ĐẠI HỌC ĐÀ NẴNG
TRƯỜNG ĐẠI HỌC KINH TẾ
KHOA THƯƠNG MẠI ĐIỆN TỬ

ỨNG DỤNG LUẬT KẾT HỢP VÀO BÀI TOÁN SẮP XẾP GIAN HÀNG

MÔN: ĐỀ ÁN THỰC HÀNH 1

Giảng viên hướng dẫn : Trần Nhật Pháp

Nhóm: 2

Sinh viên thực hiện:

Phan Thị Thảo Ngân – 46K29.2 (Nhóm trưởng)

Lê Thị Hồng Ny – 46K29.2

Phạm Thanh Lâm – 46K29.2

Đà Nẵng, ngày 17 tháng 11 năm 2022

LỜI CẢM ƠN

Để hoàn thành báo cáo này, chúng em xin gửi lời cảm ơn chân thành đến:
Ban giám hiệu trường Đại Học Kinh Tế - Đại Học Đà Nẵng vì đã tạo điều kiện về cơ sở vật chất với hệ thống thư viện hiện đại, đa dạng các loại sách, tài liệu thuận lợi cho việc tìm kiếm, nghiên cứu thông tin.
Xin cảm ơn Quý Thầy Cô khoa Thương Mại Điện Tử đã tạo điều kiện cho chúng em tham gia học tập, nghiên cứu bộ môn Đề án thực hành 1.
Xin cảm ơn quý Thầy hướng dẫn Trần Nhật Pháp đã hướng dẫn tận tình, chi tiết để em có đủ kiến thức và vận dụng chúng vào bài báo cáo này.
Do chưa có nhiều kinh nghiệm làm để tài cũng như những hạn chế về kiến thức, trong bài báo cáo của chúng em chắc chắn sẽ không tránh khỏi những thiếu sót.
Rất mong nhận được sự nhận xét, ý kiến đóng góp, phê bình từ phía Thầy để bài báo cáo của chúng em được hoàn thiện hơn.
Lời cuối cùng, nhóm xin kính chúc quý Thầy nhiều sức khỏe và hạnh phúc.

MỤC LỤC

MỞ ĐẦU

Với sự phát triển của công nghệ thông tin thì khối lượng dữ liệu lưu trữ ngày càng lớn, và giữa những lượng dữ liệu khổng lồ đó lại ẩn chứa một số thông tin được coi là chìa khóa dẫn đến thành công của mọi lĩnh vực từ hoạt động sản xuất đến kinh doanh. Việc khai thác, chiếc lọc thông tin ứng dụng vào cuộc sống của con người không chỉ dừng lại là một kĩ thuật đơn thuần, nó đòi hỏi sự ra đời của ngành khoa học mới: khoa học về phát hiện tri thức và khai phá dữ liệu.

I. Tính cấp thiết của đề tài:

Khai phá dữ liệu (Data Mining) là một lĩnh vực khoa học mới xuất hiện gần đây nhằm đáp ứng nhu cầu khai phá và phát hiện ra những thông tin , tri thức có ích đang bị che giấu trong 'núi' dữ liệu, giúp công việc của các nhà quản lý, các chuyên gia từ đó thúc đẩy khả năng sản xuất, kinh doanh và cạnh tranh với các tổ chức doanh nghiệp khác. Kết quả nghiên cứu cùng với những ứng dụng thành công trong khai phá dữ liệu cho thấy đây là một lĩnh vực rất tiềm năng, mang lại nhiều lợi ích đồng thời có ưu thế hơn hẳn so với các công cụ phân tích dữ liệu truyền thống.

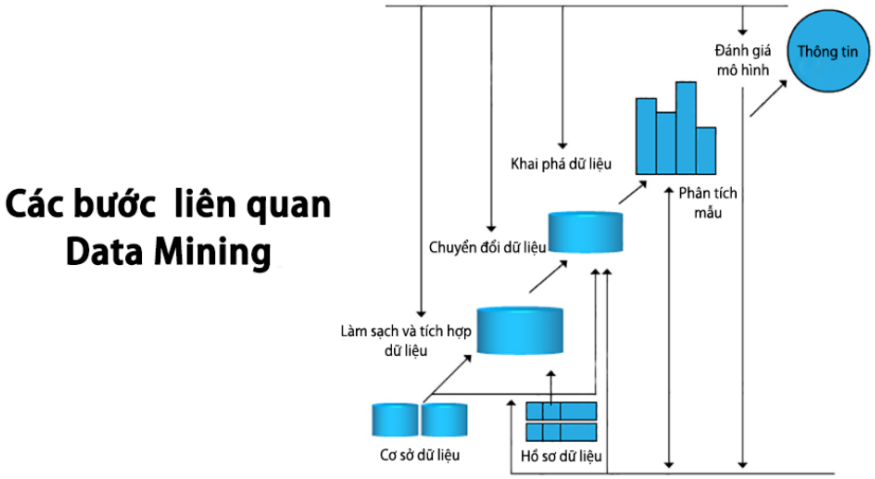
Khai phá dữ liệu được coi là quá trình trích xuất các thông tin có giá trị tiềm ẩn bên trong lượng lớn dữ liệu được lưu trữ trong các CSDL, kho dữ liệu,…Hiện

nay ngoài thuật ngữ khai phá dữ liệu, người ta còn dùng một số thuật ngữ khác có ý nghĩa tương tự như: Khám phá tri thức từ cơ sở dữ liệu (Knowledge Discovery in Database), trích lọc dữ liệu (knowledge extraction), khảo cổ dữ liệu (data archaeology), nạo vét dữ liệu(data dredging).

1. Quá trình khám phá tri thức trong CSDL:

Mục đích của quá trình khai phá tri thức là rút ra tri thức từ dữ liệu trong CSDL lớn. Quá trình gồm nhiều giai đoạn và lặp lại, mà trong đó sự lặp lại có thể xuất hiện ở bất cứ bước nào.

Quá trình có thể được mô tả như sau :



Bước 1: Hình thành và xác định bài toán. Mục đích của bước này là tìm hiểu lĩnh vự ứng dụng từ đó hình thành bài toán, xác định các nhiệm vụ cần phải hoàn thành. Bước này sẽ quyết định cho việc rút ra được các tri thức hữu ích và cho phép chọn các phương pháp khai phá dữ liệu thích hợp với mục đích tương ứng và bản chất của dữ liệu.

Bước 2: Thu thập và tiền xử lý dữ liệu. nhằm loại bỏ nhiễu (làm sạch dữ liệu), xử lý việc thiếu dữ liệu (làm giàu dữ liệu), biến đổi dữ liệu và rút gọn dữ liệu nếu cần thiết, bước này thường chiếm nhiều thời gian nhất trong toàn bộ quy trình phát hiện tri thức. Do dữ liệu được lấy từ nhiều nguồn khác nhau, không

đồng nhất,…có thể gây ra các nhầm lẫn. Sau bước này, dữ liệu sẽ nhất quán, đầy đủ, được rút gọn và rời rạc hóa.

Bước 3: Khai phá dữ liệu, rút ra các tri thức. Là khai thác dữ liệu hay nói cách khác là trích ra các mẫu hoặc mô hình ẩn dưới dữ liệu. Giai đoạn này rất quan trọng, bao gồm các công đoạn như: chức năng, nhiệm vụ và mục đích của khai phá dữ liệu, dùng phương pháp khai phá nào? Thông thường các bài toán khai phá dữ liệu bao gồm: các bài toán mang tính mô tả - đưa ra tích chất chung của dữ liệu, các bài toán dự báo,… Tùy theo bài toán cụ thể mà ta lựa ra các phương pháp khai phá phù hợp.

Bước 4: Sử dụng các tri thức phát hiện được. Là hiểu và làm sáng tỏ các mô tả và dự đoán.

