**ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

bdôca



**PHÂN TÍCH DỮ LIỆU VÀ XÂY DỰNG MÔ HÌNH DỰ ĐOÁN GIÁ LAPTOP CŨ**

|  |  |  |  |
| --- | --- | --- | --- |
| Sinh viên thực hiện: | | | |
| STT | Họ tên | MSSV | Ngành |
| 1 | Lê Thanh Tùng | 21520121 | CNTT |
| 2 | Võ Thị Tố Như | 21520386 | CNTT |
| 3 | Trần Thị Tâm Như | 21521247 | CNTT |
| 4 | Nguyễn Khánh Duy | 21522003 | CNTT |
| 5 | Trần Thành Tín | 21522684 | CNTT |

**TP. HỒ CHÍ MINH – 12/2023**

# GIỚI THIỆU

Hiện nay việc sử dụng laptop cũng như các thiết bị công nghệ khác đã trở thành một trong những nhu cầu thiết yếu của nhiều người trong hầu hết các lĩnh vực. Bên cạnh đó, thị trường đồ cũ cũng nhận được nhiều sự quan tâm của người tiêu dùng với mong muốn tìm kiếm các sản phẩm phù hợp với giá thành tốt nhất. Kết hợp hai yếu tố trên, thị trường laptop cũ trở thành một thị trường tiềm năng cho cả người bán và người tiêu dùng.

Thế nhưng các mặt hàng công nghệ có đôi chút khác biệt so với các sản phẩm thông thường, không có nhiều người tiêu dùng thật sự hiểu rõ giá trị thực sự của sản phẩm này, dẫn đến việc “mua nhầm”, một trong những mối lo ngại của người tiêu dùng trong thị trường đồ công nghệ cũ. Hiểu được tâm tư trên, nhóm chúng tôi quyết định nghiên cứu và xây dựng một mô hình dự đoán chính xác về giá cho các laptop cũ trên thị trường, dựa trên việc xác định các yếu tố quan trọng ảnh hưởng đến giá.

Để thực hiện đề tài này, chúng tôi sử dụng các công cụ như thư viện Selenium trong Python để tiến hành thu thập dữ liệu trên website “Chợ Tốt - Website Mua Bán, Rao Vặt Trực Tuyến Hàng Đầu Của Người Việt”, thư viện Numpy, Pandas, Scikit-learn để xử lý và phân tích dữ liệu, thư viện Panel và Hvplot để thực hiện trực quan hóa dữ liệu. Bộ dữ liệu và đề tài do nhóm tự phân tích thiết kế và không dựa trên đề tài nào khác. Chúng tôi cũng kết hợp với kỹ thuật Pipeline và tiến hành thử nghiệm qua các mô hình để phân tích bộ dữ liệu, tìm ra mô hình phù hợp nhất, đưa ra các dự đoán tốt nhất và phát triển mô hình đó.

Kết quả của đề tài chính là một mô hình dự đoán giá laptop cũ dựa trên bộ dữ liệu đã được thu thập và phân tích. Kết quả này sẽ cung cấp những thông tin quan trọng giúp người mua và người bán có cái nhìn rõ hơn về giá trị thực tế của sản phẩm. Điều này này này đồng thời mang lại lợi ích cho các nhà phân phối, giúp tối ưu hóa giá cả và tăng tính cạnh tranh trong thị trường công nghệ ngày nay.

# MÔ TẢ BỘ DỮ LIỆU

Bộ dữ liệu cung cấp các thông tin cần thiết phục vụ cho việc xác định các yếu tố quan trọng có ảnh hưởng đến giá cả của laptop cũ, từ đó đưa ra những dự đoán chính xác về giá cho mặt hàng này trên thị trường. Bên cạnh việc dùng để phân tích yếu tố ảnh hưởng, bộ dữ liệu cũng được dùng để thực hiện việc xây dựng mô hình dự đoán giá laptop cũ. Nhóm đã sử dụng công cụ như thư viện Selenium trong Python để tiến hành thu thập dữ liệu. Bộ dữ liệu phân tích tự thu thập tại trang web “Chợ Tốt - Website Mua Bán, Rao Vặt Trực Tuyến Hàng Đầu Của Người Việt”. Bộ dữ liệu gồm có 130 dòng, 15 cột, 9 biến phân loại, 6 biến số và 163 giá trị khuyết.

