**Hướng dẫn sử dụng folder TLCN**

**ĐỀ TÀI: XÂY DỰNG HỆ THỐNG THEO DÕI HOẠT ĐỘNG WEBSITE (WEBSITE ACTIVITY TRACKING) VỚI APACHE KAFKA**

**GIÁO VIÊN HƯỚNG DẪN: ThS. Nguyễn Minh Đạo**

**GIÁO VIÊN PHẢN BIỆN: ThS. Quách Đình Hoàng**

**Sinh viên thực hiện:**

* **LÊ CHÍ HIẾU: 18133012**
* **NGUYỄN TUẤN VŨ: 18133064**

**Mục lục**

[1. Môi trường cài đặt 3](#_Toc92702841)

[2. Cài đặt MERN 4](#_Toc92702842)

[3. Cài đặt Java 1.8 4](#_Toc92702843)

[4. Cài đặt Apache Zookeeper - Apache Kafka – Spark Streaming 5](#_Toc92702844)

[5. Cài đặt Python 3.6.8 5](#_Toc92702845)

[6. Hướng dẫn chạy chương trình 5](#_Toc92702846)

[6.1. Khởi động phần mềm 5](#_Toc92702847)

[6.1.1. Khởi động Apache Zookeeper - Apache Kafka 5](#_Toc92702848)

[6.1.2. Khởi động Jupyter notebook và chạy Spark Streaming 6](#_Toc92702849)

[6.1.3. Khởi động web MERN 8](#_Toc92702850)

[6.2. Kiểm thử chương trình 8](#_Toc92702851)

**Mục lục hình**

[Hình 1: Version node 4](#_Toc92700600)

[Hình 2: Version npm 4](#_Toc92700601)

[Hình 3: Java 1.8 4](#_Toc92700602)

[Hình 4: Thiết lập JAVA\_HOME 4](#_Toc92700603)

[Hình 5: Môi trường path Java 4](#_Toc92700604)

[Hình 6: Thiết lập môi trường Hadoop, Kafka 5](#_Toc92700605)

[Hình 7: Thiết lập môi trường cho Zookeeper 5](#_Toc92700606)

[Hình 8: Thiết lập môi trường cho Spark 5](#_Toc92700607)

[Hình 9: Môi trường path Hadoop, Spark, Zookeeper 5](#_Toc92700608)

[Hình 10: Môi trường path Kafka 5](#_Toc92700609)

[Hình 11: Chạy Zookeeper 5](#_Toc92700610)

[Hình 12: Chạy kafka 6](#_Toc92700611)

[Hình 13: Chạy Jupyter notebook 6](#_Toc92700612)

[Hình 14: Chạy code 6](#_Toc92700613)

[Hình 15: Function spark streaming 7](#_Toc92700614)

[Hình 16: Cài đặt thư viện 7](#_Toc92700615)

[Hình 17: Chạy web 7](#_Toc92700616)

[Hình 18: Giao diện web 8](#_Toc92700617)

[Hình 19: Click sản phẩm Xiaomi 8](#_Toc92700618)

[Hình 20: Function gợi ý sản phẩm 9](#_Toc92700619)

[Hình 21: Dữ liệu được lưu vào mongoDB 9](#_Toc92700620)

[Hình 22: Sản phẩm tương tự 10](#_Toc92700621)

1. Môi trường cài đặt

Nhóm em sử dụng hệ điều hành Windows và môi trường local làm môi trường cài đặt, kiểm thử phần mềm.

1. Cài đặt MERN

Để chạy Mern Stack cần phải cài đặt node. Link: <https://nodejs.org/en/download/>

Tải về và cài đặt node.

