

**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**KHOA KHOA HỌC VÀ KỸ THUẬT THÔNG TIN**

-----🙞🙜🕮🙞🙜-----



**MÔN HỌC**

KHAI THÁC DỮ LIỆU VÀ ỨNG DỤNG

**ĐỀ TÀI**

**DỰ ĐOÁN TĂNG, GIẢM TRONG PHIÊN GIAO DỊCH**

Giảng viên hướng dẫn: Th.S Nguyễn Thị Anh Thư

Lớp: CS313.K21

Nhóm sinh viên: Nhóm 2

10/4/2019

**NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN**

…………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………

…, ngày … tháng… năm 2019

**Người nhận xét**

(Ký và ghi rõ họ tên)

# **LỜI MỞ ĐẦU**

Data Science – Khoa học dữ liệu là từ khóa có thể nói là từ khóa được tìm kiếm nhiều nhất trong thập kỉ vửa qua. Chính vì sự phát triển của xã hội, công nghệ thông tin và sự bùng nổ của cuộc cánh mạng công nghiệp 4.0 mà nhu cầu sử dụng dữ liệu được đẩy lên cao hơn bao giở hết. Người ta còn ví von rằng “ Ai nắm trong tay dữ liệu lớn – Người đó sẽ nắm cả tương lai “ vì vậy đã biến việc lưu trữ, phân tích, sử dụng dữ liệu không chỉ còn là của các công ty lớn nữa mà nó đã trở thành xu thế của toàn cầu trong mọi lĩnh vực như: kinh tế, y tế, nông nghiệp, ... và nhiều lĩnh vực trọng yếu khác của mỗi quốc gia.

Như đã nói ở trên, Dữ liệu bây giờ có thể được xem là một mỏ vàng thô để khai thác tri thức nhân loại trong mọi lĩnh vưc của cuộc sống. Vậy câu hỏi đặt ra là làm sao để khai thác nó hiệu quả nhất ?. Khi đó người ta sẽ tìm đến một kĩ thuật gọi là Data Mining – Khai thác dữ liệu – Kĩ thuật được dùng để đi sâu vào dữ liệu để phân tích và tìm kiếm các chi tiết, giá trị ẩn đi trong dữ liệu, hay cụ thể là muốn xác định, muốn biết xem dữ liệu cung cấp tri thức gì từ đó có thể đưa ra quyết định có lợi. Theo Data – Flair, Data Mining là một trong những lợi thế các công ty có ngành sản xuất, kinh doang, marketing nếu họ biết cách ứng dụng hợp lí để tăng hiệu quả hoạt động. Do đó nhu cầu xây dựng một hệ thống Data Mining tiêu chuẩn ngày càng cao. Các quy trình, mô hình Data Mining phải có độ tin cậy cao và tọa điều kiện để các nhà kinh doanh – những người không nắm rõ kiến thức chuyên môn về Data Science có thể sử dụng được.

Để làm rõ vấn để này, nhóm mình đã ứng dụng kĩ thuật Data Mining để phân tích và dự đoán sự tăng, giảm của sàn giao dịch tiền tệ trong từng phiên giao dịch 5 phút được thu thập từ Forex Market.

Sau đây, chúng ta sẽ cùng tìm hiểu về “ Dự đoán giá trị tăng, giảm trong phiên giao dịch ”thông qua các chương sau:

* Chương 1: Hiểu dữ liệu
* Chương 2: Nội dung đồ án
* Chương 3: Kết quả thu được
* Chương 4: Kết luận

**MỤC LỤC**

**THÔNG TIN NHÓM**

|  |  |  |
| --- | --- | --- |
| MSSV | Họ và Tên | Ghi chú |
| 18520356 | Lê Việt Thắng | Trưởng nhóm |
| 18520936 | Trần Đăng Khoa |  |
| 18521084 | Nguyễn Thế Mạnh |  |
| 18521190 | Lê Quang Nhật |  |
| 18521294 | Hoàng Đình Quang |  |
| 18521589 | Võ Đình Tứ |  |

