

BigData

Stock Market Analysis

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

GVHD: TS. Đỗ Trọng Hợp

HVTH: Trần Văn Sơn - CH2001013

Trần Bình Hậu - CH2001004

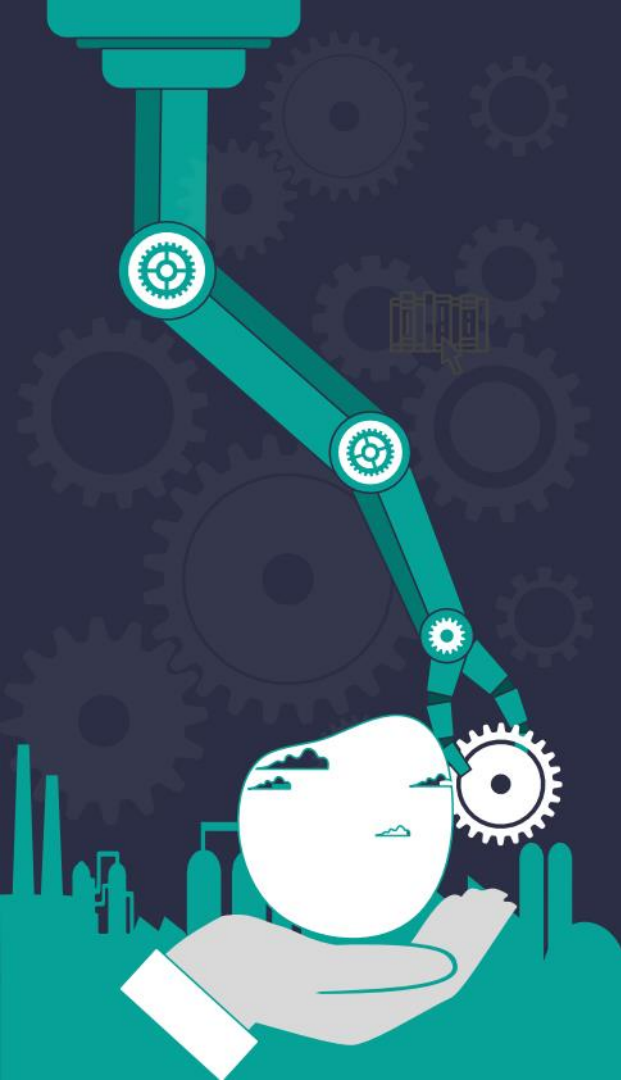
Trần Quốc Thành - CH2001035

Nguyễn Xuân Huy - CH2002007

Phan Thanh Tuấn - CH2002023

NỘI DUNG

- 1 Giới thiệu
- 2 Dữ liệu - Mô hình
- 3 Thực nghiệm
- 4 Kết luận & Tương lai
- 5 Hỏi đáp



