Nguyễn Chí Thanh-64TTNT2

2251262639

# House Price Prediction Using Linear and Lasso Regression

Bài báo nghiên cứu này tập trung vào việc dự đoán giá nhà với hiệu quả và độ chính xác cao hơn. Các tác giả bao gồm Mayank Sharma, Rahul Chauhan, Swati Devliyal từ Graphic Era Hill University và Kanegonda Ravi Chythanya từ SR University[***Abstract— This research paper aims to predict house prices***

***with more efficiency and accuracy. The research paper involves three machine learning algorithms, "linear regression, lasso regression, and ridge regression,” that are applied to the data set of Bangalore and have various information about different places in Bangalore, like BHK, number of bathrooms, area, etc. The house price prediction is quite a difficult task, as predicting the price of the house involves many factors like area, number of bathrooms, climate condition, age of the property, and many other factors. To make predictions according to all the factors, the author has applied three algorithms, one by one. From these machine learning algorithms, we have developed a model that helps predict the prices of houses with more accuracy and a lower error percentage. This model can be very useful for both buyers and sellers, as it can justify the price of the house for both of them.*]-[1]**

Việc dự đoán giá nhà là một nhiệm vụ khó khăn do liên quan đến nhiều yếu tố khác nhau như diện tích, số lượng phòng tắm, điều kiện khí hậu, tuổi thọ của tài sản và nhiều yếu tố khác[1].... Phương pháp định giá bất động sản truyền thống, dựa vào các yếu tố chủ quan và khách quan như so sánh doanh số, vị trí và tình trạng [**Facilities available.**

**Age of the property.**

**Parking area.**

**Bathroom.**

**And many other factors are there. In order to make reasonable decisions, buyers, sellers, investors, and real estate agents needs accurate real estate predictions to justify the actual price of the property.**

**Traditionally real estate involves a combination of subjective and objective factors. It considers factors such as:-**

**Comparable sales**

**Location.**

**Condition.**

**Use their other traditional methods to determine the property's value.**], được cho là tốn thời gian, đắt đỏ, dễ mắc lỗi và có thể không công bằng do yếu tố cá nhân[**However, this method is time consuming, expensive, and leads to many errors.**

**Moreover, this process of predicting a house value is not always fair or objective because it involves personal opinions and judgments, which may lead to differences in the price and unfair treatment to both buyer or seller .**

**Machine learning algorithms and data can help predict housing prices with a more of accuracy than that of traditional methods. As Machine learning algorithms are capable of identifying those patterns from the given data that humans may are no capable of or ignore those patterns.**]-[]

Các thuật toán học máy và dữ liệu mang lại tiềm năng dự đoán giá nhà chính xác hơn phương pháp truyền thống.... Chúng có thể xác định các mẫu trong dữ liệu mà con người có thể bỏ sót[2], giúp phát triển các mô hình dự đoán với độ chính xác, tốc độ và hiệu quả cao. Nhiều kỹ thuật học máy đã được áp dụng cho dự đoán giá nhà, bao gồm hồi quy tuyến tính, các cách tiếp cận trí tuệ tính toán[**By using machine learning techniques, we can develop models that predict real estate prices with high precision, speed and more accuracy.**

**There are various techniques used for house price prediction apart from traditional appraisal methods followed by real estate businesspersons.**

**Machine learning and artificial intelligence have emerged as popular alternatives because of their ability to leverage data to determine the best pricing strategies to give fair judgement .Linear regression models and computational intelligence approaches are some of the commonly used techniques for house price prediction.**], mạng nơ-ron BP, kỹ thuật hồi quy nâng cao hơn, lý thuyết giá hedonic, hồi quy k-Nearest-Neighbors và Random Forest[**More advanced methods like the BP neural network model and regression techniques have also been employed. The hedonic prices theory has been used for price modeling in real estate, and multiple linear regression models have been found effective for predicting and analyzing housing prices [8] . Additionally, in a study, k-Nearest-Neighbors regression and Random Forest regression were used to predict house prices .**

**We all know that how much house price prediction models can provide valuable impact into the real estate market, but there are some limitations that researchers should be aware of when conducting their analysis or making their model.**].

Tuy nhiên, các mô hình dự đoán cũng có hạn chế, đặc biệt là chất lượng dữ liệu (không đủ, không chính xác, có outliers). Các mô hình cũng có thể không tính đến các yếu tố bên ngoài như thay đổi thị trường hay kinh tế [**One of limitation is the quality of the data used to train the models. If the data is not sufficient or inaccurate this may lead to false or inaccurate predictions, data can be inaccurate if some outliers are present in the data which needs to be fixed first . The fact that , the models may not account for external factors that could affect or influence house prices, such as changes in the ups and downs in the market, economy or shifts in the housing market.**

**Our Solution utilizes linear regression to analyze and predict trends in a given dataset.**].-[3]

Giải pháp được đề xuất trong bài báo này sử dụng hồi quy tuyến tính để phân tích và dự đoán xu hướng trong tập dữ liệu[3].... Hồi quy tuyến tính hoạt động bằng cách dự đoán giá trị của biến phụ thuộc dựa trên biến độc lập[**Linear regression is used to predict value of dependent variable on certain independent variable. As to predict the value of dependent variable it requires the set of data .**

**So by analyzing the data and identifying patterns, we can create a model that can be used to predict future values and trends. This process involves determining the best fit line that shows the relation between independent variable and dependent variable. The result is a mathematical equation that can be used to make predictions based on the values of the independent variable . Ours Solution's use of linear regression provides a powerful tool for analyzing and predicting trends in various industries, including real estate, finance, and marketing**]-[4]. Nó sử dụng một phương trình toán học (Y=a+bX) để tìm mối quan hệ và đưa ra dự đoán[4].... Kỹ thuật này hữu ích trong nhiều lĩnh vực, bao gồm bất động sản[4].

