

**BỘ GIÁO DỤC VÀ ĐÀO TẠO  
TRƯỜNG ĐẠI HỌC BÌNH DƯƠNG  
KHOA CÔNG NGHỆ THÔNG TIN, ROBOT VÀ TRÍ TUỆ NHÂN TẠO**



**TIỂU LUẬN MÔN HỌC  
KHOA HỌC DỮ LIỆU VÀ ỨNG DỤNG  
ĐỀ TÀI: PHÂN TÍCH CHỈ SỐ SẢN PHẨM  
SHOPEE**

*Giảng viên hướng dẫn:* **Huỳnh Quang Đức**

*Sinh viên thực hiện:* **Phạm Văn Minh – 19050026**

**Bình Dương, ngày 28 tháng 10 năm 2022**

## LỜI CẢM ƠN

Trong suốt quá trình học tại trường Đại học Bình Dương, với sự nỗ lực của bản thân và sự giúp đỡ của các thầy cô giáo trong trường đặc biệt là thầy cô giáo trong khoa Công nghệ thông tin, Robot – Trí tuệ nhân tạo đã giúp chúng em có một vốn tri thức vững vàng để hoàn thành bài tiểu luận của các môn học. Trong thời gian làm tiểu luận các môn học, được sự giúp đỡ của giáo viên hướng dẫn về mọi mặt, từ nhiều phía em đã hoàn thành tốt tiểu luận đúng thời gian quy định. em xin chân thành cảm ơn đến:

Đầu tiên, Nhóm chúng em muốn gửi lời cảm ơn chân thành và biết ơn sâu sắc tới thầy Huỳnh Quang Đức là người luôn theo sát chỉ bảo hướng đi và cho chúng em những lời khuyên quý báu cũng như cung cấp các thông tin và căn cứ khoa học để chúng em định hướng tốt trong suốt quá trình thực hiện tiểu luận môn học Khoa học dữ liệu và ứng dụng.

Nhóm chúng em xin chân thành cảm ơn tới các thầy cô giáo khoa Công nghệ thông tin, Robot – Trí tuệ nhân tạo đã tận tình huấn luyện, cung cấp cho những kiến thức vô cùng quý giá và đã tạo điều kiện tốt nhất cho chúng em trong suốt quá trình học tập, nghiên cứu tại trường.

Đồng thời, chúng em xin cảm ơn tất cả những người thân yêu trong gia đình chúng em, cùng toàn thể bạn bè những người luôn giúp đỡ, động viên chúng em những khi vấp phải những khó khăn, bế tắc.

Mặc dù, tiểu luận đã được hoàn thành đúng thời gian quy định nhưng do điều kiện thời gian, tiểu luận của chúng em không tránh khỏi những thiếu sót. Vì vậy chúng em rất mong nhận được sự đóng góp ý kiến của thầy Huỳnh Quang Đức và bạn bè tạo điều kiện cho tiểu luận của chúng em được hoàn thiện hơn.

***Bình Dương tháng 10 năm 2022***

***Nhóm sinh viên thực hiện***

**Phạm Văn Minh**

## TÓM TẮT

Việc thay đổi thói quen mua sắm từ mua trực tiếp thay đổi thành mua sắm online đã thổi bùng giá trị thị trường các trang thương mại điện tử trở nên khổng lồ. Shopee là một trong những sàn thương mại điện tử lớn nhất khu vực.

Việc nắm bắt xu hướng thị trường từ lâu vốn dĩ đã quan trọng nay lại càng quan trọng hơn trong ngành bán lẻ của các ông lớn thương mại điện tử.

Vì thế em muốn xây dựng một hệ thống phân tích xu hướng thị trường của các ngành hàng trên trang thương mại điện tử Shopee.

### ❖ Các hướng tiếp cận là:

- Thu thập tài liệu bán hàng của Shopee
- Thu thập tài liệu, tìm hiểu các thuật toán phân tích dữ liệu
- Thiết kế hệ thống, xác định quy mô, xác định mục tiêu của hệ thống
- Thực hiện, xây dựng mô hình hoạt động của ứng dụng
- Thiết kế giao diện
- Kiểm thử và sửa lỗi hệ thống

### ❖ Cách giải quyết vấn đề theo yêu cầu của đề tài đặt ra:

- Đầu vào dự án : dữ liệu bán hàng Shopee
- Xử lý và phân tích
- Đầu ra kết quả: dữ liệu được trực quan hóa
- Tính ứng dụng của đề tài: giúp các nhà bán lẻ nắm bắt được xu hướng mua hàng của thị trường.

### ❖ Kết quả đạt được:

- Đảm bảo dữ liệu được phân tích và trực quan hóa chính xác
- Thiết kế giao diện thân thiện dễ sử dụng

## MỤC LỤC

LỜI CẢM ƠN.....	i
TÓM TẮT.....	ii
MỤC LỤC .....	1
PHẦN 1 – TỔNG QUAN VỀ KHOA HỌC DỮ LIỆU .....	3
1.1 Định nghĩa và ứng dụng của khoa học dữ liệu .....	3
1.1.1- Định Nghĩa .....	3
1.1.2- Ứng dụng của khoa học dữ liệu.....	3
1.2 Sàn thương mại điện tử Shopee .....	5
1.2.1- Thương mại điện tử là gì .....	5
1.2.2- Shopee .....	6
PHẦN 2- CƠ SỞ LÝ THUYẾT .....	10
1. CƠ SỞ LÝ THUYẾT .....	10
1.1 Data Analysis.....	10
.....	10
.....	10
1.2 Django.....	10
PHẦN 3 – TRIỂN KHAI ỨNG DỤNG .....	12
3.1 Thực Hiện Đề Tài .....	12
3.1.1- Thu thập dữ liệu .....	12
3.1.2- Xử lý dữ liệu.....	12
3.1.3- Phân tích dữ liệu.....	13
3.1.4- Trực quan hóa dữ liệu .....	15
PHẦN 4 – KẾT QUẢ ĐẠT ĐƯỢC .....	17
4.1 Những kiến thức đã học được thông qua đề tài .....	17
4.2 Kết luận về đề tài đã thực hiện .....	17
TÀI LIỆU THAM KHẢO .....	18