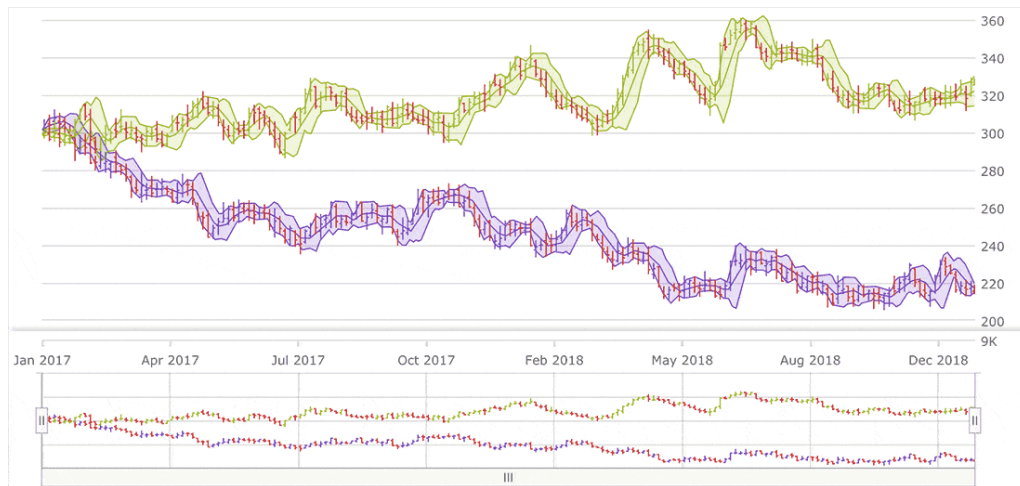
1. Giới thiệu - Phát biểu bài toán

Căn cứ: Dữ liệu chứng khoán thu thập từ API Yahoo/VNDirect/CafeF.

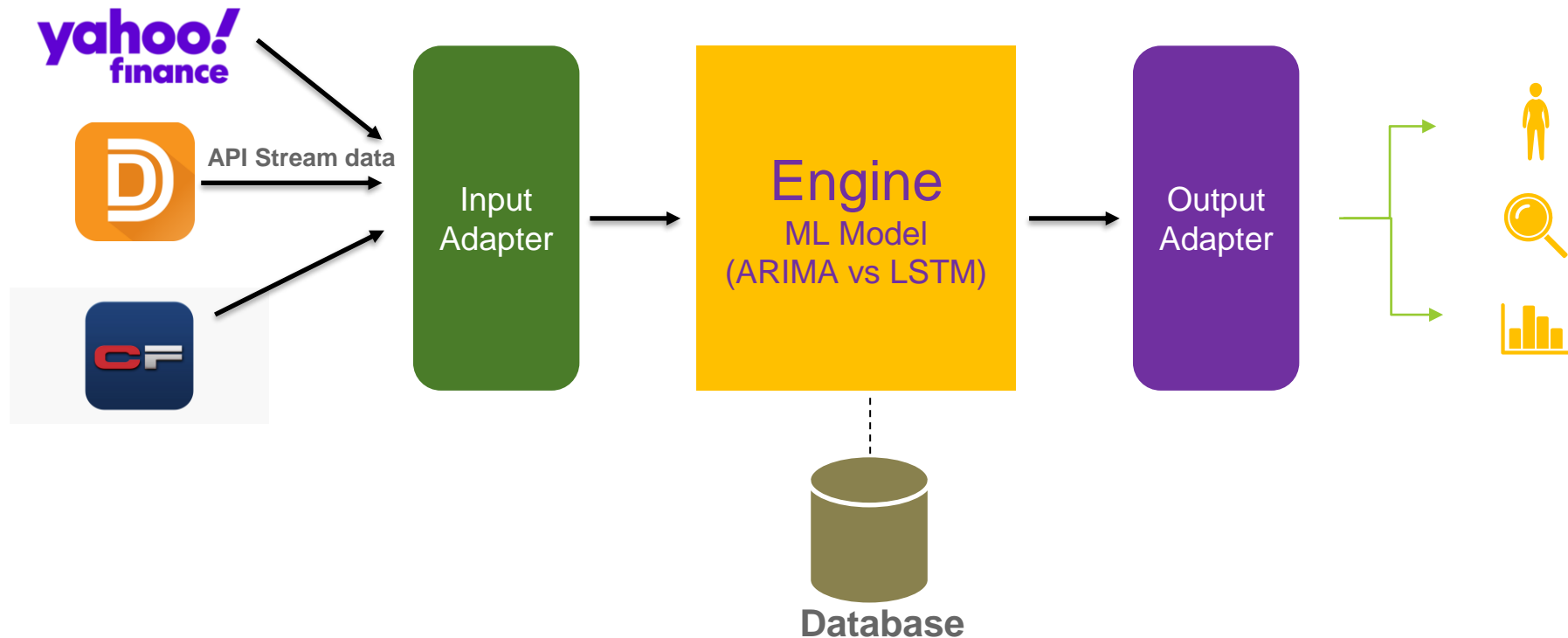
Input: dữ liệu lịch sử giao dịch của mã cổ phiếu.

