



Trường ĐH Khoa Học Tự Nhiên Tp. Hồ Chí Minh
TRUNG TÂM TIN HỌC

Đồ án tốt nghiệp Data Science

Topic: **Customer Segmentation**

Phòng LT & Mạng

https://csc.edu.vn/data-science-machine-learning/Data-Science-Capstone-Project-Hinh-thuc-2_225

2024

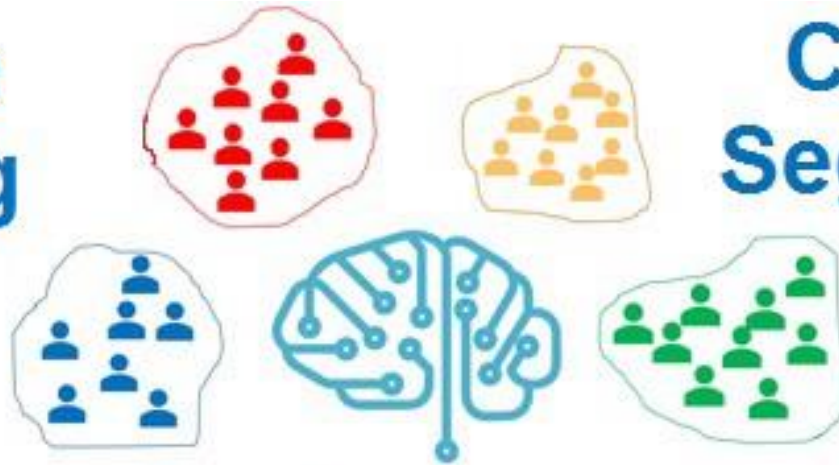


1. Customer Segmentation sử dụng RFM
2. Ví dụ
3. Giới thiệu project
4. Triển khai project theo Data Science Process

Customer Segmentation sử dụng RFM

□ Customer Segmentation

Machine
Learning



Customer
Segmentation



- “Customer Segmentation” là gì?
 - Phân khúc/nhóm/cụm khách hàng (market segmentation – còn được gọi là phân khúc thị trường) là quá trình nhóm các khách hàng lại với nhau dựa trên các đặc điểm chung. Nó phân chia và nhóm khách hàng thành các nhóm nhỏ theo đặc điểm địa lý, nhân khẩu học, tâm lý học, hành vi (geographic, demographic, psychographic, behavioral) và các đặc điểm khác.
 - Các nhà tiếp thị sử dụng kỹ thuật này để nhắm mục tiêu khách hàng thông qua việc cá nhân hóa, khi họ muốn tung ra các chiến dịch quảng cáo, truyền thông, thiết kế một ưu đãi hoặc khuyến mãi mới, và cũng để bán hàng.

Triển khai project theo Data Science Process



- Tại sao cần “Customer Segmentation”

“Selling to people who actually want to hear from you is more effective than interrupting strangers who don’t.”



Seth Godin -American author and former dot com business executive



Triển khai project theo Data Science Process



- Để xây dựng các chiến dịch tiếp thị tốt hơn
- Giữ chân nhiều khách hàng hơn: ví dụ với những khách hàng mua hàng nhiều nhất của công ty sẽ tạo ra các chính sách riêng cho họ hoặc thu hút lại những người đã mua hàng trong một khoảng thời gian.
- Cải tiến dịch vụ: hiểu rõ khách hàng cho phép bạn điều chỉnh và tối ưu hóa các dịch vụ của mình để đáp ứng tốt hơn nhu cầu và mong đợi của khách hàng => giúp cải thiện sự hài lòng của khách hàng.
- Tăng khả năng mở rộng: giúp doanh nghiệp có thể hiểu rõ hơn về những điều mà khách hàng có thể quan tâm => thúc đẩy mở rộng các sản phẩm và dịch vụ mới phù hợp với đối tượng mục tiêu của họ.

