

TRƯỜNG ĐẠI HỌC KỸ THUẬT CÔNG NGHIỆP
KHOA ĐIỆN TỬ

Bộ môn : Công nghệ thông tin

ĐỒ ÁN
MÔN HỌC
THỰC TẬP CHUYÊN NGÀNH

Sinh viên : Triệu Văn Đông

Lớp : 55KMT

Giáo viên hướng dẫn : Phùng Thị Thu Hiền

Thái Nguyên – 2023

TRƯỜNG ĐHKTCN
KHOA ĐIỆN TỬ

CỘNG HOÀ XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập - Tự do - Hạnh phúc

ĐỒ ÁN

MÔN HỌC: THỰC TẬP CHUYÊN NGÀNH

BỘ MÔN : CÔNG NGHỆ THÔNG TIN

Sinh viên: Triệu Văn Đông Lớp: 55KMT

Ngành: Kỹ thuật máy tính

Ngày giao đề: .../.../2023

Ngày hoàn thành: .../.../2023

1. Tên đề tài : Xây dựng website dự đoán giá xe ô tô cũ

2. Yêu cầu

.....

.....

.....

.....

TRƯỞNG BỘ MÔN
(Ký và ghi rõ họ tên)

GIÁO VIÊN HƯỚNG DẪN
(Ký và ghi rõ họ tên)

NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN

.....

.....

.....

.....

.....

.....

Thái Nguyên, ngày....tháng.....năm 20....

GIÁO VIÊN HƯỚNG DẪN

(Ký ghi rõ họ tên)

NHẬN XÉT CỦA GIÁO VIÊN CHẤM

.....

.....

.....

.....

.....

.....

Thái Nguyên, ngày....tháng.....năm 20....

GIÁO VIÊN CHẤM

(Ký ghi rõ họ tên)

MỤC LỤC

MỤC LỤC	4
LỜI CẢM ƠN	5
LỜI MỞ ĐẦU	6
DANH MỤC HÌNH ẢNH	7
CHƯƠNG 1: GIỚI THIỆU	8
1.1.Đoán giá	8
1.2.Ứng dụng.....	9
1.2.1.Đoán giá tài sản và bất động sản.....	9
1.2.2. Đoán giá thị trường tài chính	10
1.2.3.Đoán giá thương mại điện tử.....	10
1.2.4.Đoán giá tiếp thị và quảng cáo.....	11
1.3.Các công nghệ	12
1.4.Ứng dụng.....	13
CHƯƠNG 2 : CƠ SỞ LÝ THUYẾT	14
2.1.Xử lý dữ liệu	14
2.1.1.Crawl data	14
2.1.2.Xử lý dữ liệu	14
2.2.Model	16
CHƯƠNG 3 : XÂY DỰNG CHƯƠNG TRÌNH	17
3.1.Xây dựng giao diện	17
3.2.Code	17
KẾT LUẬN.....	20
TÀI LIỆU THAM KHẢO.....	22

LỜI CẢM ƠN

Trong suốt quá trình học tập, và hoàn thành bài tập này, em đã nhận được rất nhiều sự hướng dẫn tận tình quý báu của thầy cô, anh chị cùng các bạn. Với lòng biết ơn sâu sắc em xin được bày tỏ lời cảm ơn đến:

Ban giám hiệu nhà trường Đại học Kỹ thuật Công nghiệp Thái Nguyên khoa đã tạo mọi điều kiện để em hoàn thành được tốt đồ án của mình.

Em xin cảm ơn Thầy Nguyễn Văn Huy giảng viên khoa Điện tử , người thầy đã hướng dẫn và hết lòng giúp đỡ, bảo ban, động viên và tạo mọi điều kiện để em hoàn thành tốt được bài tập này.

Xin chân thành cảm ơn các thầy cô trong hội đồng chấm thi đã có những góp ý để em có thể hoàn thiện được luận văn này một cách tốt nhất.

Cảm ơn các anh chị khóa trên cùng các bạn đã giúp đỡ em rất nhiều trong quá trình tìm tài liệu để có thể hoàn thành bài luận....

Dưới đây là kết tinh của những gì em tìm hiểu và nghiên cứu trong suốt thời gian qua. Do tính thực tế, kiến thức và kinh nghiệm còn nhiều hạn chế nên đồ án chắc chắn em còn mắc phải nhiều sai sót rất mong các thầy cô và các bạn đóng góp ý kiến để đồ án của em được hoàn thiện hơn.

Em xin chân thành cảm ơn!

