VIỆN TRÍ TUỆ NHÂN TẠO

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ

BITCOIN PRICE STREAMING PROCESSING

Học phần: Kỹ thuật và công nghệ dữ liệu lớn

Giảng viên: TS. Trần Hồng Việt

ThS. Ngô Minh Hương

CN. Lương Sơn Bá

Tổng quan

1. Giới thiệu dự án
2. Đặt vấn đề
3. Dữ liệu sử dụng
4. Công nghệ & Công cụ
5. Triển khai
6. Kết quả
7. Khó khăn & Giải pháp
8. Thảo luận

1 Giới thiệu dự án

Thành viên:

- Phạm Thành Long - 22022604
- Nguyễn Đức Minh - 22022533
- Phan Văn Hiếu - 22022527
- Trần Tiến Nam - 22022594

Mục tiêu:

- Phát triển một hệ thống dự đoán giá Bitcoin theo thời gian thực.
- Tạo một hệ thống xử lý dữ liệu mạnh mẽ và có khả năng mở rộng, cho phép thu thập, xử lý và phân tích dữ liệu Bitcoin từ các sàn giao dịch.
- Ứng dụng mô hình học máy để dự đoán xu hướng Bitcoin trong ngắn hạn.
- Đảm bảo về tính khả dụng và khả năng mở rộng.

2 Đặt vấn đề

- Dự đoán giá Bitcoin và crypto, chứng khoán nói chung rất quan trọng.
- Dự đoán chính xác và nhanh chóng có thể giúp xác định thời điểm mua vào hoặc bán ra hay giúp quản lý các chiến lược đầu tư hợp lý.

2 Đặt vấn đề

Thách thức

Khối lượng dữ liệu lớn

Chi phí tính toán

Quản lý dữ liệu

Tốc độ xử lý dữ liệu (real-time)

Tích hợp hệ thống

Độ chính xác của mô hình

3 Dữ liệu

- Dữ liệu trong dự án này được lấy từ một bộ dữ liệu của Kaggle.
- Bộ dữ liệu gồm 10 columns với hơn 3M bản ghi.

timestamp Timestamp	open Opening price for that timestamp	high Highest price for the timestamp interval	low Lowest price for the timestamp interval	close Closing price for the timestamp interval	volume Volume for the specified timestamp range
2017-08-17 2023-08-01	2.83k 69k	2.83k 69k	2.82k 68.8k	2.82k 69k	0 5.88k
2023-08-01 13:19:00	28902.48	28902.49	28902.48	28902.49	4.68658
2023-08-01 13:18:00	28902.48	28902.49	28902.48	28902.49	4.77589
2023-08-01 13:17:00	28908.52	28908.53	28902.48	28902.49	11.52263
2023-08-01 13:16:00	28907.41	28912.74	28907.41	28908.53	15.8961
2023-08-01 13:15:00	28896.0	28907.42	28893.03	28907.41	37.74657
2023-08-01 13:14:00	28890.4	28896.0	28890.39	28895.99	9.88869
2023-08-01 13:13:00	28889.63	28890.4	28889.63	28890.39	17.87871
2023-08-01 13:12:00	28881.54	28889.64	28881.53	28889.64	13.48153
2023-08-01 13:11:00	28876.0	28881.54	28875.99	28881.54	6.85924
2023-08-01 13:10:00	28872.48	28876.0	28870.0	28876.0	10.75734

Bitcoin Price Dataset (2017-2023)

Analyzing Bitcoin Price Dynamics: High-Frequency Data from 2017 to 2023

kaggle.com

3 Dữ liệu

- Cung cấp bản ghi toàn diện về sự phát triển giá của Bitcoin, kéo dài từ tháng 8 năm 2017 đến tháng 7 năm 2023.
- Dữ liệu được thu thập từ API Binance, với dữ liệu giá được ghi lại trong khoảng thời gian một phút.
- Mỗi bản ghi bao gồm các thông tin cần thiết như giá mở cửa, giá cao nhất, giá thấp nhất và giá đóng cửa cùng với khối lượng giao dịch liên quan.