Triển khai project theo Data Science Process

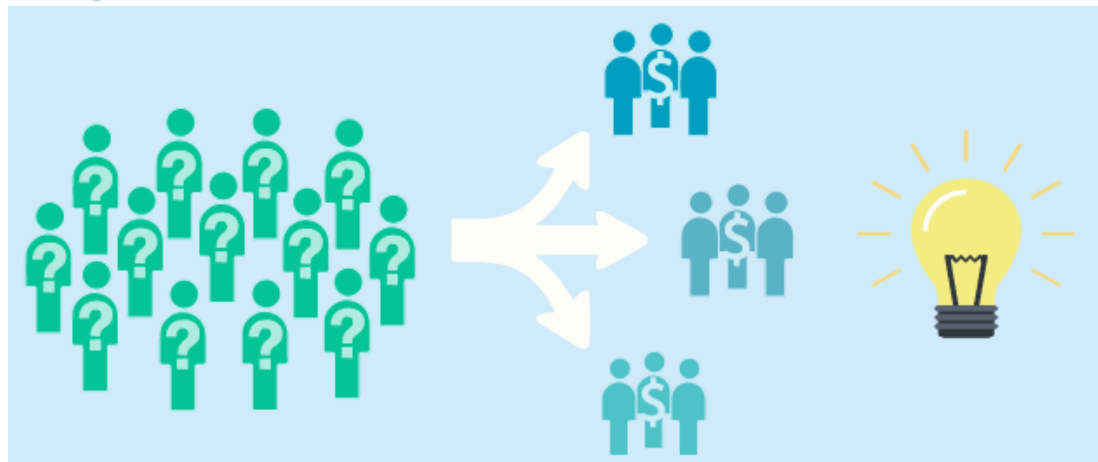


- Tối ưu hóa giá: Việc có thể xác định tình trạng xã hội và tài chính của khách hàng giúp doanh nghiệp dễ dàng xác định giá cả phù hợp cho sản phẩm hoặc dịch vụ của mình mà khách hàng sẽ cho là hợp lý.
- Tăng doanh thu: Dành ít thời gian, nguồn lực và nỗ lực tiếp thị cho các phân khúc khách hàng có lợi nhuận thấp hơn, và ngược lại. => hầu hết các phân khúc khách hàng thành công cuối cùng đều dẫn đến tăng doanh thu và lợi nhuận cũng như giảm chi phí bán hàng.

Triển khai project theo Data Science Process



- “Customer Segmentation” tốt cần:
 - Dễ dàng và rõ ràng nhận dạng
 - Đo lường được
 - Đủ lớn và có giá trị để có lãi
 - Có thể truy cập bởi các kênh quảng bá, truyền thông và phân phối



