[1 Giới thiệu 25](#_Toc163402183)

[1.1 Đặt vấn đề 25](#_Toc163402184)

[1.2 Giới thiệu 26](#_Toc163402185)

[2 Miêu tả dữ liệu 27](#_Toc163402186)

[2.1 Tổng quan: 27](#_Toc163402187)

[2.2 Phân tích đặc điểm dữ liệu: 28](#_Toc163402188)

[2.3 Các vấn đề của dữ liệu: 29](#_Toc163402189)

[3 Phương pháp thực hiện 31](#_Toc163402190)

[3.1 THUẬT TOÁN APRIORI 31](#_Toc163402191)

[Cách chạy của thuật toán: 32](#_Toc163402192)

[3.2 THUẬT TOÁN FP - GROWTH 33](#_Toc163402193)

[4 Phân tích, so sánh các kết quả: 35](#_Toc163402194)

|  |  |  |
| --- | --- | --- |
| Id | Cite | Abstract |
|  |  |  |
|  |  |  |
| 1 | @article{Gupta2023ECommerceMB,  title={Phân tích giỏ thị trường thương mại điện tử bằng thuật toán Apriori}, tác giả={Khushi Gupta và Kashyapi Shah và Ameya A Kadam}, tạp chí={Tạp chí nghiên cứu khoa học quốc tế về kỹ thuật và quản lý}, năm={ 2023}, url={https://api.semanticscholar.org/CorpusID:263659317} } | Bài viết này trình bày cách sử dụng Thuật toán Apriori để thực hiện phân tích giỏ hàng thị trường nhằm xác định mô hình mua hàng đối với các mặt hàng thường được khách hàng mua cùng nhau. Kết quả phân tích này chủ yếu được sử dụng để cải thiện doanh số bán hàng của các cửa hàng đa sản phẩm bằng cách nâng cao vị trí sản phẩm dựa trên thói quen mua sắm của người tiêu dùng. Trong trường hợp cụ thể này, chúng tôi đã sử dụng dữ liệu từ một cửa hàng thương mại điện tử trực tuyến phục vụ khách hàng trên toàn thế giới nhưng chủ yếu tập trung vào Vương quốc Anh. Từ khóa: Giỏ thị trường , Thuật toán Apriori, Quy tắc kết hợp, Thương mại điện tử, Hành vi người tiêu dùng, Khai thác dữ liệu |
| 2 | @article{Gino2023ExploratoryAO,  title={Exploratory Analysis on Market Basket Data using Network Visualization},  author={Henrique L. S. Gino and Diogenes S. Pedro and Jean R. Ponciano and Claudio D. G. Linhares and Agma J. M. Traina},  journal={Anais do XII Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2023)},  year={2023},  url={https://api.semanticscholar.org/CorpusID:259312945}  } | Phân tích giỏ thị trường là một kỹ thuật mạnh mẽ để hiểu hành vi của khách hàng và tối ưu hóa chiến lược kinh doanh dựa trên sự hiểu biết đó. Phân tích dựa trên thị trường theo thời gian bằng cách sử dụng kỹ thuật trực quan có thể cung cấp thông tin chuyên sâu về xu hướng và mối quan hệ thị trường , đơn giản hóa dữ liệu phức tạp và truyền đạt thông tin chi tiết một cách hiệu quả, có thể giúp các tổ chức đưa ra quyết định sáng suốt hơn. Bài viết này tận dụng một tập dữ liệu tập trung vào thu nhập của người dùng và các khía cạnh tạm thời của việc mua hàng trên thị trường . Chúng tôi đã mô hình hóa tập dữ liệu này thành ba mạng tạm thời riêng biệt và thực hiện đánh giá thăm dò để xác định các mẫu và điểm bất thường trong dữ liệu. Cụ thể hơn, chúng tôi đã xác định các nhóm sản phẩm có liên quan, chỉ ra các hoạt động mua hàng theo chủ đề và đánh giá tác động của các yếu tố nhân khẩu học, chẳng hạn như thu nhập, đối với chi tiêu của khách hàng. |
