

Hệ hỗ trợ quyết định phân khúc khách hàng sử dụng thẻ tín dụng

Sinh viên: Lê Thị Vân Anh
Giảng viên hướng dẫn: TS. Lê Chí Ngọc

Hà Nội, 6/2023

Nội dung

- 1 Điều tra và khảo sát
- 2 Đặc tả yêu cầu hệ thống
- 3 Dữ liệu
- 4 Phân tích thiết kế hệ thống
- 5 Thuật toán phân cụm
- 6 Demo

Nội dung

- 1 Điều tra và khảo sát
- 2 Đặc tả yêu cầu hệ thống
- 3 Dữ liệu
- 4 Phân tích thiết kế hệ thống
- 5 Thuật toán phân cụm
- 6 Demo

Điều tra và khảo sát

- Ngày nay, việc sử dụng thẻ tín dụng đã trở thành một phần không thể thiếu trong cuộc sống hàng ngày của nhiều người.
- Các ngân hàng có dữ liệu lịch sử về hành vi sử dụng thẻ tín dụng nhưng chưa thực hiện phân tích dữ liệu để khai thác giá trị của chúng.

Vì vậy các nhà quản lý cần có hệ hỗ trợ quyết định phân khúc khách hàng để đưa ra chiến lược tiếp thị phù hợp trong việc hiểu và phục vụ khách hàng một cách tốt nhất.

Nội dung

- 1 Điều tra và khảo sát
- 2 Đặc tả yêu cầu hệ thống**
- 3 Dữ liệu
- 4 Phân tích thiết kế hệ thống
- 5 Thuật toán phân cụm
- 6 Demo

Yêu cầu chức năng

- Phân cụm khách hàng.
- Đăng ký, đăng nhập.
- Hiển thị kết quả phân cụm.

Yêu cầu phi chức năng

- Chạy trên nền tảng web để truy cập mọi lúc mọi nơi.
- Có thể phục vụ nhiều người truy cập cùng lúc.
- Hệ thống có giao diện đơn giản, dễ sử dụng.

Nội dung

- 1 Điều tra và khảo sát
- 2 Đặc tả yêu cầu hệ thống
- 3 Dữ liệu**
- 4 Phân tích thiết kế hệ thống
- 5 Thuật toán phân cụm
- 6 Demo

Dữ liệu

Bộ dữ liệu về hành vi sử dụng của khoảng 9000 chủ thẻ tín dụng đang hoạt động trong 6 tháng, 18 thuộc tính.

Tổng quan

	CUST_ID	BALANCE	BALANCE_FREQUENCY	PURCHASES	ONEOFF_PURCHASES	INSTALLMENTS_PURCHASES	CASH_ADVANCE	PURCHASES_FRE
0	C10001	40.9007	0.8182	95.4	0	95.4	0	
1	C10002	3,202.4674	0.9091	0	0	0	6,442.9455	
2	C10003	2,495.1489	1	773.17	773.17	0	0	
3	C10004	1,666.6705	0.6364	1,499	1,499	0	205.788	
4	C10005	817.7143	1	16	16	0	0	
5	C10006	1,809.8288	1	1,333.28	0	1,333.28	0	
6	C10007	627.2608	1	7,091.01	6,402.63	688.38	0	
7	C10008	1,823.6527	1	436.2	0	436.2	0	
8	C10009	1,014.9265	1	861.49	661.49	200	0	
9	C10010	152.226	0.5455	1,281.6	1,281.6	0	0	

Số lượng khách hàng: (8950, 17)

Hình 1: Dữ liệu

Các trường dữ liệu

- CUSTID: ID của khách hàng.
- AccountID: ID tài khoản thẻ.
- BALANCE: Số dư còn lại trong tài khoản để thực hiện mua hàng.
- BALANCE_FREQUENCY: Tần suất cập nhật số dư, điểm số từ 0 đến 1 (1 = cập nhật thường xuyên, 0 = không cập nhật thường xuyên)