❑ Customer Segmentation sử dụng RFM

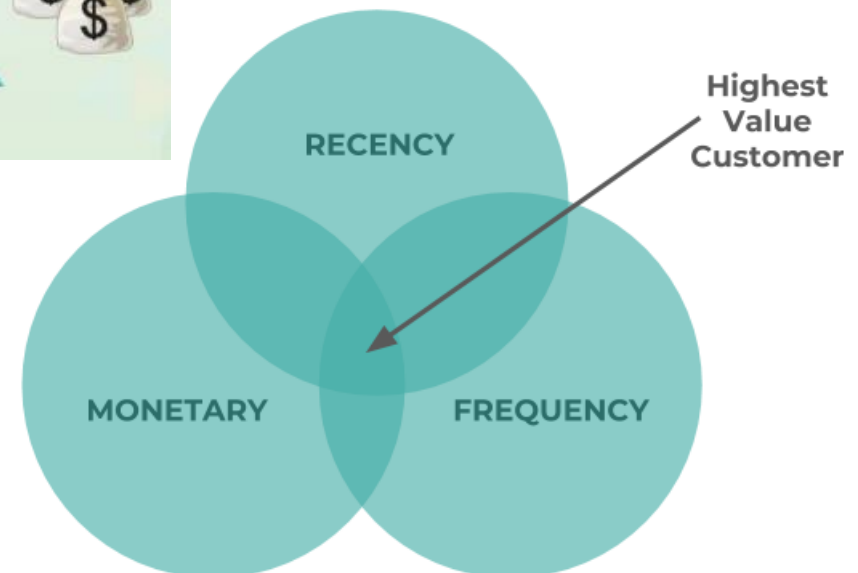
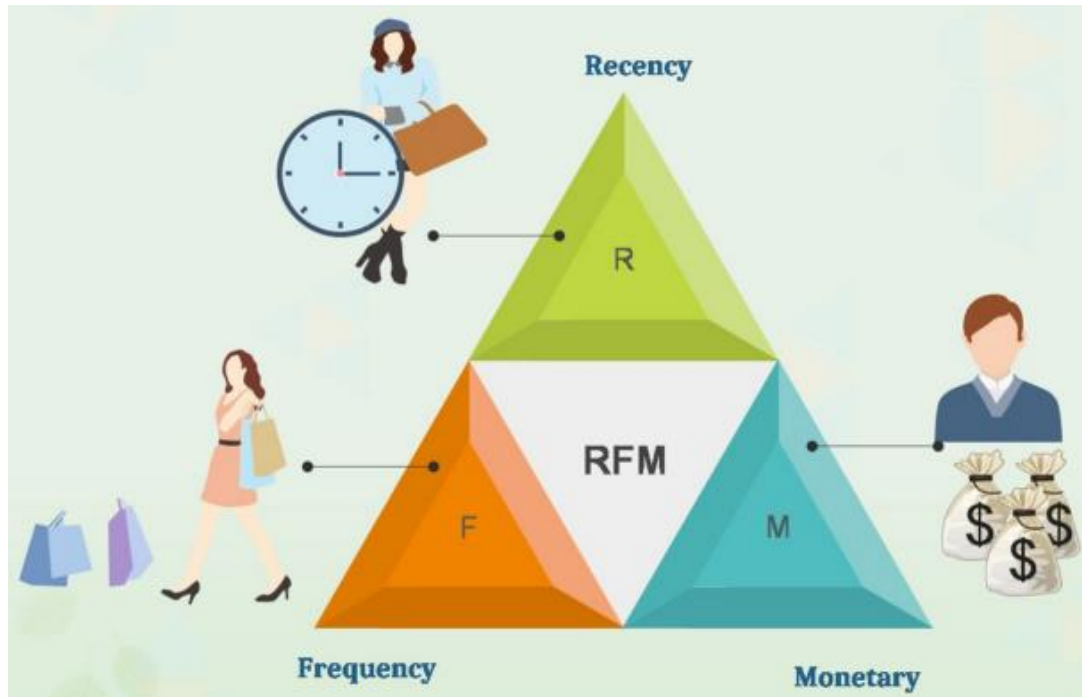
- Phân tích RFM (Recency, Frequency, Monetary) là phương pháp tiếp cận dựa trên hành vi của khách hàng thành để nhóm thành các phân khúc. RFM phân nhóm khách hàng trên cơ sở các giao dịch mua hàng trước đó của họ, nhằm mục đích phục vụ khách hàng tốt hơn.
- RFM giúp người quản lý xác định được khách hàng tiềm năng để kinh doanh có lợi nhuận hơn. Ngoài ra, nó giúp các nhà quản lý thực hiện các chiến dịch quảng cáo hiệu quả cho dịch vụ được cá nhân hóa.

Customer Segmentation sử dụng RFM



- RFM nghiên cứu hành vi của khách hàng và phân nhóm dựa trên ba yếu tố đo lường:
 - Recency (R): đo lường số ngày kể từ lần mua hàng cuối cùng (lần truy cập gần đây nhất) đến ngày giả định chung để tính toán (ví dụ: ngày hiện tại, hoặc ngày max trong danh sách giao dịch).
 - Frequency (F): đo lường số lượng giao dịch (tổng số lần mua hàng) được thực hiện trong thời gian nghiên cứu.
 - Monetary Value (M): đo lường số tiền mà mỗi khách hàng đã chi tiêu trong thời gian nghiên cứu.