Phiên bản node

Text

Description automatically generated

**Hình 1: Version node**

Phiên bản npm

Text

Description automatically generated

**Hình 2: Version npm**

1. Cài đặt Java 1.8

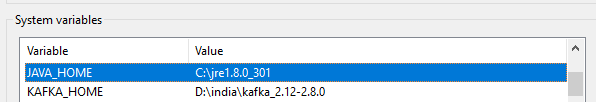
Chạy file cài đặt: (Lưu ý để thư mục cài đặt tại ổ C:\



**Hình 3: Java 1.8**

Thiết lập biến môi trường path dẫn đến folder cài đặt java:

System variables:



**Hình 4: Thiết lập JAVA\_HOME**

Path:

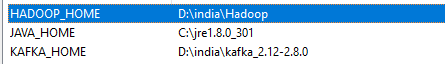


**Hình 5: Môi trường path Java**

1. Cài đặt Apache Zookeeper - Apache Kafka – Spark Streaming

Giải nén thư mục kafka\_sparkStreaming.zip tại ổ đĩa D:\ và thiết lập các biến môi trường như sau:

System variables:



**Hình 6: Thiết lập môi trường Hadoop, Kafka**

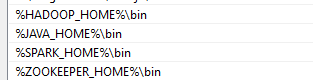


**Hình 7: Thiết lập môi trường cho Zookeeper**



**Hình 8: Thiết lập môi trường cho Spark**

Path:



**Hình 9: Môi trường path Hadoop, Spark, Zookeeper**



**Hình 10: Môi trường path Kafka**

1. Cài đặt Python 3.6.8

Cài đặt thư viện jupyter notebook và 1 số thư viện hỗ trợ nếu có phát sinh lỗi:

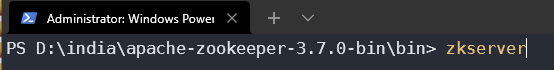
pip install notebook

1. Hướng dẫn chạy chương trình
   1. Khởi động phần mềm

Sau khi reset máy, chúng ta khởi chạy từng phần mềm theo thứ tự sau:

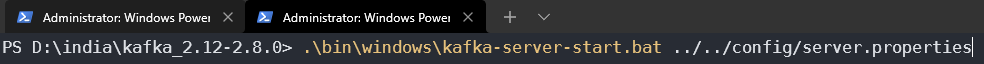
* + 1. Khởi động Apache Zookeeper - Apache Kafka

Mở cmd trò vào thư mục bin của zookeeper và chạy lệnh



**Hình 11: Chạy Zookeeper**

Mở cmd trò vào thư mục chứa kafka và chạy lệnh



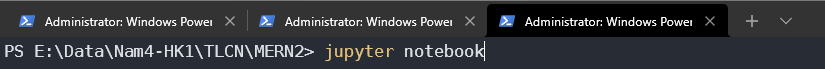
**Hình 12: Chạy kafka**

Hoặc dùng lệnh trỏ trực tiếp vào file với cú pháp sau:

D:\india\kafka\_2.12-2.8.0\bin\windows\kafka-server-start.bat D:\india\kafka\_2.12-2.8.0\config\server.properties