Bộ dữ liệu được thống kê chi tiết như sau:

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Thành phần | Số lượng | Tên cột/biến chi tiết |
| 1 | Cột | 15 | ‘company’, ‘typeName’, ‘cpu’, ‘processor’, ‘os’, ‘screen\_size’, ‘ram’, ‘card’, ‘ssd’, ‘hdd’, ‘hard\_driver’, ‘made\_in’, ‘warranty’, ‘condition’, ‘price’ |
| 2 | Dòng | 130 |  |
| 3 | Biến phân loại | 9 | ‘company’, ‘typeName’, ‘cpu’, ‘processor’, ‘os', 'card', ‘ssd’, ‘hdd’, ‘made\_in’, ‘warranty’, ‘condition’ |
| 4 | Biến số | 6 | ‘screen\_size’, ‘ram’, ‘hard\_driver’, ‘price’ |
| 5 | Giá trị khuyết | 163 |  |

Ý nghĩa của các biến/cột có trong bộ dữ liệu được thể hiện qua bảng sau:

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Tên biến/cột** | **Ý nghĩa các biến/cột** | **Kiểu dữ liệu** |
| 1 | company | Tên công ty của các hãng laptop | object |
| 2 | typeName | Tên của các mẫu laptop | object |
| 3 | cpu | Tên các cpu của laptop | object |
| 4 | processor | Tên của bộ vi xử lý | object |
| 5 | os | Hệ điều hành của laptop | object |
| 6 | screen\_size | Kích thước màn hình của laptop (inch) | float64 |
| 7 | ram | Dung lượng bộ nhớ RAM của laptop (GB) | float64 |
| 8 | card | Loại card đồ họa của laptop | object |
| 9 | ssd | Ổ cứng SSD của laptop (1: là SSD; 0: không phải) | int64 |
| 10 | hdd | Ổ cứng HDD của laptop (1: là HDD; 0: không phải) | int64 |
| 11 | hard\_driver | Dung lượng ổ đĩa cứng của laptop | float64 |
| 12 | made\_in | Nguồn gốc xuất sứ | object |
| 13 | warranty | Thời gian bảo hành của laptop | object |
| 14 | condition | Tình trạng của laptop | object |
| 15 | price | Giá bán của laptop (VNĐ) | int64 |

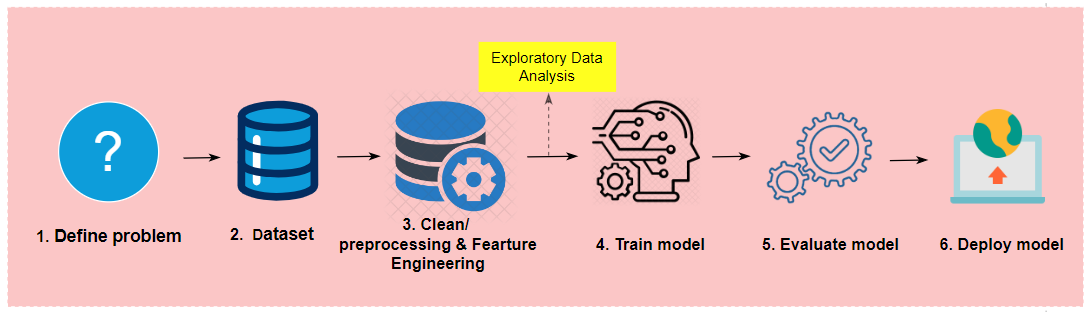
Phương pháp thu thập dữ liệu sử dụng thư viện Selenium:

* *Mục tiêu*:Thu thập những thông tin về giá cả và các thuộc tính liên quan đến laptop cũ trên trang web “**Chợ Tốt**” như: Tên, Công ty, Ổ cứng, Bộ nhớ, …
* *Ý tưởng*: Tự động hóa công việc thu thập dữ liệu bằng cách truy cập vào từng trang bán Laptop trên trang ‘Chợ tốt’ sau đó truy cập vào từng sản phẩm của trang đó rồi lấy các thông tin cần thiết.