Thái Nguyên, ngày tháng năm 2023

Sinh viên

LỜI MỞ ĐẦU

Trong bối cảnh phát triển của nền kinh tế, việc sở hữu một chiếc xe hơi dễ dàng trong giao thông đã trở nên phổ biến hơn bao giờ hết. Tuy nhiên, khi muốn bán lại chiếc xe đã qua sử dụng, chúng ta thường gặp khó khăn trong việc định giá xe một cách chính xác và hợp lý. Thậm chí, chúng ta có thể bị ép giá, khi không có đủ kiến thức về xe. Vì vậy, việc xây dựng một hệ thống dự đoán giá xe hơi trở thành một nhu cầu thiết yếu để giúp chúng ta đánh giá giá trị của chiếc xe một cách chính xác và công bằng.

Để xây dựng một hệ thống dự đoán giá xe cũ đáp ứng được yêu cầu của người dùng là một thách thức đáng kể. Để dự đoán giá trị của một chiếc xe cũ, chúng ta cần phải xem xét nhiều yếu tố khác nhau, như tuổi của xe, số dặm đã đi, tình trạng và chất lượng xe, vị trí địa lý, giá bán trên thị trường...và nhiều yếu tố khác. Do đó, việc tạo ra một hệ thống dự đoán giá xe cũ chính xác đòi hỏi sự phân tích kỹ lưỡng.

Một điểm đáng lưu ý trong việc xây dựng hệ thống dự đoán giá xe cũ là việc sử dụng các công nghệ mới nhất và các công cụ phân tích dữ liệu. Các công nghệ này cho phép chúng ta thu thập, lưu trữ và phân tích các dữ liệu về xe cũ một cách nhanh chóng và chính xác, từ đó giúp cho việc tạo ra một hệ thống dự đoán giá xe cũ hiệu quả và tin cậy hơn. Đồng thời, việc sử dụng các công nghệ này cũng cho phép chúng ta cập nhật và điều chỉnh mô hình dự đoán giá xe cũ một cách nhanh chóng và hiệu quả, đáp ứng được nhu cầu thay đổi của thị trường.

Nhận thấy điều đó tôi quyết định chọn đề tài xây dựng hệ thống dự đoán giá xe đã qua sử dụng làm đề tài nghiên cứu của mình. Người dùng cần có thể truy cập hệ thống dễ dàng và nhanh chóng, và nhận được kết quả dự đoán giá xe cũ một cách chính xác và tin cậy.

DANH MỤC HÌNH ẢNH

Hình 1: Đoán giá	8
Hình 2: Đoán giá nhà	9
Hình 3: Octoparse 8	14
Hình 4: Loại bỏ các cột không cần thiết	15
Hình 5: Xử lý dữ liệu cột fuel_type, company, name.....	15
Hình 6: Xử lý cột year, price, kms_driven.....	16
Hình 7: Mô hình Random Forest	16
Hình 8: Giao diện chương trình	19

CHƯƠNG 1: GIỚI THIỆU

1.1.Đoán giá

Đoán giá (price prediction) là quá trình ước tính giá trị tương lai của một tài sản, sản phẩm hoặc dịch vụ dựa trên các thông tin hiện có và các phương pháp phân tích dữ liệu. Nó thường được áp dụng trong lĩnh vực tài chính, thị trường chứng khoán, thương mại điện tử và nhiều lĩnh vực khác để đưa ra các dự đoán về giá cả.

Để đoán giá, các chuyên gia thường sử dụng các kỹ thuật và công cụ trong lĩnh vực khoa học dữ liệu và học máy như mô hình hồi quy, mạng neural, phân tích chuỗi thời gian và các thuật toán khác. Các yếu tố quan trọng được sử dụng để đưa ra dự đoán bao gồm các dữ liệu lịch sử, chỉ số kỹ thuật, tin tức và yếu tố tác động đến thị trường hoặc ngành công nghiệp liên quan.



Hình 1: Đoán giá

Tuy nhiên, đoán giá đi không phải là một nhiệm vụ dễ dàng vì thị trường và các yếu tố liên quan có thể bị ảnh hưởng bởi nhiều yếu tố khác nhau và có tính không chắc chắn. Kết quả dự đoán cũng không thể đảm bảo chính xác tuyệt đối và thường mang tính tương đối.

1.2.Ứng dụng

1.2.1.Đoán giá tài sản và bất động sản

Trong lĩnh vực định giá bất động sản, việc đoán giá có thể cung cấp những thông tin quan trọng về giá trị thị trường của các căn nhà, lô đất và tài sản khác. Những người mua, bán, và các nhà đầu tư có thể sử dụng thông tin này để ra quyết định thông minh trong việc mua, bán hoặc đầu tư vào bất động sản. Điều này giúp tạo ra sự minh bạch và đảm bảo tính công bằng trong giao dịch bất động sản.