Output: Dự đoán giá tương lai của mã cổ phiếu đầu vào.

Ràng buộc: Dữ liệu được cung cấp dạng Streaming qua API hoặc Database.



2. Dữ liệu - mô hình



3. Thực nghiệm

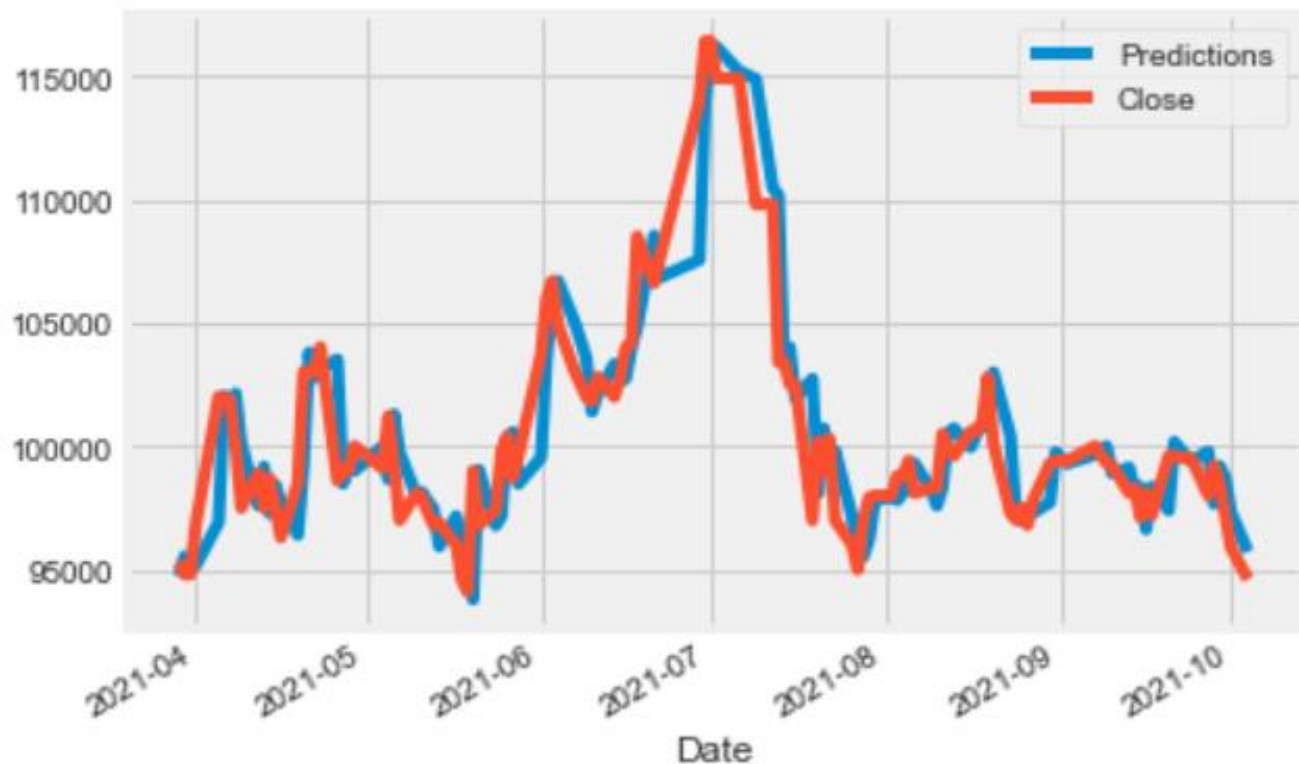
Dữ liệu
trả về từ
API
Stream
sau khi
đã xử lý,
chia tỷ lệ
train: test
là 90:10

