu

### Mô tả dữ liệu của bài toán:

1. Vector dữ liệu gồm: 105 vector dữ liệu.
2. Chiều thông tin dữ liệu : Ngày, Giá mở cửa, Số lượng mua cổ phiếu, Số lượng bán cổ phiếu, Giá đóng cửa
3. Đầu vào của bài toán gồm : Giá mở cửa, Số lượng mua cổ phiếu, Số lượng bán cổ phiếu (Open, PVolume,SVolume)
4. Thông tin đầu ra cần dự đoán : Giá đóng cửa(Close)

III. Chia dữ liệu thành 2 phần

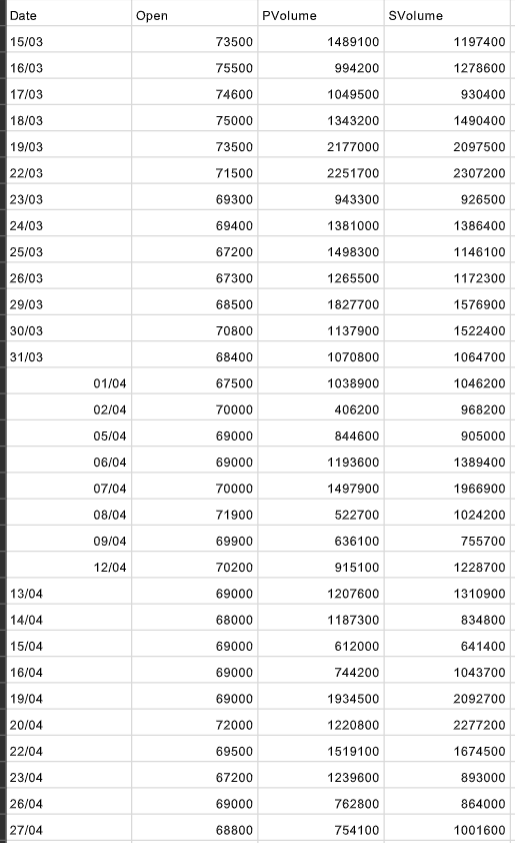
Dữ liệu được làm 2 phần bao gồm training và test  
 -Training : gồm 75 dữ liệu đầu tiên

-Test : gồm 30 dữ liệu còn lại

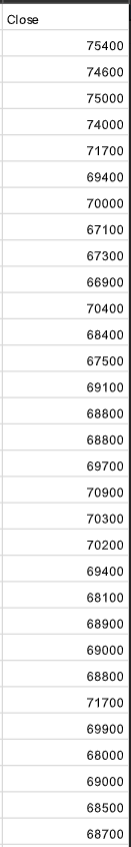
IV . Mô tả dữ liệu huấn luyện

Trích dẫn 1 phần trong dữ liệu huấn luyện từ ngày 15/3/2021 cho đến ngày 27/4/2021

* Dữ liệu huấn luyện(X)



* Vecto đầu ra (Y):