Ngoài ra, việc đoán giá tài sản và bất động sản cũng hỗ trợ trong việc dự báo xu hướng thị trường. Bằng cách sử dụng dữ liệu lịch sử và các yếu tố liên quan như kinh tế, chính trị và xã hội, các mô hình đoán giá có thể dự báo xu hướng giá và thị trường trong tương lai. Điều này giúp các nhà đầu tư và chủ sở hữu tài sản có cái nhìn rõ ràng và đưa ra các quyết định chiến lược, như mua, bán hoặc giữ tài sản, dựa trên thông tin dự báo chính xác.



Hình 2: Đoán giá nhà

Ứng dụng của đoán giá tài sản và bất động sản không chỉ giới hạn trong việc định giá và dự báo, mà còn trong việc quản lý danh mục đầu tư và đánh giá rủi ro. Các nhà đầu tư và tổ chức tài chính có thể sử dụng mô hình đoán giá để đánh giá giá trị các tài sản tài chính, từ cổ phiếu đến trái phiếu và các sản phẩm tài chính khác. Điều này giúp tối ưu hóa danh mục đầu tư và đưa ra

quyết định mua, bán hoặc giữ tài sản dựa trên dự báo và đánh giá rủi ro chính xác.

1.2.2. Đoán giá thị trường tài chính

Đoán giá trong lĩnh vực thị trường tài chính đóng một vai trò quan trọng trong việc đưa ra quyết định đầu tư thông minh. Các nhà đầu tư và nhà giao dịch sử dụng các phương pháp dự báo giá để đoán trước và ước lượng giá trị của các tài sản như cổ phiếu, tiền tệ, hàng hóa và các công cụ tài chính khác.

Việc đoán giá trong thị trường tài chính mang lại nhiều ứng dụng hữu ích. Một trong những ứng dụng quan trọng nhất của đoán giá là hỗ trợ việc ra quyết định mua hoặc bán tài sản tài chính. Bằng cách dự đoán giá cổ phiếu, nhà đầu tư có thể quyết định khi nào mua vào hoặc bán ra để tối ưu hóa lợi nhuận. Đoán giá cũng giúp định hình chiến lược đầu tư và quản lý rủi ro, cho phép nhà đầu tư đưa ra các quyết định dựa trên thông tin dự đoán và phân tích kỹ thuật.

Ngoài ra, đoán giá cũng hỗ trợ trong việc dự báo xu hướng thị trường tài chính. Bằng cách sử dụng các mô hình dự đoán, các nhà giao dịch có thể nhận biết và đánh giá xu hướng tăng giá hoặc giảm giá trên thị trường. Điều này giúp họ đưa ra quyết định mua hoặc bán dựa trên những thông tin về sự biến động của giá cả.

1.2.3. Đoán giá thương mại điện tử

Một trong những ứng dụng chính của đoán giá trong thương mại điện tử là ước lượng giá cạnh tranh. Các doanh nghiệp có thể sử dụng các phương pháp dự đoán giá để xác định giá cả của các sản phẩm hoặc dịch vụ tương tự từ các đối thủ cạnh tranh. Điều này giúp họ điều chỉnh giá cả của mình để cạnh tranh một cách hiệu quả và thu hút khách hàng.

Đoán giá cũng giúp các doanh nghiệp thương mại điện tử tối ưu hóa chiến lược giá. Bằng cách sử dụng các mô hình dự đoán, họ có thể xác định mức giá tối ưu cho các sản phẩm và dịch vụ của mình, nhằm tối đa hóa lợi nhuận hoặc doanh số bán hàng. Đoán giá cung cấp thông tin quan trọng để

đưa ra các quyết định về việc giảm giá, tăng giá hoặc áp dụng các chương trình khuyến mãi để thu hút khách hàng.

Thêm vào đó, đoán giá cũng hỗ trợ trong việc dự đoán nhu cầu của khách hàng. Các mô hình dự đoán giá có thể phân tích dữ liệu lịch sử và các yếu tố khác để ước lượng nhu cầu tiềm năng cho các sản phẩm và dịch vụ. Điều này cho phép các doanh nghiệp thương mại điện tử điều chỉnh cung cấp, quảng cáo và phân phối hàng hóa một cách hiệu quả, đồng thời tăng khả năng đáp ứng nhu cầu khách hàng.

1.2.4. Đoán giá tiếp thị và quảng cáo

Ứng dụng của đoán giá trong lĩnh vực tiếp thị và quảng cáo là vô cùng quan trọng để đảm bảo hiệu quả và thành công của chiến dịch tiếp thị. Đoán giá cung cấp các thông tin quan trọng về xu hướng thị trường và giá trị của sản phẩm hoặc dịch vụ, từ đó giúp các nhà tiếp thị và doanh nghiệp xác định mức giá phù hợp và xây dựng chiến lược tiếp thị mạnh mẽ.

