BỘ CÔNG THƯƠNG

TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP. HCM



LẬP TRÌNH PYTHON

BÀI BÁO CÁO 11 + 12

NHÓM 1

Giảng viên : CAO VĂN KIÊN

Sinh viên :

Trần Công Hòa 20017691

Đinh Ngọc Thiện và Nguyễn Hữu Thắng chỉ làm ngày 23/10 sau đó thì nhắn tin không phản hồi

TP.HCM – 2022

Câu 1: Viết chương trình thu thập dữ liệu từ web làm database để huấn luyện mô hình AI (tự chọn). Thực hiện các bước tiền xử lý dữ liệu (nếu có) trước khi lưu thành file .csv cuối cùng.

Nguồn dữ liệu: <https://data.binance.vision/?prefix=data/spot/daily/klines/DOGEUSDT/>

Sau khi lấy về thực hiện xử lý dữ liệu có được file csv

Sau khi tải dữ liệu gốc:

Graphical user interface

Description automatically generated with medium confidence

Lấy ví dụ mẫu một file dữ liệu gốc: “DOGEUSDT-1m-2022-10-16.csv”

Graphical user interface, application, table, Excel

Description automatically generated

* Thì ở đây quan tâm đến các cột dữ liệu ở cột A thời gian tính theo phút
* Giá trị cần quan tâm là giá mở đầu của 1 phút thì cột E là giá bắt đầu của 1 phút tiếp theo
* Giá trị cần quan tâm là giá cuối cùng của một phút thì cột B là giá cuối cùng của 1 phút tiếp theo đó

Tiếp theo là tổng hợp các file lại và xử lý lưu thành file csv làm dữ liệu để train:

* File tổng hợp và điều chỉnh cho dễ quan sát

Graphical user interface, application, table, Excel

Description automatically generated

* Khi xử lý thì chỉ cần quan tâm đến giá trị kết thúc 1 phút là bao nhiêu nên chỉ lấy giá trị của cột “PRICE\_LAST” theo thời gian của cột “DATE”

File csv:

* Dữ liệu theo từng phút vì dữ liệu khá nhiều nên lấy một vài dữ liệu làm mẫu

Table

Description automatically generated

* Giải thích:

x5: là kết quả của 4 phút trước

x4: là kết quả của 3 phút trước

x3: là kết quả của 2 phút trước

x2: là kết quả của 1 phút trước

x1: là kết quả của hiện tại

y1: là kết quả 1 phút sau

y2: là kết quả 2 phút sau

Câu 2: Viết chương trình huấn luyện mô hình AI sử dụng các công cụ có sẵn trên tensorflow. Đọc giá trị database trong file .csv ở câu 1. Kết quả huấn luyện phải lưu lại thành file.

Code:

Text

Description automatically generated

Run Code: Do qua trình train khá lâu nên chỉ chụp một vài đoạn để thể hiện

* Bắt đầu train

Text

Description automatically generated

* Kết thúc train

Text

Description automatically generated with medium confidence

File lưu:

Graphical user interface, application

Description automatically generated

Câu 3: Thực hiện các bước phân tích dữ liệu, so sánh với kết quả dự đoán của mô hình NN.

CODE: Để vẽ đồ thị của 7 ngày

Text

Description automatically generated

Đồ thị: Dữ liệu 7 ngày

Chart, line chart

Description automatically generated

* Để nhìn thấy rõ hơn về dữ liệu gốc và dữ liệu dự đoán khác biệt nhau thì thực hiện vẽ đồ thị 1 ngày

CODE:

Graphical user interface, text, application

Description automatically generated

Đồ thị: Dữ liệu 1 ngày

Chart, line chart, histogram

Description automatically generated

Phân tích:

* Đồ thị trên thì trục hoành thể hiện thời gian đơn vị tính theo phút, trục tung thì thì thể hiện cho giá trị của đồng “Dogcoin” đơn vị tính theo đơn vị (USDT).
* Đường màu cam là giá trị gốc của đồng “Dogcoin”, đường màu xanh là giá trị dự đoán của đồng “Dogcoin”.

So sánh:

* Kết quả dự đoán của giá “Dogcoin” thể hiện trên đồ thị cho thấy dự đoán có sự trên lệch nhau nhưng không nhiều. Kết quả dự đoán có phần giá trị cao hơn so với kết quả gốc nhưng không cao hơn bao nhiêu.