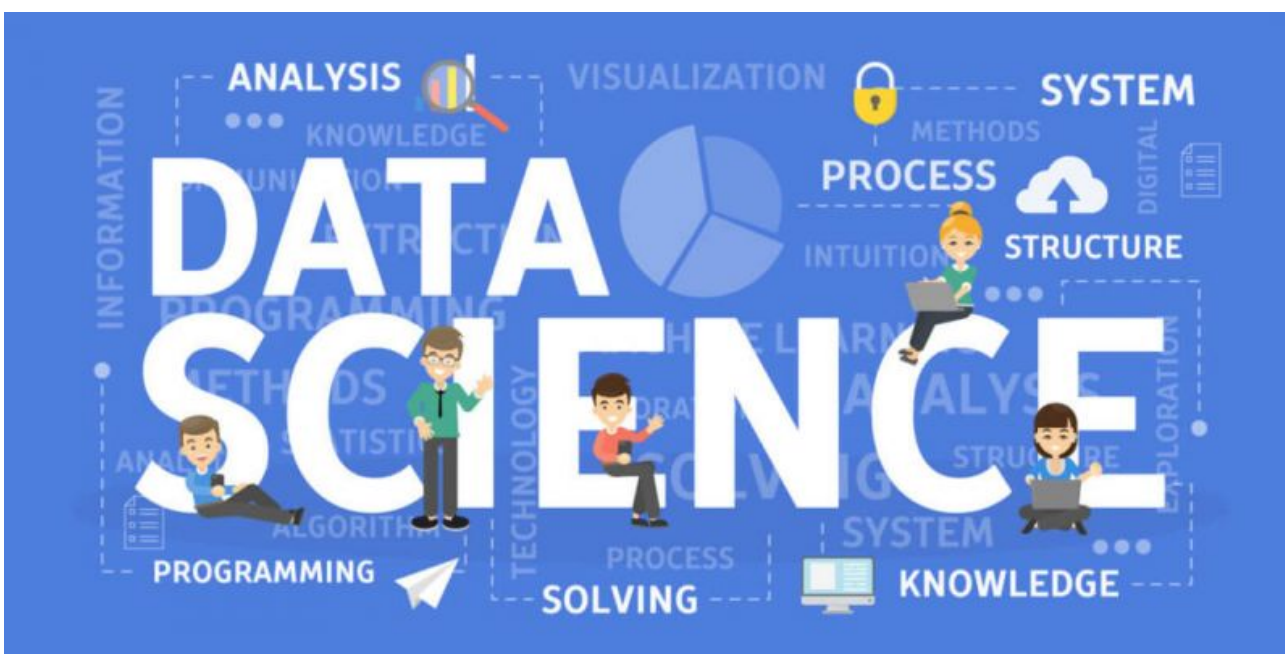


# PHẦN 1 – TỔNG QUAN VỀ KHOA HỌC DỮ LIỆU

## 1.1 Định nghĩa và ứng dụng của khoa học dữ liệu

### 1.1.1- Định Nghĩa

Khoa học dữ liệu (Data science) là ngành khoa học về việc khai phá, quản trị và phân tích dữ liệu để dự đoán các xu hướng trong tương lai và đưa ra các quyết định, chiến lược hành động. Khoa học dữ liệu (Data science) gồm ba phần chính: tạo và quản trị dữ liệu, phân tích dữ liệu, và áp dụng kết quả phân tích thành những hành động có giá trị. Việc phân tích và sử dụng dữ liệu dựa vào ba nguồn tri thức: toán học (thống kê toán học - Mathematical Statistics), công nghệ thông tin (máy học - Machine Learning) và tri thức của lĩnh vực ứng dụng cụ thể.



### 1.1.2- Ứng dụng của khoa học dữ liệu

#### Nhận dạng mẫu

Tương tự như vậy, xác định các mẫu trong tập dữ liệu là một nhiệm vụ khoa học dữ liệu cốt lõi. Ví dụ, nhận dạng mẫu giúp các nhà bán lẻ và công ty thương mại điện tử phát hiện ra xu hướng trong hành vi mua hàng của khách hàng. Việc cung cấp sản phẩm phù hợp và đảm bảo độ tin cậy của chuỗi cung ứng là điều quan trọng đối với các tổ chức muốn giữ cho khách hàng của họ hài lòng – và thay vào đó ngăn họ mua hàng từ các đối thủ cạnh tranh.

## **Phát hiện bất thường**

Một ứng dụng mạnh mẽ của khoa học dữ liệu là việc sử dụng phân tích thống kê để phát hiện những điểm bất thường trong tập dữ liệu, đặc biệt là những tập dữ liệu lớn. Mặc dù có thể là một bài tập hơi đơn giản để sắp xếp dữ liệu vào các cụm hoặc nhóm và sau đó xác định các ngoại lệ khi xử lý một lượng nhỏ dữ liệu, nhiệm vụ này trở nên khó khăn hơn đáng kể đối với các tổ chức phải phân tích lượng petabyte hoặc exabyte dữ liệu.