Các trường dữ liệu

- PURCHASES: Số tiền mua hàng từ tài khoản.
- ONEOFF_PURCHASES: Số tiền mua hàng lớn nhất trong một lần mua.
- INSTALLMENTS_PURCHASES: Số tiền mua hàng được trả góp
- CASH_ADVANCE: Số tiền tạm ứng được cung cấp bởi người dùng

Các trường dữ liệu

- PURCHASES_FREQUENCY: Tần suất mua hàng
- ONEOFF_PURCHASES_FREQUENCY: Tần suất mua hàng lớn trong một lần mua
- PURCHASES_INSTALLMENTS_FREQUENCY: Tần suất mua hàng trả góp

Các thuộc tính tần số này sẽ được đánh giá theo thang điểm từ 0 đến 1 (1 = mua hàng thường xuyên, 0 = không mua hàng thường xuyên)

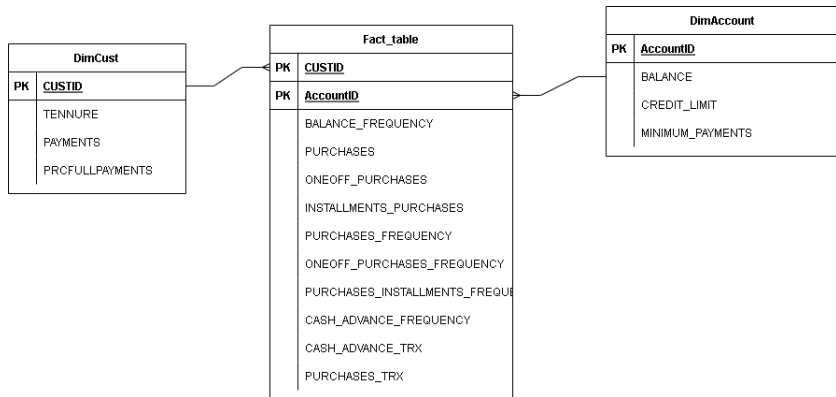
Các trường dữ liệu

- CASH_ADVANCE_FREQUENCY: Tần suất thanh toán tiền tạm ứng
- CASH_ADVANCE_TRX: Số giao dịch được thực hiện bằng "Tiền tạm ứng"
- PURCHASES_TRX: Số giao dịch mua hàng được thực hiện
- CREDIT_LIMIT: Hạn mức thẻ tín dụng cho người dùng

Các trường dữ liệu

- PAYMENTS: Số tiền thanh toán của người dùng
- MINIMUM_PAYMENTS: Số tiền thanh toán tối thiểu của người dùng
- PRCFULLPAYMENT: Phần trăm thanh toán đầy đủ của người dùng
- TENURE: Thời gian sử dụng thẻ tín dụng cho người dùng

Cơ sở dữ liệu quan hệ

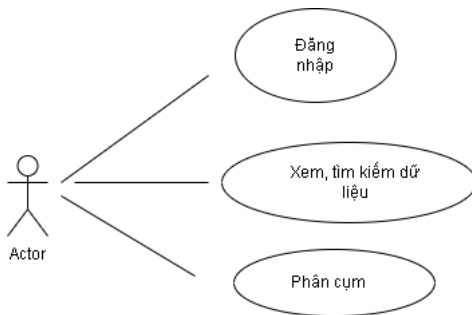


Hình 2: Cơ sở dữ liệu quan hệ

Nội dung

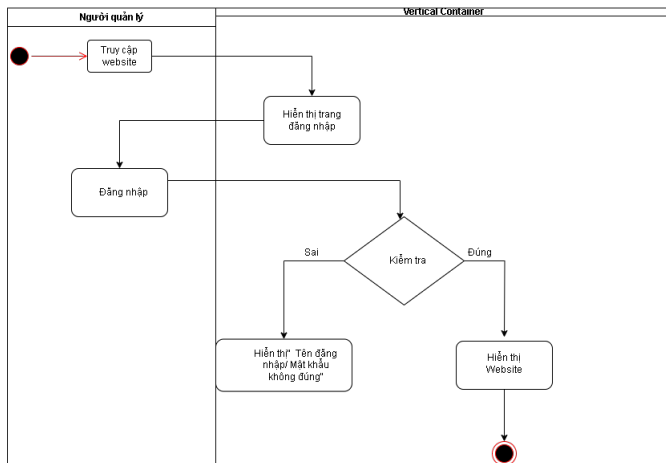
- 1 Điều tra và khảo sát
- 2 Đặc tả yêu cầu hệ thống
- 3 Dữ liệu
- 4 Phân tích thiết kế hệ thống**
- 5 Thuật toán phân cụm
- 6 Demo