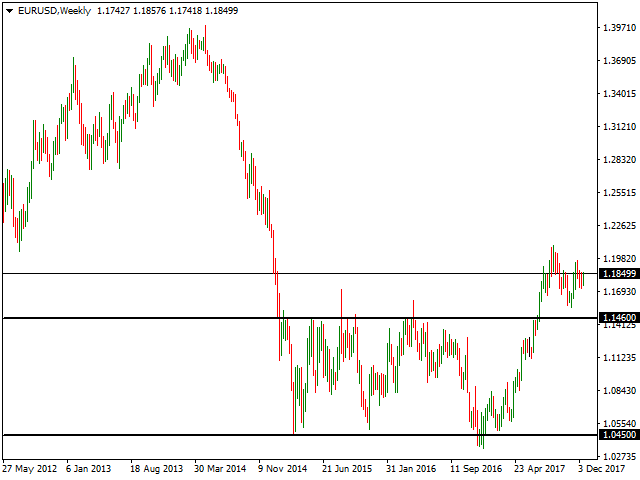
**BẢNG PHÂN CÔNG CÔNG VIỆC**

**1 Thông tin về tập dữ liệu sử dụng:**

Tên tập dữ liệu: Stock Market Dataset

Dữ liệu được thu thập từ **Forex Market** để dự đoán tương lai. Mục tiêu chính của chúng ta là dự đoán giá giao dịch nhiều nhất trên thế giới của 2 đơn vị tiền tệ lớn là EUR/USD. Ta chỉ cần dự đoán giá trị **Tăng** hay **Giảm** của sàn giao dịch hiện tại với dữ liệu được đưa ra trong khoảng thời gian 01/01/2008 – 19/03/2018 được lấy từ www.dukascopy.com Historical Data Feed được nhóm tổng hợp từ cuộc thi 2019 International Data Science Competition

Xem thêm về Original Description tại [file](Description.docx).

****

Tập dữ liệu gồm có 709314 mẫu và 212 thuộc tính mô tả một phiên giao dịch 5 phút của 2 đơn vị tiền tệ EUR/USD từ 01/01/2018 đến 19/03/2018 thông qua các feature như giá, biến động, thông tin phiên giao dịch và một số feature khác. Để hiểu kĩ hơn ta sẽ tham khảo mô tả các thuộc tính bên dưới:

**Gmt time:** Thời gian mở phiên, thời gian bắt đầu của phiên giao dịch 5 phút

**Open:** Giá mở đầu của phiên giao dịch

**High:** Giá cao nhất của phiên giao dịch

**Close:** Giá đóng phiên giao dịch

**Volume:** Số lượng giao dịch trong phiên ( triệu giao dịch)

**Body:** chiều dài thân của biểu đồ nến

**Upper\_tail:** đầu trên của biểu đồ nến

**Lower\_tail:** đầu dưới của biểu đồ nến

**SMA\_50:** đường trung bình động trong 50 phiên giao dịch trước

**SMA\_20:** đường trung bình động trong 20 phiên giao dịch trước

**ATR:** Chỉ số chuyên ngành ATR của 50 phiên giao dịch trước (biểu diễn biến động trong phiên)

**CCI :** Chỉ số chuyên ngành CCI của 20 phiên giao dịch trước

**SAR:** Chỉ số chuyên ngành SAR (biển diễn trend )

**Hour:** giờ dữ liệu được thu thập

**Min:** phút dữ liệu được thu thập

**Dayofweek:** Ngày dữ liệu được thu thập

**JPY:** nếu JPY mở trading

**AUD:** nếu AUD mở trading

**EUR:** nếu EUR mở trading

**GBP:** nếu GBP mở trading

**USD:** nếu USD mở trading

**Return\_1, return\_2, return\_3,..., return\_96:** tiền thu về trong 5 phút, 10 phút, 15 phút,..., 8 tiếng vừa qua.

**Lag\_return\_1, Lag\_return\_2, Lag\_return\_3,..., Lag\_return\_96:** tiền thu về trong 5 phút, 10 phút, 15 phút,..., 8 tiếng trước phiên giao dịch.

