TRƯỜNG ĐẠI HỌC KỸ THUẬT CÔNG NGHIỆP THÁI NGUYÊN

**KHOA: ĐIỆN TỬ**

**BỘ MÔN: CÔNG NGHỆ THÔNG TIN**

****

**ĐỒ ÁN**

**MÔN HỌC**

**THỰC TẬP CHUYÊN NGÀNH**

Sinh viên: Nguyễn Đình Đức

Lớp: K55KMT

Giáo viên hướng dẫn: TS.Nguyễn Văn Huy

**Thái Nguyên - 2023**

**TRƯỜNG ĐHKTCN CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM**

**KHOA: ĐIỆN TỬ *Độc lập - Tự do - Hạnh phúc***

**ĐỒ ÁN**

**MÔN HỌC: THỰC TẬP CHUYÊN NGÀNH**

**BỘ MÔN: CÔNG NGHỆ THÔNG TIN**

*Sinh viên: Nguyễn Đình Đức. Lớp: K55KMT.*

*Nghành: Kỹ Thuật Máy Tính.*

*Giáo viên hướng dẫn: TS.Nguyễn Văn Huy.*

*Ngày giao đề:…/…/…… Ngày hoàn thành: …/…/……*

1. Tên đề tài: Dự đoán giá chung cư khu vực Hà Nội.
2. Yêu cầu:

* Thu thập dữ liệu từ website.
* Phân tích dữ liệu.
* Chọn mô hình để huấn luyện.
* Xây dựng website dự đoán giá chung cư.

1. Các bản vẽ, chương trình và đồ thị: Kiểm thử chương trình.

**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 và ghi rõ họ, tên)*

**NHẬN XÉT CỦA GIÁO VIÊN CHẤM**

.......................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................... .................................................................................................................................. .................................................................................................................................. .................................................................................................................................. .................................................................................................................................. .................................................................................................................................. .................................................................................................................................. ..................................................................................................................................

*Thái Nguyên, ngày… tháng … năm 20…*

**GIÁO VIÊN CHẤM**

*( Ký và ghi rõ họ, tên)*

**MỤC LỤC**

[**MỤC LỤC** 5](#_Toc137727585)

[**MỤC LỤC HÌNH ẢNH** 5](#_Toc137727586)

[**LỜI NÓI ĐẦU** 7](#_Toc137727587)

[**CHƯƠNG I: GIỚI THIỆU CHUNG** 8](#_Toc137727588)

[**1.** **Đoán giá là gì?** 8](#_Toc137727589)

[**2.** **Ứng dụng** 8](#_Toc137727590)

[**3.** **Công nghệ** 9](#_Toc137727591)

[**4.** **Áp dụng** 11](#_Toc137727592)

[**CHƯƠNG II: CƠ SỞ LÝ THUYẾT** 12](#_Toc137727593)

[**1.** **Xử lý dữ liệu.** 12](#_Toc137727594)

[**a)** **Crawl data.** 12](#_Toc137727595)

[**b)** **Xử lý data trên Excel.** 13](#_Toc137727596)

[**c)** **Xử lý data với python.** 14](#_Toc137727597)

[**2.** **Model Random Forest** 18](#_Toc137727598)

[**CHƯƠNG III: XÂY DỰNG CHƯƠNG TRÌNH** 20](#_Toc137727599)

[**1.** **Framework Django.** 20](#_Toc137727600)

[**2.** **Xây dựng web.** 20](#_Toc137727601)

[**a)** **Front-end.** 20](#_Toc137727602)

[**b)** **Back-end.** 21](#_Toc137727603)

[**c)** **Kiểm thử.** 22](#_Toc137727604)

[**CHƯƠNG IV: KẾT LUẬN** 23](#_Toc137727605)

[**1.** **Tổng quan** 23](#_Toc137727606)

[**2.** **Ưu, nhược điểm** 23](#_Toc137727607)

[**a)** **Ưu điểm** 23](#_Toc137727608)

[**b)** **Nhược điểm.** 23](#_Toc137727609)

[**3.** **Hướng phát triển.** 23](#_Toc137727610)

[**TÀI LIỆU THAM KHẢO** 24](#_Toc137727611)

**MỤC LỤC HÌNH ẢNH**

[Hình 1: Đoán giá là gì? 8](#_Toc137727613)

[Hình 2: https://batdongsan.com.vn/. 11](#_Toc137727614)

[Hình 3: Dữ liệu crawl trên Octoparse 8. 12](#_Toc137727615)

[Hình 4: Dữ liệu thôi được đưa về Excel xử lý. 13](#_Toc137727616)

[Hình 5: Dữ liệu đã qua xử lý trên Excel. 13](#_Toc137727617)

[Hình 6: Đọc file CSV bằng thư viện Pandas. 14](#_Toc137727618)

[Hình 7: Thông tin kiểu dữ liệu. 14](#_Toc137727619)

[Hình 8: Dữ liệu duy nhất. 15](#_Toc137727620)

[Hình 9: Dữ liệu bị trống (Nan). 15](#_Toc137727621)

[Hình 10: Điền dữ liệu vào dữ liệu bị trống bằng hàm mean(). 16](#_Toc137727622)

[Hình 11: Tạo hàm Outlier. 16](#_Toc137727623)

[Hình 12: Tìm các dữ liệu khác thường. 17](#_Toc137727624)