| 3 | @article{Bagaskara2023AnalisisDI, title={Phân tích và Phân tích giỏ thị trường thực hiện (MBA) Thuật toán bổ sung dựa trên dữ liệu trực quan hóa của Dukungan}, tác giả={Septembri Rio Bagaskara và Dwi Hosanna Bangkalang}, tạp chí={Jurnal Sistem Komputer và Informatika (JSON)} , năm={2023}, url={https://api.semanticscholar.org/CorpusID:260034709} } | Văn hóa Cà phê MSME là một trong những MSME hoạt động trong lĩnh vực ẩm thực và đang gặp phải sự cạnh tranh kinh doanh. Cần có một chiến lược tiếp thị cùng với quy trình ra quyết định đúng đắn để doanh nghiệp có thể tồn tại và phát triển. UMKM Culture Coffee sử dụng ứng dụng điểm bán hàng để đáp ứng quy trình giao dịch và ghi lại các giao dịch. Dữ liệu lịch sử khách hàng có thể được xử lý làm cơ sở để đưa ra quyết định cho các chiến lược tiếp thị nhằm tăng doanh số bán hàng một cách hiệu quả. Tuy nhiên, dữ liệu giao dịch chưa được sử dụng một cách tối ưu. Cần phải phân tích dữ liệu lịch sử khách hàng để có thể tạo ra thông tin hình thành các chiến lược tiếp thị. Phân tích giỏ thị trường (MBA) là một trong những phương pháp khai thác dữ liệu được sử dụng để biết các sản phẩm có xu hướng được khách hàng mua cùng nhau được gọi là Quy tắc kết hợp. Quy tắc kết hợp tạo ra sản phẩm dưới dạng gói hoặc gói được sử dụng làm chiến lược tiếp thị. Chiến lược tiếp thị thu được được hỗ trợ bởi trực quan hóa dữ liệu chứa thông tin từ dữ liệu. Thuật toán Apriori được sử dụng để tạo ra các luật kết hợp. Kết quả của nghiên cứu này là một quy tắc kết hợp trên dữ liệu lịch sử mua hàng của khách hàng MSME Culture Coffee. Dựa trên những quy định này, có thể đưa ra những khuyến nghị về việc bán gói thực đơn cho khách hàng. Mục đích của nghiên cứu này là tìm ra các mô hình mua hàng của khách hàng làm cơ sở cho việc ra quyết định xác định doanh số bán thực đơn. Kết quả cho thấy 2 gói sản phẩm là cốm và khoai tây chiên kèm xúc xích và khoai tây chiên có độ hỗ trợ và độ tin cậy lần lượt là 12,5% và 37,6% với 10,8% và 29%. Kết quả của nghiên cứu này có thể được sử dụng làm cơ sở cho chiến lược bán hàng và tiếp thị của các MSME về Cà phê Văn hóa nhằm tăng doanh thu kinh doanh. |
| 4 | @article{Nafi2023MarketBA, title={Phân tích giỏ thị trường cho giao dịch bán hàng trong cửa hàng mua sắm}, tác giả={Mohd Noor Azam Nafi và Azni Sharlina Zakaria và Nur Izzati Mohamad Arif và Siti Nurhafizah Mohd Shafie và Nasuhar Ab. Aziz và Omar Kairan}, tạp chí={Tạp chí quốc tế về nghiên cứu học thuật về khoa học xã hội và kinh doanh}, Year={2023}, url={https://api.semanticscholar.org/CorpusID:257119151} } | Hệ thống Phân tích Giỏ Thị trường (MBA) là một kỹ thuật được các nhà tiếp thị sử dụng rộng rãi, đặc biệt là để phân tích khai thác dữ liệu vô hướng. MBA còn được gọi là phân tích liên kết sản phẩm và kết quả của phân tích này được gọi là luật kết hợp. Kết quả có thể được sử dụng để lên kế hoạch cho các chiến lược tiếp thị hoặc quảng cáo cũng như thiết kế danh