## **Phân loại và tạo nhóm**

Các công cụ khoa học dữ liệu đã cho thấy khả năng thực sự để sắp xếp thông qua khối lượng lớn dữ liệu và phân loại hoặc phân nhóm các đối tượng dựa trên các đặc điểm đã học. Điều này đặc biệt hữu ích với dữ liệu phi cấu trúc. Trong khi dữ liệu có cấu trúc có thể dễ dàng tìm kiếm và truy vấn thông qua một lược đồ, thì dữ liệu phi cấu trúc khó xử lý và phân tích hơn nhiều. Email, tài liệu, hình ảnh, video, tệp âm thanh, văn bản và thông tin nhị phân thuộc mọi loại là các dạng dữ liệu phi cấu trúc. Cho đến gần đây, việc khai thác dữ liệu đó để có những thông tin chi tiết có giá trị được chứng minh là một thách thức.

## **Mô hình dự đoán**

Ngoài việc phát hiện các mẫu và ngoại lệ, khoa học dữ liệu còn nhằm mục đích làm cho mô hình dự đoán chính xác hơn. Mặc dù phân tích dự đoán đã tồn tại trong nhiều thập kỷ, khoa học dữ liệu áp dụng học máy và các phương pháp tiếp cận thuật toán khác cho các tập dữ liệu lớn để cải thiện khả năng ra quyết định bằng cách tạo ra các mô hình dự đoán tốt hơn hành vi của khách hàng, rủi ro tài chính, xu hướng thị trường và hơn thế nữa.

## **Công cụ đề xuất và hệ thống cá nhân hóa**

Khách hàng và người dùng thường hài lòng nhất khi sản phẩm và dịch vụ phù hợp với nhu cầu hoặc sở thích của họ – đặc biệt là nếu họ có thể nhận được đúng sản phẩm vào đúng thời điểm trên kênh phù hợp, với ưu đãi phù hợp được truyền đạt bằng cách sử dụng đúng thông điệp và quyền mức độ dịch vụ và sự chú ý. Và giữ cho khách hàng hài lòng và gắn bó có nghĩa là họ có thể sẽ tiếp tục quay lại.

## **Phân tích hành vi và cảm xúc**

Dựa trên khả năng phân tích dữ liệu của hệ thống học máy và học sâu, các nhà khoa học dữ liệu đang đào sâu qua hàng loạt dữ liệu để hiểu được cảm xúc của khách hàng hoặc người dùng và hành vi của họ.

Thông qua các ứng dụng phân tích tình cảm và phân tích hành vi, khoa học dữ liệu cho phép các tổ chức xác định hiệu quả hơn các mô hình mua và sử dụng cũng như biết mọi người nghĩ gì về sản phẩm và dịch vụ cũng như mức độ hài lòng của họ với trải nghiệm của họ. Các ứng dụng này cũng có thể phân loại tình cảm và hành vi của khách hàng và theo dõi cách họ thay đổi theo thời gian.

### **Hệ thống tự trị**

Nhắc đến ô tô, một trong những giấc mơ từ lâu của những người đam mê AI là ô tô tự lái. Sẽ thật tuyệt nếu bạn ngồi trên một chiếc ô tô hoặc xe tải và để nó lái trong khi bạn làm những việc khác mà không cần phải chú ý đến những gì đang xảy ra trên đường? Khoa học dữ liệu đang đóng một vai trò lớn trong sự phát triển không ngừng của các phương tiện tự hành, cũng như rô bốt điều khiển bằng AI và các máy móc thông minh khác

### **Hệ thống hội thoại**

Một trong những ứng dụng đầu tiên của học máy là sự phát triển của một chatbot có thể có các cuộc trò chuyện giống như thật mà không cần sự can thiệp của con người. Trên thực tế, Bài kiểm tra Turing, được nhà tiên phong về máy tính Alan Turing nghĩ ra vào năm 1950, sử dụng một định dạng hội thoại để chỉ ra liệu một hệ thống có thể bắt chước trí thông minh của con người hay không. Do đó, không có gì ngạc nhiên khi các tổ chức đang tìm kiếm chatbots và các hệ thống đàm thoại khác để giúp tăng cường quy trình công việc hiện có và đảm nhận một số nhiệm vụ do con người xử lý trước đây.

## **1.2 Sàn thương mại điện tử Shopee**

### **1.2.1- Thương mại điện tử là gì**

Thương mại điện tử là gì? Đây là việc tiến hành một phần hay toàn bộ hoạt động kinh doanh bằng các phương tiện điện tử. Một cách dễ hiểu hơn thì thương mại điện tử chính là việc mua bán sản phẩm hay dịch vụ thông qua internet và các phương tiện điện tử khác. Các giao dịch này bao gồm tất cả các hoạt động như: giao dịch, mua bán, thanh toán, đặt hàng, quảng cáo và giao hàng...

“Bản chất cốt lõi để web và internet phát triển trong tương lai chính là thương mại. Các



trung tâm thương mại trên mạng internet sẽ xuất hiện. Nó sẽ giúp các nhà cung cấp sản phẩm tiếp cận trực tiếp và nhanh nhất với người tiêu dùng”, người nắm giữ chức CEO của Apple tại thời điểm năm 1996 chia sẻ trong hội thảo với chủ đề “Steve Jobs: Điều vĩ đại tiếp theo”. Thật vậy, hiện nay ngành Thương mại điện tử đang có tốc độ phát triển rất mạnh. Hầu hết các công ty bán hàng lập nên ngày nay đa số đều là các công ty thương mại điện tử và mua sắm qua mạng đã trở thành thói quen hàng ngày của nhiều người Việt Nam.