VI, So sánh kết quả dự đoán và kết quả thực

Ngày 01/07 ,Giá Đoán: 81,911 VNĐ ,Giá Thực tế: 83,200 VNĐ ,Độ lệch: 1,289 VNĐ

Ngày 02/07 ,Giá Đoán: 82,834 VNĐ ,Giá Thực tế: 83,200 VNĐ ,Độ lệch: 366 VNĐ

Ngày 05/07 ,Giá Đoán: 82,986 VNĐ ,Giá Thực tế: 83,700 VNĐ ,Độ lệch: 714 VNĐ

Ngày 06/07 ,Giá Đoán: 85,324 VNĐ ,Giá Thực tế: 78,000 VNĐ ,Độ lệch: -7,324 VNĐ

Ngày 07/07 ,Giá Đoán: 81,613 VNĐ ,Giá Thực tế: 83,400 VNĐ ,Độ lệch: 1,787 VNĐ

Ngày 08/07 ,Giá Đoán: 84,683 VNĐ ,Giá Thực tế: 83,500 VNĐ ,Độ lệch: -1,183 VNĐ

Ngày 09/07 ,Giá Đoán: 85,737 VNĐ ,Giá Thực tế: 87,900 VNĐ ,Độ lệch: 2,163 VNĐ

Ngày 12/07 ,Giá Đoán: 89,980 VNĐ ,Giá Thực tế: 82,200 VNĐ ,Độ lệch: -7,780 VNĐ

Ngày 13/07 ,Giá Đoán: 84,582 VNĐ ,Giá Thực tế: 81,000 VNĐ ,Độ lệch: -3,582 VNĐ

Ngày 14/07 ,Giá Đoán: 82,225 VNĐ ,Giá Thực tế: 78,000 VNĐ ,Độ lệch: -4,225 VNĐ

Ngày 15/07 ,Giá Đoán: 77,656 VNĐ ,Giá Thực tế: 80,800 VNĐ ,Độ lệch: 3,144 VNĐ

Ngày 16/07 ,Giá Đoán: 81,129 VNĐ ,Giá Thực tế: 82,100 VNĐ ,Độ lệch: 971 VNĐ

Ngày 19/07 ,Giá Đoán: 81,109 VNĐ ,Giá Thực tế: 79,300 VNĐ ,Độ lệch: -1,809 VNĐ

Ngày 20/07 ,Giá Đoán: 80,900 VNĐ ,Giá Thực tế: 81,400 VNĐ ,Độ lệch: 500 VNĐ

Ngày 21/07 ,Giá Đoán: 82,368 VNĐ ,Giá Thực tế: 82,000 VNĐ ,Độ lệch: -368 VNĐ

Ngày 22/07 ,Giá Đoán: 84,402 VNĐ ,Giá Thực tế: 86,400 VNĐ ,Độ lệch: 1,998 VNĐ

Ngày 23/07 ,Giá Đoán: 93,685 VNĐ ,Giá Thực tế: 92,400 VNĐ ,Độ lệch: -1,285 VNĐ

Ngày 26/07 ,Giá Đoán: 92,854 VNĐ ,Giá Thực tế: 96,500 VNĐ ,Độ lệch: 3,646 VNĐ

Ngày 27/07 ,Giá Đoán: 98,865 VNĐ ,Giá Thực tế: 93,000 VNĐ ,Độ lệch: -5,865 VNĐ

Ngày 28/07 ,Giá Đoán: 93,564 VNĐ ,Giá Thực tế: 93,700 VNĐ ,Độ lệch: 136 VNĐ