Þ Sử dụng thư viện Selenium trong python để tự động hóa các công việc

* *Chuẩn bị:* Cài đặt thư viện Selenium và cài đặt Webdriver tương ứng cho trình duyệt đang sử dụng.
* *Các bước thực hiện:*
* *Bước 1*: Import thư viện. Import các thư viện cần thiết như: Selenium, Numpy, Pandas
* *Bước 2*: Khai báo các list cần thiết để lưu trữ dữ liệu thu thập được. Sử dụng các list để lưu trữ dữ liệu thu thập được, mỗi list sẽ đại diện cho một cột trong dataframe.
* *Bước 3*: Thực hiện code lấy dữ liệu. Sau khi chuẩn bị các công cụ, thư viện, nơi lưu trữ thì bắt đầu vào thực hiện code thu thập dữ liệu. Các bước thực hiện được tóm tắt như sau:
  + *Bước 3.1*: Truy cập vào trang web cần lấy dữ liệu (trang web “**Chợ Tốt**”)
  + *Bước 3.2*: Sử dụng vòng lặp truy cập qua các trang và các sản phẩm trong trang, đồng thời xử lý dữ liệu nhận được. Khi truy cập vào trang web sẽ tự động truy cập vào từng sản phẩm của trang đó và lấy thông tin mô tả của từng sản phẩm. Đồng thời xử lý các dữ liệu và thêm vào các list đã tạo sẵn. Sau khi lấy được các thông tin của sản phẩm đầu tiên thì sẽ trở về và tiếp tục truy cập vào sản phẩm kế tiếp cho đến khi hết sản phẩm. Tiếp theo, trang web sẽ được tự động chuyển sang trang tiếp theo cho đến khi nút click chuyển trang không còn nữa thì vòng lặp sẽ dừng lại.
* *Bước 4*: Tạo dataframe và xử lý sơ bộ dữ liệu. Thực hiện tạo dataframe từ bộ dữ liệu thu thập được, đồng thời cũng xử lý sơ bộ dữ liệu giúp quan sát dễ dàng và trực quan hơn trong việc xử lý sau này.
* *Bước 5*: Xuất bộ dữ liệu. Sau khi đã thực hiện lấy dữ liệu, tạo bộ dataset và xử lý sơ bộ dữ liệu thì xuất bộ dữ liệu đó ra để sử dụng lâu dài.

# PHƯƠNG PHÁP PHÂN TÍCH



*Hình 1: Quy trình phân tích dữ liệu*

## Xác định vấn đề

Để xây dựng một mô hình dự đoán giá thành của các mặt hàng laptop cũ, nhóm chúng tôi cần xác định được những yếu tố quan trọng tác động đến giá của laptop, ảnh hưởng của từng thuộc tính của sản phẩm đến giá, và hơn nữa là mong muốn có thể dự đoán được giá thành tốt nhất có thể mua được laptop với các tiêu chí xác định.

Mục tiêu: Xây dựng một mô hình dự đoán giá laptop cũ dựa trên các thông số kỹ thuật và giá ban đầu của những chiếc laptop để đưa ra dự đoán giá cho những sản phẩm laptop khác với các thông số của nó.

## Bộ dữ liệu

Để có thể xây dựng một mô hình dự đoán giá laptop cũ, trước tiên ta cần có một bộ dữ liệu để có thể đưa vào mô hình để phân tích và huấn luyện, từ đó hoàn thiện mô hình và sẵn sàng đưa ra dự đoán giá cho các laptop cũ.

### Thực hiện thu thập dữ liệu

Bộ dữ liệu về laptop cũ được thu thập từ trang web bán laptop cũ là: Chợ Tốt.

### Hiểu về bộ dữ liệu

Trước khi bắt đầu thực hiện thì cần phải hiểu về bộ dữ liệu của mình. Bao gồm các thông tin cơ bản như: số lượng cột, số lượng dòng, tên các cột, các giá trị trong từng cột, số lượng giá trị khuyết trong từng cột và số lượng dòng trùng nhau trong bộ dữ liệu.