2. Các kĩ thuật khai phá dữ liệu:
    a. *Mô tả khái niệm* (concept description): thiên về mô tả, tổng hợp và tóm tắt khái niệm. **Ví dụ:** tóm tắt văn bản.
    b. *Luật kết hợp* (association rules): là dạng luật biểu diễn tri thứ ở dạng khá đơn giản. **Ví dụ:** "60 % nam giới vào siêu thị nếu mua bia thì có tới 80% trong số họ sẽ mua thêm thịt bò khô". Luật kết hợp được ứng dụng nhiều trong lĩnh vực kính doanh, y học, tin-sinh, tài chính & thị trường chứng khoán, .v.v.
    c. *Phân lớp và dự đoán* (classification & prediction): xếp một đối tượng vào một trong những lớp đã biết trước. **Ví dụ:** phân lớp vùng địa lý theo dữ liệu thời tiết. Hướng tiếp cận này thường sử dụng một số kỹ thuật của machine learning như cây quyết định (decision tree), mạng nơ ron nhân tạo (neural network), .v.v. Người ta còn gọi phân lớp là học có giám sát (học có thầy).
    d. *Phân cụm (clustering):* xếp các đối tượng theo từng cụm (số lượng cũng như tên của cụm chưa được biết trước. Người ta còn gọi phân cụm là học không giám sát (học không thầy).

e. *Khai phá chuỗi* (sequential/temporal patterns): tương tự như khai phá luật kết hợp nhưng có thêm tính thứ tự và tính thời gian. Hướng tiếp cận này được ứng dụng nhiều trong lĩnh vực tài chính và thị trường chứng khoán vì nó có tính dự báo cáo.

3. Ứng dụng của khai phá dữ liệu:

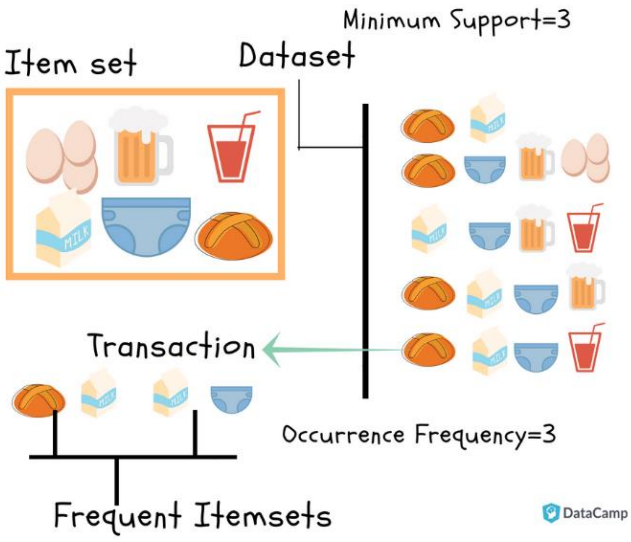
Khai phá dữ liệu là một lĩnh vực được quan tâm và ứng dụng rộng rãi. Một số ứng dụng điển hình trong khai phá dữ liệu có thể liệt kê:

a. Phân tích dữ liệu và hỗ trợ ra quyết định.
b. Điều trị y học
c. Phát hiện văn bản
d. Tin sinh học
e. Tài chính và thị trường chứng khoán.
f. …

II. Xác định bài toán:

Hiện nay, khai phá dữ liệu trở nên khá phổ biến trong lĩnh vực bán lẻ và là phương pháp phân tích hiệu quả cho phát hiện thông tin hữu ích và chưa biết trong dữ liệu bán lẻ. Việc sắp xếp tổ chức hàng hoá và các hoạt động kinh doanh có liên quan nhằm nâng cao sự hài lòng của khách hàng là một trong những công việc rất quan trọng. Nghiên cứu này sẽ tập trung phân tích, khai phá và tìm ra luật kết hợp dựa trên dữ liệu của quá khứ, từ đó đề xuất một số kiến nghị để hỗ trợ cho hoạt động kinh doanh.

Khi một người đi mua sắm, thường sẽ nghĩ trong đầu 1 vài sản phẩm dự định mua. Và giỏ hàng của mỗi người sẽ khác nhau tùy vào sở thích và nhu cầu sử dụng.



Hẳn ai ban đầu cũng sẽ phải ngạc nhiên khi biết siêu thị Walmart bày trí những quầy bia và quầy bỉm cạnh nhau. Nghe chừng đây có vẻ là ý tưởng hết sức quái đản. Tuy nhiên chiến thuật này lại thực sự thành công và ảnh hưởng mạnh tới doanh số. Bạn hãy tưởng tượng một tình huống thú vị là một ông bố vào siêu thị mua đồ bỉm sữa cho con và rồi quay sang bên cạnh lại đụng món khoái khẩu và trường hợp ngược lại. WalMart tìm được quy luật này dựa trên nghiên cứu dữ liệu hành vi mua hàng của họ **"Có 60% đàn ông đi siêu thị vào tối thứ 6, nếu họ mua bỉm cho trẻ em thì sẽ mua cả bia."** Từ đó ta có thể nhận thấy rằng chúng ta hoàn toàn có thể tìm ra được mối quan hệ tương quan, kết hợp giữa các đối tượng để tạo ra giá trị cao hơn.

Một tiệm bánh ở Edinburgh muốn tăng số lượng sản phẩm bán của họ bằng cách sắp xếp lại các sản phẩm có khả năng được mua cùng nhau nhiều nhất. Khi một khách hàng A mua bánh mì thì họ có muốn mua thêm trà hay bánh ngọt nữa hay không? Và một khách hàng sẽ không thể biết hết những sản phẩm mà cửa hàng bán và cũng rất tốn thời gian để họ đi hết 1 vòng cửa hàng để xem qua tất cả các loại sản phẩm hiện có Chính vì vậy mà sắp xếp lại gian hàng cũng như lên các chiến lược quảng cáo các sản phẩm mới của cửa hàng là việc hoàn toàn cấp thiết .

Đứng ở góc độ là một nhà quản lý cửa hàng, nhân viên báng hàng sẽ quan tâm đến một khách hàng là nam / nữ , trẻ / lớn tuổi ,… đến cửa hàng sẽ mua những gì? Tiếp đến sau khi đã có thứ mình cần, họ có mua thêm sản phẩm khác hay không nếu có thì họ đã có sẵn list những gì cần mua hay chưa, hoặc thể nhận thấy cần thiết mua thêm tại thời điểm mua hàng? Và việc phát sinh nhu cầu mới, sản phẩm mới có liên kết gì so với nhu cầu ban đầu khi bước vào cửa hàng? Bên cạnh đó khi nghiên cứu dữ liệu bán hàng còn giúp cho nhân viên có sự chuẩn bị tốt về nguyên vật liệu cũng như số lượng bán phù hợp tránh tình trạng thiếu sản phẩm này hay là dư thừa quá nhiều các sản phẩm không được ưu chuộng.

III. Phân tích giỏ hàng

1. Phân tích giỏ hàng là gì?

Phân tích giỏ hàng (Market Basket Analysis) là một kỹ thuật phân tích hành vi mua dựa trên lịch sử giao dịch của họ. Kỹ thuật này thường được sử dụng để nắm được thị hiếu, thói quen tiêu dùng của khách hàng. Từ đó đưa ra những chiến lược Marketing một cách hợp lý.

2. Lợi ích của phân tích giỏ hàng

- Thấu hiểu hành vi khách hàng

Việc hiểu hành vi giúp cho các nhà quản trị đưa ra chiến lược marketing hiệu quả. Phân tích giỏ hàng hữu ích trong việc sắp đặt vị trí, định giá hợp lý cho các mặt hàng trên kệ tại cửa hàng. Nó còn mang lại lợi ích cho hoạt động khác như tối ưu hóa hàng tồn kho cho từng cửa hàng hoặc nhà kho

- Tăng thị phần, tạo ra lợi thế cạnh tranh

Một khi doanh nghiệp đạt mức tăng trưởng tối đa, việc xác định các chiến lược phát triển mới để tăng thị phần sẽ trở nên khó khăn. Phân tích giỏ hàng có thể được sử dụng để tập hợp dữ liệu nhân khẩu học.

- Tối ưu hóa hiệu quả của hoạt động bán hàng

Phân tích giỏ hàng là cơ sở để chỉ ra sản phẩm nào tiềm năng trong các dòng sản phẩm mà doanh nghiệp đang kinh doanh.

- Tăng trưởng doanh thu

Các nền tảng website thương mại điện tử như Amazon, Taobao, Shopee, Tiki,… Hay là website bán hàng của một tổ chức được hưởng lợi từ phân tích giỏ hàng. Họ sẽ đề xuất các sản phẩm khách hàng xem thường xuyên hoặc có liên quan đến sản phẩm đang xem.