Thế giới ngày càng bùng nổ về mạng máy tính, con người có thể nhanh chóng giao tiếp và kết nối với nhau dễ dàng hơn thông qua nhiều loại dịch vụ internet. Đây chính là điều kiện thuận lợi cho lĩnh vực Thương mại điện tử ngày càng phát triển hơn. Hiện nay, Thương mại điện tử đã trở thành một phương tiện giao dịch quen thuộc của các công ty thương mại lớn trên thế giới. Thương mại điện tử có khả năng giúp ích rất nhiều cho những doanh nghiệp cả lớn lẫn nhỏ và người hưởng lợi nhất thường là khách hàng. Khách hàng sẽ mua được sản phẩm rẻ hơn, nhanh hơn, hiệu quả hơn và thuận lợi hơn, còn doanh nghiệp có thể đưa sản phẩm của mình đến với thị trường một cách nhanh nhất, bán hàng thuận lợi hơn.



### 1.2.2- Shopee

**Shopee là gì? Shopee của nước nào?**

Shopee là trang thương mại điện tử mua sắm, nó được lập ra bởi tập đoàn SEA của Forrest Li ở Singapore vào năm 2015. Nó chính là một cái chợ Online, là trung gian kết nối giữa người mua và người bán, giúp hoạt động kinh doanh online trở nên dễ dàng hơn. Ở đó người bán đăng tải các thông tin về sản phẩm và dịch vụ mà không cần người tư vấn hay vận chuyển, đồng thời người mua cũng tiếp cận được các thông tin ấy một cách trực quan mà không cần đến cửa hàng.

Shopee hiện nay đã có mặt tại 7 quốc gia ở khu vực châu Á đó là: Singapore, Malaysia, Indonesia, Đài Loan, Philippines và đặc biệt ngày 8/8/2016, Shopee đã chính thức ra mắt tại Việt Nam.



*Shopee là gì mà được nhiều người dùng yêu thích đến vậy*

### **Giới thiệu về công ty Shopee**

Trên các sàn thương mại điện tử Shopee có rất nhiều mặt hàng sản phẩm khác nhau với mẫu mã đa dạng: quần áo, đồ chơi, đồ gia dụng, thiết bị điện tử,... Kể từ khi ra mắt doanh thu của Shopee tăng trưởng theo cấp số nhân và ngày càng phát triển mạnh mẽ. Shopee hiện có hơn 160 triệu người dùng đang hoạt động với khoảng 6 triệu người bán, bao gồm hơn 7.000 thương hiệu (Brand) và nhà phân phối hàng đầu.



Shopee hoạt động dựa trên mô hình kinh doanh Win – Win đôi bên cùng có lợi. Và giờ đây nó đang dần trở thành kênh Marketing thay thế cho các kênh bán hàng truyền thống như Facebook hay Youtube.

Nền tảng thương mại điện tử này luôn cập nhật thêm các tính năng mới để người bán cũng như người mua có thể dễ dàng tiếp cận với nhau nhất. Một số tính năng nổi bật trên Shopee có thể kể đến như:

- Được xây dựng trên nền tảng thiết bị di động

- Trò chuyện trực tiếp trên Shopee

- Đảm bảo thanh toán an toàn với nhiều lựa chọn hình thức thanh toán

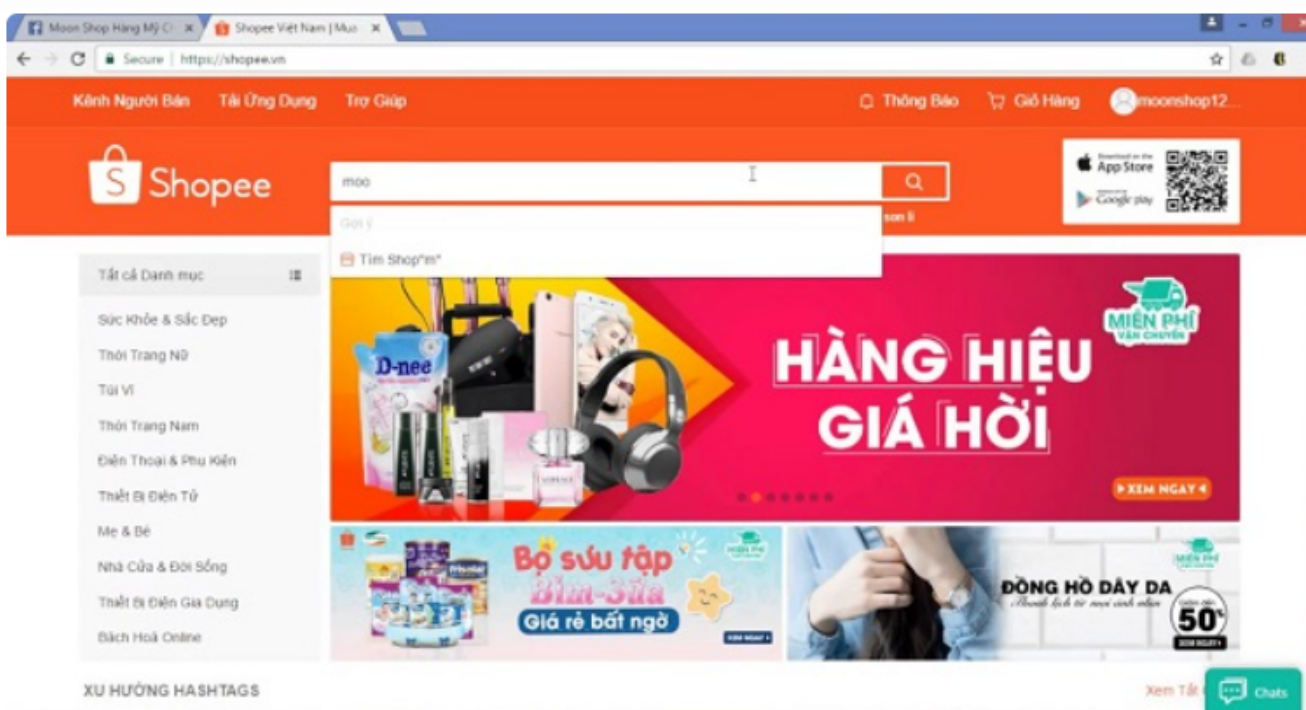
- Sử dụng miễn phí

Vận chuyển an toàn và nhanh chóng, bạn có thể lựa chọn bên vận chuyển mà mình mong muốn.

**Trên Shopee có bán những mặt hàng gì?**

Có rất nhiều loại mặt hàng trên Shopee từ vật dụng gia đình, đời sống sức khỏe, làm đẹp, thời trang, thể thao, sản phẩm ăn uống đến các đồ điện tử, từ đồ bình dân đến những đồ xa xỉ. Dường như tất cả mọi thứ bạn muốn đều có thể mua trên Shopee.

Ở trên Shopee có phần phản hồi về sản phẩm nên nếu có bất kỳ phản hồi nào về hàng giả, hàng nhái thì các chủ shop bán hàng sẽ bị ngưng hợp tác vĩnh viễn. Do đó người mua có thể yên tâm hơn khi mua hàng trên sàn thương mại điện tử này. Bên cạnh đó bạn cũng có thể tham khảo những phản hồi công khai trước khi quyết định mua một sản phẩm nào đó.



## PHẦN 2- CƠ SỞ LÝ THUYẾT

### 1. CƠ SỞ LÝ THUYẾT

#### 1.1 Data Analysis

Phân tích dữ liệu là môn khoa học phân tích dữ liệu thô (raw data) để đưa ra được kết luận về thông tin đó. Các chuyên gia Phân tích dữ liệu tìm ra các xu hướng và số liệu trong các khối thông tin mà có thể bị bỏ sót nếu không sử dụng kỹ thuật hay công cụ phân tích. Thông tin thu được có thể được sử dụng để tối ưu hóa các quy trình làm tăng hiệu quả tổng thể của một doanh nghiệp hoặc một hệ thống



#### 1.2 Django

- Django là một trong số những web framework bậc cao miễn phí, là mã nguồn mở được tạo ra bởi ngôn ngữ Python dựa trên mô hình mô hình MTV (gồm Model-Template-Views). Hiện framework này được phát triển, quản lý bởi Django Software Foundation.





Hình. 1 Django

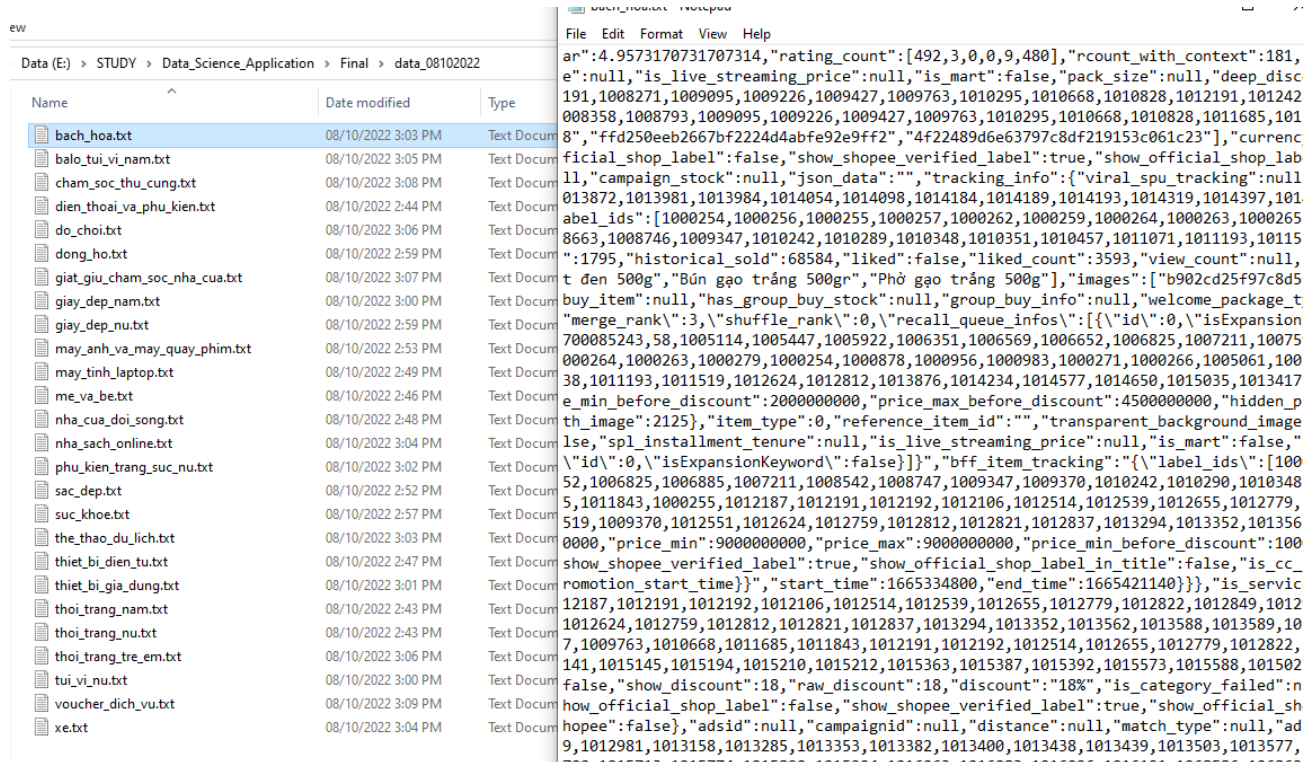
Django ra đời với mục tiêu hỗ trợ thiết kế các website phức tạp dựa trên những CSDL có sẵn. Nó hoạt động dựa theo nguyên lý ‘cắm’ các thành phần và tái sử dụng để tạo nên các website với ít code, ít khớp nối, có khả năng phát triển và không bị trùng lặp.