## Tiền xử lý dữ liệu và trích xuất đặc trưng

### Tiền xử lý dữ liệu

Để thực hiện việc tiền xử lý dữ liệu, nhóm thực hiện các công việc như sau:

* Đầu tiên, thực hiện kiểm tra và loại bỏ các dòng trùng nhau trong bộ dữ liệu. Trong bộ dữ liệu sau khi kiểm tra cho thấy được có 6 dòng trùng nhau, thực hiện bỏ 6 dòng trùng nhau này ta thu được bộ dữ liệu có 124 dòng.
* Tiếp theo, thực hiện định dạng dữ liệu, nhóm thực hiện cụ thể như sau:
* Tại cột “cpu” và “card” thực hiện thay các giá trị “Khác” thành np.nan.
* Thay các giá trị “AMD” không thuộc trường “card” thành np.nan.
* Tại cột “hard\_driver” thực hiện thay các giá trị “250”, “500” thành “256”, “512” để phù hợp với thực tế.
* Tại cột “os” thực hiện đổi tên các giá trị “iOS” thành “MacOS”
* Tại cột “made\_in” thực hiện chuẩn hóa lần lượt các giá trị “Việt Nam”, “Mỹ”, “Nhật Bản”, “Đài Loan” thành các giá trị “VietNam”, “US”, “Japan” và “Taiwan”.
* Tại cột “warranty” thực hiện chuẩn hóa bằng cách loại bỏ từ “tháng” trong các giá trị, thay các giá trị “Hết bảo hành” thành 0 và các giá trị như “Bảo hành hãng”, “Còn bảo hành” thành np.nan.
* Sau khi thực hiện định dạng dữ liệu, nhóm tiến hành làm sạch dữ liệu như sau:
* Sử dụng công cụ SimpleImputer trong thư viện scikit-learn của Python để thực hiện điền các giá trị khuyết.
* Tại cột “cpu”, “processor”, “card” thực hiện điền các giá trị np.nan bằng giá trị xuất hiện nhiều nhất (most\_frequent).
* Tại cột “screen\_size”, “ram”, “hard\_driver” thực hiện điền các giá trị np.nan bằng giá trị xuất hiện nhiều nhất (most\_frequent).
* Tại cột “made\_in” thực hiện điền các giá trị np.nan thành giá trị “Unknown”.
* Loại bỏ cột “condition” vì các laptop được thu thập đều là laptop cũ nên giá trị tại cột này ảnh hưởng đến biến mục tiêu “price” như nhau.
* Tại cột “waranty” thực hiện điền các giá trị np.nan bằng giá trị xuất hiện nhiều nhất (most\_frequent).
* Gộp hai cột “ssd” và “hdd” thành một cột “ssd/hdd”.

### Trích xuất đặc trưng

Trong thư viện scikit-learn của Python, sử dụng công cụ OneHotEncoder và công cụ StandardScaler để chuẩn hóa dữ liệu theo phân phối chuẩn. Bên cạnh đó nhóm còn sử dụng lớp ColumnTransformer và lớp Pipeline để tổ chức và thực hiện tuần tự các bước tiền xử lý dữ liệu hoặc huấn luyện mô hình.

## Huấn luyện mô hình

Tiến hành huấn luyện nhiều mô hình với các thuật toán khác nhau và chọn ra mô hình tốt nhất với tham số tốt nhất để dự đoán:

* Mô hình Random Forest Regressor
* Mô hình Decision Tree Regressor
* Mô hình Logistic Regression
* Mô hình Linear Regression
* Mô hình Kneighbors Regression
* Mô hình Ridge

## Đánh giá mô hình

Sử dụng các thang đo đánh giá:

* R2 score: đo lường mức độ phù hợp của mô hình đối với dữ liệu. Giá trị R2 cho biết tỷ lệ phương sai của biến phụ thuộc mà mô hình có thể giải thích. Nếu R2 gần 0, mô hình không giải thích được dữ liệu; còn nếu R2 gần 1, mô hình giải thích tốt phần lớn biến độc lập.
* MAE (Mean absolute error): MAE tính trung bình của giá trị tuyệt đối của sự chênh lệch giữa giá trị dự đoán và giá trị thực tế. MAE thường được sử dụng để đo lường sự chính xác trung bình của mô hình trong việc dự đoán. Một giá trị MAE nhỏ hơn cho thấy mô hình dự đoán gần với giá trị thực tế.

Bên cạnh đó, để có thể xác thực rằng mô hình máy học đã áp dụng có thực sự là đáng tin cậy hay không thì nhóm đã thực hiện kiểm chứng chéo K-Fold Cross Validation để tính giá trị R2 score và MAE trung bình.