[Hình 13: Mã hoá các dữ liệu object. 17](#_Toc137727625)

[Hình 14: Dữ liệu đã xử lý xong, chuẩn bị training. 17](#_Toc137727626)

[Hình 15: Xây dựng mô hình bằng phương pháp Random Forest Regressor. 18](#_Toc137727627)

[Hình 16: Sơ đồ dự đoán giá so với dữ liệu đầu vào của mô hình được xây dựng bằng phương pháp Random Forest Regressor. 19](#_Toc137727628)

[Hình 17: Chỉ số RMSE và MSE của mô hình xây dựng bằng phương pháp Random Forest Regressor. 19](#_Toc137727629)

[Hình 18: HTML phần giao diện web. 21](#_Toc137727630)

[Hình 19: Hàm Home() để đưa dữ liệu nhập từ web vào mô hình dự đoán và đưa kết quả ra web. 21](#_Toc137727631)

[Hình 20: Giao diện dự đoán và thử dự đoán. 22](#_Toc137727632)

[Hình 21: Giá của chung cư trên web chính. 22](#_Toc137727633)

**LỜI NÓI ĐẦU**

Trong bối cảnh tăng trưởng đô thị và phát triển kinh tế, giá nhà trở thành một trong những yếu tố quan trọng ảnh hưởng đến quyết định mua bán và đầu tư trong lĩnh vực bất động sản. Việc dự đoán giá mua chung cư trở thành một thách thức lớn đối với các nhà phát triển, chính phủ, và cả những người đang tìm kiếm mua nhà.

Để tìm hiểu sâu hơn về lĩnh vực giá trên, tôi chọn đề tài “Dự đoán giá chung cư khu vực Hà Nội”. Để làm đề tài này tôi dùng kiến thức của khoa học dữ liệu, học máy và lập trình web.

Đề này nhằm mục tiêu sử dụng dữ liệu được lấy trên website để phân tích và xây dựng mô hình dự đoán giá chung cư khu vực Hà Nội. Chúng ta sẽ áp dụng các kỹ thuật và phương pháp trong lĩnh vực khoa học dữ liệu để tìm hiểu các yếu tố ảnh hưởng đến giá chung cư và xây dựng mô hình dự đoán chính xác.

**CHƯƠNG I: GIỚI THIỆU CHUNG**

1. **Đoán giá là gì?**



Hình 1: Đoán giá là gì?

"Đoán giá" có thể ám chỉ việc ước lượng hoặc dự đoán giá trị của một sản phẩm, tài sản hoặc dịch vụ. Việc đoán giá thường được thực hiện dựa trên các yếu tố như dữ liệu thị trường, thông tin liên quan, xu hướng và mô hình phân tích. Các phương pháp và kỹ thuật trong việc đoán giá có thể được áp dụng từ nhiều lĩnh vực, bao gồm kinh tế, tài chính, bất động sản, hàng hóa và các thị trường tài sản khác.

Tuy nhiên, để có dự đoán giá chính xác, thông tin cụ thể về sản phẩm hoặc dịch vụ cần được xem xét, cũng như các yếu tố kinh tế và thị trường liên quan. Trong một số trường hợp, các mô hình dự đoán hoặc thuật toán máy học có thể được sử dụng để phân tích dữ liệu và đưa ra dự đoán về giá trị trong tương lai.

1. **Ứng dụng**

Có nhiều ứng dụng của việc đoán giá trong các lĩnh vực khác nhau. Dưới đây là một số ví dụ phổ biến về các ứng dụng của đoán giá:

* Tài chính và chứng khoán: Trong lĩnh vực tài chính, việc đoán giá có thể được sử dụng để dự đoán giá cổ phiếu, hàng hóa, ngoại tệ và các tài sản tài chính khác. Các nhà đầu tư và các công ty tài chính thường sử dụng các mô hình và thuật toán dự đoán để đưa ra quyết định giao dịch và đầu tư.
* Bất động sản: Đoán giá cũng được áp dụng trong lĩnh vực bất động sản để ước lượng giá trị các tài sản như căn hộ, nhà đất hoặc tòa nhà. Các nhà phát triển, nhà môi giới và các chuyên gia bất động sản sử dụng các mô hình đoán giá để đưa ra quyết định về mua, bán hoặc đầu tư vào bất động sản.
* Thương mại điện tử: Trong lĩnh vực thương mại điện tử, việc đoán giá có thể được sử dụng để ước lượng giá sản phẩm và dịch vụ trên các nền tảng mua sắm trực tuyến. Các công ty thương mại điện tử thường sử dụng các thuật toán dự đoán giá để đề xuất giá cả cạnh tranh và tùy chỉnh cho khách hàng.
* Ngành y tế: Đoán giá cũng có thể được áp dụng trong ngành y tế để ước lượng giá trị của các dịch vụ y tế và sản phẩm dược phẩm. Các công ty bảo hiểm y tế, bệnh viện và nhà nghiên cứu y tế sử dụng các mô hình đoán giá để đưa ra quyết định về giá cả, chi phí và quản lý tài chính.
* Du lịch và khách sạn: Trong ngành du lịch và khách sạn, việc đoán giá có thể được sử dụng để ước lượng giá phòng khách sạn, vé máy bay, tour du lịch và các dịch vụ du lịch khác. Các công ty du lịch, đại lý du lịch và các trang web đặt phòng sử dụng các mô hình đoán giá để quản lý giá cả, tối ưu hóa doanh thu và cung cấp ưu đãi cho khách hàng.