Đầu tiên, đoán giá giúp xác định mức giá phù hợp cho sản phẩm hoặc dịch vụ. Bằng cách sử dụng các mô hình dự đoán và phân tích dữ liệu, các doanh nghiệp có thể định giá sản phẩm dựa trên các yếu tố như giá thành, giá trị đối với khách hàng và mức độ cạnh tranh. Điều này giúp tạo ra mức giá cạnh tranh và hấp dẫn cho khách hàng, từ đó thu hút sự quan tâm và mua hàng.

Thứ hai, đoán giá định hướng chiến dịch quảng cáo. Dự đoán giá cung cấp thông tin quan trọng về xu hướng và giá trị của từng mục tiêu tiếp thị. Các doanh nghiệp có thể sử dụng dự đoán giá để xác định mức đầu tư và phân bổ ngân sách quảng cáo một cách hiệu quả. Điều này giúp họ định hình chiến dịch quảng cáo và tiếp thị đúng mục tiêu, tăng khả năng tiếp cận khách hàng và nâng cao tỷ lệ chuyển đổi.

Thứ ba, đoán giá hỗ trợ trong việc đối tác và cộng tác tiếp thị. Các doanh nghiệp có thể sử dụng dự đoán giá để đánh giá giá trị của các đối tác tiếp thị và hợp đồng quảng cáo. Bằng cách dự đoán giá cả và giá trị trên thị trường, các doanh nghiệp có thể lựa chọn đối tác tiếp thị phù hợp và đảm bảo mối quan hệ hợp tác mang lại lợi ích kinh tế.

Cuối cùng, đoán giá cung cấp cơ sở dữ liệu để theo dõi và đánh giá hiệu quả tiếp thị và quảng cáo. Bằng cách so sánh dự đoán giá với thực tế, các doanh nghiệp có thể đánh giá hiệu quả chiến dịch và điều chỉnh chiến lược tiếp thị một cách linh hoạt. Điều này giúp tối ưu hóa kế hoạch tiếp thị, nắm bắt nhu cầu và sự quan tâm của khách hàng, và tăng cường thành công của chiến dịch tiếp thị và quảng cáo.

1.3.Các công nghệ

Học máy (Machine Learning): Học máy là một phương pháp sử dụng các thuật toán và mô hình để học từ dữ liệu và dự đoán kết quả trong tương lai. Các thuật toán phổ biến trong lĩnh vực đoán giá bao gồm hồi quy tuyến tính, hồi quy logistic, máy vector hỗ trợ (SVM), cây quyết định và các thuật toán học sâu như mạng neural.

Học sâu (Deep Learning): Học sâu là một lĩnh vực của trí tuệ nhân tạo (AI) tập trung vào việc xây dựng và huấn luyện các mạng neural sâu để hiểu và giải quyết các bài toán phức tạp. Trong lĩnh vực đoán giá, mạng neural sâu, chẳng hạn như mạng neural tích chập (CNN) hoặc mạng neural hồi quy (RNN), có thể được sử dụng để học từ dữ liệu và dự đoán giá trị.

Kỹ thuật tăng cường (Ensemble Techniques): Kỹ thuật tăng cường là một phương pháp kết hợp nhiều mô hình dự đoán để tạo ra một dự đoán tốt hơn. Ví dụ: Kỹ thuật Bagging (Bootstrap Aggregating) sử dụng nhiều mô hình học máy độc lập và kết hợp kết quả để đưa ra dự đoán cuối cùng.

Xử lý ngôn ngữ tự nhiên (Natural Language Processing - NLP): Trong một số trường hợp, đoán giá có thể dựa trên thông tin và mô tả liên quan đến sản phẩm hoặc dịch vụ. Xử lý ngôn ngữ tự nhiên là một lĩnh vực của AI tập trung vào việc xử lý và hiểu ngôn ngữ tự nhiên của con người. Trong đoán giá, NLP có thể được sử dụng để phân tích và rút trích thông tin từ các mô tả sản phẩm hoặc nhận xét của người dùng để đưa ra dự đoán về giá trị.

Kỹ thuật trích xuất đặc trưng (Feature Extraction): Trích xuất đặc trưng là quá trình chuyển đổi dữ liệu đầu vào thành một tập hợp các đặc trưng có ý nghĩa và dễ dùng cho việc dự đoán. Trong đoán giá, kỹ thuật trích xuất đặc trưng có thể bao gồm việc chuyển đổi dữ liệu định tính thành dữ liệu số, trích

xuất thông tin từ hình ảnh hoặc văn bản, và xử lý các biến đầu vào để tạo ra các đặc trưng mới..

Những công nghệ này có thể được kết hợp và tùy chỉnh phù hợp với bài toán đoán giá cụ thể để đạt được kết quả tốt nhất. Việc lựa chọn công nghệ phụ thuộc vào tính chất của dữ liệu, độ phức tạp của bài toán, và tài nguyên có sẵn.

