

PHÂN LOẠI KHÁCH HÀNG BẰNG RFM VÀ KMEANS

Nhằm phân loại khách hàng, ta sẽ dùng phương pháp RFM và K-Means. RFM là phương pháp phân loại bằng các chỉ số Recency (khoảng thời gian từ lần giao dịch gần nhất, tính bằng ngày), Frequency (tổng số lần giao dịch) và Monetary (tổng giá trị giao dịch). K-Means là thuật toán dùng để nhóm các khách hàng thành k cụm. Và những khách hàng trong cùng một cụm sẽ có các đặc điểm tương tự nhau. Kết hợp hai phương pháp RFM và K-Means, ta sẽ phân loại khách hàng theo các cụm (loại) dựa trên các đặc điểm về Recency, Frequency và Monetary.

Tóm tắt model:

1. Tính các chỉ số R-F-M của từng khách hàng
2. Vẽ biểu đồ phân phối của R-F-M và kiểm tra dữ liệu có chuẩn chưa (Normal Distribution), nếu chuẩn thì sang bước 4
3. Tiến hành chuẩn hóa dữ liệu bằng các hàm log, sqrt, boxcox
4. Scale dữ liệu đã được chuẩn hoá
5. Áp dụng dữ liệu đã được scale vào mô hình K-Mean, tìm k bằng phương pháp elbow
6. Xác định cụm thích hợp của từng khách hàng
7. Tìm giá trị trung bình của R-F-M trong từng cụm và xác định loại khách hàng.

Kết quả sau bước 7:

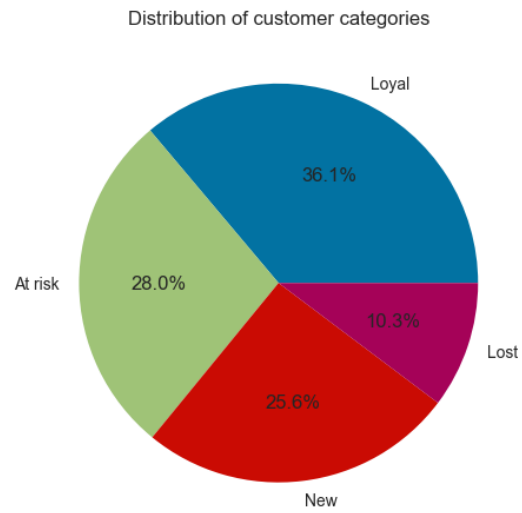
Bảng: Giá trị trung bình của R-F-M trong từng cụm sau khi áp dụng model

Cluster	Recency	Frequency	Monetary
0	7.35	118.38	563.03
1	0.88	86.13	379.34
2	12.42	48.80	210.63
3	1.43	514.00	2437.29

- + Cluster 0: R khá cao → Khá lâu rồi chưa mua hàng, F và M cũng khá cao → Mua sắm nhiều
⇒ At risk customer
- + Cluster 1: R rất thấp → Mua hàng gần đây, F và M lại không cao → Mua sắm ít
⇒ New customer
- + Cluster 2: R rất cao → Lâu rồi họ chưa mua hàng, F và M rất thấp → Mua sắm rất ít
⇒ Lost customer
- + Cluster 3: R thấp → Mua hàng khá gần đây, F và M rất cao → Mua sắm rất nhiều
⇒ Loyal customer

Bảng: Tóm tắt hành vi của từng tệp khách hàng

Nhóm	Hành vi
At risk customer	Đã từng mua sắm với tần suất trung bình, tuy nhiên gần đây không hoạt động
New customer	Mới bắt đầu mua hàng
Lost customer	Khách hàng đã không hoạt động, lượng mua sắm và tần suất trước đây cũng ít
Loyal customer	Nhóm khách hàng trung thành, mua sắm nhiều và đang hoạt động gần đây



Hình: Biểu đồ tỉ trọng của tệp khách hàng