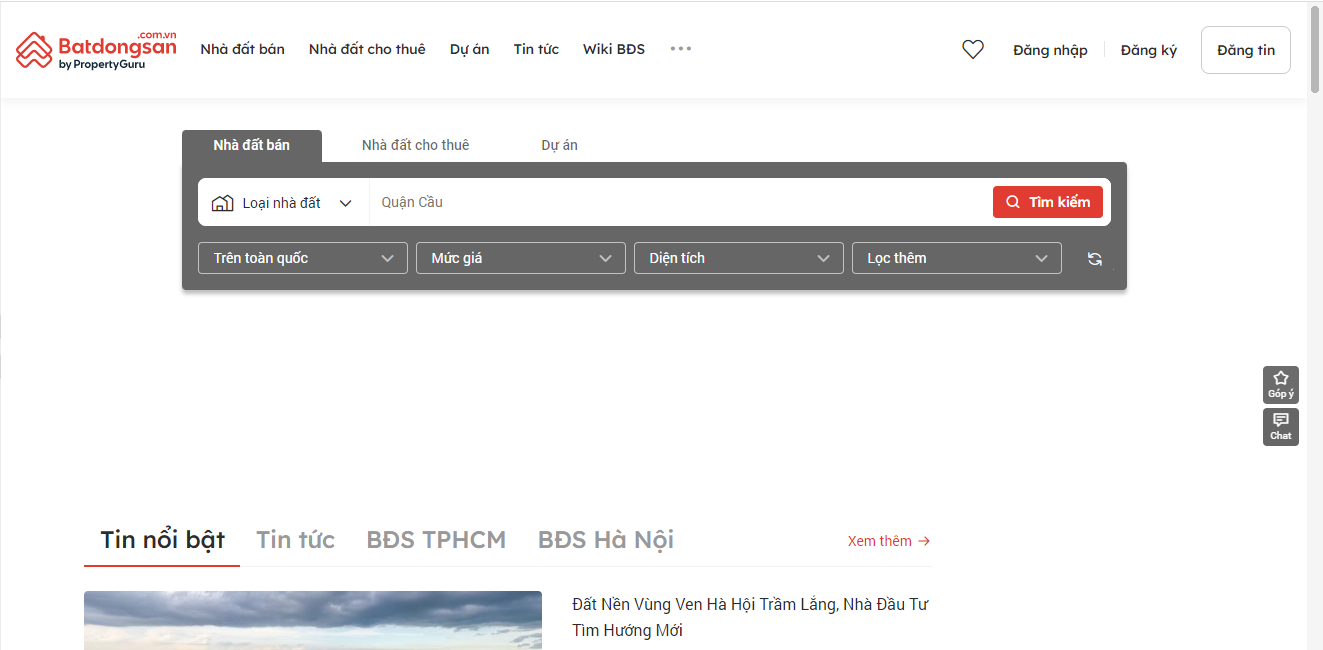
1. **Công nghệ**

Công nghệ đóng vai trò quan trọng trong việc đoán giá và các ứng dụng liên quan. Dưới đây là một số công nghệ thường được sử dụng trong việc đoán giá:

* 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. **Áp dụng**



Hình 2: https://batdongsan.com.vn/.

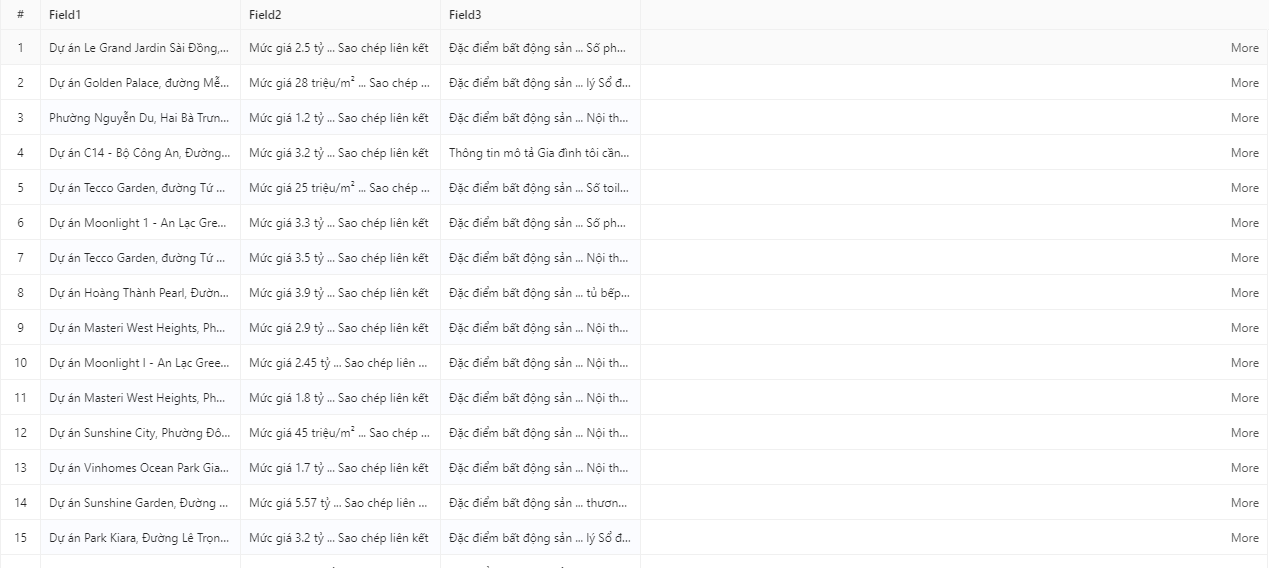
Đề tài “Dự đoán giá chung cư khu vực Hà Nội” là một đề tài trong lĩnh vực bất động sản. Để có thể hoàn thành đề tài này, thì đầu tiên tôi đã lấy dữ liệu về giá chung cư trên trang <https://batdongsan.com.vn/>. Sau khi lấy được dữ liệu thô, thông qua excel để xử lý dữ liệu thô xong rồi đưa dữ liệu đã được xử lý vào bài làm để xây dựng mô hình. Phương pháp được sử dụng để xây dựng mô hình đề sử dụng ở đây là Random Forest. Đây là một phương pháp trong lĩnh vực Học máy và thuộc vào kỹ thuật tăng cường.

