Chủ đề của bài thuyết trình

Phân tích dữ liệu thị trường xe Ấn Đô & Dự đoán giá và các đặc trưng ảnh hưởng

Bố cục bài thuyết trình

1 Lý do chọn đề tài

* 1. Tăng trưởng kinh tế và dân số trẻ
  2. Số lượng nên tảng giao dịch trực tuyến tăng
  3. Ý nghĩa cỉa dự án

1. Giới thiệu về dataset
   1. PAS

- P (Target population): Tất cả xe cũ đang được rao bán tại Ấn Độ.

- A (Access frame): Xe cũ được rao bán trên trang web www.cardekho.com và www.cars24.com.

- S (Sample): Hơn 15.000 bản ghi thông tin xe cũ được chọn ngẫu nhiên từ dữ liệu trên hai trang web www.cardekho.com và www.cars24.com.

* 1. Các yếu tố ảnh hưởng đến tính đúng đắn của dữ liệu

- Time: Dữ liệu xe cũ có thể trở nên lỗi thời nhanh chóng theo thời gian.

• Giá xe thay đổi theo thời gian do nhiều yếu tố như sự biến động của thị trường, lạm phát, xuất hiện mẫu xe mới, ...

- Place: Giá xe cũ có thể khác nhau giữa các khu vực do sự chênh lệch về nhu cầu, thuế, …

• Culture:

 Sở thích về thương hiệu, kiểu dáng của mỗi vùng là khác nhau. Ví dụ, xe hơi cỡ nhỏ có thể được ưa chuộng ở các thành phố đông đúc, trong khi xe bán tải lại phổ biến ở các vùng nông thôn.

 Tôn giáo và tín ngưỡng cũng gián tiếp đến việc lựa chọn màu sắc và ngày màu, tác động đến giá xe.

• Population:

 Mật độ dân số, thu nhập bình quân: Khu vực đông dân, mức sống cao hơn mặt bằng chung thì có nhu cầu tiêu thụ xe nhiều hơn.

 Cấu trúc tuổi: Nhu cầu thay đổi theo độ tuổi. Ví dụ, người trẻ ưu chuộng các dòng xe thể thao hơn là người lớn tuổi.

• Environment:

 Điều kiện giao thông: Hệ thống đường xá, cơ sở giao thông cùng với đó là môi trường ô nhiễm ảnh hưởng đến giá xe.

• Legality:

 Chính sách về xe cũ như thuế, kiểm định, …

 Các quy định nghiêm ngặt về nguồn gốc, giấy tờ liên quan. Các quy định về môi trường như khí thải,.. cũng ảnh hưởng đến giá xe.

* 1. Dataset 1: Car price prediction Dataset
     1. Thông tin cơ bản như tác giả, số lượng mẫu, số lượng đặc trưng, nguồn gốc, định dạng
     2. Nội dung của dữ liệu
     3. Biến mục tiêu
     4. Danh sách các cột
  2. Dataset 2: User Car Prices
     1. Thông tin cơ bản như tác giả, số lượng mẫu, số lượng đặc trưng, nguồn gốc, định dạng
     2. Nội dung của dữ liệu
     3. Biến mục tiêu
     4. Danh sách các cột
  3. Dataset 3: Car price prediction dataset
     1. Thông tin cơ bản như tác giả, số lượng mẫu, số lượng đặc trưng, nguồn gốc, định dạng
     2. Nội dung của dữ liệu
     3. Biến mục tiêu
     4. Danh sách các cột

1. Tiền xử lý dữ liệu và phân tích khám phá dữ liệu
   1. Hình minh họa quá trình tiền xử lý và khám phá dữ liệu
   2. Tiền xử lí dữ liệu
   3. Phân tích và khám phá dữ liệu
      1. Biểu đồ phân phối của dữ liệu (1 bức ảnh)
      2. Biểu đồ phân bố các đặc trưng (1 bức ảnh)

A group of pie charts

AI-generated content may be incorrect.