**Labels:**

**Up\_down:** 1 nếu giá trị sàn giao dịch tăng, 0 nếu giá trị sàn giao dịch giảm.

**2 Bảng mô tả giá trị các thuộc tính tập trong dữ liệu:**

Bảng mô tả giá trị các thuộc tính trong tập dữ liệu:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tên thuộc tính | Kiểu dữ liệu | Trung bình/ số giá trị phân biệt | Phương sai/ số giá trị duy nhất | Số mẫu bị thiếu |
| Open | Numeric | 1.36531417 | 0.0067936 | 0(0%} |
| High | Numeric | 1.365578008 | 0.006798722 | 0(0%} |
| Low | Numeric | 1.365039135 | 0.00678722 | 0(0%} |
| Close | Numeric | 1.365314736 | 0.00679351 | 0(0%} |
| Volume | Numeric | 743.5132221 | 1331260.381 | 0(0%} |
| body | Numeric | 0.005655464 | 21.20722913 | 0(0%} |
| upper\_tail | Numeric | 1.379340461 | 3.900779969 | 0(0%} |
| lower\_tail | Numeric | 1.496966647 | 4.343564139 | 0(0%} |
| SMA\_50 | Numeric | 1.365321353 | 0.006791751 | 0(0%} |
| SMA\_20 | Numeric | 1.365317304 | 0.006792802 | 0(0%} |
| ATR | Numeric | 0.000541021 | 2.00E-07 | 0(0%} |
| CCI | Numeric | -0.164692475 | 10446.49052 | 0(0%} |
| SAR | Numeric | 1.365311019 | 0.006793079 | 0(0%} |
| hour | Numeric | 11.50021993 | 47.91752378 | 0(0%} |
| min | Numeric | 27.49991541 | 297.9156093 | 0(0%} |
| dayofweek | Numeric | 3.002020262 | 4.001686284 | 0(0%} |
| JPY | Numeric | 0.374973566 | 0.234368391 | 0(0%} |
| AUD | Numeric | 0.333340382 | 0.222224572 | 0(0%} |
| EUR | Numeric | 0.374997533 | 0.234374383 | 0(0%} |
| GBP | Numeric | 0.375014451 | 0.234378612 | 0(0%} |
| USD | Numeric | 0.37501586 | 0.234378965 | 0(0%} |
| lag\_return\_1 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_2 | Numeric | -2.81E-07 | 2.32E-07 | 0(0%} |
| lag\_return\_2 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_3 | Numeric | -4.23E-07 | 3.45E-07 | 0(0%} |
| lag\_return\_3 | Numeric | -1.40E-07 | 1.18E-07 | 0(0%} |
| return\_4 | Numeric | -5.66E-07 | 4.57E-07 | 0(0%} |
| lag\_return\_4 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_5 | Numeric | -7.08E-07 | 5.69E-07 | 0(0%} |
| lag\_return\_5 | Numeric | -1.40E-07 | 1.18E-07 | 0(0%} |
| return\_6 | Numeric | -8.52E-07 | 6.80E-07 | 0(0%} |
| lag\_return\_6 | Numeric | -1.40E-07 | 1.18E-07 | 0(0%} |
| return\_7 | Numeric | -9.95E-07 | 7.93E-07 | 0(0%} |
| lag\_return\_7 | Numeric | -1.40E-07 | 1.18E-07 | 0(0%} |
| return\_8 | Numeric | -1.14E-06 | 9.05E-07 | 0(0%} |
| lag\_return\_8 | Numeric | -1.40E-07 | 1.18E-07 | 0(0%} |
| return\_9 | Numeric | -1.28E-06 | 1.02E-06 | 0(0%} |
| lag\_return\_9 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_10 | Numeric | -1.42E-06 | 1.13E-06 | 0(0%} |
| lag\_return\_10 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_11 | Numeric | -1.56E-06 | 1.24E-06 | 0(0%} |
| lag\_return\_11 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_12 | Numeric | -1.70E-06 | 1.36E-06 | 0(0%} |
| lag\_return\_12 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_13 | Numeric | -1.85E-06 | 1.47E-06 | 0(0%} |
| lag\_return\_13 | Numeric | -1.40E-07 | 1.18E-07 | 0(0%} |
| return\_14 | Numeric | -1.99E-06 | 1.58E-06 | 0(0%} |
| lag\_return\_14 | Numeric | -1.40E-07 | 1.18E-07 | 0(0%} |
| return\_15 | Numeric | -2.13E-06 | 1.69E-06 | 0(0%} |
| lag\_return\_15 | Numeric | -1.40E-07 | 1.18E-07 | 0(0%} |
| return\_16 | Numeric | -2.28E-06 | 1.81E-06 | 0(0%} |
| lag\_return\_16 | Numeric | -1.40E-07 | 1.18E-07 | 0(0%} |
| return\_17 | Numeric | -2.42E-06 | 1.92E-06 | 0(0%} |