## Triển khai mô hình

Dựa vào kết quả của thang đo đánh giá mô hình thì nhóm chọn mô hình Ridge, mô hình này có R2 score cao nhất và MAE tương đối thấp so với những giá trị còn lại. Đây là mô hình tốt nhất trong 6 mô hình được chọn để huấn luyện.

Sau khi chọn mô hình tốt nhất, nhóm thực hiện xây dựng mô hình dự đoán giá laptop cũ bằng cách sử dụng các widget Select trong thư viện Panel để giúp chọn các thuộc tính của laptop và sau đó dự đoán giá dựa trên mô hình đã được huấn luyện trước, cụ thể:

* Đầu tiên, các widget Select được khởi tạo cho từng thuộc tính của laptop, các tùy chọn cho từng thuộc tính được lấy từ dataframe có sẵn dữ liệu.
* Sau đó, khởi tạo một danh sách với số lượng phần tử tương ứng với số lượng thuộc tính.
* Tiếp theo, định nghĩa một hàm để thực hiện việc cập nhật giá trị được chọn từ widget Select vào danh sách đã khởi tạo.
* Kết nối callback với sự kiện ‘value’ của từng widget Select để cập nhật giá trị vào danh sách đã khởi tạo.
* Hiển thị danh sách đã khởi tạo và tạo một dataframe dựa trên danh sách này, dataframe vừa tạo được sử dụng để chuẩn bị dữ liệu đầu vào cho mô hình dự đoán.
* Sử dụng trình tiền xử lý đã được khởi tạo trước để biến đổi dữ liệu đầu vào.
* Cuối cùng, sử dụng mô hình đã chọn để dự đoán giá laptop dựa trên dữ liệu đầu vào và hiển thị kết quả dự đoán