**CHƯƠNG II: CƠ SỞ LÝ THUYẾT**

1. **Xử lý dữ liệu.**
2. **Crawl data.**

Crawl data là quá trình thu thập dữ liệu và thông tin website nhằm phục vụ nhiều mục tiêu khác nhau. Theo đó, các bot của công cụ tìm kiếm (Search Engine) như Google, Bing,… sẽ lần lượt truy cập vào tất cả trang trên website cũng như liên kết liên quan để thống kê dữ liệu.

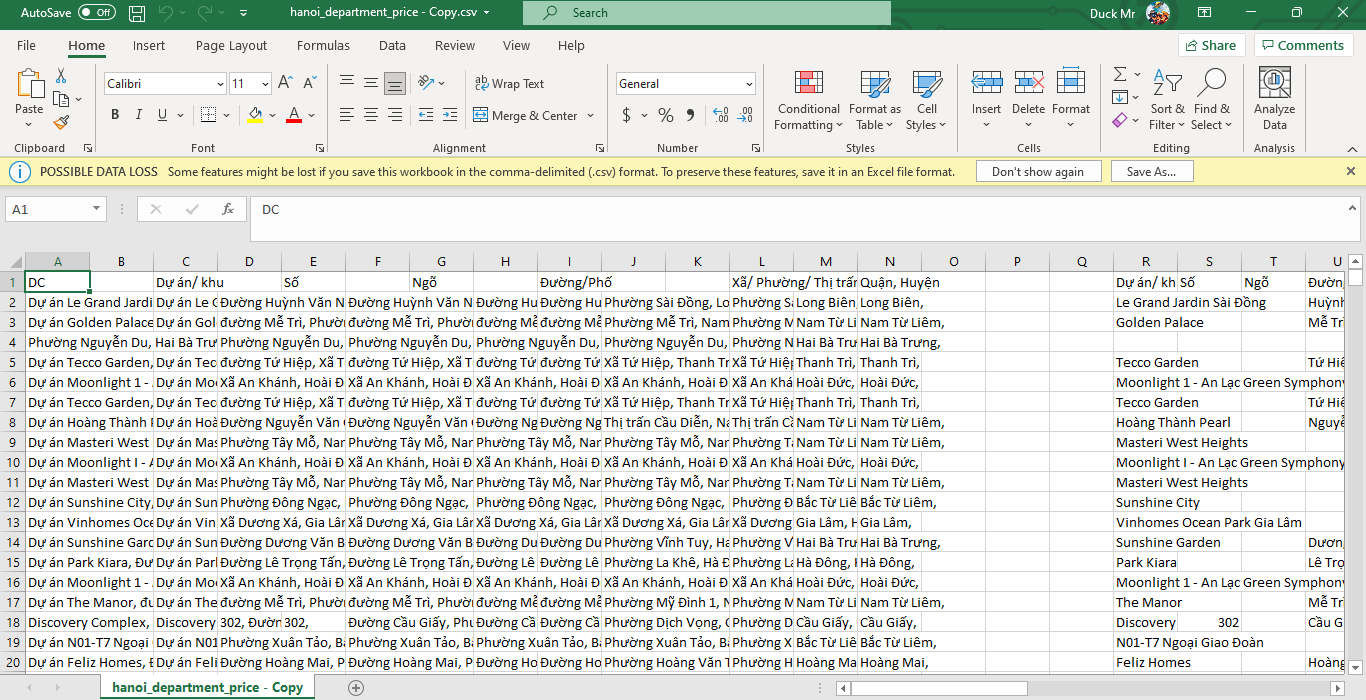
Ở đây tôi sử dụng Octoparse 8 để có thể thu thập thông tin từ <https://batdongsan.com.vn/>.



Hình 3: Dữ liệu crawl trên Octoparse 8.

