

Tổng quan bài toán doanh nghiệp

- Bài toán: Phân tích chân dung Khách hàng;
- Phân tích chân dung Khách hàng là phương pháp phân tích thông tin chi tiết về khách hàng mục tiêu của một công ty. Phân nhóm khách hàng tiềm năng giúp doanh nghiệp điều chỉnh sản phẩm của mình dựa trên khách hàng mục tiêu từ các đối tượng khách hàng khác nhau;
- Thực hiện bài toán phân cụm khách hàng để có cái nhìn tổng quan về nhóm khách hàng có đặc điểm tương tự, từ đó có thể tối ưu hóa chiến lược tiếp thị và dịch vụ, tăng cường tương tác cá nhân hóa và nâng cao trải nghiệm của khách hàng.

Mô tả bộ dữ liệu

- Bộ dữ liệu là thông tin tổng hợp của các khách hàng trong một khoảng thời gian cố định, bao gồm thông tin cá nhân, chi tiêu và hành vi mua hàng;
- Bộ dữ liệu gồm 3069 bản ghi (bao gồm 8 trường category và 23 trường numeric), được thu thập và tổng hợp trong 2 năm.

Bộ dữ liệu chi tiết tại **đây**.

Các trường thông tin

PEOPLE

- ID: Mã số duy nhất của khách hàng;
- Year_Of_Birth: Năm sinh của khách hàng;
- Gender: Giới tính của khách hàng;
- Phone_Number: Số điện thoại của khách hàng;
- Phone: Số điện thoại của khách hàng;
- Payment_Method: Hình thức thanh toán nhiều nhất;
- Academic_Level: Trình độ học vấn của khách hàng;

Các trường thông tin

PEOPLE

- Living_With: Tình trạng hôn nhân và số con trong nhà;
- Income: Thu nhập hàng năm của khách hàng;
- Registration_Time: Thời gian khách hàng đăng ký với công ty;
- Year_Register: Năm đăng ký thẻ thành viên;
- Month_Register: Tháng đăng ký thẻ thành viên;
- Recency: Số ngày kể từ lần mua hàng cuối cùng của khách hàng;
- Complain: 1 nếu khách hàng phàn nàn trong vòng 2 năm qua, 0 nếu không.

Các trường thông tin

PRODUCTS

- Liquor: Số tiền đã chi tiêu cho rượu trong 2 năm qua;
- Vegetables: Số tiền đã chi tiêu cho rau củ trong 2 năm qua;
- Pork: Số tiền đã chi tiêu cho thịt lợn trong 2 năm qua;
- Seafood: Số tiền đã chi tiêu cho cá trong 2 năm qua;
- Candy: Số tiền đã chi tiêu cho đồ ngọt trong 2 năm qua;
- Jewellery: Số tiền đã chi tiêu cho trang sức trong 2 năm qua.

Các trường thông tin

PROMOTION

- Num_Deals_Purchases: Số lượng giao dịch mua hàng với giảm giá;
- Promo_10: 1 nếu khách hàng chấp nhận ưu đãi trong chiến dịch thứ nhất, 0 nếu không;
- Promo_20: 1 nếu khách hàng chấp nhận ưu đãi trong chiến dịch thứ hai, 0 nếu không;

Các trường thông tin

PROMOTION

- Promo_30: 1 nếu khách hàng chấp nhận ưu đãi trong chiến dịch thứ ba, 0 nếu không;
- Promo_40: 1 nếu khách hàng chấp nhận ưu đãi trong chiến dịch thứ tư, 0 nếu không;
- Promo_50: 1 nếu khách hàng chấp nhận ưu đãi trong chiến dịch thứ năm, 0 nếu không.

Các trường thông tin

PLACE

- Num_Web_Purchases: Số lượng giao dịch mua hàng thông qua trang website;
- Num_Catalog_Purchases: Số lượng giao dịch mua hàng sử dụng danh mục trên website;
- Num_Store_Purchases: Số lượng giao dịch mua hàng trực tiếp tại cửa hàng;
- Num_Web_Visits_Month: Số lần truy cập vào trang website trong tháng qua;
- Total_Purchase: Tổng số lượng giao dịch mua hàng của công ty.