| lag\_return\_17 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_18 | Numeric | -2.56E-06 | 2.03E-06 | 0(0%} |
| lag\_return\_18 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_19 | Numeric | -2.70E-06 | 2.14E-06 | 0(0%} |
| lag\_return\_19 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_20 | Numeric | -2.85E-06 | 2.25E-06 | 0(0%} |
| lag\_return\_20 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_21 | Numeric | -2.99E-06 | 2.36E-06 | 0(0%} |
| lag\_return\_21 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_22 | Numeric | -3.13E-06 | 2.47E-06 | 0(0%} |
| lag\_return\_22 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_23 | Numeric | -3.27E-06 | 2.59E-06 | 0(0%} |
| lag\_return\_23 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_24 | Numeric | -3.42E-06 | 2.70E-06 | 0(0%} |
| lag\_return\_24 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_25 | Numeric | -3.56E-06 | 2.81E-06 | 0(0%} |
| lag\_return\_25 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_26 | Numeric | -3.70E-06 | 2.92E-06 | 0(0%} |
| lag\_return\_26 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_27 | Numeric | -3.84E-06 | 3.03E-06 | 0(0%} |
| lag\_return\_27 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_28 | Numeric | -3.98E-06 | 3.15E-06 | 0(0%} |
| lag\_return\_28 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_29 | Numeric | -4.13E-06 | 3.26E-06 | 0(0%} |
| lag\_return\_29 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_30 | Numeric | -4.27E-06 | 3.37E-06 | 0(0%} |
| lag\_return\_30 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_31 | Numeric | -4.41E-06 | 3.48E-06 | 0(0%} |
| lag\_return\_31 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_32 | Numeric | -4.55E-06 | 3.59E-06 | 0(0%} |
| lag\_return\_32 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_33 | Numeric | -4.69E-06 | 3.71E-06 | 0(0%} |
| lag\_return\_33 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_34 | Numeric | -4.84E-06 | 3.82E-06 | 0(0%} |
| lag\_return\_34 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_35 | Numeric | -4.98E-06 | 3.93E-06 | 0(0%} |
| lag\_return\_35 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_36 | Numeric | -5.12E-06 | 4.04E-06 | 0(0%} |
| lag\_return\_36 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_37 | Numeric | -5.26E-06 | 4.16E-06 | 0(0%} |
| lag\_return\_37 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_38 | Numeric | -5.40E-06 | 4.27E-06 | 0(0%} |
| lag\_return\_38 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_39 | Numeric | -5.55E-06 | 4.38E-06 | 0(0%} |
| lag\_return\_39 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_40 | Numeric | -5.69E-06 | 4.49E-06 | 0(0%} |
| lag\_return\_40 | Numeric | -1.40E-07 | 1.18E-07 | 0(0%} |
| return\_41 | Numeric | -5.83E-06 | 4.60E-06 | 0(0%} |
| lag\_return\_41 | Numeric | -1.40E-07 | 1.18E-07 | 0(0%} |
| return\_42 | Numeric | -5.98E-06 | 4.71E-06 | 0(0%} |
| lag\_return\_42 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_43 | Numeric | -6.12E-06 | 4.82E-06 | 0(0%} |
| lag\_return\_43 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_44 | Numeric | -6.26E-06 | 4.94E-06 | 0(0%} |
| lag\_return\_44 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_45 | Numeric | -6.40E-06 | 5.05E-06 | 0(0%} |
| lag\_return\_45 | Numeric | -1.40E-07 | 1.18E-07 | 0(0%} |