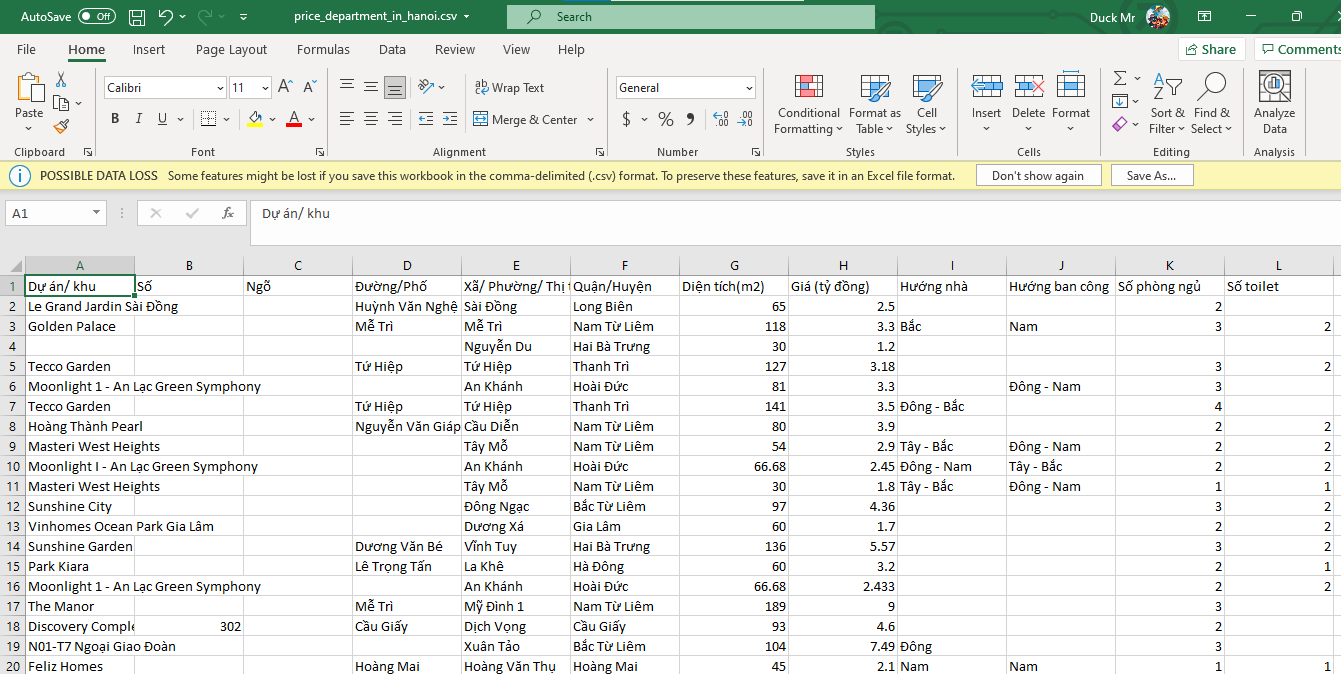
Sau khi Crawl data xong thì sẽ export data về file excel, csv … để bắt đầu quá trình xử lý data thô.

1. **Xử lý data trên Excel.**



Hình 4: Dữ liệu thôi được đưa về Excel xử lý.

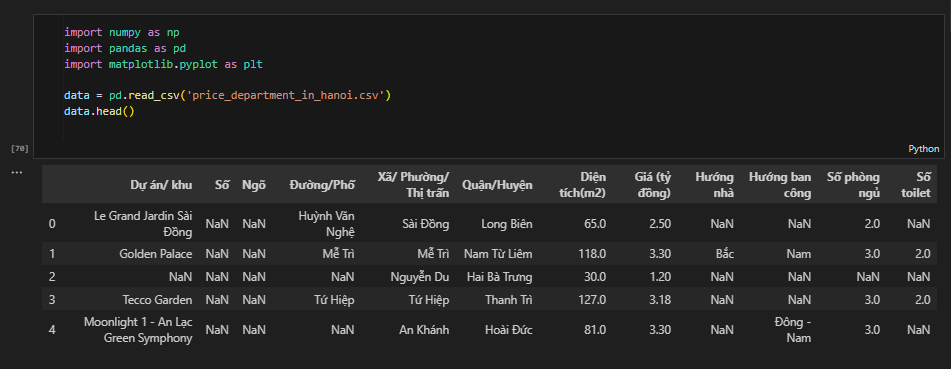
Với dữ liệu thô như hình trên, thông qua quá dùng các hàm có trên Excel như TRIM(), SUBSTITUTE(), LEFT(), RIGHT(), MID(),… thì dữ liệu đã được tách ra làm các trường phù hợp để có thể vào công đoạn xử lý data cho mô hình.



Hình 5: Dữ liệu đã qua xử lý trên Excel.

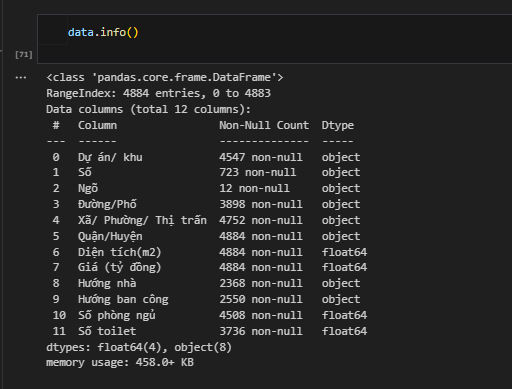
Qua xử lý dữu liệu trên Excel thì dữ liệu còn khoảng hơn 5000 dòng.