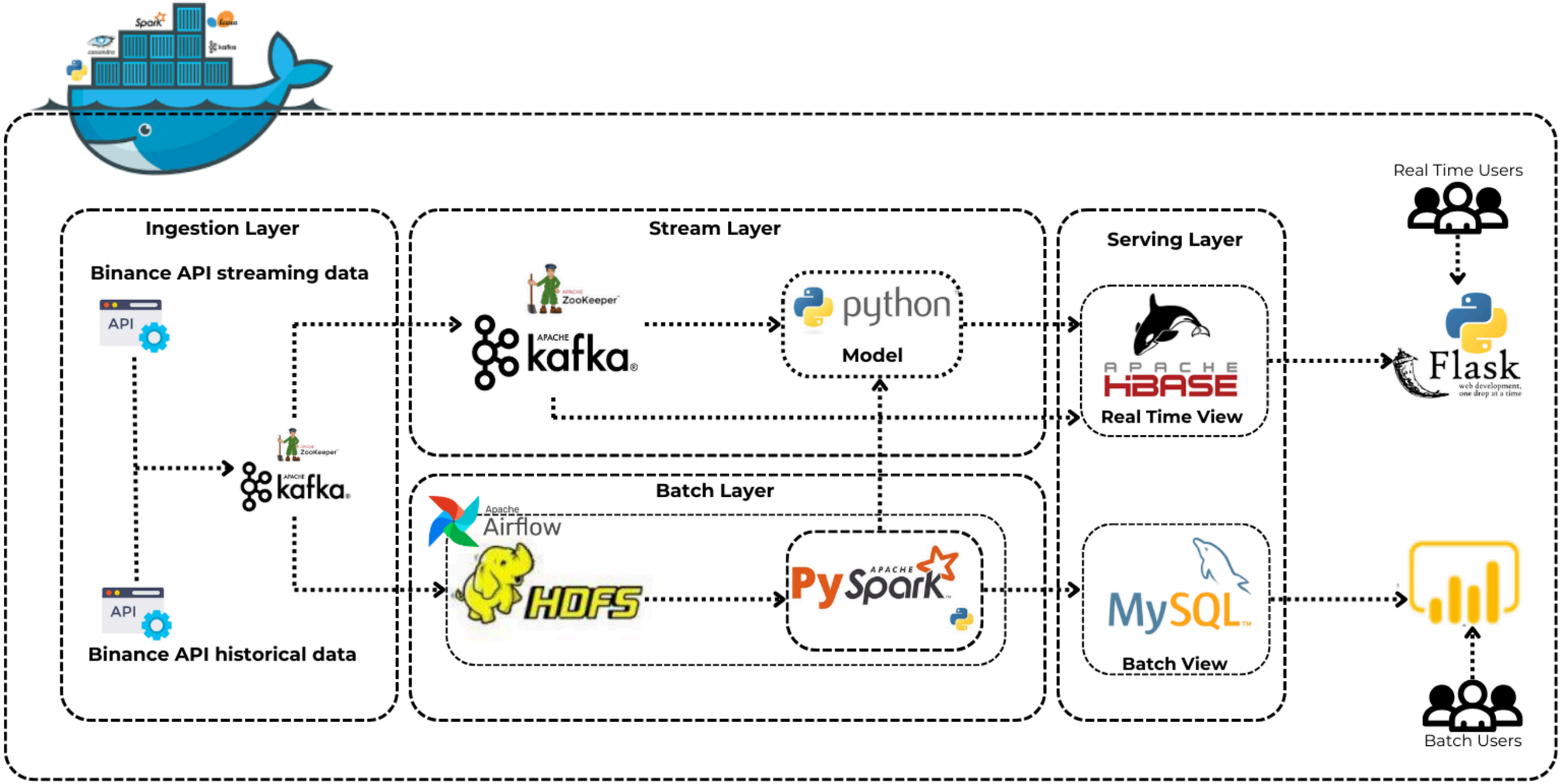
Bitcoin Price Dataset (2017-2023)