1.4. Ứng dụng

Trong bài tôi có Crawl dữ liệu từ [chợ tốt](#). Và dùng mô hình RandomForest được sử dụng để dự đoán giá của xe dựa trên các đặc trưng như tên xe, công ty sản xuất, năm sản xuất, số kilomet đã đi và loại nhiên liệu.

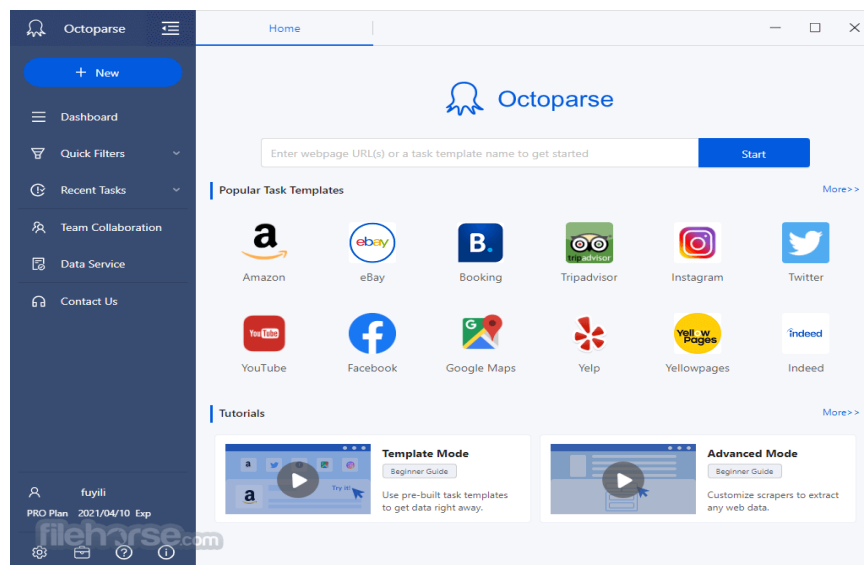
CHƯƠNG 2 : CƠ SỞ LÝ THUYẾT

2.1.Xử lý dữ liệu

2.1.1.Crawl data

Quá trình crawl data là quá trình quan trọng trong việc thu thập thông tin và dữ liệu từ các trang web. Các công cụ tìm kiếm hàng đầu như Google, Bing và nhiều công cụ khác sử dụng các bot để tự động truy cập vào các trang web và thu thập dữ liệu để thống kê và cung cấp thông tin cho người dùng.

Để thực hiện việc crawl dữ liệu, tôi đã sử dụng Octoparse 8 - một công cụ mạnh mẽ giúp tôi thu thập thông tin từ trang web <https://xe.chotot.com/mua-ban-oto?page=13>.



Hình 3: Octoparse 8

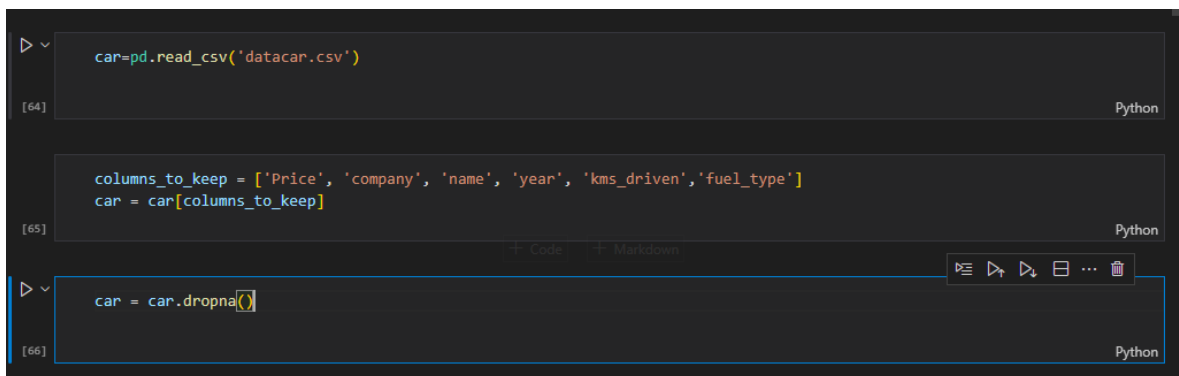
Sau khi quá trình crawl dữ liệu hoàn tất, tôi đã xuất dữ liệu về dạng file excel hoặc csv để bắt đầu quá trình xử lý dữ liệu thô. Bằng cách này, tôi có thể tiếp tục phân tích và trích xuất thông tin hữu ích từ dữ liệu đã thu thập được, phục vụ cho các mục đích khác nhau như nghiên cứu thị trường, phân tích dữ liệu, hay xây dựng các ứng dụng liên quan đến lĩnh vực mua bán ô tô.