Bài báo áp dụng ba thuật toán học máy: hồi quy tuyến tính (linear regression), hồi quy Lasso (Lasso regression), và hồi quy Ridge (Ridge regression)... trên tập dữ liệu của Bangalore[1].... Tập dữ liệu này chứa thông tin về diện tích, địa điểm, số phòng, số phòng tắm và giá cả tại các địa điểm khác nhau[***A. Dataset A data set is an organized or Ordered collection of data.*Our solution uses the data set of Bangalore which includes various data of Bangalore like data of different places which includes the area, location ,number of rooms, number of bathroom prices as parameters.**].

•

Hồi quy Lasso là một biến thể của hồi quy tuyến tính, sử dụng khái niệm "co rút" (shrinkage)[**This type of regression is one of the most popular and widely used regression techniques in the case where the data is highly corelated. So because of this correlation of data the model may make less-reliable or unworthy prediction when it is given a new dataset.**

**Below is the equation or mathematical formula for lasso regression :-**

***D. Ridge Regression Ridge regression is one of the most popular machine***

**learning algorithms for prediction on the base of given set of data. Ridge regression is a fashion used to estimate the portions of multiple retrogression models in a situation where the independent variables parade high correlation. It serves to develop a terse model if the number of predictor variables in the set is greater than the number of consensus, or if the database shows multicollinearity in the relationship between predictor variables. Ridge regression offers a precious approach for assaying multiple regression on data that parade multicollinearity, which refers to the presence of nearly direct connections among independent variables**

].-[5]

•

Hồi quy Ridge được sử dụng phổ biến khi các biến độc lập có mối tương quan cao (đa cộng tuyến - multicollinearity)[5]...

Bài báo sử dụng các thước đo như Root Mean Square Error (RMSE) để đo lường sự khác biệt giữa giá trị thực và dự đoán (RMSE thấp hơn là tốt hơn)[***E. Root Mean Square Error The root mean square error is used to measure the***

**difference between the value of the true value and the value observed or predicted. The value is predicted by the chosen model. It is a useful method to determine whether a model is suitable or not. The lower the root mean square error value; the better perdition will occur.**]-[6]..., và R-Squared để đánh giá độ chính xác (R-Squared cao hơn là hiệu quả hơn)[6]. Kỹ thuật One Hot Encoding cũng được sử dụng để chuyển đổi các giá trị phân loại thành giá trị số mà thuật toán học máy có thể xử lý[6]...

Sau khi áp dụng và so sánh ba thuật toán[**This model of house price prediction uses 3 machine learning algorithms for the prediction that are linear, ridge and lasso machine learning algorithms. Below is the given table for the result of above 3 algorithms based which are tested on different parameters. The below table shows the scores achieved by all algorithms. Table 3. Score**

**Below are some graphs showing actual vs predicted value using different algorithms:-*G. Linear Regre***], kết quả điểm số (scores) được trình bày trong Bảng III[**S.no Model score 1. Linear regression 0.823399994270895 2. Lasso 0.8128285650772719 3. Ridge regression 0.823414663331264**

**machine learning algorithm do not works on categorical values. The working of one hot encoding is, it assigns the value one to the column which is currently in use and make the value of other to 0 due to this for each variable a different binary code is assigned.**]. Dựa trên bảng này, các mô hình dựa trên Hồi quy tuyến tính và Hồi quy Ridge hoạt động tốt hơn mô hình Lasso. Cụ thể hơn, mô hình Hồi quy Ridge cho hiệu suất tốt hơn một chút so với Hồi quy tuyến tính (Ridge R-squared là 0.823414663331264, Linear R-squared là 0.823399994270895)[***I. Ridge Regression: From the result of above table, it shows that Linear***

**regression and Ridge regression based models outperform the lasso based model .**

**Also in comparison between linear regression and ridge regression based models the ridge regression slightly give better performance in comparison of linear regression as ridge**

**Fig. 8. Predicition Lasso Regression**

**regression is capable of achieving the as R-squared of ‘0.823414663331264’ where as lasso is able to achieve the score of ‘0.823399994270895’ which makes ridge slightly better than linear regression based model.**].

Bài báo kết luận rằng họ đã thành công trong việc xây dựng một mô hình có khả năng dự đoán giá nhà với độ chính xác và hiệu quả cao hơn, hữu ích cho lĩnh vực bất động sản. Mô hình này có thể được kết hợp với giao diện web để người dùng sử dụng và có thể được cập nhật dữ liệu trong tương lai để theo kịp biến động giá[**So, after creating a model which is evaluated on different machine learning algorithms we are successfully created a model which is that is capable of predicting the prices of houses with more accuracy and efficiency which can help in real-estate.**

**After creating the model this model is combined with web development to create a user interface so that users can come to check the house prices which may help them to plan their journey accordingly.**

**The prices of houses may vary in future so the model can be updated in future by updating the dataset so that users can trust this model and are able to check prices after future changes as well.**]

# Code :[dudoan.ipynb](https://colab.research.google.com/drive/1Abi9YiRv27g_boNOsRxG3XDzuyrASvxV?usp=sharing)

Predicting Criticality of Patients Based on Disease Symptoms Using Machine Learning