Việc này giúp thu hút và giữ chân người dùng ở lại lâu hơn trên website. Đồng thời mở ra nhiều cơ hội bán hàng, tăng doanh thu một cách hiệu quả.

IV. Giới thiệu kĩ thuật:

ASSOCIATION ANALYSIS

1. Luật kết hợp:

Được giới thiệu từ năm 1993, bài toán khai phá luật kết hợp nhận được rất nhiều sự quan tâm của nhiều nhà khoa học. Ngày nay việc khai thác các luật như thế

vẫn là một trong những phương pháp khai thác mẫu phổ biến nhất trong việc khai phá tri thức và khai phá dữ liệu. Luật kết hợp là mối quan hệ giữa các tập thuộc tính trong cơ sở dữ liệu; là phương tiện hữu ích để khám phá các mối liên kết trong dữ liệu. Một luật kết hợp là một mệnh đề kéo theo có dạng **X → Y**, trong đó X, Y ⊆ I, thỏa mãn điều kiện X giao Y = rỗng. Tập X gọi là nguyên nhân [Tiền đề], tập Y gọi là hệ quả.

Có 3 độ đo quan trọng đối với luật kết hợp:

- **Độ hỗ trợ (support):** Đề cập đến mức độ phổ biến mặc định của một mặt hàng.

$$\textbf{Support (X)} = \frac{P(X \cup Y)}{N}$$

- **Độ tin cậy (confidence)** (Nếu vế trái xảy ra thì có bao nhiêu % vế phải xảy ra theo).

$$\textbf{Confidence (X → Y) = P(X|Y)} = \frac{P(X \cup Y)}{P(X)}$$

- **Lift:** đề cập đến sự gia tăng tỷ lệ bán B khi A được bán. Là tỉ số giữa confidence và tỉ lệ trường hợp hiện diện X.

$$\textbf{Lift (X → Y)} = \frac{confidence\,(X \rightarrow Y)}{P(Y)} = \frac{P(X \cup Y)}{P(X) * P(Y)}$$

→ Hai tiêu chí rất quan trọng trong việc đánh giá luật kết hợp đó là độ hỗ trợ (support) và độ tin cậy (confidence).

2. Một số định nghĩa và kí hiệu:

a. Tập mục:

- Gọi I = { $x_1, x_2, .., x_n$ } là tập có n mục (**item**) . Một tập X ⊆ I được gọi là một tập mục (**itemset**).

- Nếu X có k mục (tức |X| = k) thì X được gọi là **k–itemset.**
- **Ví dụ:** Tập tất cả các mặt hàng thực phẩm trong siêu thị: I = {sữa, trứng, đường, bánh mỳ, mật ong, mứt, bơ, thịt bò, giá, . . . }.

b. Giao dịch:

- Ký hiệu D = $\{T_1, T_2, \ldots, T_m\}$ là cơ sở dữ liệu gồm m giao dịch (**transaction**). Mỗi giao dịch Ti ∈ D là một tập mục, tức Ti ⊆ I.
- Tập mục I là các sản phẩm trong siệu thị, Cơ sở giao dịch là những đơn mua của khách hàng.

  T1 = {sữa, trứng, đường, bánh mỳ}

  T2 = {sữa, mật ong, mứt, bơ}

  T3 = {trứng, mì tôm, thịt bò, giá}

c. Tập phổ biến:

- Cho tập mục X (⊆ I)
- Độ hỗ trợ của X (sup(X, D)): là số lượng giao dịch trong D chứa tập X:

$$\sup(X, D) = |\{T | T \subseteq D \text{ và } X \subseteq T\}$$

- Độ hỗ trợ tương đối của X (rsup(X, D)) : là số phần trăm các giao dịch trong D chứa X:

$$\text{rsup}(X, D) = \sup(X, D)/|D|$$

- Độ hỗ trợ tối thiểu kí hiệu là minsup là một giá trị cho trước bởi người sử dụng được xác định trước khi sinh ra luật. Nếu tập mục X có ***sup(X)*** ≥ ***minsup*** thì ta nói X **là tập mục phổ biến**. Một tập mục phổ biến được sử dụng như một tập đáng quan tâm trong các thuật toán, ngược lại, những tập không phải tập phổ biến là những tập không đáng quan tâm.

Ví dụ: Các tập phổ biến (với minsup = 3) từ cơ sở dữ liệu D (tức số lần xuất hiện của tập trong 6 giao dịch ≥ 3) D = {T1, T2, T3, T4, T5, T6} trong đó:

1. T1 = {A, B, D, E}
2. T2 = {B, C, E}
3. T3 = {A, B, D, E}
4. T4 = {A, B, C, E}
5. T5 = {A, B, C, D, E}
6. T6 = {B, C, D}

   → Ta có tập các tập phổ biến là:

   F = {A, B, C, D, E, AB, AD, AE, BC, BD, BE, CE, DE, ABD, ABE, ADE, BCE, BDE, ABDE }

   a. F(1) = {A, B, C, D, E} - k độ dài

   b. F(2) = {AB, AD, AE, BC, BD, BE, CE, DE}

   c. F(3) = {ABE, ABD, ADE, BCE, BDE}

   d. F(4) = {ABDE}

Tuy nhiên, không phải bất cứ luật kết hợp nào có mặt trong tập các luật có thể được sinh ra cũng đều có ý nghĩa trên thực tế. Mà các luật đều phải thỏa mãn một ngưỡng hỗ trợ và tin cậy cụ thể. Với một tập các giao dịch D, bài toán phát hiện luật kết hợp là sinh ra tất cả các luật kết hợp mà có độ tin cậy conf lớn hơn độ tin cậy tối thiểu minconf và độ hỗ trợ sup lớn hơn độ hỗ trợ tối thiểu minsup tương ứng do người dùng xác định . Khai phá luật kết hợp được phân thành hai bài toán con:

a. Tìm tất cả các tập mục mà có độ hỗ trợ lớn hơn độ hỗ trợ tối thiểu. Các tập mục thõa mãn được gọi là các tập mục phổ biến.

b. Dùng các tập mục phổ biến để sinh ra các luật mong muốn.

3. Một số hướng tiếp cận trong luật khai phá hết hợp:

Có một số hướng chính sau đây:

a. Luật kết hợp nhị phân (Binary association rule): là hướng nghiên cứu đầu tiên của luật kết hợp. Theo dạng luật kết hợp này thì các items chỉ được quan tâm là có hay không xuất hiện trong cơ sở dữ liệu giao tác (Transaction database) chứ không quan tâm về mức độ hay tần xuất xuất hiện. Thuật toán tiêu biểu nhất của khai phá dạng luật này là thuật toán Apriori.

b. Luật kết hợp có thuộc tính số và thuộc tính hạng mục (Quantitative and categorial association rule): các cơ sở dữ liệu thực tế thường có các thuộc tính đa dạng (như nhị phân, số, mục (categorial),...) chứ không nhất quán ở một dạng nào cả. Vì vậy để khai phá luật kết hợp với các cơ sở dữ liệu này các nhà nghiên cứu đề xuất một số phương pháp rời rạc hóa nhằm chuyển dạng luật này về dạng nhị phân để có thể áp dụng các thuật toán đã có.

c. Luật kết hợp tiếp cận theo hướng tập thô (mining association rule base on rough set): tìm kiếm luật kết hợp dựa trên lí thuyết tập thô.

d. Luật kết hợp nhiều mức (multi-level association ruls): với cách tiếp cận luật kết hợp thế này sẽ tìm kiếm thêm những luật có dạng: mua máy tính PC → mua hệ điều hành Window AND mua phần mềm văn phòng Microsoft Office,…

e. Luật kết hợp mờ (fuzzy association rule): Với những khó khăn gặp phải khi rời rạc hóa các thuộc tính số, các nhà nghiên cứu đề xuất luật kết hợp mờ khắc phục hạn chế đó và chuyển luật kết hợp về một dạng gần gũi hơn.

f. Luật kết hợp với thuộc tính được đánh trọng số (association rules with weighted items): Các thuộc tính trong cơ sở dữ liệu thường không có vai trò như nhau. Có một số thuộc tính quan trọng và được chú trọng hơn các thuộc

tính khác. Vì vậy trong quá trình tìm kiếm luật các thuộc tính được đánh trọng số theo mức độ xác định nào đó.

g. Khai thác luật kết hợp song song (parallel mining of association rule): Nhu cầu song song hóa và xử lý phân tán là cần thiết vì kích thước dữ liệu ngày càng lớn nên đòi hỏi tốc độ xử lý phải được đảm bảo.Trên đây là những biến thể của khai phá luật kết hợp cho phép ta tìm kiếm luật kết hợp một cách linh hoạt trong những cơ sở dữ liệu lớn. Bên cạnh đó các nhà nghiên cứu còn chú trọng đề xuất các thuật toán nhằm tăng tốc quá trình tìm kiếm luật kết hợp trong cơ sở dữ liệu.