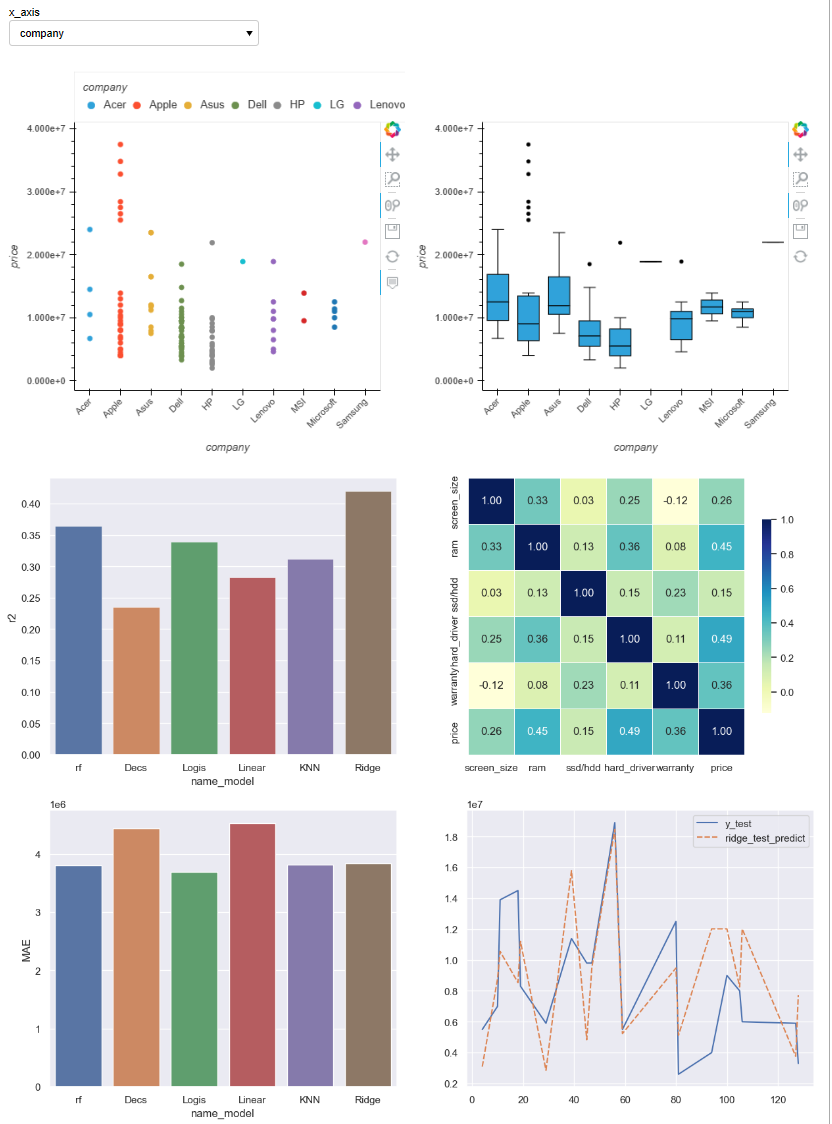
# PHÂN TÍCH THĂM DÒ

Sau khi thực hiện phân tích thăm dò nhóm đã đưa ra một số kết luận chính về bộ dữ liệu của mình như sau:

* Hầu hết giá tiền của những chiếc laptop sẽ có giá trong khoảng từ 3-13 triệu đồng. Tuy nhiên cũng có những chiếc laptop có giá cao hơn nhưng chỉ chiếm thiểu số trong bộ dữ liệu
* Sự mất cân bằng giữa các biến trong các cột là tương đối đáng kể. Các biến xuất hiện phổ biến trong các cột được thể hiện như sau:

|  |  |
| --- | --- |
| **Tên cột** | **Biến phổ biến** |
| company | Dell |
| cpu | Intel |
| processor | Core i5 |
| ssd/hdd | Ssd |
| os | Windows |
| card | Onboard |
| made\_in | Unknown |
| screen\_size | 14.0 |
| ram | 8.0 |
| hard\_driver | 256.0 |
| warranty | 3 |

* Các cấu hình về máy được nêu ở trên là chiếm nhiều trong bộ dữ liệu. Tuy nhiên điều đó không phải là những chiếc laptop có những cấu hình như thế sẽ có giá cao mà những chiếc laptop đó sẽ nằm ở mức giá trung bình từ 6-12 triệu đồng.



*Hình 2: Các biểu đồ trực quan hóa dữ liệu*

Giải thích: Do những chiếc laptop với những cấu hình được nêu ở trên hiện đang rất phổ biến và bán rất nhiều trên trang "Chợ Tốt''. Vì vậy đã dẫn đếnviệc mất cân bằng dữ liệu giữa các biến với nhau. Tuy nhiên, điều này là không thể không tránh khỏi khi thu thập dữ liệu.

Nhóm chúng tôi cũng đã thực hiện xử lí và loại bỏ các giá trị ngoại lai. Tuy nhiên, khi loại bỏ thì những chiếc laptop thuộc hãng Apple sẽ bị loại bỏ, điều này tương đối ảnh hưởng tới việc sau khi train model thì sẽ không thể dự đoán tương đối được giá của những chiếc laptop khác cùng hãng Apple khi người dùng nhập vào bởi vì không có những thông tin tương tự về cấu hình của những chiếc laptop Apple trong lúc train model.

# KẾT QUẢ PHÂN TÍCH

Qua nghiên cứu và phân tích bộ dữ liệu, nhóm chúng tôi đã xây dựng một mô hình dự đoán giá laptop cũ dựa trên tập các đặc trưng quan trọng ảnh hưởng đến giá.

Mô hình của nhóm đã trải qua quá trình tiền xử lý dữ liệu kỹ lưỡng, chuẩn hóa dữ liệu, và tạo ra các đặc trưng mới nhằm cải thiện hiệu suất dự đoán. Các thuật toán học máy như Linear Regression, Random Forest, đã được thử nghiệm và đánh giá thông qua K-Fold Cross-Validation. Trong đó, mô hình Ridge đã cho kết quả tốt nhất trên tập dữ liệu kiểm tra với mức R-squared đạt được là 42% và giá trị MAE là 0.38e^7.

Điều đó cho thấy mô hình dự đoán đạt mức tương đối cho việc dự đoán giá laptop cũ với những dữ liệu thu thập được trên trang ‘Chợ Tốt’.

# KẾT LUẬN

Mô hình nhóm xây dựng góp phần cung cấp khả năng dự đoán giá các laptop cũ một cách chính xác trên thị trường, đặc biệt là trang Chơ tốt. Bên cạnh đó, nghiên cứu còn giúp xác định các đặc trưng quan trọng nhất đóng vai trò quan trọng trong việc ảnh hưởng đến giá của laptop cũ. Các đặc trưng đó là: ….