1. **Xử lý data với python.**



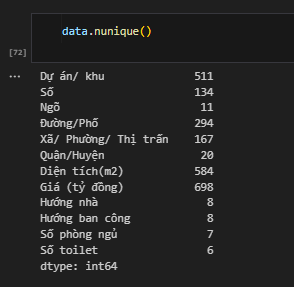
Hình 6: Đọc file CSV bằng thư viện Pandas.

Đầu tiên sử dụng thư viện pandas để đọc file csv. Sau khi xem qua dữ liệu đầu vào, sẽ xem các thông số thuộc tính,… để từ đó đánh giá dữ liệu đầu vào nên làm những gì để phù hợp với mô hình.



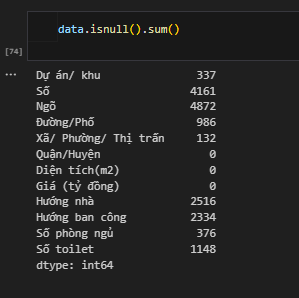
Hình 7: Thông tin kiểu dữ liệu.

Ở đây dữ liệu đầu vào có hai kiểu là object và float64. Float là kiểu dữ liệu số dạng thập phân, còn object là kiểu dữ liệu gốc và là cơ sở cho tất cả các kiểu dữ liệu khác. Trong dữ liệu đầu vào có 3 cột Đường/Phố, Xã/Phường/Thị Trấn, Quận/Huyện là thuộc kiểu object, cụ thể ở đây là kiểu text.



Hình 8: Dữ liệu duy nhất.

Sau khi xem kiểu dữ liệu của dữ liệu đầu vào, dùng nunique() để xem có bao nhiêu dữ liệu duy nhất trong dư liệu đầu vào.



Hình 9: Dữ liệu bị trống (Nan).

Dùng vòng lặp để xem số các dòng dữ liệu của tùng cột thuộc tính không có giá trị, hay còn là giá trị Nan (Các dữ liệu trống đó sẽ được gọi là bị missing value). Do số lượng dữ liệu trống khá lớn nên sẽ phải xử lý các dữ liệu đó.

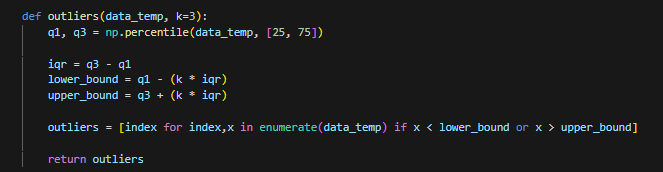
Đối với trường hợp missing value ít so với tập dữ liệu thì các dòng đó sẽ bị xoá bằng hàm dropna(). Nhưng ở đây dữ liệu trống so với tập dữ liệu khá lớn nên sẽ phải dùng cách khác, đó là dùng giá trị trung bình của cột để điền vào các ô bị thiếu dữ liệu. Đối với dữ liệu kiểu object thì tôi điền là không có.



Hình 10: Điền dữ liệu vào dữ liệu bị trống bằng hàm mean().

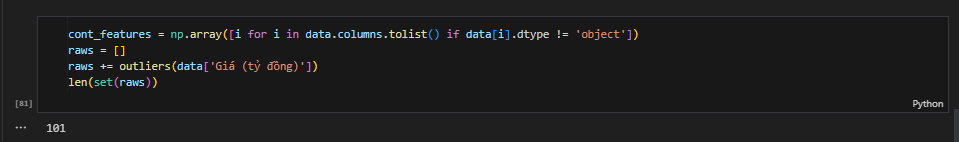
Sau khi xử lý các dữ liệu bị thiếu xong thì chuyển qua xử lý dữ liệu bị outliers, tức là các dữ liệu khác thường trong dữ liệu đầu vào.

Để tìm các dòng dữ liệu bị outliers, tôi sử dụng hàm outliers dựa trên phương pháp IQR để xác định ngưỡng cho outliers.



Hình 11: Tạo hàm Outlier.

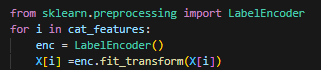
Hàm này sẽ tính các phân vị q1, q3 sau đó sẽ tìm giới hạn trên và dưới từ các phân vị và tham số k. Cuối cùng đưa ra mảng các dữ liệu nhỏ hơn dữ liệu dưới và lớn hơn dữ liệu trên. Sau khi lấy được mảng sẽ loại bỏ các dữ liệu đó.



Hình 12: Tìm các dữ liệu khác thường.

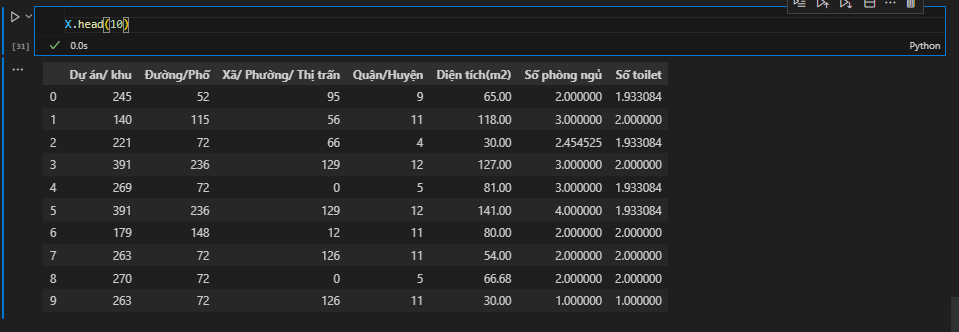
Trong bài thì tôi dùng tìm các outlier cho cột giá và tìm được 101 dữ liệu outlier.