2.1.2.Xử lý dữ liệu

Đầu tiên ta đọc dữ liệu từ tệp 'datacar.csv' và tạo một DataFrame có tên là "car". Tiếp theo, danh sách các cột cần giữ lại được chỉ định bao gồm 'Price', 'company', 'name', 'year', 'kms_driven', 'fuel_type'. Ta lọc DataFrame

"car" chỉ giữ lại các cột bao gồm 'Price', 'company', 'name', 'year', 'kms_driven', 'fuel_type'.

Sau khi lọc, các dòng chứa giá trị NaN (không xác định) trong bất kỳ cột nào của DataFrame "car" được loại bỏ bằng cách sử dụng câu lệnh "car.dropna()". Kết quả là DataFrame "car" chỉ chứa các hàng không chứa giá trị NaN, đảm bảo rằng dữ liệu đã được làm sạch và sẵn sàng cho việc xử lý và phân tích tiếp theo.



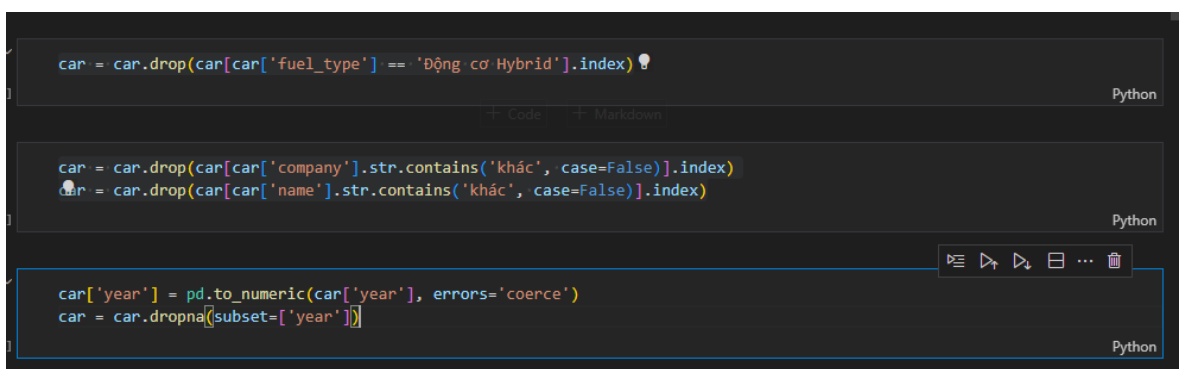
```
[64] car=pd.read_csv('datacar.csv')

[65] columns_to_keep = ['Price', 'company', 'name', 'year', 'kms_driven', 'fuel_type']
car = car[columns_to_keep]

[66] car = car.dropna()
```

Hình 4: Loại bỏ các cột không cần thiết

Ta sẽ loại bỏ các dòng không cần thiết dựa trên các điều kiện về "fuel_type", "company" và "name", chuyển đổi cột "year" thành dạng số và loại bỏ các dòng có giá trị thiếu trong cột "year".



```
car = car.drop(car[car['fuel_type'] == 'Động cơ Hybrid'].index)

car = car.drop(car[car['company'].str.contains('khác', case=False)].index)
car = car.drop(car[car['name'].str.contains('khác', case=False)].index)

car['year'] = pd.to_numeric(car['year'], errors='coerce')
car = car.dropna(subset=['year'])
```

Hình 5: Xử lý dữ liệu cột fuel_type, company, name

Tiếp theo ta thực hiện việc chuyển đổi kiểu dữ liệu và xử lý các cột trong DataFrame "car". Nó chuyển đổi cột "year" thành kiểu số nguyên, loại bỏ dấu chấm trong cột "Price" và giữ lại chỉ giá trị số, chuyển đổi cột "Price" thành kiểu số nguyên, chuyển đổi cột "kms_driven" thành kiểu số nguyên, và cuối cùng trả về kích thước của DataFrame "car". Các xử lý này giúp đảm bảo