mục cho các cách bố trí cửa hàng khác nhau. Việc khám phá mô hình từ thói quen mua hàng của khách hàng tại các cửa hàng mua sắm đã được thu thập trong giao dịch mua hàng của họ. Nghiên cứu này nhằm mục đích so sánh mặt hàng mà người trả lời mua giữa Cửa hàng A và Cửa hàng B và tìm ra những sản phẩm tiềm năng nhất mà khách hàng đã mua cùng với một danh mục sản phẩm cụ thể. Lấy mẫu phi xác suất thuận tiện được thực hiện bằng bảng câu hỏi có cấu trúc về các mặt hàng trong cửa hàng được thu thập để phân tích dữ liệu. Phân tích liên kết được sử dụng bằng cách phân tích kết quả từ mức hỗ trợ, độ tin cậy và mức tăng. Các phát hiện cho thấy có 13 quy tắc liên kết thú vị được tiết lộ trong nghiên cứu này. Hơn nữa, kết quả cũng cho thấy hầu hết các sản phẩm được mua cùng nhau là khăn giấy, gia vị, thực phẩm ăn liền, dầu ăn, thịt, bánh quy, đồ khô, đồ uống và sản phẩm tẩy rửa. |
| 5 | @article{Dwiputra2023EvaluatingTP, title={Đánh giá hiệu suất của các quy tắc liên kết trong thuật toán tăng trưởng Apriori và FP: Phân tích giỏ thị trường để khám phá các quy tắc kết hợp mặt hàng}, tác giả={Dedy Dharmadi Cakra Dwiputra và Agung Mulyo Widodo và Habibullah Akbar và Gerry Firmansyah }, tạp chí={Tạp chí Khoa học Thế giới}, năm={2023}, url={https://api.semanticscholar.org/CorpusID:261602737} } | Nghiên cứu này tập trung vào việc áp dụng các kỹ thuật khai thác dữ liệu, đặc biệt là khai thác luật kết hợp bằng thuật toán Apriori và FP-GROWTH để phân tích giỏ thị trường trên PT. XYZ là một công ty dược phẩm ở Indonesia. Phương pháp định lượng sử dụng tập dữ liệu gồm 100.498 giao dịch có nguồn gốc từ 432.356 hàng dữ liệu từ tháng 7 đến tháng 12 năm 2022 trong khu vực JABODETABEK. Thuật toán Apriori và FP-GROWTH được áp dụng để khai thác luật kết hợp. Kết quả cho thấy FP-GROWTH có thời gian thực hiện nhanh nhất là 84.655 giây. Tuy nhiên, mức sử dụng bộ nhớ cho thuật toán Apriori là thấp nhất ở mức 482,32 MiB, với mức tăng là: 0,21 MiB. Đối với các quy tắc được tạo, cả hai thuật toán, cả Apriori và FP-GROWTH, đều tạo ra cùng một số quy tắc và giá trị về mức độ hỗ trợ, độ tin cậy, mức tăng, Hỗ trợ hai chiều, Độ tin cậy kép và Bi-Lift. Tóm lại, Apriori được khuyến nghị cho các tập dữ liệu bán hàng nếu việc sử dụng bộ nhớ và tính dễ triển khai là quan trọng. Tuy nhiên, nếu xét đến tốc độ thời gian thực hiện và lượng dữ liệu lớn thì FP-GROWTH là lựa chọn tốt hơn vì thời gian thực hiện nhanh hơn đối với lượng lớn dữ liệu. Tuy nhiên, việc lựa chọn thuật toán phụ thuộc vào mục tiêu phân tích cụ thể, kích thước tập mục, quy mô dữ liệu và khả năng tính toán. Kết quả từ việc khai thác quy tắc kết hợp cung cấp bằng chứng về mức độ phổ biến của sản phẩm, mô hình mua hàng và cơ hội tiếp thị chiến lược và quản lý hàng tồn kho. Những phát hiện này có thể giúp PT. XYZ cải thiện hiệu quả kinh doanh, hiểu hành vi của khách hàng và tăng lợi nhuận. |