4. Phương pháp khai phá luật kết hợp:

Thuật toán Apriori được công bố bởi R. Agrawal và R. Srikant vào năm 1994 vì để tìm các tập phổ biến trong một bộ dữ liệu lớn. Tên của thuật toán là Apriori vì nó sử dụng kiến thức đã có từ trước (prior) về các thuộc tính, vật phẩm thường xuyên xuất hiện trong cơ sở dữ liệu. Để cải thiện hiệu quả của việc lọc các mục thường xuyên theo cấp độ, một thuộc tính quan trọng được sử dụng gọi là thuộc tính Apriori giúp giảm phạm vi tìm kiếm của thuật toán.

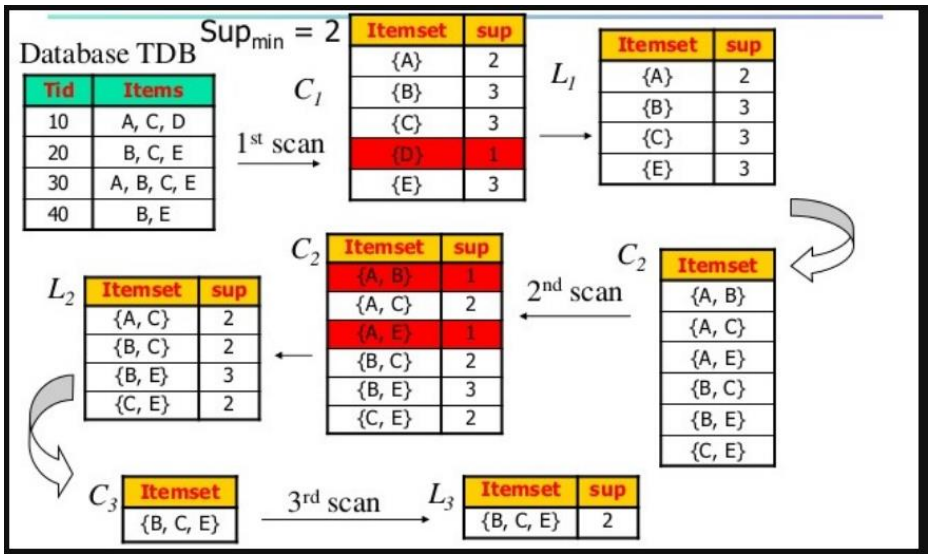
Tất cả các tập hợp con không rỗng của tập thường xuyên cũng phải thường xuyên. Khái niệm chủ chốt này của thuật toán Apriori nhằm chống lại đơn điệu của phương pháp tính theo độ hỗ trợ (surport). Apriori cho rằng:

Tất cả các tập con của một tập hợp thường xuyên phải là thường xuyên (thuộc tính Apriori).Trong một vật phẩm không thường xuyên, tất cả các tập cha của nó sẽ không thường xuyên. Hãy xem xét các tập dữ liệu sau đây và chúng ta sẽ tìm thấy các tập thường xuyên và tạo quy tắc kết hợp cho chúng.

❖ **Mục đích của thuật toán Apriori:**

Thuật toán giúp tìm ra mối quan hệ giữa các đối tượng trong khối lượng lớn dữ liệu. Việc thuật toán Apriori có thể làm là nhìn vào quá khứ và khẳng định rằng nếu một việc gì đó xảy ra thì sẽ có tỉ lệ bao nhiêu phần trăm sự việc tiếp theo sẽ xảy ra. Nó giống như nhìn vào quá khứ để dự đoán tương lai và việc này rất có ích cho các nhà kinh doanh.

❖ **Trình tự thực hiện của thuật toán:**



1. Duyệt (Scan) toàn bộ transaction database để có được support S của 1-itemset, so sánh S với min_sup, để có được 1-itemset (L_1)

2. Sử dụng L_{k-1} nối (join) L_{k-1} để sinh ra candidate k-itemset. Loại bỏ các itemsets không phải là frequent itemsets thu được k-itemset

3. Scan transaction database để có được support của mỗi candidate k-itemset, so sánh S với min_sup để thu được frequent k –itemset (L_k)

4. Lặp lại từ bước 2 cho đến khi Candidate set (C) trống (không tìm thấy frequent itemsets)

5. Với mỗi frequent itemset I, sinh tất cả các tập con s không rỗng của I

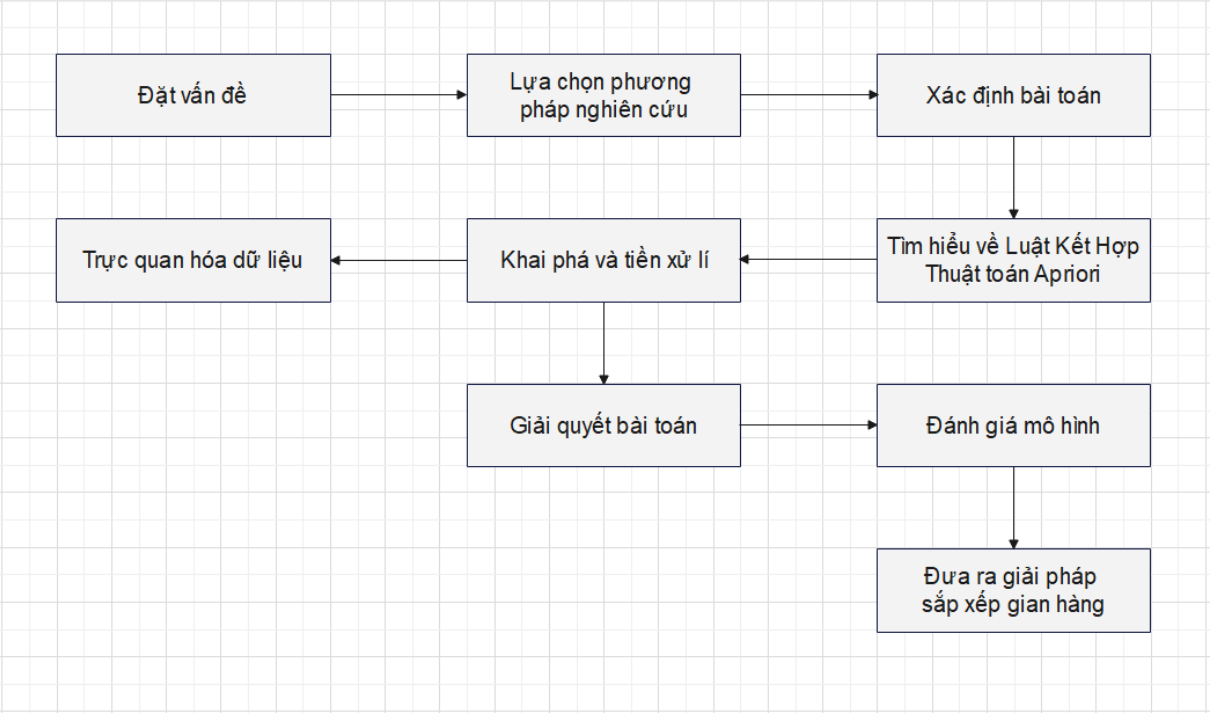
6. Với mỗi tập con s không rỗng của I, sinh ra các luật **s => (I-s)** nếu độ tin cậy (Confidence) của nó > =min_conf

Thuật toán Apriori có thể chậm. Hạn chế chính là thời gian cần thiết để chứa một số lượng lớn các tập ứng viên với nhiều tập phổ biến, hỗ trợ tối thiểu thấp hoặc tập phổ biến, tức là nó không phải là một cách tiếp cận hiệu quả cho số lượng lớn tập dữ liệu. Ví dụ: nếu có 10^4 từ tập phổ biến 1 mục, nó cần tạo hơn 10^7 ứng cử viên thành 2 độ dài, đến lượt chúng sẽ được kiểm tra và tích lũy. Hơn nữa, để phát hiện mẫu phổ biến có kích thước 100 tức là v_1, v_2… ,v_{100}, nó phải tạo ra 2^{100} tập mục ứng viên dẫn đến tốn kém và lãng phí thời gian cho việc tạo ứng viên. Vì vậy, nó sẽ kiểm tra nhiều bộ từ các tập ứng viên, nó cũng sẽ quét cơ sở dữ liệu nhiều lần để tìm các tập mục ứng viên. Apriori sẽ rất thấp và kém hiệu quả khi dung lượng bộ nhớ hạn chế với số lượng giao dịch lớn.