Ngày 29/07 ,Giá Đoán: 94,675 VNĐ ,Giá Thực tế: 96,000 VNĐ ,Độ lệch: 1,325 VNĐ

Ngày 30/07 ,Giá Đoán: 96,198 VNĐ ,Giá Thực tế: 95,500 VNĐ ,Độ lệch: -698 VNĐ

Ngày 02/08 ,Giá Đoán: 96,641 VNĐ ,Giá Thực tế: 97,100 VNĐ ,Độ lệch: 459 VND

Ngày 03/08 ,Giá Đoán: 98,188 VNĐ ,Giá Thực tế: 96,900 VNĐ ,Độ lệch: -1,288 VNĐ

Ngày 04/08 ,Giá Đoán: 97,346 VNĐ ,Giá Thực tế: 99,500 VNĐ ,Độ lệch: 2,154 VNĐ

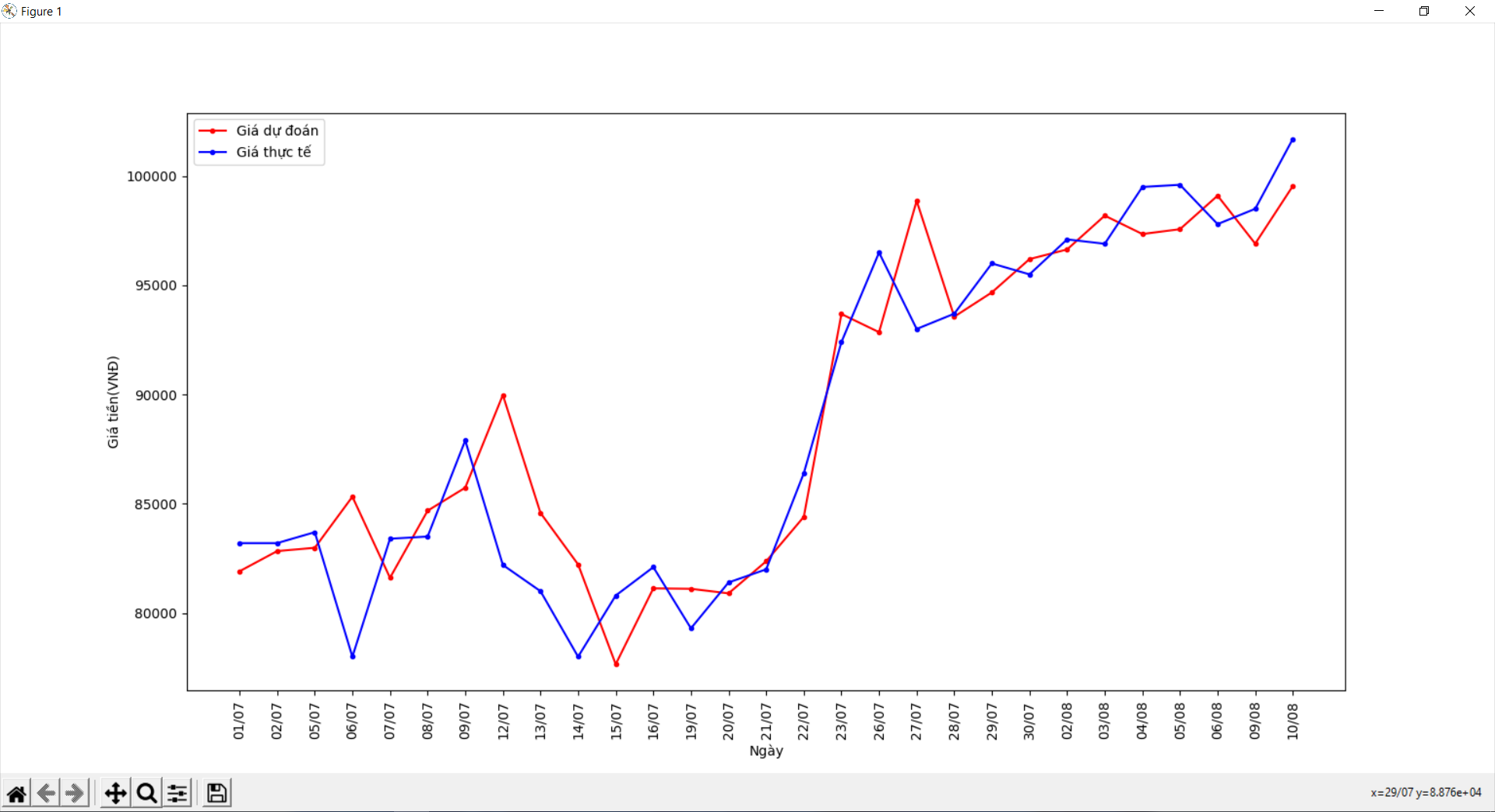
Ngày 05/08 ,Giá Đoán: 97,568 VNĐ ,Giá Thực tế: 99,600 VNĐ ,Độ lệch: 2,032 VNĐ

Ngày 06/08 ,Giá Đoán: 99,102 VNĐ ,Giá Thực tế: 97,800 VNĐ ,Độ lệch: -1,302 VNĐ

Ngày 09/08 ,Giá Đoán: 96,906 VNĐ ,Giá Thực tế: 98,500 VNĐ ,Độ lệch: 1,594 VNĐ

Ngày 10/08 ,Giá Đoán: 99,544 VNĐ ,Giá Thực tế: 101,700 VNĐ ,Độ lệch: 2,156 VNĐ

-Biểu đồ giữ kết quả dự đoán và kết quả thực tế của bài toán



VII, Kết Luận

## Ứng dụng của phương pháp hồi quy tuyến tính (Linear Regression)

Sử dụng phương pháp hồi quy tuyến tính giúp ta dễ dàng dự đoán được các giá trị tuyến tính dựa vào tập huấn luyện cho trước. Ưu điểm của hồi quy tuyến tính là tính đơn giản của hàm, khi giá trị x tăng hoặc giảm thì giá trị y cũng tăng hoặc giảm, tuyến tính theo f(x).

## Hạn chế của phương pháp hồi quy tuyến tính (Linear Regression)

Hạn chế đầu tiên của Linear Regression là nó rấtnhạy cảm với nhiễu. Trong ví dụ về mối quan hệ giữa diện tích và giá căn phòng bên trên, nếu có chỉ một cặp dữ liệu nhiễu thì kết quả sẽ sai khác đi rất nhiều. Vì vậy, trước khi thực hiện Linear Regression, các nhiễu (*outlier*) cần phải được loại bỏ, bước 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.