Customer Segmentation sử dụng RFM



Customer Segmentation sử dụng RFM



- Về mặt lý thuyết, chúng ta sẽ có các segment :
 - Low Value: Những khách hàng ít hoạt động hơn những khách hàng khác, không phải là khách thường xuyên và tạo ra doanh thu rất thấp, (doanh thu có khi bằng không / có thể âm).
 - Mid Value: Giá trị trung bình ở các tiêu chí. Nhóm khách hàng đông, thường xuyên, tạo ra doanh thu vừa phải.
 - High Value: Nhóm khách hàng đem lại doanh thu cao, tần suất mua hàng thường xuyên. Nhóm khách hàng doanh nghiệp muốn “giữ chân” nhiều nhất.

Customer Segmentation sử dụng RFM



- Về cách triển khai, chúng ta cần tính các giá trị R, F, M => sau đó áp dụng Unsupervised Machine learning (Kmeans, Hierarchical Clustering, GMM...) để xác định các nhóm/cụm khác nhau (cluster).
- Với mỗi (R, F, M) sẽ bao gồm bốn nhóm, tạo ra 64 (4x4x4) phân khúc khách hàng khác nhau; được thực hiện theo Quantile (tứ phân vị) cho từng R, F, M:
 - Tính R, F, M cho từng khách hàng.
 - Tính giá trị bin (theo quantile) và đưa vào RFM
 - Sắp xếp RFM score của khách hàng tăng dần

1. Customer Segmentation sử dụng RFM
2. Ví dụ
3. Giới thiệu project
4. Triển khai project theo Data Science Process

❑ Thực hiện việc phân cụm khách hàng áp dụng RFM và Clustering Algorithm

Demo: e-commerce

1. Customer Segmentation sử dụng RFM
2. Ví dụ
3. Giới thiệu project
4. Triển khai project theo Data Science Process

❑ Business Objective/Problem

- Công ty X chủ yếu bán các sản phẩm là quà tặng dành cho những dịp đặc biệt. Nhiều khách hàng của công ty là khách hàng bán buôn.
- Công ty X mong muốn có thể bán được nhiều sản phẩm hơn cũng như giới thiệu sản phẩm đến đúng đối tượng khách hàng, chăm sóc và làm hài lòng khách hàng.

❑ Các kiến thức/ kỹ năng cần để giải quyết vấn đề này:

- Hiểu vấn đề
- Import các thư viện cần thiết và hiểu cách sử dụng
- Đọc dữ liệu (dữ liệu project này được cung cấp)
- Thực hiện EDA cơ bản
- Tiền xử lý dữ liệu: làm sạch, tạo tính năng mới, lựa chọn tính năng cần thiết...

Giới thiệu project



- Trực quan hóa dữ liệu
- Lựa chọn thuật toán cho bài toán phân cụm
- Xây dựng model
- Đánh giá model
- Báo cáo kết quả

1. Customer Segmentation sử dụng RFM
2. Ví dụ
3. Giới thiệu project
4. Triển khai project theo Data Science Process

Triển khai project theo Data Science Process



- Thư viện sử dụng

- numpy, pandas, matplotlib, seaborn, plotly
- pandas_profiling / dataprep
- scipy
- squarify
- scikit-learn (sklearn) – LDS6
- pyspark (LDS9)
- ...

□ Triển khai dự án

● Bước 1: Business Understanding

■ Dựa vào mô tả nói trên => xác định vấn đề:

- Tìm ra giải pháp giúp cải thiện hiệu quả quảng bá, từ đó giúp tăng doanh thu bán hàng, cải thiện mức độ hài lòng của khách hàng.

=> Mục tiêu/ vấn đề: Xây dựng hệ thống phân cụm khách hàng dựa trên các thông tin do công ty cung cấp từ đó có thể giúp công ty xác định các nhóm khách hàng khác nhau để có chiến lược kinh doanh, chăm sóc khách hàng phù hợp.