Ưu điểm của thuật toán Apriori: Nó được sử dụng để tính toán các tập phổ biến lớn. Đơn giản để hiểu và áp dụng.

V. Ứng dụng vào bài toán phân tích dữ liệu bán hàng:

1. Xây dựng framework:



2. Giới thiệu dữ liệu :

https://www.kaggle.com/code/akashdeepkuila/market-basket-analysis/data?select=bread+basket.csv

a. Dữ liệu ở 1 tiệm bánh ở Edinburgh, Scotland.
b. Thời gian 30/10/2016 - 9/4/2017.
c. Dữ liệu gồm 20507 dòng và 5 cột.

- **Mô tả dữ liệu gốc:**

Thứ tự	Tên cột	Kiểu dữ liệu	Giải thích
1	**Transaction**	Int	Mã định danh duy nhất cho mọi giao dịch đơn lẻ.
2	**Item**	Object	Các mặt hàng đã mua.
3	**DateTime**	Object	Ngày và giờ của các giao dịch.
4	**Period_day**	Object	Phân loại thời gian bán trong ngày.
5	**DayType**	Object	Phân loại giao dịch đã được thực hiện vào cuối tuần hay các ngày trong tuần.

d. Không chứa dữ liệu null.

e. Gồm 94 loại sản phẩm.

f. Khai phá dữ liệu: Từ cột data_time khai phá thêm các cột như là : 'YEAR' , 'MONTH', 'DAY', 'HOUR' .

	Transaction	Item	date_time	period_day	weekday_weekend	Year	Month	Week Day	Hour
0	1	Bread	30-10-2016 09:58	morning	weekend	2016	October	Sunday	9
1	2	Scandinavian	30-10-2016 10:05	morning	weekend	2016	October	Sunday	10
2	2	Scandinavian	30-10-2016 10:05	morning	weekend	2016	October	Sunday	10
3	3	Hot chocolate	30-10-2016 10:07	morning	weekend	2016	October	Sunday	10
4	3	Jam	30-10-2016 10:07	morning	weekend	2016	October	Sunday	10
5	3	Cookies	30-10-2016 10:07	morning	weekend	2016	October	Sunday	10
6	4	Muffin	30-10-2016 10:08	morning	weekend	2016	October	Sunday	10
7	5	Coffee	30-10-2016 10:13	morning	weekend	2016	October	Sunday	10
8	5	Pastry	30-10-2016 10:13	morning	weekend	2016	October	Sunday	10
9	5	Bread	30-10-2016 10:13	morning	weekend	2016	October	Sunday	10

❖ Year: năm bán được in trong hóa đơn.

❖ Month: Tháng bán được in trong hóa đơn.

❖ WeekDay: Ngày trong tuần

❖ Hour: Khung giờ bán .

- Ví dụ giờ bán là 9h40 thì vẫn được tính ở khung 9h

3. Trực quan hóa dữ liệu:

Công cụ sử dụng: Tableau

Link public:
https://public.tableau.com/app/profile/phan.th.th.o.ng.n/viz/New_bakery_/Dashboard1?publish=yes

4. Xây dựng luật:

Bước đầu tiên để tạo các Luật kết hợp là xác định **ngưỡng hỗ trợ (Min Support)** và **ngưỡng độ tin cậy (Min Confidence)**. Nếu đặt các giá trị này quá thấp thì thuật toán sẽ chạy mất thời gian hơn và đồng thời tạo ra nhiều luật kết hợp không có hoặc có ít giá trị.

Với minsup = 0.01, minconf = 0.4 : Sử dụng thuật toán Apriori sinh ra 61 tập mục phổ biến và 42 luật kết hợp các sản phẩm với nhau. Cụ thể:

Các tập phổ biến có 1 item:

	support	itemsets	length
6	0.478394	(Coffee)	1
2	0.327205	(Bread)	1
26	0.142631	(Tea)	1
4	0.103856	(Cake)	1
19	0.086107	(Pastry)	1
21	0.071844	(Sandwich)	1
16	0.061807	(Medialuna)	1
12	0.058320	(Hot chocolate)	1
8	0.054411	(Cookies)	1
3	0.040042	(Brownie)	1
9	0.039197	(Farm House)	1
15	0.038563	(Juice)	1
18	0.038457	(Muffin)	1
0	0.036344	(Alfajores)	1
23	0.034548	(Scone)	1
24	0.034443	(Soup)	1
28	0.033597	(Toast)	1
22	0.029054	(Scandinavian)	1
29	0.020285	(Truffles)	1
7	0.019440	(Coke)	1
25	0.018172	(Spanish Brunch)	1
1	0.016059	(Baguette)	1
27	0.015425	(Tiffin)	1
13	0.015003	(Jam)	1
10	0.015003	(Fudge)	1
17	0.014157	(Mineral water)	1
14	0.013207	(Jammie Dodgers)	1
5	0.012995	(Chicken Stew)	1
11	0.010565	(Hearty & Seasonal)	1
20	0.010460	(Salad)	1

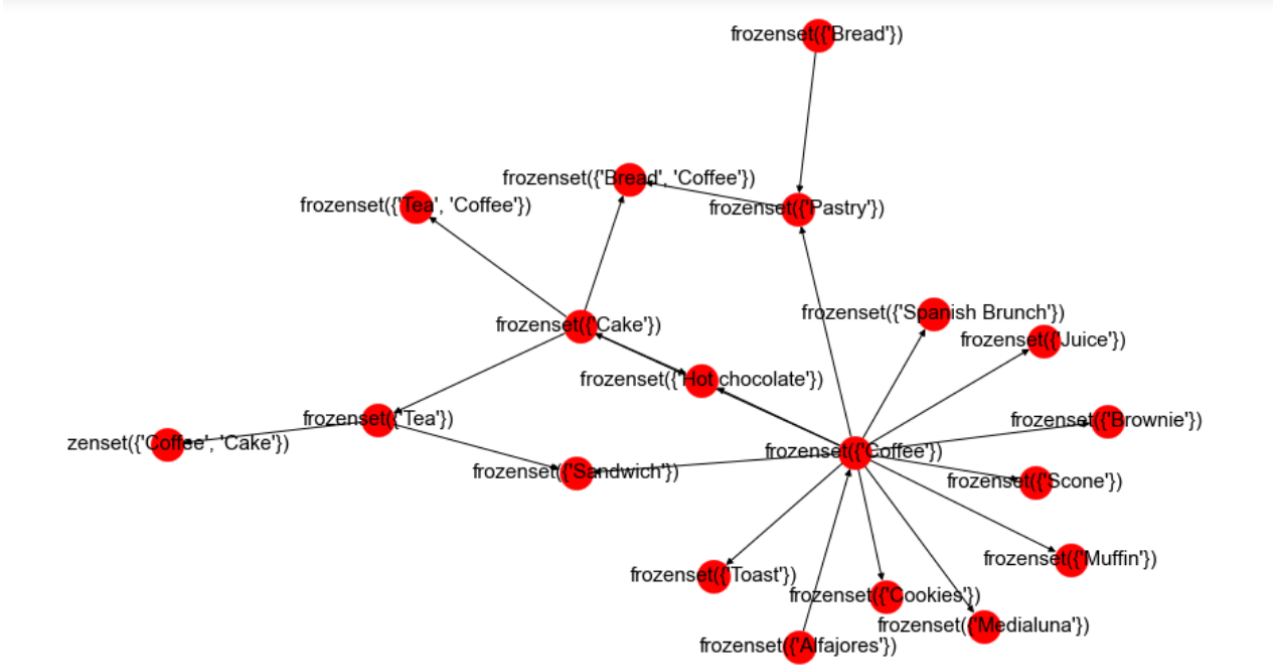
Các tập phổ biến có 2 item:

	support	itemsets	length
34	0.090016	(Bread, Coffee)	2
42	0.054728	(Cake, Coffee)	2
55	0.049868	(Tea, Coffee)	2
50	0.047544	(Pastry, Coffee)	2
51	0.038246	(Sandwich, Coffee)	2
48	0.035182	(Medialuna, Coffee)	2
46	0.029583	(Hot chocolate, Coffee)	2
38	0.029160	(Bread, Pastry)	2
45	0.028209	(Cookies, Coffee)	2
40	0.028104	(Bread, Tea)	2
44	0.023772	(Cake, Tea)	2
56	0.023666	(Toast, Coffee)	2
33	0.023349	(Bread, Cake)	2
47	0.020602	(Juice, Coffee)	2
31	0.019651	(Alfajores, Coffee)	2
41	0.019651	(Brownie, Coffee)	2
49	0.018806	(Muffin, Coffee)	2
52	0.018067	(Scone, Coffee)	2
39	0.017010	(Bread, Sandwich)	2
37	0.016904	(Bread, Medialuna)	2
53	0.015848	(Soup, Coffee)	2
35	0.014474	(Cookies, Bread)	2
57	0.014369	(Sandwich, Tea)	2
36	0.013418	(Bread, Hot chocolate)	2
43	0.011410	(Hot chocolate, Cake)	2
54	0.010882	(Spanish Brunch, Coffee)	2
32	0.010777	(Bread, Brownie)	2
30	0.010354	(Bread, Alfajores)	2

Các tập phổ biến có 3 item:

	support	itemsets	length
59	0.011199	(Bread, Pastry, Coffee)	3
58	0.010037	(Bread, Cake, Coffee)	3
60	0.010037	(Cake, Tea, Coffee)	3

Mô hình mạng nơ-ron thể hiện sự kết hợp giữa các sản phẩm với độ hỗ trợ tối thiểu là 0.01:



Sắp xếp các luật được sinh ra theo sự giảm dần của chỉ số confidence.

Giải thích bảng:

- Antecedents : những sản phẩm được mua nằm trong dự tính của khách hàng khi bước vào tiệm – sản phẩm tiền tố.
- Consequents: những sản phẩm được phát sinh thêm hay là sản phẩm sau khi mua ‘antecedents’ – sản phẩm hậu tố.

- Antecedent support: độ hỗ trợ của các sản phẩm ‘antencedent’ – là tần suất xuất hiện ‘antencedent’ có trong cơ sở dữ liệu giao dịch.
- Consequent support: độ hỗ trợ của các sản phẩm ‘Consequent’ – là tần suất xuất hiện ‘Consequent’ có trong cơ sở dữ liệu giao dịch.
- Support: là tần xuất xuất hiện giao dịch có chứa cả ‘ Antencedent - Consequent’ trong tập cơ sở dữ liệu giao dịch.
- Confidence: là xác suất xuất hiện ‘Consequent’ khi đã xuất hiện ‘Antencedent’ .

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
30	(Toast)	(Coffee)	0.033597	0.478394	0.023666	0.704403	1.472431	0.007593	1.764582
28	(Spanish Brunch)	(Coffee)	0.018172	0.478394	0.010882	0.598837	1.251766	0.002189	1.300235
18	(Medialuna)	(Coffee)	0.061807	0.478394	0.035182	0.569231	1.189878	0.005614	1.210871
22	(Pastry)	(Coffee)	0.086107	0.478394	0.047544	0.552147	1.154168	0.006351	1.164682
0	(Alfajores)	(Coffee)	0.036344	0.478394	0.019651	0.540698	1.130235	0.002264	1.135648
16	(Juice)	(Coffee)	0.038563	0.478394	0.020602	0.534247	1.116750	0.002154	1.119919
24	(Sandwich)	(Coffee)	0.071844	0.478394	0.038246	0.532353	1.112792	0.003877	1.115384
6	(Cake)	(Coffee)	0.103856	0.478394	0.054728	0.526958	1.101515	0.005044	1.102664
26	(Scone)	(Coffee)	0.034548	0.478394	0.018067	0.522936	1.093107	0.001539	1.093366
12	(Cookies)	(Coffee)	0.054411	0.478394	0.028209	0.518447	1.083723	0.002179	1.083174
14	(Hot chocolate)	(Coffee)	0.058320	0.478394	0.029583	0.507246	1.060311	0.001683	1.058553
4	(Brownie)	(Coffee)	0.040042	0.478394	0.019651	0.490765	1.025860	0.000495	1.024293
20	(Muffin)	(Coffee)	0.038457	0.478394	0.018806	0.489011	1.022193	0.000408	1.020777
3	(Pastry)	(Bread)	0.086107	0.327205	0.029160	0.338650	1.034977	0.000985	1.017305
10	(Cake)	(Tea)	0.103856	0.142631	0.023772	0.228891	1.604781	0.008959	1.111865
39	(Tea, Coffee)	(Cake)	0.049868	0.103856	0.010037	0.201271	1.937977	0.004858	1.121962
32	(Sandwich)	(Tea)	0.071844	0.142631	0.014369	0.200000	1.402222	0.004122	1.071712
8	(Hot chocolate)	(Cake)	0.058320	0.103856	0.011410	0.195652	1.883874	0.005354	1.114125
38	(Cake, Coffee)	(Tea)	0.054728	0.142631	0.010037	0.183398	1.285822	0.002231	1.049923

11	(Tea)	(Cake)	0.142631	0.103856	0.023772	0.166667	1.604781	0.008959	1.075372
37	(Pastry)	(Bread, Coffee)	0.086107	0.090016	0.011199	0.130061	1.444872	0.003448	1.046033
36	(Bread, Coffee)	(Pastry)	0.090016	0.086107	0.011199	0.124413	1.444872	0.003448	1.043749
7	(Coffee)	(Cake)	0.478394	0.103856	0.054728	0.114399	1.101515	0.005044	1.011905
34	(Bread, Coffee)	(Cake)	0.090016	0.103856	0.010037	0.111502	1.073621	0.000688	1.008606
9	(Cake)	(Hot chocolate)	0.103856	0.058320	0.011410	0.109868	1.883874	0.005354	1.057910
33	(Tea)	(Sandwich)	0.142631	0.071844	0.014369	0.100741	1.402222	0.004122	1.032134
23	(Coffee)	(Pastry)	0.478394	0.086107	0.047544	0.099382	1.154168	0.006351	1.014740
35	(Cake)	(Bread, Coffee)	0.103856	0.090016	0.010037	0.096643	1.073621	0.000688	1.007336
40	(Cake)	(Tea, Coffee)	0.103856	0.049868	0.010037	0.096643	1.937977	0.004858	1.051779
2	(Bread)	(Pastry)	0.327205	0.086107	0.029160	0.089119	1.034977	0.000985	1.003306
25	(Coffee)	(Sandwich)	0.478394	0.071844	0.038246	0.079947	1.112792	0.003877	1.008807
19	(Coffee)	(Medialuna)	0.478394	0.061807	0.035182	0.073542	1.189878	0.005614	1.012667
41	(Tea)	(Cake, Coffee)	0.142631	0.054728	0.010037	0.070370	1.285822	0.002231	1.016827
15	(Coffee)	(Hot chocolate)	0.478394	0.058320	0.029583	0.061837	1.060311	0.001683	1.003749
13	(Coffee)	(Cookies)	0.478394	0.054411	0.028209	0.058966	1.083723	0.002179	1.004841
31	(Coffee)	(Toast)	0.478394	0.033597	0.023666	0.049470	1.472431	0.007593	1.016699
17	(Coffee)	(Juice)	0.478394	0.038563	0.020602	0.043065	1.116750	0.002154	1.004705
1	(Coffee)	(Alfajores)	0.478394	0.036344	0.019651	0.041078	1.130235	0.002264	1.004936
5	(Coffee)	(Brownie)	0.478394	0.040042	0.019651	0.041078	1.025860	0.000495	1.001080
21	(Coffee)	(Muffin)	0.478394	0.038457	0.018806	0.039311	1.022193	0.000408	1.000888
27	(Coffee)	(Scone)	0.478394	0.034548	0.018067	0.037765	1.093107	0.001539	1.003343
29	(Coffee)	(Spanish Brunch)	0.478394	0.018172	0.010882	0.022747	1.251766	0.002189	1.004682

So sánh conf của từng dòng luật với min_conf để tìm ra các luật mạnh.