## PHẦN 3 – TRIỂN KHAI ỨNG DỤNG

### 3.1 Thực Hiện Đề Tài

#### 3.1.1- Thu thập dữ liệu

Dữ liệu được thu thập từ các trang danh mục của Shopee. Các dữ liệu thu về có dạng text như sau:



The screenshot displays a file explorer window with a directory named 'Data (E:) > STUDY > Data\_Science\_Application > Final > data\_08102022'. The directory contains numerous text files, each named with a product category and a date, such as 'bach\_hoa.txt', 'balo\_tui\_vi\_nam.txt', 'cham\_soc\_thu\_cung.txt', etc. The files are listed with their names, modification dates, and types (all are Text Documents). To the right, a preview of one of the text files is shown, displaying a large block of JSON data. The JSON data includes various product attributes and metrics, such as 'rating\_count', 'is\_live\_streaming\_price', 'show\_shopee\_verified\_label', and 'show\_official\_shop\_label'.

#### 3.1.2- Xử lý dữ liệu

Các dữ liệu thu về chủ yếu ở dạng text và không thể sử dụng ngay được. Trước tiên phải đưa được bộ dữ liệu vào thuật toán và tiến hành quá trình xử lý. Dưới đây là ảnh ví dụ một hàm xử lý:

```
final_preprocessing.py X
E: > STUDY > Data_Science_Application > Final > final_preprocessing.py

22 def get_item_info(item_split_text):
23     result_item = {}
24     for i in range(len(item_split_text)):
25         # get name
26         if "name" in item_split_text[i]:
27             if "name" in result_item:
28                 continue
29             else:
30                 if item_split_text[i].split(":")[1] != '{"name":':
31                     result_item["name"] = item_split_text[i].split(":")[1].replace("'", '')
32         # get price
33         elif "price_min" in item_split_text[i]:
34             min_price = int(item_split_text[i].split(":")[1][:-5:])
35         elif "price_max" in item_split_text[i]:
36             max_price = int(item_split_text[i].split(":")[1][:-5:])
37             result_item["price"] = (min_price+max_price)/2
38
39         # historical sold
40         elif "historical_sold" in item_split_text[i]:
41             result_item["historical_sold"] = int(item_split_text[i].split(":")[1])
42
43         # shop_location
44         elif "shop_location" in item_split_text[i]:
45             result_item["shop_location"] = item_split_text[i].split(":")[1].replace("'", '')
46
47         # cmt_count
48         elif "cmt_count" in item_split_text[i]:
49             result_item["cmt_count"] = int(item_split_text[i].split(":")[1])
50
51         # rating_star
52         elif "rating_count" in item_split_text[i]:
53             result_item["one_star_rating"] = int(item_split_text[i+1])
```