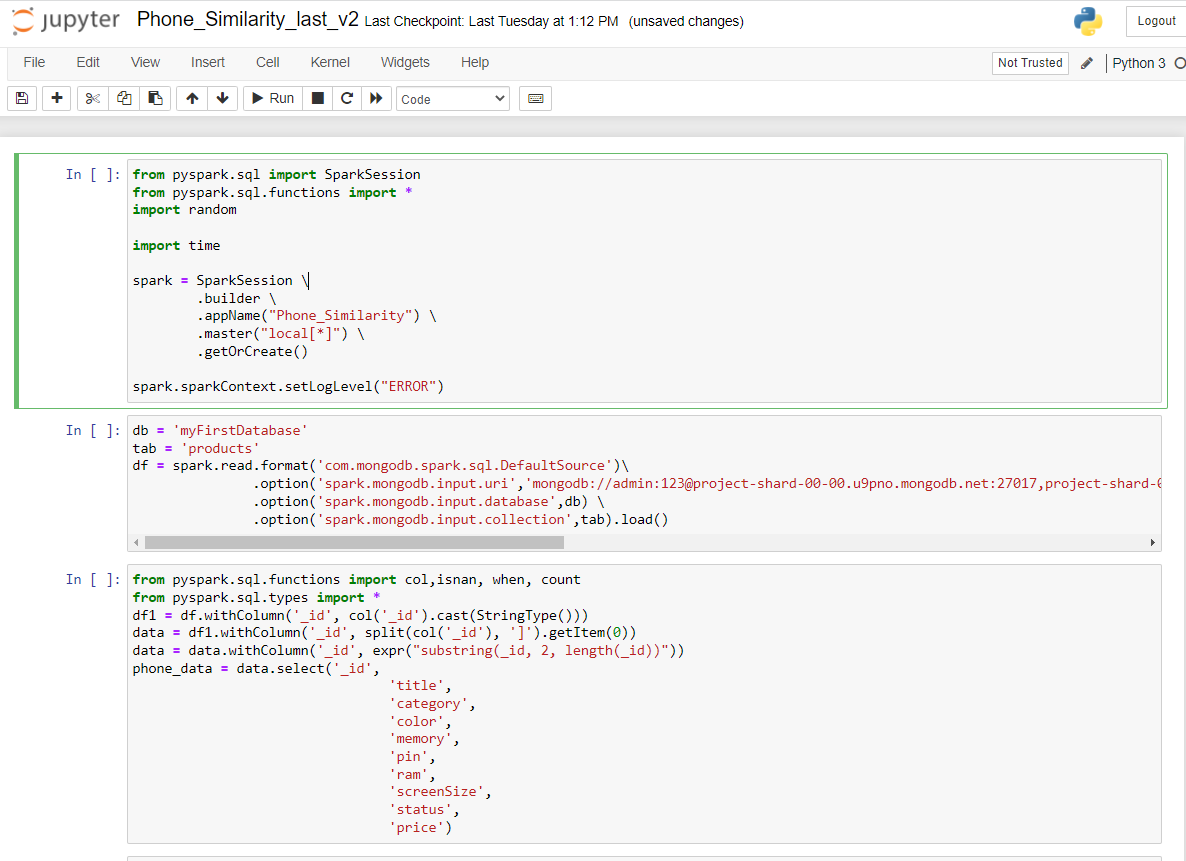
* + 1. Khởi động Jupyter notebook và chạy Spark Streaming

Mở cmd trỏ vào thư mục chứa source web và chạy lệnh



**Hình 13: Chạy Jupyter notebook**

Sau khi jupyter khởi động lên trình duyệt web, vào thư mục Kafka\_Spark\_Similarity và bấm vào file Phone\_Similarity\_last\_v2.ipynb



**Hình 14: Chạy code**

Khởi chạy các cell lệnh của Spark Streaming đến for\_batch\_function (Lưu ý không chạy query3.stop() vì hàm này sẽ dừng Spark Streaming:



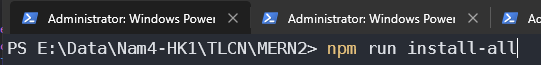
**Hình 15: Function spark streaming**

, trong quá trình chạy, nếu có lỗi thiếu thư viện xuất hiện, hãy cài đặt theo cú pháp:

pip install “tên thư viện”

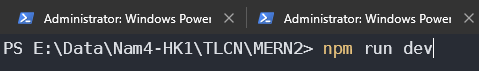
* + 1. Khởi động web MERN

Mở cmd trỏ vào thư mục … và chạy lệnh npm run install-all để cài đặt các thư viện cần thiết