Ở đây ta sẽ chỉ quan tâm đến các luật có độ Lift > 1 điều này nhấn mạnh rằng các luật được sinh ra không phải là những luật được phát sinh ngẫu nhiên mà đã có ý định từ trước.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
31	(Toast)	(Coffee)	0.033597	0.478394	0.023666	0.704403	1.472431	0.007593	1.764582
29	(Spanish Brunch)	(Coffee)	0.018172	0.478394	0.010882	0.598837	1.251766	0.002189	1.300235
19	(Medialuna)	(Coffee)	0.061807	0.478394	0.035182	0.569231	1.189878	0.005614	1.210871
23	(Pastry)	(Coffee)	0.086107	0.478394	0.047544	0.552147	1.154168	0.006351	1.164682
0	(Alfajores)	(Coffee)	0.036344	0.478394	0.019651	0.540698	1.130235	0.002264	1.135648
16	(Juice)	(Coffee)	0.038563	0.478394	0.020602	0.534247	1.116750	0.002154	1.119919
24	(Sandwich)	(Coffee)	0.071844	0.478394	0.038246	0.532353	1.112792	0.003877	1.115384
6	(Cake)	(Coffee)	0.103856	0.478394	0.054728	0.526958	1.101515	0.005044	1.102664
27	(Scone)	(Coffee)	0.034548	0.478394	0.018067	0.522936	1.093107	0.001539	1.093366
12	(Cookies)	(Coffee)	0.054411	0.478394	0.028209	0.518447	1.083723	0.002179	1.083174
14	(Hot chocolate)	(Coffee)	0.058320	0.478394	0.029583	0.507246	1.060311	0.001683	1.058553
4	(Brownie)	(Coffee)	0.040042	0.478394	0.019651	0.490765	1.025860	0.000495	1.024293
21	(Muffin)	(Coffee)	0.038457	0.478394	0.018806	0.489011	1.022193	0.000408	1.020777

Luật kết hợp có dạng X → Y (antecedents → consequents)

Luật 1: Toast → Coffee có support = 0.02 và confidencr = 0.7 có nghĩa: Có hơn 2% số giao dịch mua đồng thời cả toast và coffee. Và khoảng 70% khách hàng sau khi mua toast thì sẽ mua thêm coffee.

Luật 2: Spanish Brunch → Coffee có support = 0.01 và confidencr = 0.6 có nghĩa: Có hơn 1% số giao dịch mua đồng thời cả toast và coffee. Và khoảng 60% khách hàng sau khi mua toast thì sẽ mua thêm coffee.

Luật 3: Medialuna → Coffee có support = 0.035 và confidencr = 0.57 có nghĩa: Có hơn 3.5% số giao dịch mua đồng thời cả toast và coffee. Và khoảng 57% khách hàng sau khi mua toast thì sẽ mua thêm coffee.

Luật 4: Pastry → Coffee có support = 0.047 và confidencr = 0.55 có nghĩa: Có hơn 4.7% số giao dịch mua đồng thời cả toast và coffee. Và khoảng 55% khách hàng sau khi mua toast thì sẽ mua thêm coffee.

Luật 5: Alfajores → Coffee có support = 0.02 và confidencr = 0.54 có nghĩa: Có hơn 2% số giao dịch mua đồng thời cả toast và coffee. Và khoảng 54% khách hàng sau khi mua toast thì sẽ mua thêm coffee.

Luật 6: Juice → Coffee có support = 0.02 và confidencr = 0.53 có nghĩa: Có hơn 2% số giao dịch mua đồng thời cả toast và coffee. Và khoảng 53% khách hàng sau khi mua toast thì sẽ mua thêm coffee.

Luật 7: Sandwich → Coffee có support = 0.04 và confidencr = 0.53 có nghĩa: Có hơn 4% số giao dịch mua đồng thời cả toast và coffee. Và khoảng 53% khách hàng sau khi mua toast thì sẽ mua thêm coffee.

Luật 8: Cake → Coffee có support = 0.054 và confidencr = 0.53 có nghĩa: Có hơn 5.4% số giao dịch mua đồng thời cả toast và coffee. Và khoảng 53% khách hàng sau khi mua toast thì sẽ mua thêm coffee.

Luật 9: Scone → Coffee có support = 0.02 và confidencr = 0.52 có nghĩa: Có hơn 2% số giao dịch mua đồng thời cả toast và coffee. Và khoảng 52% khách hàng sau khi mua toast thì sẽ mua thêm coffee.

Luật 10: Cookies → Coffee có support = 0.03 và confidencr = 0.52 có nghĩa: Có hơn 3% số giao dịch mua đồng thời cả toast và coffee. Và khoảng 53% khách hàng sau khi mua toast thì sẽ mua thêm coffee.

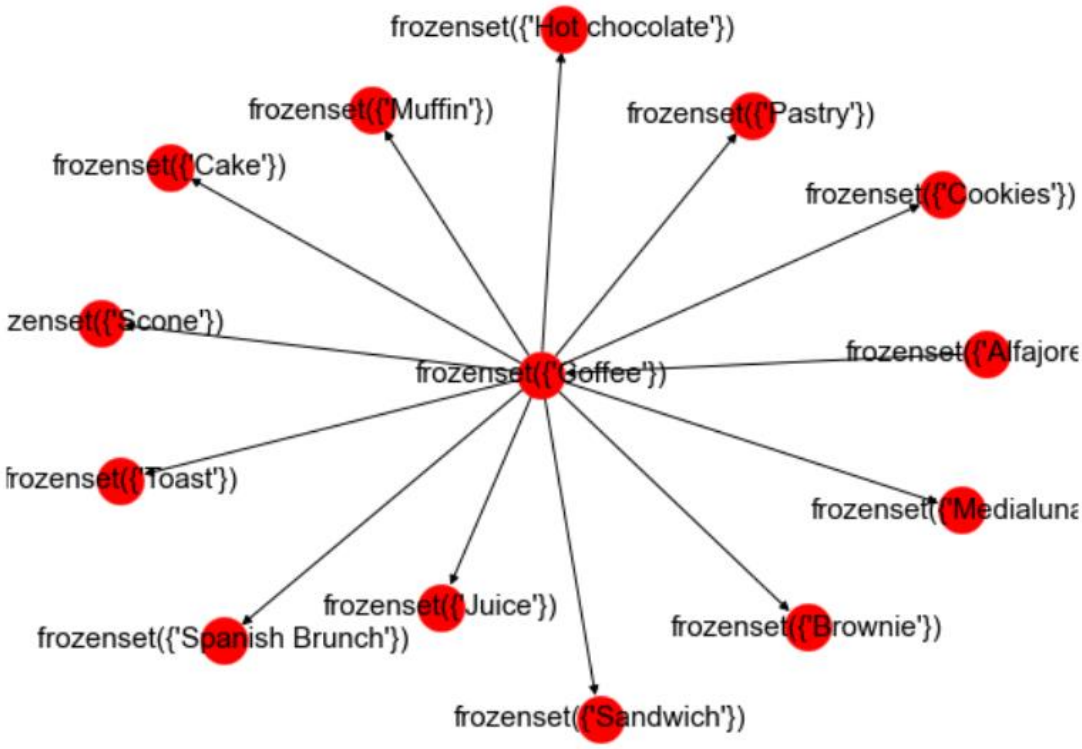
Luật 11: Hot Chocolate → Coffee có support = 0.03 và confidencr = 0.51 có nghĩa: Có hơn 3% số giao dịch mua đồng thời cả toast và coffee. Và khoảng 51% khách hàng sau khi mua toast thì sẽ mua thêm coffee.

Luật 12: Brownie → Coffee có support = 0.02 và confidencr = 0.5 có nghĩa: Có hơn 2% số giao dịch mua đồng thời cả toast và coffee. Và khoảng 50% khách hàng sau khi mua toast thì sẽ mua thêm coffee.