* + 1. Biểu đồ số mẫu xe cũ theo năm sản xuất (2000–2020) (1 bức ảnh)

A graph of a number of people

AI-generated content may be incorrect.

* + 1. Biểu đồ phân bố số lượng xe theo từng loại nhiên liêu (2000-2020) (1 bức ảnh)

A graph of blue and white bars

AI-generated content may be incorrect.

* + 1. Biểu đồ phân bố số lượng xe theo từng loại hộp số (2000–2020) (1 bức ảnh)

A graph of blue and white bars

AI-generated content may be incorrect.

* + 1. Biểu đồ phân bố số lượng xe theo số đời chủ sở hữu (2000–2020) (1 bức ảnh)

A graph of a bar graph

AI-generated content may be incorrect.

* + 1. Biểu đồ thống kê số lượng xe của các hãng phổ biến theo năm sản xuất (1 bức ảnh)

A graph with different colored lines

AI-generated content may be incorrect.

* + 1. Biểu đồ phân tích giá bán theo năm (2000–2020) (1 bức ảnh)

A group of different colored bars

AI-generated content may be incorrect.

* + 1. Biểu đồ phân tích quãng đường đã đi theo năm (2000–2020) (1 bức ảnh) A group of colored bars

       AI-generated content may be incorrect.
    2. Biểu đồ phân tích mức tiêu hao nhiên liệu (km/l) theo năm (2000–2020) (1 bức ảnh) A group of colored bars

       AI-generated content may be incorrect.
    3. Biểu đồ phân tích dung tích động cơ theo năm (2000–2020) (1 bức ảnh)

A group of colored bars

AI-generated content may be incorrect.

* + 1. Biều đồ phân tích công suất tối đa theo năm (2000–2020) (1 bức ảnh)

A group of colored bars

AI-generated content may be incorrect.

* + 1. Biểu đồ phân tích số chỗ ngồi của xe theo năm (2000–2020) (1 bức ảnh)

A group of colored bars

AI-generated content may be incorrect.

* + 1. Biểu đồ so sánh giá xe trung bình giữa xe chạy bằng nguyên liệu Diesel và Petrol (2000–2020) (1 bức ảnh)

A graph of a bar graph

AI-generated content may be incorrect.

* + 1. Biểu đồ so sánh giá xe trung bình giữa xe dùng hộp số từ động và thủ công (2000–2020) (1 bức ảnh)

A graph with blue and orange bars

AI-generated content may be incorrect.

* + 1. Biểu đồ so sánh giá xe trung bình của các hãng theo năm (2000–2020) (1 bức ảnh)
    2. Biểu đồ Skewness (1 bức ảnh) A graph with different colored squares

       AI-generated content may be incorrect.
    3. Ma trận tương quan (1 bức ảnh) (1 bức ảnh)

A blue and white grid with black text

AI-generated content may be incorrect.

* + 1. Biểu đồ mức độ tương quan của các đặc trưng với biến mục tiêu (1 bức ảnh) A graph with blue squares

       AI-generated content may be incorrect.

1. Xây dựng mô hình
   1. Các bước xây dựng mô hình
   2. Danh sách các mô hình được chọn

Linear Regression

Decision Tree

Random Forest

Gradient Boosting

XGBoost

* 1. Lựa chọn mô hình

1. Đánh giá mô hình
2. Phân tích và đánh giá
   1. Mô hình nào hiệu quả nhất

Biểu đồ MSE của các mô hình

Biểu đồ phân tán thể hiện mối quan hệ giữa dự đoán và thực tế( 2 biểu đồ)

Kết luận

* 1. Mức độ ảnh hưởng của các đặc trưng

1 biểu đồ

1 đoạn giải thích lý do

1 đoạn nêu vai trò của các yếu tố khác

1. Kết luận và hướng tương lai
   1. Lợi ích
   2. Định hướng để cải thiện

8 Tài liệu tham khảo