Biểu đồ usecase



Hình 3: Biểu đồ usecase

Quá trình đăng nhập



Hình 4: Quá trình đăng nhập

Nội dung

- 1 Điều tra và khảo sát
- 2 Đặc tả yêu cầu hệ thống
- 3 Dữ liệu
- 4 Phân tích thiết kế hệ thống
- 5 Thuật toán phân cụm**
- 6 Demo

Thuật toán giảm chiều dữ liệu PCA

1 Tìm vector kỳ vọng của toàn bộ dữ liệu

$$\bar{X} = \frac{1}{N} \sum_{n=1}^N x_n$$

2 Trừ mỗi điểm dữ liệu đi vector kỳ vọng của toàn bộ dữ liệu

$$\hat{x}^n = x_n - \bar{X}$$

Thuật toán giảm chiều dữ liệu PCA

3 Tìm ma trận hiệp phương sai:

$$S = \frac{1}{N} \hat{X} \hat{X}^T$$

4 Tính các trị riêng và vector riêng ma trận S, sắp xếp chúng theo thứ tự giảm dần của trị riêng.

Thuật toán giảm chiều dữ liệu PCA

- 5 Chọn K vector riêng ứng với K trị riêng lớn nhất để xây dựng ma trận U_k có các cột tạo thành một hệ trực giao.
- 6 Chiếu dữ liệu ban đầu đã chuẩn hóa xuống không gian con tìm được.
- 7 Dữ liệu mới chính là tọa độ của các điểm dữ liệu trong không gian mới.

$$Z = U_k^T X$$

Thuật toán phân cụm K-Means

- 1 Chọn số lượng cụm K, K điểm làm trung tâm.
- 2 Tính khoảng cách các điểm còn lại với các điểm trung tâm và gán điểm đó với điểm trung tâm gần nhất.
- 3 Tính toán lại điểm trung tâm.
- 4 Lặp lại bước 2 và 3 cho tới khi số lần lặp đạt ngưỡng hoặc không có sự thay đổi trong việc gán dữ liệu.

Nội dung

- 1 Điều tra và khảo sát
- 2 Đặc tả yêu cầu hệ thống
- 3 Dữ liệu
- 4 Phân tích thiết kế hệ thống
- 5 Thuật toán phân cụm
- 6 Demo

Giao diện

Đăng nhập

Tên đăng nhập

admin

Mật khẩu

...



Đăng nhập

Hình 5: Giao diện đăng nhập

Giao diện

Dữ liệu

Tổng quan Đồ thị

Upload file CSV



Drag and drop file here
Limit 200MB per file • CSV

Browse files

Hình 6: Upload data

Giao diện

Tìm kiếm

Nhập ID khách hàng:

	CUST_ID	BALANCE	BALANCE_FREQUENCY	PURCHASES	ONEOFF_PURCHASES	INSTALLMENTS_PURCHASES	CASH_ADVANCE	PURCHASES
20	C10021	2,016.6847	1	176.68	0	176.68	0	

Hình 7: Tìm kiếm thông tin

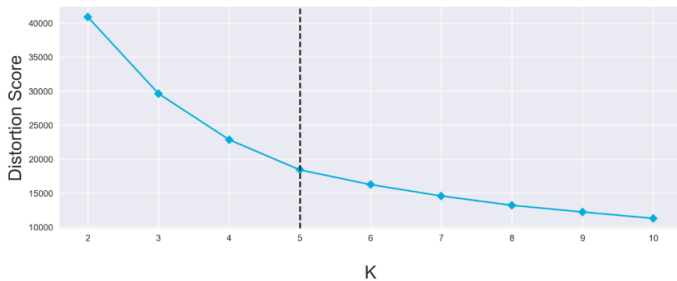
Lựa chọn số thành phần chính

Explained variance

	n_component	Explained variance
0	1	56.7063
1	2	76.775
2	3	88.8358
3	4	95.2861
4	5	99.3284
5	6	99.9998
6	7	100

