

PHÒNG THÍ NGHIỆM TÍNH TOÁN HIỆU NĂNG CAO
CÂU LẠC BỘ BIG DATA
BÁO CÁO BÀI TẬP LỚN SỐ 3 - NHÓM 8



DỰ ĐOÁN GIÁ MỘT SỐ MÃ CHỨNG
KHOẢN TẠI VIỆT NAM BẰNG MÔ HÌNH
TIME-SERIES FORECASTING

Sinh viên thực hiện

Huỳnh Tấn Lộc

K20

Phan Phước Minh

K20

Nguyễn Tuấn Minh

K21

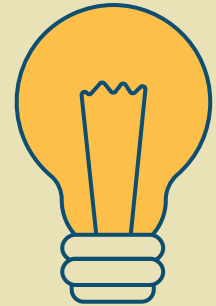
Phạm Phú Khang

K21

Back

Activities

Next



Nội dung

01

Giới thiệu về
Time-Series
Forecasting

02

Time-Series
Forecasting
trong dự đoán
chứng khoán

03

Mô tả Dataset

04

Các thuật toán
sử dụng

05

Kết quả mô
hình



01

Giới thiệu về Time-Series Forecasting

Back



Next

Giới thiệu về Time-Series Forecasting

Time-Series Forecasting (dự đoán chuỗi thời gian) là một lĩnh vực quan trọng của Machine Learning vì có rất nhiều bài toán dự đoán liên quan đến các thành phần của thời gian. Tuy nhiên Time-Series Forecasting thường bị bỏ qua bởi vì chính các thành phần của thời gian làm cho bài toán về chuỗi thời gian trở nên khó khăn hơn để có thể xử lý được



Time-Series

Một dataset (tập dữ liệu) thông thường trong Machine Learning là tập hợp của các observation (quan sát).

Observation #1
Observation #2
Observation #3

Tuy nhiên một time-series dataset (tập dữ liệu chuỗi thời gian) thì lại khác. Time-series thêm vào một sự ràng buộc thứ tự giữa các observation với nhau: một chiều không gian. Chiều không gian này vừa là một hạn chế vừa là một cấu trúc để cung cấp nguồn thông tin bổ sung.

Hay nói đơn giản, một time-series (chuỗi thời gian) là một chuỗi các observation được sắp xếp theo thứ tự thời gian một cách liên tục.

Time-series bao gồm 2 thành phần: Thời gian và observation của mốc thời gian đó.


Back



Next




Time-Series Analysis



Trong Time-Series analysis, một Time-Series sẽ được mô hình hóa để xác định các thành phần của nó như xu hướng, quan hệ với các yếu tố bên ngoài,...

Time-Series analysis trả lời cho câu hỏi “vì sao” của một Time-Series dataset

Mục tiêu chính của Time-Series analysis là phát triển các mô hình toán học cung cấp các mô tả hợp lý từ dữ liệu mẫu.



Back



Next

Các thành phần của Time-Series

Giá trị đường cơ sở của chuỗi nếu nó là đường thẳng

Level (mức độ)

Xu hướng của dữ liệu (tăng hoặc giảm)

Trend (xu hướng)

Sự lặp lại một hành vi theo chu kỳ của chuỗi theo thời gian

Seasonality (tính mùa vụ)

Sự thay đổi ngẫu nhiên của các observation không thể giải thích bằng mô hình, chỉ ra bất thường của dữ liệu.

Noise (nhiều)

Back



Next

Time-Series Forecasting

Time-Series forecasting sử dụng các thông tin từ Time-Series (có thể có các thông tin thêm khác) để dự đoán giá trị tương lai của Time-Series đó.

Dự đoán liên quan đến việc lấy các mô hình phù hợp với observation trong quá khứ và dùng chúng để đưa ra các observation trong tương lai.

Một điểm khác biệt quan trọng trong dự báo là tương lai là hoàn toàn không có sẵn và chỉ được ước tính từ những gì đã xảy ra.

Back



Next



Những điều cần quan tâm


Trong lúc dự đoán, hiểu được mục đích cuối cùng là rất quan trọng. Sử dụng Socratic method và hỏi nhiều câu hỏi có thể giúp làm rõ các chi tiết cụ thể của bài toán

- ① Frequency (tần suất): Có thể dữ liệu được cung cấp ở một tần suất quá cao hoặc có khoảng cách không đều theo thời gian lấy mẫu trong một số mô hình.
- ① Outliers (ngoại lệ): có thể có các giá trị bị hỏng hoặc ngoại lệ quá lớn cần được xác định và xử lý
- ① Missing (mất mát): có thể có những khoảng trống hoặc dữ liệu bị mất cần được lấp đầy.

Back



Next



02

Time-Series

■ Forecasting trong dự đoán chứng khoán



Back



Next

Time-Series Forecasting trong dự đoán chứng khoán

Dự báo thị trường chứng khoán là một hành vi nhằm xác định giá trị tương lai của cổ phiếu doanh nghiệp hoặc các công cụ tài chính khác được giao dịch trên các sàn giao dịch. Dự báo thành công về giá cổ phiếu trong tương lai có thể tạo ra lợi nhuận đáng kể



Back

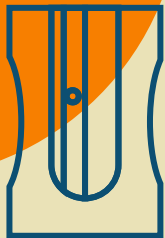


Next



03

Mô tả Dataset



Back



Next

Dữ liệu về chứng khoán của nhóm được lấy từ trang
<https://www.investing.com/indices/vn> với 5 bộ data như sau

HNX



Back



HNX30



Next

Dữ liệu về chứng khoán của nhóm được lấy từ trang
<https://www.investing.com/indices/vn> với 5 bộ data như sau