Sau khi xử lý xong các dữ liệu kiểu float xong thì tôi đi vào xử lý dữ liệu kiểu object. Đối với các dữ liệu này tôi dùng hàm LabelEncoder() để mã hoá các biến thành các số nguyên vì mô hình cần dữ liệu là số để xây dựng.



Hình 13: Mã hoá các dữ liệu object.

Dưới đấy là dữ liệu đã được xử lý qua tất các công đoạn ở trên:

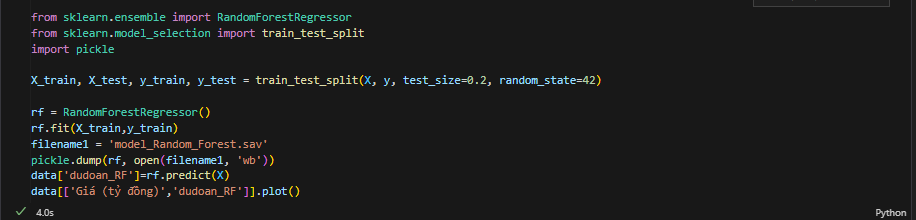


Hình 14: Dữ liệu đã xử lý xong, chuẩn bị training.

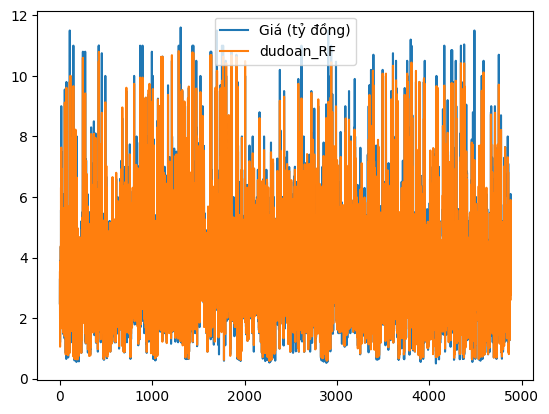
1. **Model Random Forest**

Random Forest Regressor là một mô hình học máy trong scikit-learn được dùng để huấn luyện và dự đoán trên dữ liệu dạng hồi quy (regression). Nó sử dụng phương pháp Random Forest (rừng ngẫu nhiên) để xây dựng nhiều cây quyết định (decision trees) và kết hợp kết quả của chúng để đưa ra dự đoán cuối cùng.

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.

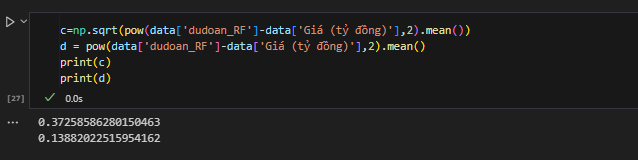


Hình 15: Xây dựng mô hình bằng phương pháp Random Forest Regressor.



Hình 16: Sơ đồ dự đoán giá so với dữ liệu đầu vào của mô hình được xây dựng bằng phương pháp Random Forest Regressor.

Các dường dự đoán đã khả quan hơn so với hồi quy tuyến tính.



Hình 17: Chỉ số RMSE và MSE của mô hình xây dựng bằng phương pháp Random Forest Regressor.

Kết quả với chỉ số MSE là 0.138 và RMSE là 0.3725. Hai chỉ số này đều khá gần 0 nên có thể nói rằng mô hình dự đoán khá chính xác.

**CHƯƠNG III: XÂY DỰNG CHƯƠNG TRÌNH**

1. **Framework Django.**

Django là một khung Web Python cấp cao, khuyến khích phát triển nhanh chóng và thiết kế thực dụng, gọn gàng.Được xây dựng bởi các nhà phát triển có kinh nghiệm, nó xử lý nhiều rắc rối của việc phát triển Web, vì vậy bạn có thể tập trung vào viết ứng dụng của mình mà không cần phải phát minh lại bánh xe. Nó có nguồn mở và miễn phí.

Với django bạn có thể lấy các ứng dụng web từ ý tưởng để khởi chạy trong vài phút. Và để làm được điều này thì django có một vài tính năng nhẹ như sau:

* + Nhanh: Django được thiết kế để giúp các nhà phát triển đưa các ứng dụng từ ý tưởng đến hoàn thành càng nhanh càng tốt..
  + Có đầy đủ các thư viện/module cần thiết: Django bao gồm hàng tá các tính năng bổ sung mà bạn có thể sử dụng để xử lý các tác vụ phát triển Web phổ biến. Django chăm sóc xác thực người dùng, quản trị nội dung, bản đồ trang web, nguồn cấp dữ liệu RSS và nhiều tác vụ khác - ngay lập tức.
  + Đảm bảo về tính bảo mật: Django rất coi trọng vấn đề bảo mật và giúp các nhà phát triển tránh được nhiều lỗi bảo mật phổ biến, chẳng hạn như SQL SQL, kịch bản chéo trang, giả mạo yêu cầu chéo trang và nhấp chuột. Hệ thống xác thực người dùng của nó cung cấp một cách an toàn để quản lý tài khoản và mật khẩu người dùng.
  + Khả năng mở rộng tốt: Một số địa điểm bận rộn nhất trên hành tinh sử dụng khả năng có thể mở rộng nhanh chóng và linh hoạt của django để đáp ứng nhu cầu giao thông nặng nhất.
  + Tính linh hoạt: Các công ty, tổ chức và chính phủ đã sử dụng Django để xây dựng tất cả mọi thứ - từ hệ thống quản lý nội dung đến mạng xã hội đến nền tảng điện toán khoa học.