| return\_46 | Numeric | -6.55E-06 | 5.16E-06 | 0(0%} |
| lag\_return\_46 | Numeric | -1.40E-07 | 1.18E-07 | 0(0%} |
| return\_47 | Numeric | -6.69E-06 | 5.27E-06 | 0(0%} |
| lag\_return\_47 | Numeric | -1.40E-07 | 1.18E-07 | 0(0%} |
| return\_48 | Numeric | -6.83E-06 | 5.38E-06 | 0(0%} |
| lag\_return\_48 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_49 | Numeric | -6.98E-06 | 5.49E-06 | 0(0%} |
| lag\_return\_49 | Numeric | -1.38E-07 | 1.18E-07 | 0(0%} |
| return\_50 | Numeric | -7.12E-06 | 5.60E-06 | 0(0%} |
| lag\_return\_50 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_51 | Numeric | -7.26E-06 | 5.71E-06 | 0(0%} |
| lag\_return\_51 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_52 | Numeric | -7.40E-06 | 5.83E-06 | 0(0%} |
| lag\_return\_52 | Numeric | -1.38E-07 | 1.18E-07 | 0(0%} |
| return\_53 | Numeric | -7.54E-06 | 5.94E-06 | 0(0%} |
| lag\_return\_53 | Numeric | -1.38E-07 | 1.18E-07 | 0(0%} |
| return\_54 | Numeric | -7.69E-06 | 6.05E-06 | 0(0%} |
| lag\_return\_54 | Numeric | -1.38E-07 | 1.18E-07 | 0(0%} |
| return\_55 | Numeric | -7.83E-06 | 6.16E-06 | 0(0%} |
| lag\_return\_55 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_56 | Numeric | -7.97E-06 | 6.27E-06 | 0(0%} |
| lag\_return\_56 | Numeric | -1.38E-07 | 1.18E-07 | 0(0%} |
| return\_57 | Numeric | -8.11E-06 | 6.38E-06 | 0(0%} |
| lag\_return\_57 | Numeric | -1.38E-07 | 1.18E-07 | 0(0%} |
| return\_58 | Numeric | -8.25E-06 | 6.49E-06 | 0(0%} |
| lag\_return\_58 | Numeric | -1.38E-07 | 1.18E-07 | 0(0%} |
| return\_59 | Numeric | -8.40E-06 | 6.60E-06 | 0(0%} |
| lag\_return\_59 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_60 | Numeric | -8.54E-06 | 6.71E-06 | 0(0%} |
| lag\_return\_60 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_61 | Numeric | -8.68E-06 | 6.83E-06 | 0(0%} |
| lag\_return\_61 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_62 | Numeric | -8.82E-06 | 6.94E-06 | 0(0%} |
| lag\_return\_62 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_63 | Numeric | -8.96E-06 | 7.05E-06 | 0(0%} |
| lag\_return\_63 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_64 | Numeric | -9.11E-06 | 7.16E-06 | 0(0%} |
| lag\_return\_64 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_65 | Numeric | -9.25E-06 | 7.27E-06 | 0(0%} |
| lag\_return\_65 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_66 | Numeric | -9.39E-06 | 7.39E-06 | 0(0%} |
| lag\_return\_66 | Numeric | -1.40E-07 | 1.18E-07 | 0(0%} |
| return\_67 | Numeric | -9.53E-06 | 7.50E-06 | 0(0%} |
| lag\_return\_67 | Numeric | -1.40E-07 | 1.18E-07 | 0(0%} |
| return\_68 | Numeric | -9.68E-06 | 7.61E-06 | 0(0%} |
| lag\_return\_68 | Numeric | -1.40E-07 | 1.18E-07 | 0(0%} |
| return\_69 | Numeric | -9.82E-06 | 7.72E-06 | 0(0%} |
| lag\_return\_69 | Numeric | -1.40E-07 | 1.18E-07 | 0(0%} |
| return\_70 | Numeric | -9.96E-06 | 7.83E-06 | 0(0%} |
| lag\_return\_70 | Numeric | -1.40E-07 | 1.18E-07 | 0(0%} |
| return\_71 | Numeric | -1.01E-05 | 7.94E-06 | 0(0%} |
| lag\_return\_71 | Numeric | -1.40E-07 | 1.18E-07 | 0(0%} |
| return\_72 | Numeric | -1.03E-05 | 8.06E-06 | 0(0%} |
| lag\_return\_72 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_73 | Numeric | -1.04E-05 | 8.17E-06 | 0(0%} |
| lag\_return\_73 | Numeric | -1.38E-07 | 1.18E-07 | 0(0%} |
| return\_74 | Numeric | -1.05E-05 | 8.28E-06 | 0(0%} |
| lag\_return\_74 | Numeric | -1.38E-07 | 1.18E-07 | 0(0%} |