Sau khi trải qua quá trình đọc và xử lý dữ liệu sẽ được lưu vào file csv.

```
final_preprocessing.py X  shopee_sell_item_v1.csv X
E: > STUDY > Data_Science_Application > Final > shopee_sell_item_v1.csv

3007 Trục nhôm cho bánh xe v-slot độ chính xác cao,xs,2750.0,6428,TP. Hồ Chí Minh,229,1,0,0,0,228
3008
3009 [CÓ VIDEO] Combo vệ sinh sên MC-40 400ml,xs,95000.0,18237,TP. Hồ Chí Minh,5751,24,16,85,206,5420
3090
3091 Decal carbon dán xe ,xs,16700.0,10967,Hà Nội,1985,8,9,25,73,1870
3092
3093 Vệ sinh buồng đốt,xs,33700.0,5749,TP. Hồ Chí Minh,1622,6,4,17,56,1539
3094
3095 Đèn xe đạp thể thao siêu sáng có còi pin sạc usb led T6 chống nước 7588 loại cao cấp nhất,xs,63000.0,11889,Hà Nội,3983,20,11,29,154,3769
3096
3097 Cảm Biến Áp Suất Lốp Ô Tô TPMS Phiên Bản QUỐC TẾ Màn Hình Màu dùng Năng Lượng Mặt Trời (Loại gắn ngoài van),xs,574000.0,2392,TP. Hồ Chí Minh,763
3098
3099 Little Trees Cây Thông Thơm Khứ Mùi Treo Ô Tô,xs,22000.0,18173,TP. Hồ Chí Minh,1405,6,4,15,30,1350
3100
3101 Pin Khuy Cúc Áo Panasonic CR2032 - CR2025 - CR2016 - CR1632 - CR1620 - CR1616 - CR1220 3V Lithium,xs,12000.0,6637,TP. Hồ Chí Minh,698,2,1,6,12,6
3102
3103 Bọc Vỏ Lốp Sparco Italia Cao Cấp HÀNG CHÍNH HÃNG cho các dòng ô tô xe hơi phổ thông đường kính 37 - 38cm (Đủ mẫu),xs,389000.0,3905,TP. Hồ Chí M
3104
3105 Dầu thắng DOT 4 Liqui Moly,xs,112000.0,1267,TP. Hồ Chí Minh,521,0,0,1,16,504
3106
3107 Ví Bóp đựng giấy tờ ô tô xe hơi có đủ logo các dòng xe chất liệu bền đẹp,xs,16450.0,10501,Hà Nội,2025,12,4,17,60,1932
3108
3109 Bạt phủ xe máy loại dây áo trùm phủ bọc che nắng mưa xe máy các dòng phân khúc từ nhỏ tới lớn vải trắng nhôm cao cấp,xs,75999.0,21617,Hà Nội,544
3110
3111 Khăn Lau Xe Hơi Ô Tô Cao Cấp Gia Dụng Đa Năng Siêu Thấm Microfiber,xs,11500.0,11172,Ninh Bình,1195,10,2,10,28,1145
3112
3113 Khăn Lau Đa Năng Microfiber Rửa Ô Tô Xe Hơi Xe Máy Nhà Bếp 3M Mềm Mịn Thấm Hút Nước Cực Tốt (màu ngẫu nhiên),xs,16500.0,27588,TP. Hồ Chí Minh,28
3114
3115 Khăn Lau Đa Năng Microfiber 2 Mặt Vàng - Xám Siêu Thấm Cực Dày Không Đổ Lông Chuyên Dụng Cho Ô Tô Xe Hơi,xs,25000.0,24818,TP. Hồ Chí Minh,3313,7
3116
3117 Viên sỏi rửa kính ô tô,xs,1000.0,54657,Hà Nội,617,2,3,4,28,580
3118
```

### 3.1.3- Phân tích dữ liệu



Sau khi dữ liệu được làm sạch và sẵn sàng sử dụng. Hệ thống sẽ tiến hành tổng hợp và phân tích các chỉ số quan trọng.

```
def data_for_chart_1(df, real_category_value):
    doanh_thu_value_list = []
    doanh_so_value_list = []
    for category in real_category_value:
        doanh_thu_value = 0
        doanh_so_value = 0
        for i in range(len(df["category"])):
            if df["category"][i] == category:
                doanh_so_value += df["historical_sold"][i]
                doanh_thu_value += doanh_so_value*df["price"][i]
        doanh_thu_value_list.append(doanh_thu_value)
        doanh_so_value_list.append(doanh_so_value)
    return doanh_so_value_list, doanh_thu_value_list
```

```
cmt_count_value_list = []
one_star_rating_value_list = []
two_star_rating_value_list = []
three_star_rating_value_list = []
four_star_rating_value_list = []
five_star_rating_value_list = []

for category in real_category_value:
    cmt_count_value = 0
    one_star_rating_value = 0
    two_star_rating_value = 0
    three_star_rating_value = 0
    four_star_rating_value = 0
    five_star_rating_value = 0
    for i in range(len(df["category"])):
        if df["category"][i] == category:
            cmt_count_value += df["cmt_count"][i]
            one_star_rating_value += df["one_star_rating"][i]
            two_star_rating_value += df["two_star_rating"][i]
            three_star_rating_value += df["three_star_rating"][i]
            four_star_rating_value += df["four_star_rating"][i]
            five_star_rating_value += df["five_star_rating"][i]
    cmt_count_value_list.append(cmt_count_value)
    one_star_rating_value_list.append(one_star_rating_value)
    two_star_rating_value_list.append(two_star_rating_value)
    three_star_rating_value_list.append(three_star_rating_value)
    four_star_rating_value_list.append(four_star_rating_value)
    five_star_rating_value_list.append(five_star_rating_value)
return cmt_count_value_list, one_star_rating_value_list, two_star_rating_value_list, three_star_rating_value_list, four_star_rating_value_list, five_star_rating_value_list
```

```

def data_chart_3(df, location_list):
    location_historical_list = []
    historical_sold_raito = []
    result_list = []
    for location in location_list:
        count_value = 0
        for i in range(len(df["shop_location"])):
            if df["shop_location"][i] == location:
                count_value += df["historical_sold"][i]
            location_historical_list.append(count_value)

        diff_count = 0
        for i in range(len(df["shop_location"])):
            if df["shop_location"][i] != 'TP. Hồ Chí Minh' and df["shop_location"][i] != 'Hà Nội' and df["shop_location"][i] != 'Đà Nẵng':
                diff_count += 1

        location_historical_list.append(diff_count)

    for i in location_historical_list:
        historical_sold_raito.append(i/sum(location_historical_list)*100)

    historical_sold_raito[-1] += 100 - sum(historical_sold_raito)

    return historical_sold_raito

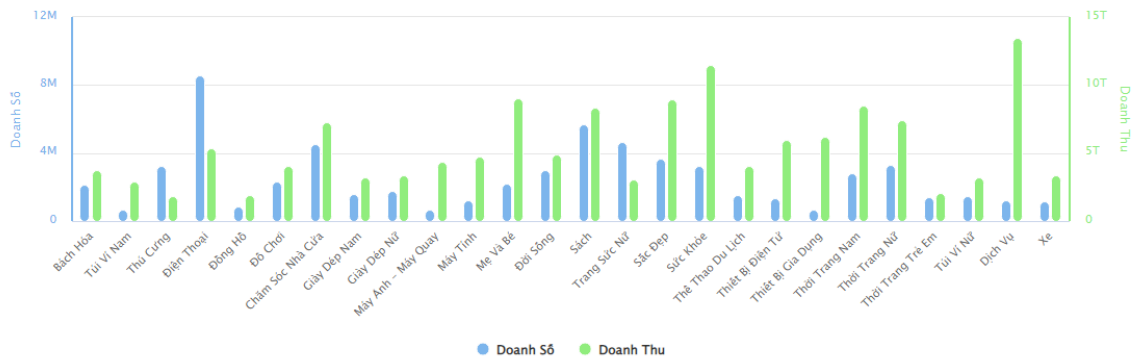
```