Luật 13 : Muffin → Coffee có support = 0.019 và confidencr = 0.49 có nghĩa: Có hơn 1.9 % số giao dịch mua đồng thời cả toast và coffee. Và khoảng 49 % khách hàng sau khi mua toast thì sẽ mua thêm coffee.

⇨ Trong 42 luật được sinh ra thì có 13 luật được xem là các luật mạnh.

⇨ Tất cả các luật sinh ra được coi là luật mạnh đều là những giao dịch có hậu tố là Coffee. Đồng nghĩa với việc Coffee là sản phẩm đường mua kèm nhiều nhất.



Vậy các giao dịch mà Coffee đóng vai trò là tiền tố thì những sản phẩm nào sẽ được mua thêm?

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
7	(Coffee)	(Cake)	0.478394	0.103856	0.054728	0.114399	1.101515	0.005044	1.011905
22	(Coffee)	(Pastry)	0.478394	0.086107	0.047544	0.099382	1.154168	0.006351	1.014740
25	(Coffee)	(Sandwich)	0.478394	0.071844	0.038246	0.079947	1.112792	0.003877	1.008807
18	(Coffee)	(Medialuna)	0.478394	0.061807	0.035182	0.073542	1.189878	0.005614	1.012667
15	(Coffee)	(Hot chocolate)	0.478394	0.058320	0.029583	0.061837	1.060311	0.001683	1.003749
13	(Coffee)	(Cookies)	0.478394	0.054411	0.028209	0.058966	1.083723	0.002179	1.004841
30	(Coffee)	(Toast)	0.478394	0.033597	0.023666	0.049470	1.472431	0.007593	1.016699
17	(Coffee)	(Juice)	0.478394	0.038563	0.020602	0.043065	1.116750	0.002154	1.004705
1	(Coffee)	(Alfajores)	0.478394	0.036344	0.019651	0.041078	1.130235	0.002264	1.004936
5	(Coffee)	(Brownie)	0.478394	0.040042	0.019651	0.041078	1.025860	0.000495	1.001080
20	(Coffee)	(Muffin)	0.478394	0.038457	0.018806	0.039311	1.022193	0.000408	1.000888
26	(Coffee)	(Scone)	0.478394	0.034548	0.018067	0.037765	1.093107	0.001539	1.003343
28	(Coffee)	(Spanish Brunch)	0.478394	0.018172	0.010882	0.022747	1.251766	0.002189	1.004682

⇨ Các sản phẩm được mua kèm sau khi mua Coffee đều là sản phẩm quen thuộc. Nhưng đa số Coffee sẽ được mua 1 mình.

Trong các luật sinh ra không kể mạnh hay không thì 3 sản phẩm nào là nổi bậc thường được mua kèm nhiều nhất?

38	(Coffee, Tea)	(Cake)	0.049868	0.103856	0.010037	0.201271	1.937977	0.004858	1.121962
33	(Sandwich)	(Tea)	0.071844	0.142631	0.014369	0.200000	1.402222	0.004122	1.071712
8	(Hot chocolate)	(Cake)	0.058320	0.103856	0.011410	0.195652	1.883874	0.005354	1.114125
39	(Coffee, Cake)	(Tea)	0.054728	0.142631	0.010037	0.183398	1.285822	0.002231	1.049923
10	(Tea)	(Cake)	0.142631	0.103856	0.023772	0.166667	1.604781	0.008959	1.075372
37	(Pastry)	(Coffee, Bread)	0.086107	0.090016	0.011199	0.130061	1.444872	0.003448	1.046033
36	(Coffee, Bread)	(Pastry)	0.090016	0.086107	0.011199	0.124413	1.444872	0.003448	1.043749
6	(Coffee)	(Cake)	0.478394	0.103856	0.054728	0.114399	1.101515	0.005044	1.011905
34	(Coffee, Bread)	(Cake)	0.090016	0.103856	0.010037	0.111502	1.073621	0.000688	1.008606

Dựa vào kết quả của mô hình Coffee, Cake, Tea là 3 sản phẩm thường được mua cùng nhau. Từ đó chúng ta có thể dùng chúng để sắp xếp lại gian hàng, xây dựng các chiến lược bán kèm, bán khuyến mãi,…

VI. ÁP DỤNG KẾT QUẢ CỦA MÔ HÌNH :

1. Sắp xếp lại gian hàng:

Với sản phẩm được bán chạy nhất là Coffee và được bán kèm theo với nhiều loại sản phẩm khác thì ta nên đặc biệt chú ý vào nó. Bằng nhiều cách ví dụ như đặt nó ở trung tâm nơi dễ tìm kiếm nhất hoặc là để ở cuối dãy để khi khách hàng đi đến cuối dãy để lấy Coffee thì khách hàng sẽ đi ngang qua rất nhiều sản phẩm trong tiệm để rồi sẽ phát sinh mua thêm nhiều sản phẩm.

Bên cạnh đó thì những sản phẩm mới, những sản phẩm cần kích thích để được bán nhiều thì nên ưu tiên đặt ở những vị trí trong tầm mắt của khách hàng khi bước vào tiệm ví dụ như đầu các kệ các dãy trưng bày hàng hoặc là kế quầy tính tiền.

2. Tối ưu hóa hoạt động marketing và các chiến dịch chăm sóc khách hàng.

3. Đưa ra các chiến dịch bán chéo:

Đây là cách khiến khách hàng phải trả thêm để mua những sản phẩm chưa có trong ý định từ trước hoặc là những sản phẩm liên quan đến sản phẩm trong dự định mua. Nhằm tăng thêm doanh thu, giới thiệu các sản phẩm mới,… Bằng cách giới thiệu các sản phẩm có tác dụng hỗ trợ, tương đồng hay là dùng chung sẽ tăng trải nghiệm cho khách hàng sử dụng.

Kích thích khách hàng chi tiêu nhiều hơn khi thường xuyên xây dựng các chiến dịch khuyến mãi giảm giá khi mua nhiều, hoặc là mua thêm sản phẩm với tổng mức giá ưu đãi nào đó. Thiết kế các combo đáp ứng thị hiếu, tạo các combo mới bao gồm những sản phẩm được bán nhiều với sản phẩm mới, hoặc là sản phẩm ít được chú ý.

Với tiệm bánh thì có thể thiết kế một số combo như:

- Combo 1: Cake với coffee , cake với trà – phù hợp cho những bữa ăn nhanh chóng và tiện lợi.

- Combo 2: Salar + toast + tea : những thực phẩm tốt cho sức khỏe bên cạnh đó cũng tăng lượng bán cho Salar.
- Combo 3: …

  Bên cạnh việc bán được nhiều sản phẩm hơn thì việc phân tích giỏ hàng sử dụng luật kết hợp cũng giúp sắp xếp lại quầy bán cho phù hợp , tối ưu hóa được chi phí quảng cáo, chuẩn bị hàng hóa cho phù hợp không quá nhiều cũng không bị thiếu,…

VII. Tổng kết :

Khai phá luật kết hợp là một trong những kỹ thuật quan trọng, mang tính thời sự đối với nền công nghệ thông tin hiện nay. Việc khai phá phục vụ tiến trình kinh doanh đã được sử dụng trong bài. Bài báo cáo hướng người đọc bước đầu từ khái niệm, định nghĩa đến khai phá luật kết hợp. Sự bùng nổ thông tin cùng với sự phát triển và ứng dụng ngày càng rộng rãi của công nghệ thông tin trong mọi lĩnh vực đã khiến nhu cầu xử lí của những khối dữ liệu khổng lồ để xuất ra những thông tin, tri thức hữu ích cho người sử dụng một cách tự động, nhanh chóng và chính xác, trở thành nhân tố quan trọng hang đầu cho mọi thành công của các tổ chức và cá nhân. Khai phá dữ liệu đang được áp dụng một cách rộng rãi trong nhiều lĩnh vực kinh doanh và đời sống. Trong thực tế, có rất nhiều tổ chức và công ty lớn trên thế giới đã áp dụng kỹ thuật KPDL vào các hoạt động sản xuất - kinh doanh của mình và thu được những lợi ích to lớn.

- Note:

  Link dữ liệu :

  https://www.kaggle.com/code/akashdeepkuila/market-basket-analysis/data?select=bread+basket.csv

  Dashboard :

  https://public.tableau.com/app/profile/phan.th.th.o.ng.n/viz/New_bakery_/Dashboard1?publish=yes