1. **Xây dựng web.**
2. **Front-end.**

****

Hình 18: HTML phần giao diện web.

1. **Back-end.**

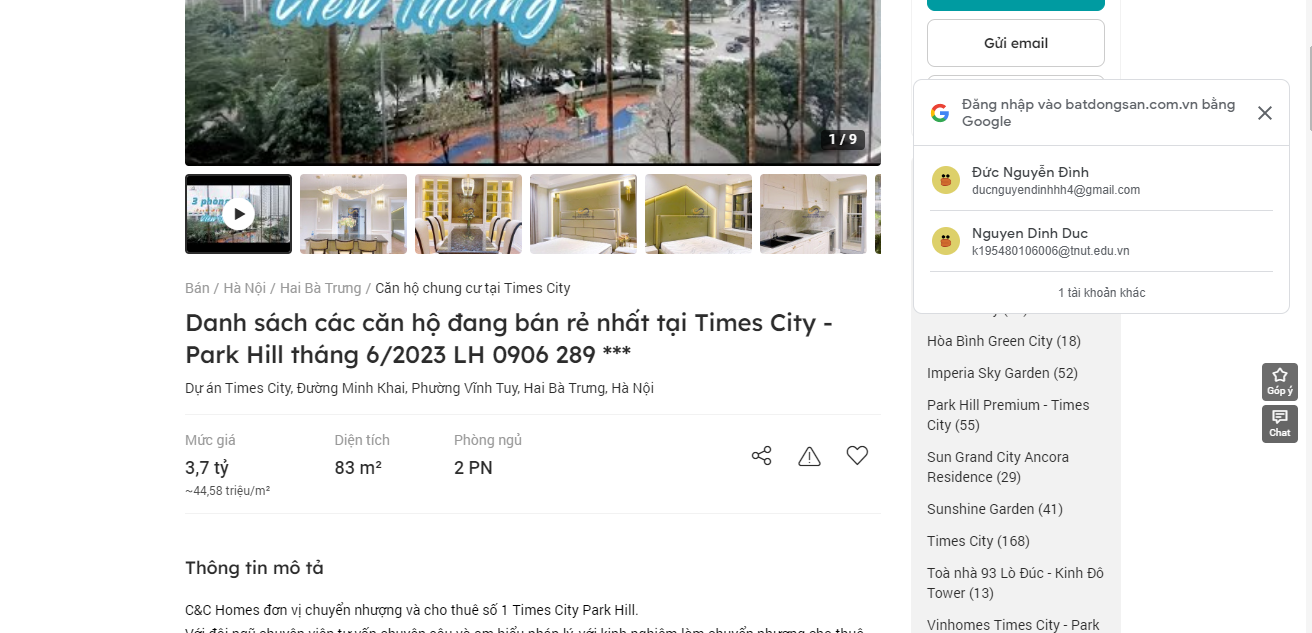
****

Hình 19: Hàm Home() để đưa dữ liệu nhập từ web vào mô hình dự đoán và đưa kết quả ra web.

1. **Kiểm thử.**



Hình 20: Giao diện dự đoán và thử dự đoán.



Hình 21: Giá của chung cư trên web chính.

**CHƯƠNG IV: KẾT LUẬN**

1. **Tổng quan**

Chương trình đã có thể thu thập các thông tin được nhập qua giao diện web của streamlit rồi đưa vào model đã được xây dựng để dự đoán giá chung cư trong khu vực Hà Nội.

1. **Ưu, nhược điểm**
2. **Ưu điểm**

* Giao diện thân thiện dễ sử dụng.
* Tốc độ dự đoán nhanh chóng, không bị gián đoạn.

1. **Nhược điểm.**

* Vẫn có một số giá khi dự đoán còn cách xa giá cũ.
* Dữ liệu còn ít.
* Dữ liệu còn ít thuộc tính.

1. **Hướng phát triển.**

Chương trình sẽ thu thập nhiều dữ liệu, bổ sung thêm các thuộc tính có tính tương quan với giá nhà, giảm độ lệch sai số thấp hơn so với kết quả hiện tại.

**TÀI LIỆU THAM KHẢO**

**[1]** [**https://www.kaggle.com/**](https://www.kaggle.com/)

**[2]** [**https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html**](https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html)

**[3]** [**https://scikit-learn.org/stable/modules/generated/sklearn.linear\_model.LinearRegression.html**](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html)

**[4]** [**https://machinelearningcoban.com/2016/12/28/linearregression/**](https://machinelearningcoban.com/2016/12/28/linearregression/)

**[5]** [**https://solieu.vip/mse-va-rmse-la-gi-va-cach-tinh-tren-stata/**](https://solieu.vip/mse-va-rmse-la-gi-va-cach-tinh-tren-stata/)