Hình 8: Explained variance

Elbow



Hình 9: Biểu đồ Elbow

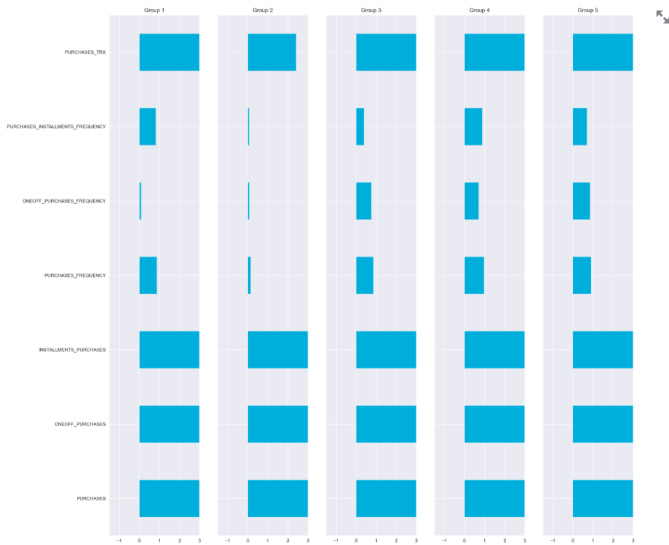
Giao diện

Số lượng khách hàng từng cụm



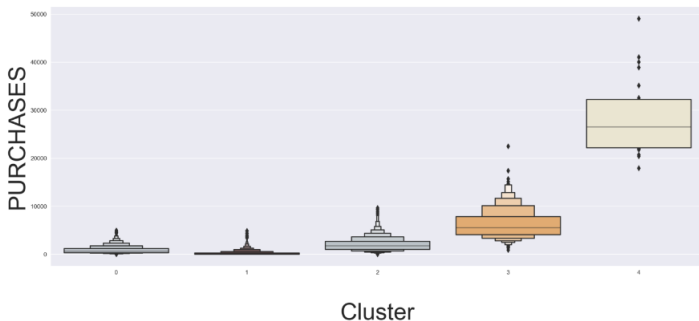
Hình 10: Biểu đồ số lượng mỗi cụm khách hàng

Giao diện



Giao diện

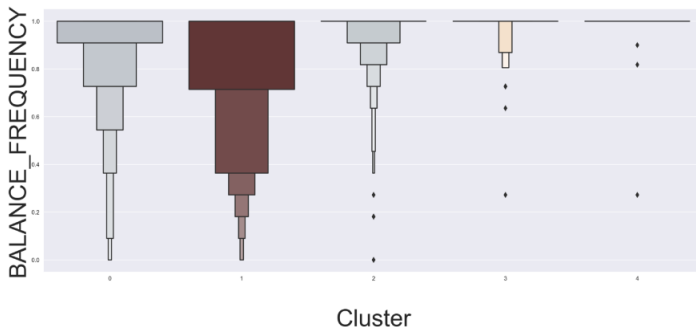
Đặc điểm tổng số tiền khách hàng đã chi tiêu cho các giao dịch mua hàng



Hình 12: Biểu đồ đặc điểm tổng số tiền chi tiêu

Giao diện

Đặc điểm tần suất nạp tiền vào thẻ của khách hàng



Hình 13: Biểu đồ đặc điểm tần suất nạp tiền vào thẻ

Kết luận

Từ hệ hỗ trợ quyết định, nhà quản lý có thể:

- Hiểu rõ hành vi mua sắm và sử dụng dịch vụ của khách hàng.
- Xác định nhóm khách hàng tiềm năng.
- Tối ưu hóa chiến lược tiếp thị và phân phối.
- Tăng cường sự tương tác và gắn kết khách hàng.

Xin chân thành cảm ơn thầy cô và các bạn đã chú ý lắng nghe.