### 3.1.4- Trực quan hóa dữ liệu

Sau khi phân tích được các chỉ số cần thiết. Hệ thống sẽ trực quan dữ liệu lên hệ thống Web dưới dạng biểu đồ và đưa ra các nhận định sau khi thuật toán.

# SHOPEE ANALYSIS DASHBOARD

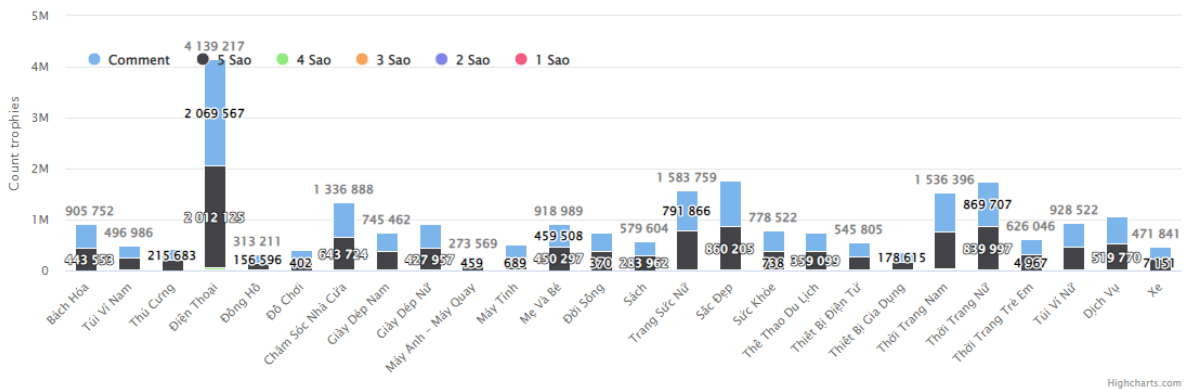
DOANH THU VÀ DOANH SỐ CÁC DANH MỤC



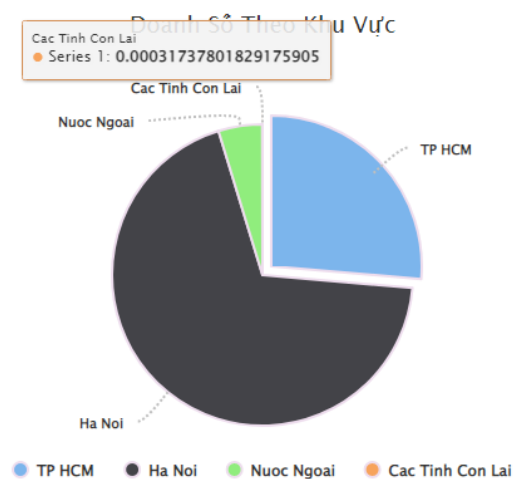
**Danh mục có doanh thu cao nhất là Dịch Vụ**  
**Danh mục có doanh số cao nhất là Điện Thoại**

MỨC ĐỘ ĐÁNH GIÁ SẢN PHẨM

MỨC ĐỘ ĐÁNH GIÁ SẢN PHẨM



**Điện Thoại là danh mục có đánh giá tích cực!**



## **PHẦN 4 – KẾT QUẢ ĐẠT ĐƯỢC**

### **4.1 Những kiến thức đã học được thông qua đề tài**

- Hiểu được cách thu thập và tổng hợp dữ liệu
- Nắm bắt được các cơ chế phân tích dữ liệu
- Hiểu rõ các xu hướng thị trường bán lẻ

### **4.2 Kết luận về đề tài đã thực hiện**

- Trong thời gian thực hiện đề tài. Em đã nắm bắt được khoa học dữ liệu và cách thức ứng dụng các cơ chế phân tích dữ liệu để áp dụng vào nghiên cứu thị trường bán lẻ Shopee.

## TÀI LIỆU THAM KHẢO

### Tiếng Anh

<https://www.simplilearn.com/data-science-roadmap-article>

[https://en.wikipedia.org/wiki/Data\\_analysis](https://en.wikipedia.org/wiki/Data_analysis)

<https://www.simplilearn.com/data-analysis-methods-process-types-article>

<https://www.similarweb.com/website/shopee.vn/>

<https://medium.com/@appseed.us/django-dashboards-open-source-and-free-projects-1d8e64919e6d>

<https://arctype.com/blog/analytics-dashboard-django/>