Triển khai project theo Data Science Process



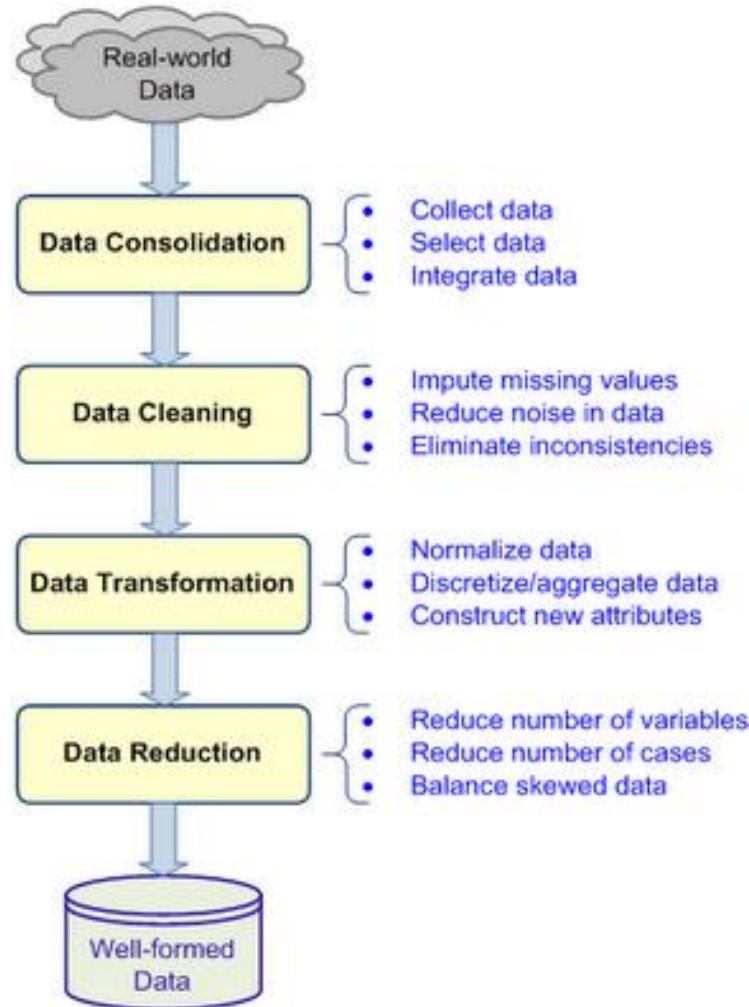
- Bước 2: Data Understanding/ Acquire

- Từ mục tiêu/ vấn đề đã xác định: Xem xét các dữ liệu cần thiết:
 - Toàn bộ dữ liệu được lưu trữ trong tập tin OnlineRetail.csv với 541.909 record chứa tất cả các giao dịch xảy ra từ ngày 01/12/2010 đến 09/12/2011 đối với bán lẻ trực tuyến.
 - Mô tả dữ liệu:
<https://archive.ics.uci.edu/ml/datasets/online+retail>

Triển khai project theo Data Science Process



● Bước 3: Data preparation/ Prepare



Triển khai project theo Data Science Process



- Bước 4&5: Modeling & Evaluation/ Analyze & Report
 - Xây dựng giải pháp phân cụm khách hàng theo RFM.
 - Xây dựng model phân cụm khách hàng theo RFM phối hợp với thuật toán phân cụm:
 - RFM + Kmeans (LDS6)
 - RFM + Hierarchical Clustering (LDS6)
 - RFM + Kmeans (LDS9)
 - Các đề xuất khác (+++)
 - Thực hiện so sánh/đánh giá các kết quả

Triển khai project theo Data Science Process



- Bước 6: Deployment & Feedback/ Act
 - Đưa ra chiến dịch quảng cáo, bán hàng, chăm sóc khách hàng phù hợp cho mỗi nhóm.

Triển khai project theo Data Science Process



❑ Các công việc cần thực hiện

- Hãy triển khai project trên với các bước theo Data Science Process
- Tiền xử lý dữ liệu, tìm các insight liên quan đến khách hàng của công ty X
- Áp dụng RFM & các thuật toán phân cụm
 - Xây dựng phân cụm khách hàng theo RFM thủ công
 - Xây dựng model phân cụm khách hàng theo RFM phối hợp với thuật toán phân cụm:
 - RFM + KMeans (LDS6)
 - RFM + Hierarchical Clustering (LDS6)
 - RFM + KMeans (LDS9)
 - Các đề xuất khác (+++)
- Đánh giá và report các kết quả