Analyzing Bitcoin Price Dynamics: High-Frequency Data from 2017 to 2023

k kaggle.com

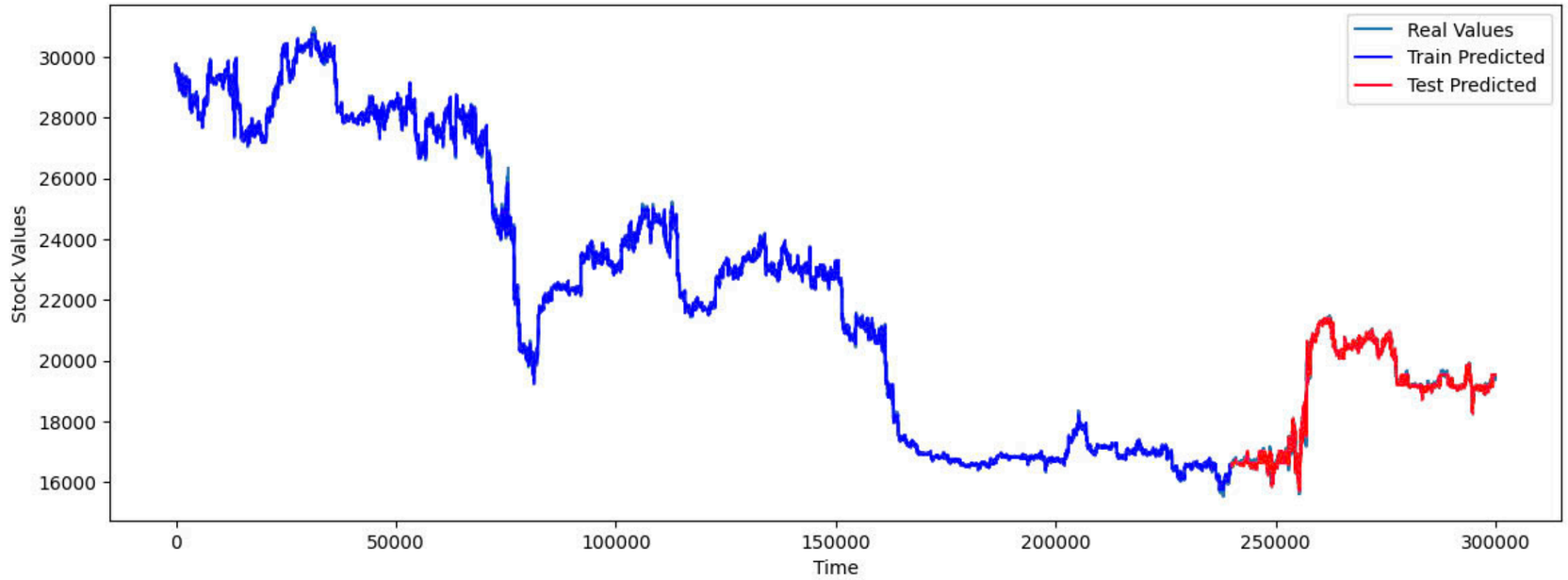
4 Công nghệ & Công cụ

5 Triển khai


Ingestion Layer
Binance API streaming data
API
ZooKeeper
kafka
API
Binance API historical data
Stream Layer
ZooKeeper
kafka
python
Model
Batch Layer
Airflow
HDFS
PySpark
Serving Layer
HBASE
Real Time View
MySQL
Batch View
Real Time Users
Flask
Batch Users

6 Kết quả

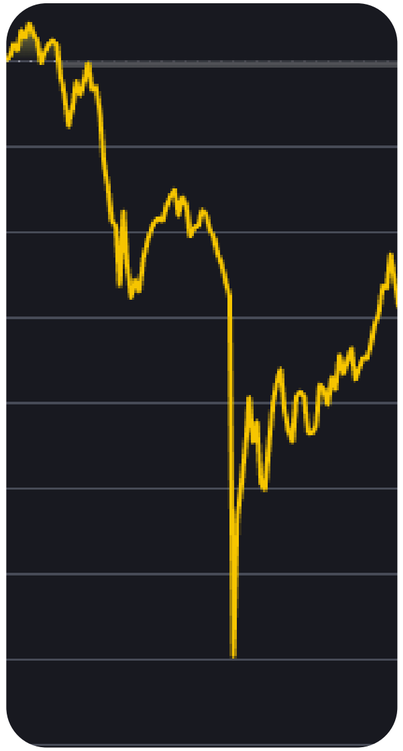
- Hệ thống hoạt động ổn định, có thể xử lý đồng thời nhiều sự kiện.
- Dữ liệu được trực quan hóa với các biểu đồ xu hướng giá theo thời gian.



Mô hình dự đoán giá Bitcoin với RMSE khoảng 5-15$.

7 Khó khăn - Thách thức

- Xử lý các outlier, mô hình dữ liệu phức tạp
- Kiến thức chuyên môn còn chưa cao
- Phạm vi dự án còn giới hạn với các loại dữ liệu Crypto và các loại dữ liệu tương tự (dữ liệu chứng khoán, ...)
- Mô hình mới chỉ đáp ứng được dự đoán giá với thời gian ngắn hạn, chưa mở rộng để dự đoán với các mốc thời gian dài hạn.



8 Thảo luận - Phương hướng phát triển

- Nâng cấp, cải tiến mô hình, kết hợp các mô hình học sâu hiện đại để dự đoán dài hạn.
- Nắm bắt, tìm hiểu kiến thức để dự đoán với độ chính xác cao.
- Phát triển dự án thành sản phẩm hữu ích cho người đầu tư, giúp xác định thời điểm đầu tư, đưa ra kế hoạch đầu tư dài hạn.

THANKS FOR WATCHING