| return\_75 | Numeric | -1.07E-05 | 8.39E-06 | 0(0%} |
| lag\_return\_75 | Numeric | -1.38E-07 | 1.18E-07 | 0(0%} |
| return\_76 | Numeric | -1.08E-05 | 8.51E-06 | 0(0%} |
| lag\_return\_76 | Numeric | -1.38E-07 | 1.18E-07 | 0(0%} |
| return\_77 | Numeric | -1.10E-05 | 8.62E-06 | 0(0%} |
| lag\_return\_77 | Numeric | -1.38E-07 | 1.18E-07 | 0(0%} |
| return\_78 | Numeric | -1.11E-05 | 8.74E-06 | 0(0%} |
| lag\_return\_78 | Numeric | -1.38E-07 | 1.18E-07 | 0(0%} |
| return\_79 | Numeric | -1.12E-05 | 8.85E-06 | 0(0%} |
| lag\_return\_79 | Numeric | -1.38E-07 | 1.18E-07 | 0(0%} |
| return\_80 | Numeric | -1.14E-05 | 8.97E-06 | 0(0%} |
| lag\_return\_80 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_81 | Numeric | -1.15E-05 | 9.08E-06 | 0(0%} |
| lag\_return\_81 | Numeric | -1.38E-07 | 1.18E-07 | 0(0%} |
| return\_82 | Numeric | -1.17E-05 | 9.20E-06 | 0(0%} |
| lag\_return\_82 | Numeric | -1.38E-07 | 1.18E-07 | 0(0%} |
| return\_83 | Numeric | -1.18E-05 | 9.31E-06 | 0(0%} |
| lag\_return\_83 | Numeric | -1.38E-07 | 1.18E-07 | 0(0%} |
| return\_84 | Numeric | -1.19E-05 | 9.43E-06 | 0(0%} |
| lag\_return\_84 | Numeric | -1.38E-07 | 1.18E-07 | 0(0%} |
| return\_85 | Numeric | -1.21E-05 | 9.54E-06 | 0(0%} |
| lag\_return\_85 | Numeric | -1.38E-07 | 1.18E-07 | 0(0%} |
| return\_86 | Numeric | -1.22E-05 | 9.65E-06 | 0(0%} |
| lag\_return\_86 | Numeric | -1.38E-07 | 1.18E-07 | 0(0%} |
| return\_87 | Numeric | -1.24E-05 | 9.77E-06 | 0(0%} |
| lag\_return\_87 | Numeric | -1.38E-07 | 1.18E-07 | 0(0%} |
| return\_88 | Numeric | -1.25E-05 | 9.89E-06 | 0(0%} |
| lag\_return\_88 | Numeric | -1.38E-07 | 1.18E-07 | 0(0%} |
| return\_89 | Numeric | -1.26E-05 | 1.00E-05 | 0(0%} |
| lag\_return\_89 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_90 | Numeric | -1.28E-05 | 1.01E-05 | 0(0%} |
| lag\_return\_90 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_91 | Numeric | -1.29E-05 | 1.02E-05 | 0(0%} |
| lag\_return\_91 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_92 | Numeric | -1.31E-05 | 1.03E-05 | 0(0%} |
| lag\_return\_92 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_93 | Numeric | -1.32E-05 | 1.05E-05 | 0(0%} |
| lag\_return\_93 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_94 | Numeric | -1.33E-05 | 1.06E-05 | 0(0%} |
| lag\_return\_94 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_95 | Numeric | -1.35E-05 | 1.07E-05 | 0(0%} |
| lag\_return\_95 | Numeric | -1.39E-07 | 1.18E-07 | 0(0%} |
| return\_96 | Numeric | -1.36E-05 | 1.08E-05 | 0(0%} |
| lag\_return\_96 | Numeric | -1.37E-07 | 1.18E-07 | 0(0%} |

**3 Nhận xét về chất lượng tập dữ liệu sử dụng:**

Dựa vào bảng mô tả giá trị các thuộc tính của tập dữ liệu ở Chương I – mục 2 ta có thể thấy rằng dữ liệu đã được chuẩn hóa tương đối đồng đều về tỉ lệ giữa các giá trị, không có thuộc tính nào có dấu hiệu mất, thiếu giá trị. Tuy nhiên theo ý kiến chủ quan của nhóm thì dữ liệu được sử dụng vẫn chưa được “sạch” vì qua quan sát sơ qua các thuộc tính thì dữ liệu vẫn còn tồn tại rất nhiều các outlier – ngoại lệ vẫn chưa được loại bỏ. Đối với các thuộc tính dữ liệu đưa ra rất chi tiết và đã được xử lí qua time series một cách rõ ràng từng phút. Mối liên hệ giữa các thuộc tính đưa ra đều có mối quan hệ phù hợp để có thể đánh giá sự tăng, giảm của một phiên giao dịch tiền tệ thông thường.

1. **Nội dung đồ án**
2. **Kết quả đạt được**
3. **Kết luận**

**TÀI LIỆU THAM KHẢO**