3

Thể lệ

Trình bày và nộp bài Vòng 2.1: Case Study

Quy định nộp bài

- Thời gian nộp bài: **Trước 21h00 ngày 26/05/2024;**
- Leader đội thi tham gia nộp bài qua Google Form do Ban Tổ chức cung cấp;
- Các sản phẩm yêu cầu:
 - 01 file Notebook;
 - 01 bản Báo cáo Phân tích (tối đa 05 trang).
- Tổng hợp vào một folder và chỉnh chế độ "**Công khai**", người truy cập có quyền "**Viewer**";
- Cách thức đặt tên folder, tên sản phẩm:
[TÊN ĐỘI THI]_[TÊN SẢN PHẨM]_VÒNG 2.1: CASE STUDY

Quy định trình bày file notebook

- Định dạng: .ipynb;
- Code bằng ngôn ngữ Python;
- Có giải thích code bằng tiếng Anh (giải thích dưới dạng comment);
- File chạy lại được, không lỗi;
- Toàn bộ các biểu đồ phân tích insights trong Báo cáo Phân tích phải xuất hiện ở trong Notebook.

Quy định trình bày báo cáo phân tích

HÌNH THỨC TRÌNH BÀY

- Định dạng: .pdf;
- Không quá 05 trang A4;
- Trình bày dưới dạng Báo cáo, khổ A4 đứng;
- Font chữ: Time News Roman;
- Size: 12;
- Các biểu đồ sẽ được trình bày trong phạm vi $\frac{1}{3}$ trang, bên trái báo cáo;
- Các luận điểm báo cáo sẽ được trình bày trong $\frac{2}{3}$ phạm vi còn lại của trang;
- Mẫu Báo cáo Phân tích được đính kèm trong file sau:

[DAZONE2024] Mẫu Báo cáo Phân tích.pdf

Quy định trình bày báo cáo phân tích

NỘI DUNG

Xử lý Dữ liệu:

- Data Cleaning: xác định và sửa lỗi cho bộ dữ liệu, như dữ liệu khuyết thiếu, ngoại lai, hoặc dữ liệu trùng lặp;
- Data Transformation: biến đổi dữ liệu thành các định dạng phù hợp để có thể phân tích và visualize.

Khám phá và Giải thích Dữ liệu:

- EDA: tìm ra các xu hướng, mẫu thử hoặc kiểm tra các giả định trong dữ liệu;
- Data Visualization: trực quan hoá dữ liệu, trình bày insight từ bộ dữ liệu.

Quy định trình bày báo cáo phân tích

NỘI DUNG

Đề xuất giải pháp kỹ thuật (01 trang):

- Feature Selection & Feature Engineering;
- Thuật toán;
- Phương pháp thử nghiệm.

Lưu ý: BTC khuyến khích đội thi có thể mở rộng các mục nội dung nhỏ phía trên. Tuy nhiên, đội thi cần lưu ý Vòng 2.1 sẽ tập trung vào **Preprocess và Insights**.

Quy định trình bày báo cáo phân tích

TRÍCH DẪN

Đối với các thông tin bên ngoài đề thi, các đội thi phải trích nguồn của các dữ liệu bên ngoài.

- Trích dẫn tài liệu tham khảo theo hệ thống **Harvard (Harvard Referencing System)**;
- Các trích dẫn được đính kèm trong phần **Appendix**.



4

Tiêu chí

Chấm điểm Vòng 2.1: Case Study

TIÊU CHÍ ĐÁNH GIÁ NỘI DUNG BÀI LÀM CỦA THÍ SINH

Phần	Tỷ trọng	Lưu ý
Tổng quan Nghiệp vụ	10%	Đối với các bài làm trình bày không đúng quy định của BTC, đội thi sẽ bị trừ 10% số điểm phần hình thức trình bày trong barem chấm điểm.
Xử lý Dữ liệu	20%	
Khám phá và Giải thích Dữ liệu	30%	
Đề xuất Giải pháp kỹ thuật	30%	
Hình thức trình bày	10%	