dữ liệu được chuẩn hóa và sẵn sàng để thực hiện các phân tích hoặc mô hình hóa tiếp theo trên DataFrame.



```
[187] car['year'] = car['year'].astype(int)
Python

[188] car['Price'] = car['Price'].str.replace('.', '').str.extract('(\d+)')
      car['Price'] = np.int64(car['Price'])
Python

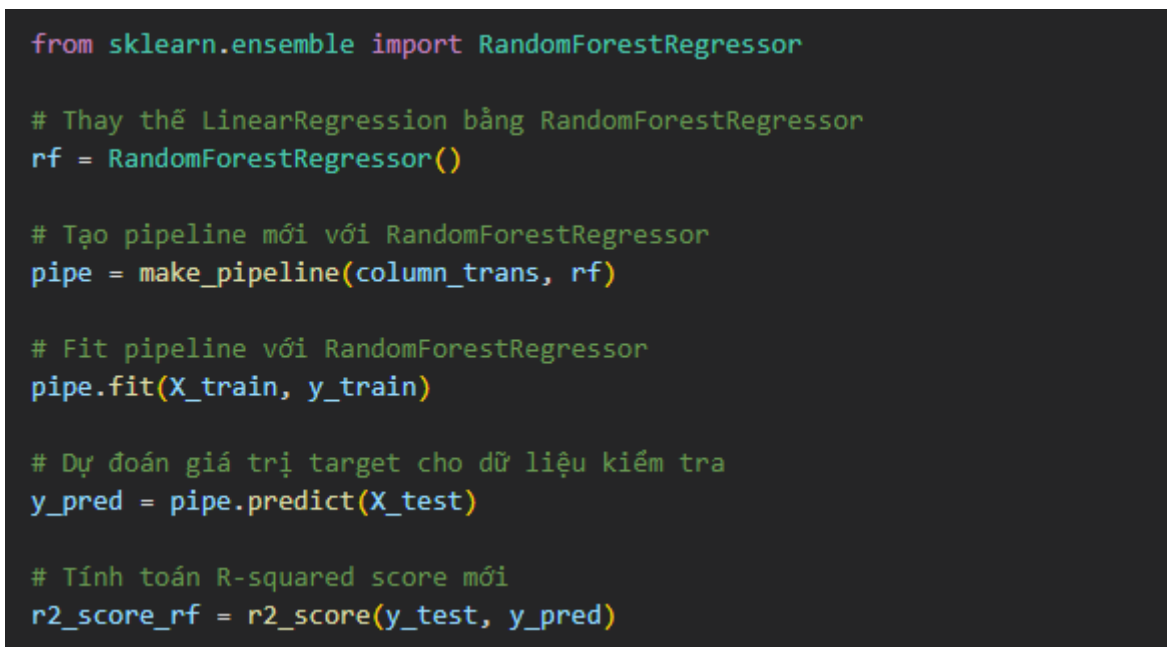
[189] car['kms_driven'] = car['kms_driven'].astype(int)
      car.shape
Python

... (5031, 6)
```

Hình 6: Xử lý cột year, price, kms_driven

2.2.Model

Mô hình Random Forest Regressor là một tập hợp các cây quyết định độc lập với nhau. Mỗi cây quyết định được huấn luyện trên một tập con của dữ liệu và sử dụng một số thuộc tính ngẫu nhiên để tạo ra các quyết định. Kết quả cuối cùng của Random Forest Regressor là sự kết hợp của dự đoán từ tất cả các cây quyết định.



```
from sklearn.ensemble import RandomForestRegressor

# Thay thế LinearRegression bằng RandomForestRegressor
rf = RandomForestRegressor()

# Tạo pipeline mới với RandomForestRegressor
pipe = make_pipeline(column_trans, rf)

# Fit pipeline với RandomForestRegressor
pipe.fit(X_train, y_train)

# Dự đoán giá trị target cho dữ liệu kiểm tra
y_pred = pipe.predict(X_test)

# Tính toán R-squared score mới
r2_score_rf = r2_score(y_test, y_pred)
```

Hình 7: Mô hình Random Forest

CHƯƠNG 3 : XÂY DỰNG CHƯƠNG TRÌNH

3.1.Xây dựng giao diện

Ta sử dụng Flask để xây dựng website dự đoán giá xe. Website cho phép người dùng chọn hãng xe, mẫu xe, năm sản xuất, kiểu nhiên liệu và số kilômét đã đi, sau đó dự đoán giá xe dựa trên mô hình học máy đã được huấn luyện.

Trang chủ của web hiển thị một biểu mẫu cho phép người dùng chọn các thông tin về xe và sau đó gửi yêu cầu dự đoán giá xe. Biểu mẫu được tạo bằng HTML và được render bởi Flask. Dữ liệu cho các trường chọn, chẳng hạn như hãng xe, mẫu xe, năm sản xuất và kiểu nhiên liệu, được truyền từ server đến client để hiển thị các tùy chọn cho người dùng.

Khi người dùng gửi biểu mẫu, thông tin được gửi đến đường dẫn "/predict". Tại đây, các giá trị đã được người dùng chọn được trích xuất từ yêu cầu và sử dụng để tạo một DataFrame mới. Mô hình học máy đã được nạp từ tập tin "LinearRegressionModel2.pkl" được sử dụng để dự đoán giá xe dựa trên dữ liệu đã cho. Giá dự đoán được trả về dưới dạng một phản hồi từ server và hiển thị cho người dùng.

Ứng dụng này tận dụng các công nghệ như Flask để tạo ra một giao diện tương tác dễ dùng cho việc dự đoán giá xe. Nó kết hợp giữa HTML, CSS và JavaScript để hiển thị trang web và thực hiện các yêu cầu dự đoán giá xe một cách linh hoạt và thuận tiện.