Mặc dù mô hình đạt được kết tương đối tốt trên tập dữ liệu kiểm tra, việc áp dụng mô hình này vào thực tế có thể cần xem xét thêm với dữ liệu mới để đảm bảo tính tổng quát hóa và khả năng áp dụng trong thực tế.

Để có thể áp dụng mô hình này một cách hiệu quả vào thực tế, nhóm nhận thấy một số hạn chế cần khắc phục như sau:

* Số lượng mẫu dữ liệu còn ít và chưa đảm bảo tính đại diện cho thị trường laptop cũ hiện nay.
* Bộ dữ liệu còn thiếu một số thông tin về các đặc trưng quan trọng.
* Bộ dữ liệu còn tĩnh, không thể cập nhật thông tin trong thời gian thực, do đó khó có thể dự đoán được xu hướng hiện tại và tương lai.

Sau khi thực hiện xong các bước xử lí và bước huấn luyện model, nhóm đưa ra định hướng phát triển cho đề tài của nhóm như sau:

* Sử dụng bộ dữ liệu động từ trang web để cung cấp thông tin liên tục và thời gian thực về sự thay đổi, cập nhật thông tin nhanh chóng.
* Tiến hành thu thập trên các thị trường đồ cũ khác để làm đầy đủ và đa dạng hơn bộ dữ liệu.
* Triển khai mô hình hoàn thiện trên trang web, qua đó giúp người tiêu dùng dự đoán, nhận định được rằng chiếc laptop mình dự định mua có thực sự khả thi không. Đồng thời từ các thông tin về cấu hình của những chiếc laptop thì sau đó trang web, app có thể gợi ý cho người tiêu dùng những sản phẩm laptop khác có cấu hình tương tự mà giá chênh lệch không quá cao trong khoảng mà mô hình đã dự đoán.

**TÀI LIỆU THAM KHẢO**

Chợ Tốt – Chợ Mua Bán, Rao Vặt Trực Tuyến Hàng Đầu Của Người Việt. Link: <https://www.chotot.com/> (25/10/2023)

scikit-learn Machine Learning in Python. Link: <https://scikit-learn.org/stable/> (30/10/2023)

seaborn: statistical data visualization. Link: <https://seaborn.pydata.org/> (28/11/2023)

hvPlot. Link: <https://hvplot.holoviz.org/> (28/11/2023)

HoloViz Panel – The powerful data exploration & web app framework for Python. Link: <https://panel.holoviz.org/> (28/11/2023)

**PHỤ LỤC PHÂN CÔNG NHIỆM VỤ**

|  |  |  |
| --- | --- | --- |
| **STT** | **Thành viên** | **Nhiệm vụ** |
| 1 | Lê Thanh Tùng | * Kiểm tra bộ dữ liệu. * Viết nội dung phần 2: Mô tả dataset (Mô tả dataset + giải thích các feature). * Trực quan hóa dữ liệu. * Hỗ trợ xây dựng dashboard. * Viết nội dung phần 3.3.1: Tiền xử lý dữ liệu. |
| 2 | Võ Thị Tố Như | * Kiểm tra bộ dữ liệu. * Thực hiện làm sạch dữ liệu. * Viết nội dung phần 3.3.1: Tiền xử lý dữ liệu. * Viết nội dung phần 3.6: Triển khai mô hình. * Làm powerpoint. |
| 3 | Trần Thị Tâm Như | * Chọn biến xây dựng bộ dữ liệu. * Định dạng dữ liệu * Viết nội dung phần 4: Phân tích thăm dò * Viết nội dung phần 5. Kết quả phân tích * Viết nội dung phần 6. Kết luận |
| 4 | Nguyễn Khánh Duy | * Thu thập dữ liệu * Trực quan hóa dữ liệu * Xây dựng model * Vẽ dashboard * Chỉnh sửa lại báo cáo * Cập nhật source code |
| 5 | Trần Thành Tín | * Kiểm tra bộ dữ liệu. * Viết nội dung phần 1.1: Giới thiệu. * Chọn biến để xây dựng bộ dữ liệu. * Hỗ trợ trực quan dữ liệu. |