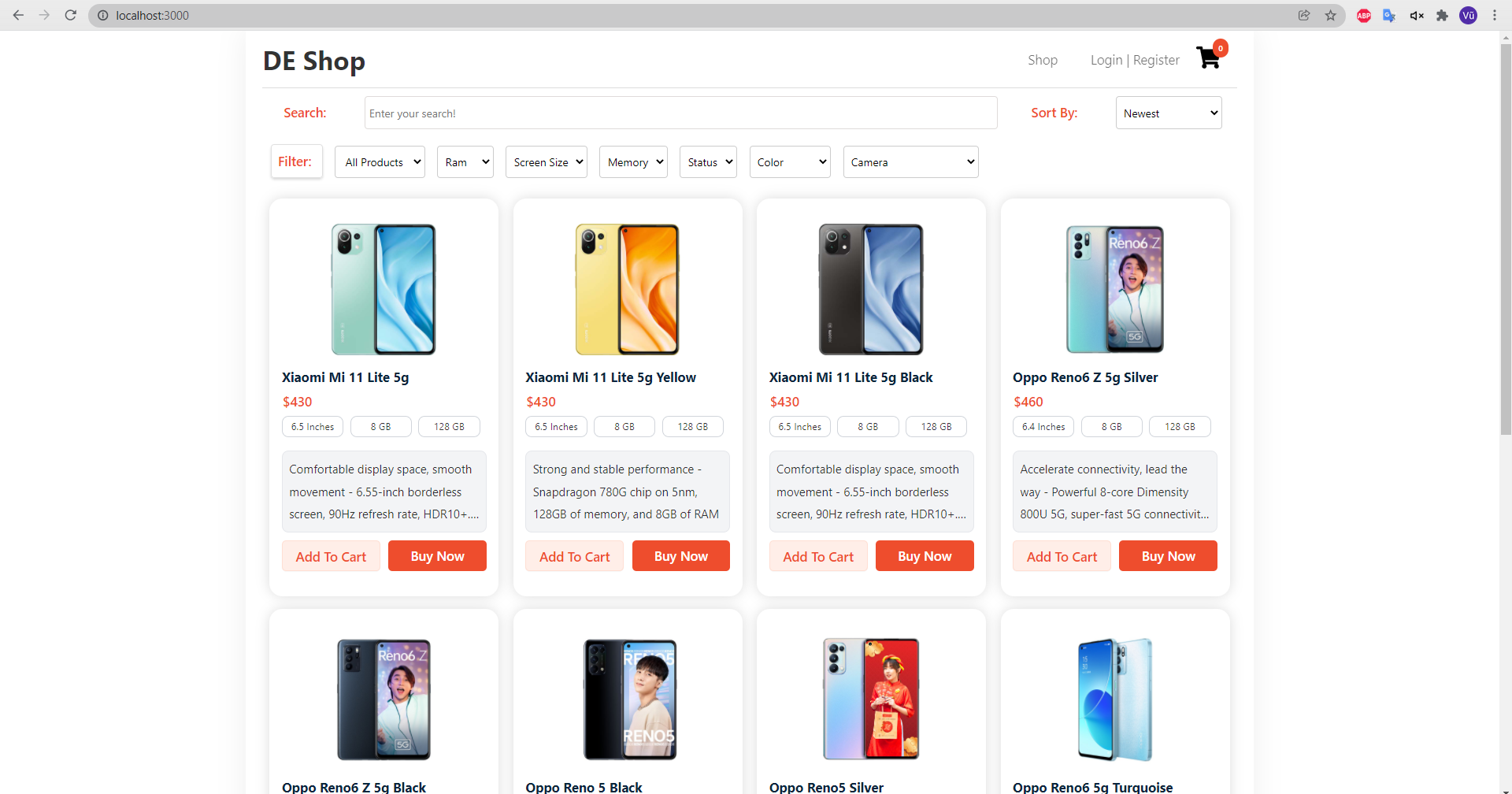
**Hình 16: Cài đặt thư viện**

Sau đó chạy lệnh npm run dev



**Hình 17: Chạy web**

Kết quả giao diện web



**Hình 18: Giao diện web**

* 1. Kiểm thử chương trình

Đăng nhập với tài khoản:

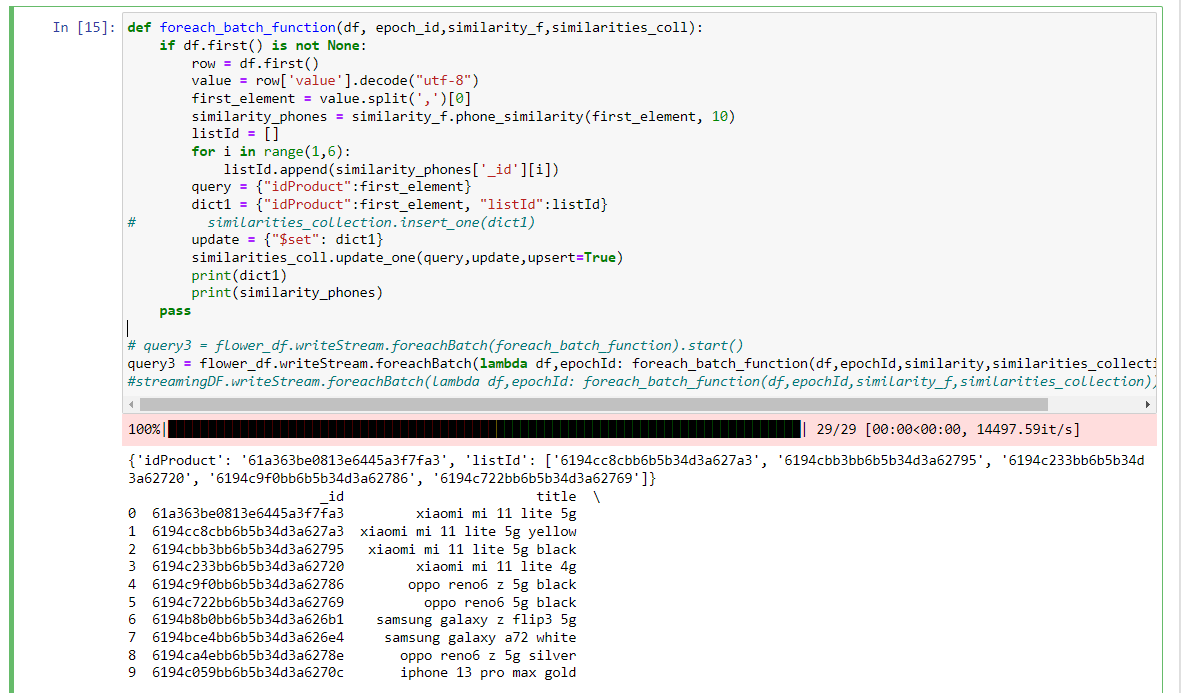
Username: gvpb@gmail.com

Password: 123456

Click thử 1 sản phẩm, id sản phẩm sẽ được gửi xuống lớp kafka, sau đó spark streaming sẽ lắng nghe các id được gửi tới và thực thi hàm đề xuất các sản phẩm tương tự dựa trên RMSE