3.2.Code

```
from flask import Flask,render_template,request,redirect
from flask_cors import CORS,cross_origin
import pickle
import pandas as pd
import numpy as np

app=Flask(__name__)
cors=CORS(app)
model=pickle.load(open('LinearRegressionModel2.pkl','rb'))
```

```
car=pd.read_csv('Cleaned_Car_data.csv')

@app.route('/',methods=['GET','POST'])
def index():
    companies=sorted(car['company'].unique())
    car_models=sorted(car['name'].unique())
    year=sorted(car['year'].unique(),reverse=True)
    fuel_type=car['fuel_type'].unique()

    companies.insert(0,'Chọn tên hãng')
    return render_template('index.html',companies=companies,
car_models=car_models, years=year,fuel_types=fuel_type)
@app.route('/predict',methods=['POST'])
@cross_origin()
def predict():
    company=request.form.get('company')
    car_model=request.form.get('car_models')
    year=request.form.get('year')
    fuel_type=request.form.get('fuel_type')
    driven=request.form.get('kilo_driven')
    prediction=model.predict(pd.DataFrame(columns=['name', 'company',
'year', 'kms_driven', 'fuel_type'],

data=np.array([car_model,company,year,driven,fuel_type]).reshape(1, 5)))
    print(prediction)
    return str(np.round(prediction[0],2))
if __name__=='__main__':
    app.run()
```

DỰ ĐOÁN GIÁ XE

Để dự đoán giá xe bạn cần điền các thông tin dưới đây

Hãng xe:

BMW

Mẫu xe:

BMW 2 Series

Năm sản xuất:

2022

Kiểu nhiên liệu:

Diesel

Số km:

34000

Dự đoán

Hình 8: Giao diện chương trình

KẾT LUẬN

Tôi đã thực hiện xử lý dữ liệu trên tập dữ liệu ô tô và xây dựng một mô hình dự đoán giá xe dựa trên thông tin như công ty sản xuất, tên xe, năm sản xuất, số kilômét đã đi và loại nhiên liệu.

Với mô hình này, Tôi tạo ra một trang web dự đoán giá xe ô tô cho người dùng. Trang web sẽ cho phép người dùng nhập thông tin về công ty sản xuất, tên xe, năm sản xuất, số kilômét đã đi và loại nhiên liệu của xe ô tô mà họ quan tâm. Sau khi người dùng nhập thông tin, mô hình sẽ áp dụng dự đoán và hiển thị kết quả giá trị dự đoán của xe ô tô đó.

Trang web dự đoán giá xe ô tô sẽ giúp người dùng có cái nhìn trước về giá trị của xe ô tô mà họ quan tâm dựa trên các thông tin cơ bản. Điều này có thể hữu ích cho người mua xe ô tô khi họ đang tìm kiếm một xe mới hoặc muốn biết giá trị hiện tại của xe cũ của mình

➤ Ưu điểm :

- Chương trình đã có thể truyền dữ liệu từ web vào model và đưa ra kết quả dự đoán.
- Tốc độ nhanh không bị gián đoạn.
- Giao diện thân thiện dễ sử dụng.

➤ Nhược điểm :

- Độ sai lệch còn lớn
- Dữ liệu còn ít thuộc tính

➤ Hướng phát triển :

- Mở rộng dữ liệu: Thu thập dữ liệu từ nhiều nguồn khác nhau để mở rộng phạm vi dự đoán giá xe ô tô. Điều này có thể bao gồm việc thu thập thông tin từ các trang web mua bán ô tô, các báo cáo thị trường và dữ liệu từ các nhà sản xuất ô tô. Điều này giúp cung cấp một tập dữ liệu đa dạng và phong phú hơn cho mô hình dự đoán.
- Cải thiện mô hình dự đoán: Thử nghiệm và đánh giá các mô hình dự đoán khác nhau để tìm ra mô hình tốt nhất cho bài toán dự đoán giá xe ô tô. Các mô hình nâng cao, chẳng hạn như mạng neural hồi quy

(RNN) hoặc mô hình học sâu (deep learning), có thể được áp dụng để cải thiện độ chính xác dự đoán.

- Tăng cường tính năng: Mở rộng tính năng của trang web hoặc ứng dụng bằng cách cung cấp thêm thông tin hữu ích cho người dùng, chẳng hạn như đánh giá xe ô tô, so sánh giá trị giữa các mẫu xe khác nhau, xu hướng giá, hoặc dự báo giá trị tương lai.

TÀI LIỆU THAM KHẢO

- [1]. <https://www.kaggle.com/code/venkatganes98/old-car-price-prediction-randomforestreg-acc-85/notebook#3.Data-Pre-Processing>
- [2]. https://scikitlearn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html
- [3]. <https://machinelearningcoban.com/2016/12/28/linearregression/>