VNI



Back



VN30



Next

Dữ liệu về chứng khoán của nhóm được lấy từ trang
<https://www.investing.com/indices/vn> với 5 bộ data như sau

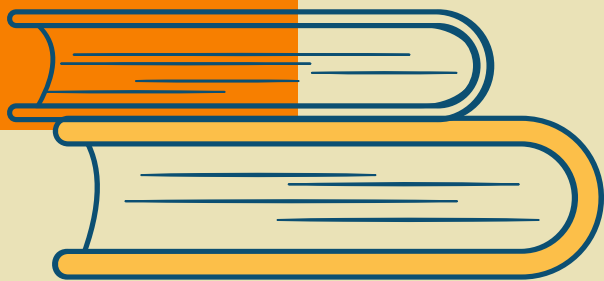
VN100



Back



Next



04

Các thuật toán sử dụng

Back

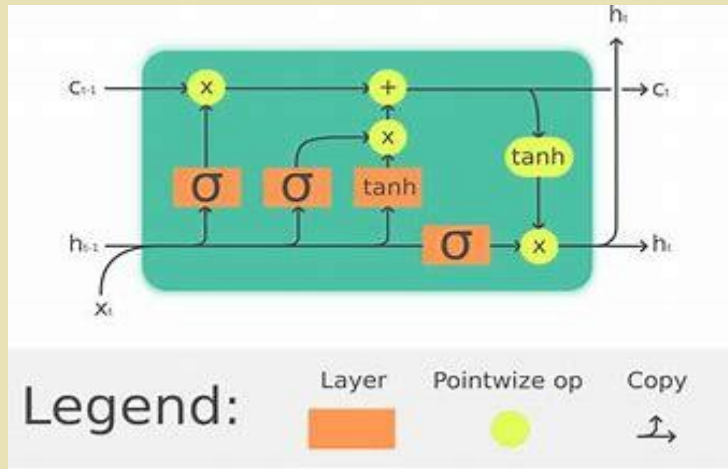


Next

Thuật toán sử dụng cho dự án

LSTM

ARIMA



ARIMA

Autoregressive Moving Averages

$ARIMA(p,d,q)(P,D,Q)m$

- p = non-seasonal AR order
- d = non-seasonal differencing
- q = non-seasonal MA order
- P = seasonal AR order
- D = seasonal differencing
- Q = seasonal MA order
- m = number of periods/season


$ARIMA(1,0,1)(4,0,7,24)$



Back



Next



05

Kết quả mô hình

Back



Next

LSTM MODEL



Back



Next

ARIMA MODEL



LSTM MODEL



```
[951] rmse = sqrt(mean_squared_error(y_test, predictions))  
      print('Test RMSE - LSTM: %.3f' % rmse)
```

Test RMSE - LSTM: 26.699

ARIMA MODEL

```
[958] ## ARIMA MODEL  
      rmse = sqrt(mean_squared_error(test, predictions))  
      print('Test RMSE - ARIMA: %.3f' % rmse)
```

Test RMSE - ARIMA: 15.700





06

Sản phẩm web



Back



Next

Sản phẩm

Hi, we are Group 8 from Big Data Club 🙋

Time Series Forecasting - Stock Prediction

This product is used for Stock Prediction.

[See our poster >](#)

About our project

[See our report >](#)

[See our slide >](#)

Import data:

Please import as CSV file



Drag and drop file here

Limit 200MB per file

[Browse files](#)

Back



Next

Sử dụng web

Hi, we are Group 8 from Big Data Club 🙋

Time Series Forecasting - Stock Prediction

This product is used for Stock Prediction.

[See our poster >](#)

About our project

[See our report >](#)

[See our slide >](#)

Import data:

Please import as CSV file



Drag and drop file here

Limit 200MB per file

Browse files

Back




Next


Sử dụng web

Import data:

Please import as CSV file

 Drag and drop file here
Limit 200MB per file

Browse files

 HNX30.csv 131.1KB ×

Load data successful 23:44:07

Data visualization

Click here to view data

Click here to visualize your data

Predict price

Click here to start model

Back



Next

Feedback

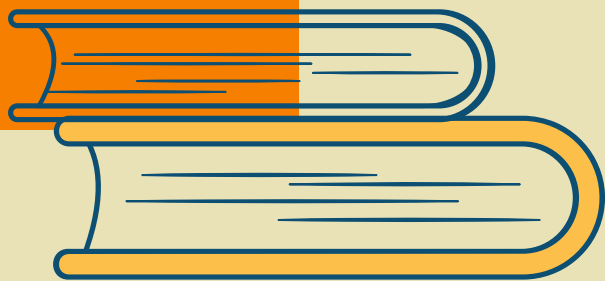
Give us your feedback!

Send

Back



Next



07

Định hướng phát triển

Back



Next

Định hướng phát triển trong tương lai

Để giúp tối ưu được sản phẩm, nhóm cũng đã có một vài phương án cho tương lai:

- ① Xem xét sử dụng một số phương pháp kết hợp LSTM và ARIMA để tận dụng tính ưu việt và ưu điểm để nâng cao độ chính xác của mô hình.
- ② Nghiên cứu thêm các mô hình có khả năng tiếp nhận các sự kiện thực tế và cũng như có khả năng nội suy các yếu tố tác động đến giá cả thị trường chứng khoán.
- ③ Triển khai thêm các phương pháp để nhận dữ liệu liên tục và đưa ra các dự đoán động so với dữ liệu tĩnh hiện nay.

Back



Next



Thanks for watching