	Open	High	Low	Close	Adj Close	Volume	company_name
Date							
2021-09-28	97800.0	99300.0	97500.0	99200.0	99200.0	484500	Vietcombank
2021-09-29	99500.0	99500.0	97500.0	98500.0	98500.0	565000	Vietcombank
2021-09-30	98500.0	98600.0	97100.0	97200.0	97200.0	1199900	Vietcombank
2021-10-01	97300.0	98200.0	95800.0	95900.0	95900.0	1611900	Vietcombank
2021-10-04	96000.0	96500.0	94400.0	94600.0	94600.0	91340000	Vietcombank



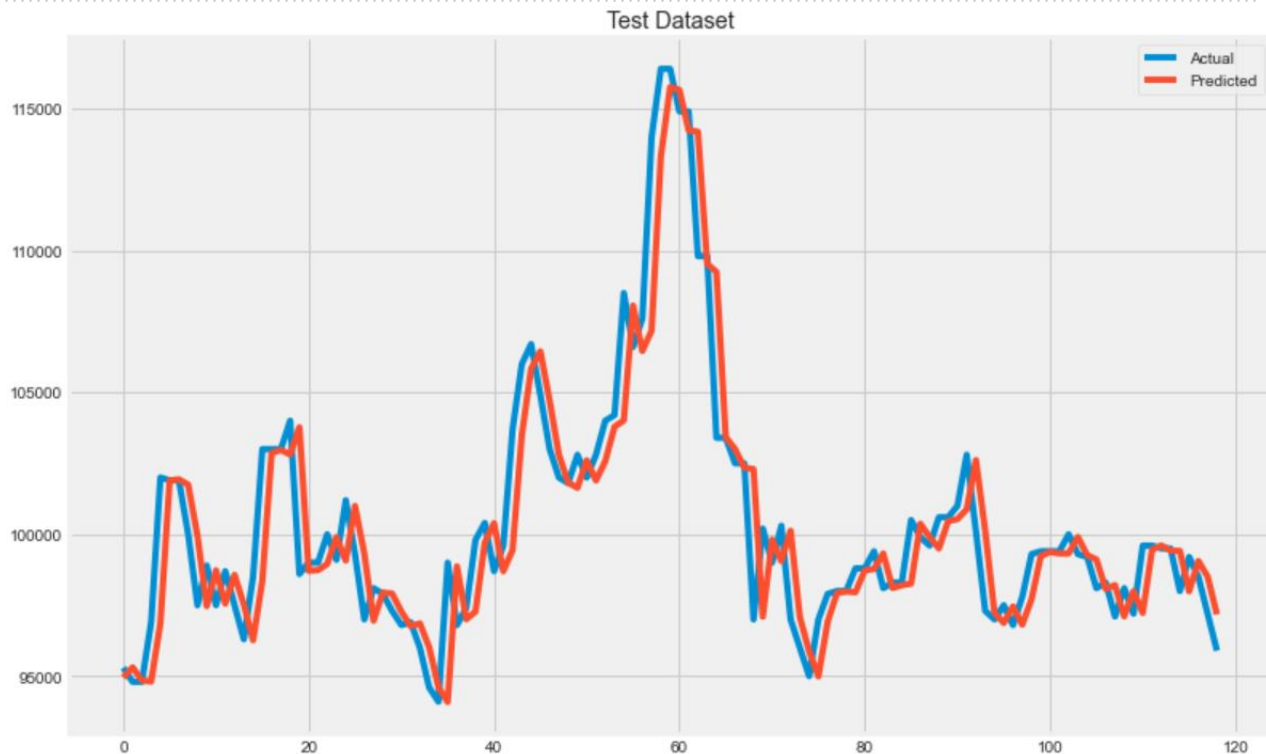
3. Thực nghiệm - ARIMA

Kết quả dự đoán của SARIMAX rất khớp với dữ liệu thực tế



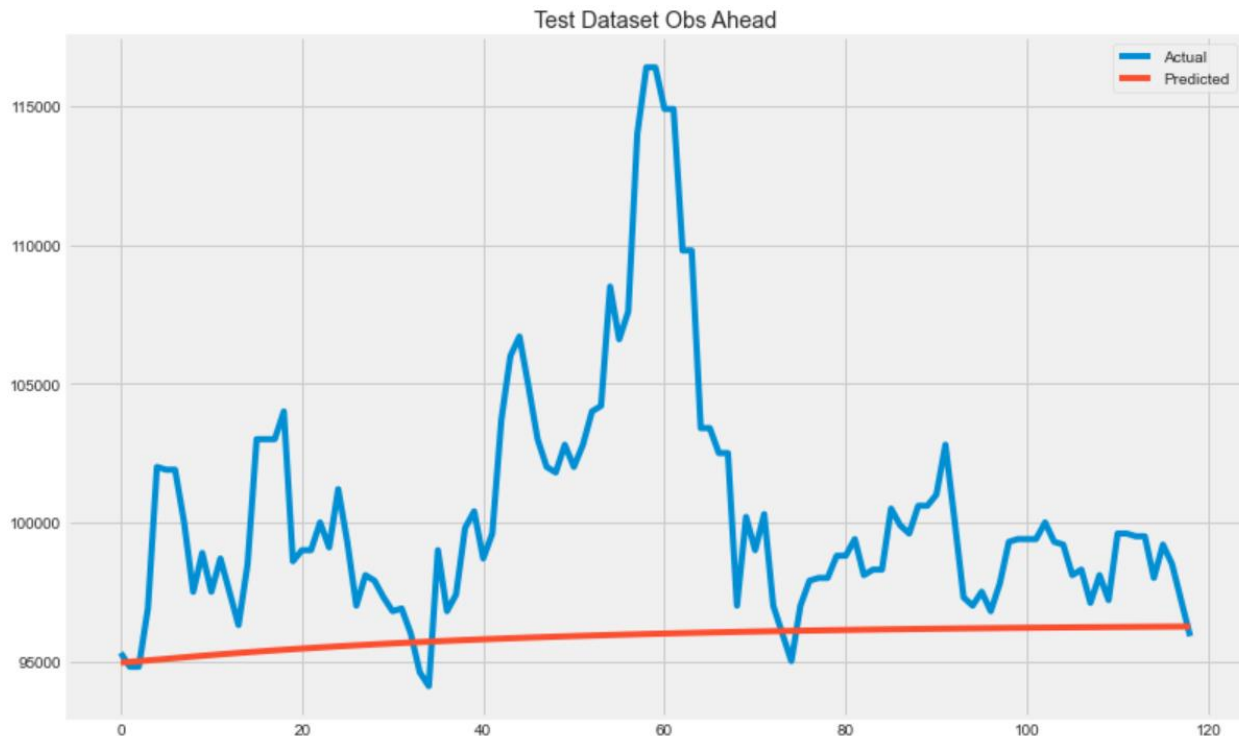
3. Thực nghiệm - LSTM

Kết quả dự đoán
cũng rất khớp với
dữ liệu thực tế