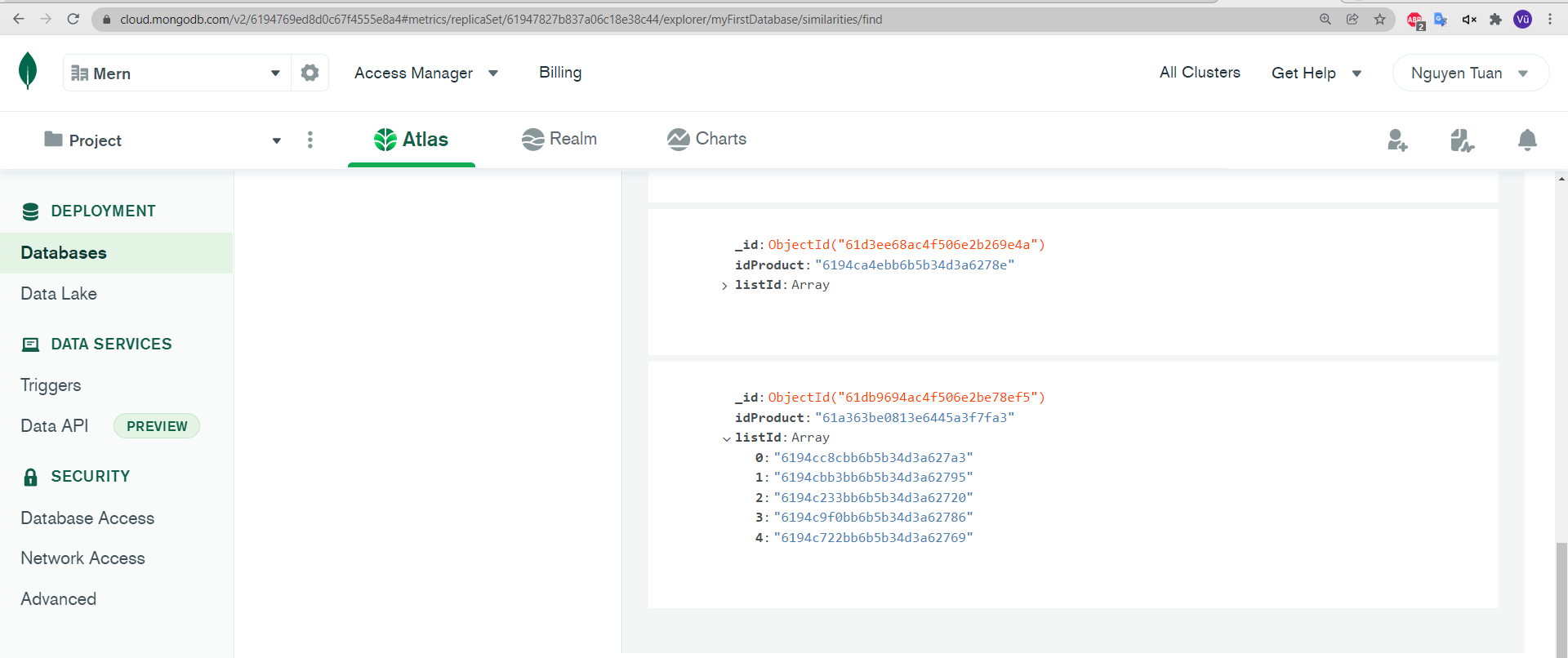


**Hình 19: Click sản phẩm Xiaomi**



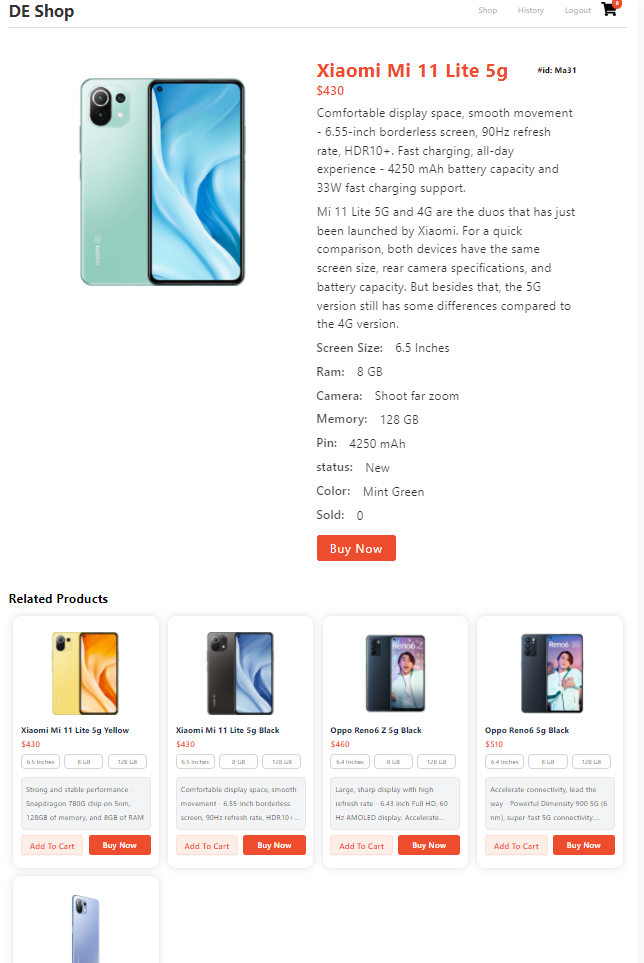
**Hình 20: Function gợi ý sản phẩm**

Sau đó Spark Streaming sẽ gửi kết quả về mongoDB



**Hình 21: Dữ liệu được lưu vào mongoDB**

Sau khi mongoDB nhận được, kết quả đề xuất sẽ hiển thị lên giao diện web



**Hình 22: Sản phẩm tương tự**