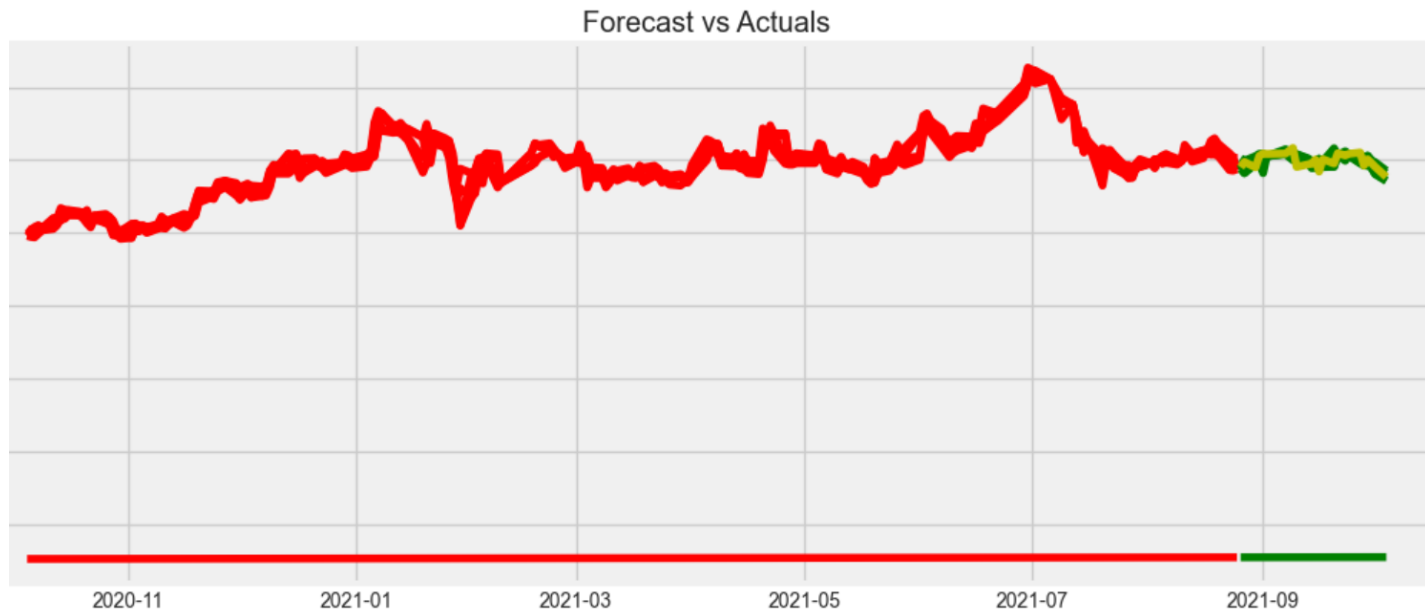
3. Thực nghiệm

Dự đoán những ngày sắp tới nếu dùng **LSTM**, sẽ cho dữ liệu dạng tuyến tính, rất khác với thực tế



3. Thực nghiệm

Nếu dùng **ARIMA**
thì dự đoán
những ngày sắp
tới tốt hơn, do nó
không cần nhận
dữ liệu giả lập



3. Thực nghiệm

So sánh độ lỗi ARIMA
thấp hơn LSTM

```
#####RMSE#####  
from sklearn.metrics import mean_squared_error  
from statsmodels.tools.eval_measures import rmse  
  
# RMSE for ARIMA model  
err_ARIMA = rmse(test["Close"], predictions)  
print('RMSE with ARIMA', err_ARIMA)  
  
#RMSE for LSTM Model  
err_LSTM = rmse(test["Close"], model.predict(X_test))  
print('RMSE with LSTM', np.average(err_LSTM))
```

```
RMSE with ARIMA 1958.2228616563912  
RMSE with LSTM 100346.65292759417
```

4. Kết luận & hướng phát triển



Khi tăng lượng dữ liệu train, thì LSTM sẽ tốt dần, tốc độ nhanh hơn ARIMA.

Khi tính toán ARIMA yêu cầu nhiều tham số, còn LSTM thì không. Mô hình ARIMA là tuyến tính, còn LSTM là phi tuyến.



Lý do cổ phiếu Việt Nam biến động trong biên độ hẹp, nên gần như là tuyến tính, vì thế áp dụng ARIMA hiệu quả hơn.

Nếu dùng mô hình này cho Crypto currency hay Forex thì ARIMA không hiệu quả.

Có thể dùng ARIMA và LSTM trong dự báo một số lĩnh vực khác.

5. Hỏi & Giải đáp

Thanks Question,
Problem