| 6 | @article{Mooy2023ContextualMB,  title={Contextual Market Basket Analysis during Covid-19},  author={Gisela Christy Mooy and Sani Muhamad Isa},  journal={Journal of Social Science},  year={2023},  url={https://api.semanticscholar.org/CorpusID:258945921}  } | Một dạng ứng dụng Khai thác dữ liệu để phân tích Phân tích giỏ thị trường . Phân tích giỏ thị trường giúp xác định các mô hình mua hàng được hình thành từ các giao dịch đồng thời. Một trong những vấn đề với Phân tích giỏ thị trường là nhu cầu của khách hàng thay đổi tùy theo mùa và thời gian trong ngày, đặc biệt là trong mùa Covid-19 này. Với mục đích này, bằng cách sử dụng Phương pháp tiếp cận Mạng thần kinh nhân tạo (ANN) được kết nối với Phân tích giỏ thị trường , nó có thể phân tích và so sánh các mô hình mua hàng, đồng thời có thể xác định các quy tắc được hình thành trước và sau Covid-19; một số thay đổi quy tắc đã được tìm thấy do những thay đổi trong mô hình hành vi của mọi người. |
| 7 | @article{Purnomo2023PenerapanDM, title={Khai thác dữ liệu Penerapan Dalam Menganalisis Pola Belanja Konsumen Menggunakan Phân tích giỏ thị trường}, tác giả={Sari Asih Kencono Purnomo và Heny Ispur Pratiwi và Muhammad Ibnu Sa'ad}, tạp chí={METIK JURNAL}, năm= {2023}, url={https://api.semanticscholar.org/CorpusID:266702253} } | Hiện nay, hầu hết mọi hoạt động đều liên quan đến dữ liệu. Trong lĩnh vực kinh doanh, dữ liệu giao dịch bán hàng hàng ngày được lưu trữ trong hệ thống cơ sở dữ liệu sẽ luôn tăng và tích lũy. Dữ liệu hiện có chỉ được chủ cửa hàng sử dụng làm kho lưu trữ nên có ảnh hưởng đến các chiến lược bán hàng chưa được triển khai tốt, mặc dù dữ liệu hiện có có thể được xử lý thành thông tin xác định cách bố trí hàng hóa để có tác động về việc tăng tần suất mua hàng ngẫu hứng, tăng hoặc duy trì doanh thu và giảm thiểu lãng phí sản phẩm. tích lũy cho đến khi hết hạn, điều này có thể gây bất lợi cho cửa hàng. Mục đích của nghiên cứu này là tìm ra các mô hình mua sắm của người tiêu dùng bằng cách sử dụng Phân tích giỏ đánh dấu . Phương pháp nghiên cứu này gọi là phân tích giỏ thị trường hay còn gọi là luật kết hợp, là một kỹ thuật khai thác dữ liệu nhằm tìm ra các mẫu thường xuất hiện đồng thời trong dữ liệu giao dịch, từ đó có thể sử dụng làm phương pháp tìm kiếm thông tin về loại hàng hóa nào thường xuyên được sử dụng. đã sử dụng. được người tiêu dùng mua cùng một lúc. Kết quả nghiên cứu này dựa trên phân tích dữ liệu bằng ứng dụng Rapidminer đã tìm thấy 25 mối quan hệ hoặc quy tắc liên kết có giá trị tỷ lệ nâng lớn hơn 1, các quy tắc này trở thành tài liệu tham khảo trong việc xác định cách bố trí hàng hóa. Việc đưa ra các đề xuất thay đổi bố cục nhằm mục đích giúp người tiêu dùng mua sắm dễ dàng hơn, tăng khả năng mua sắm ngẫu hứng của người tiêu dùng và tối đa hóa việc trưng bày sản phẩm, từ đó giảm thiểu việc tích tụ hàng hóa trong Kho của Cửa hàng Purnama. |
| 8 | @article{Wahyudi2023MarketBA, title={Phân tích giỏ thị trường với Phương pháp cải tiến Apriori và FP-Growth Pada Data Transaksi XYZ}, tác giả={Rizki Nofrian Wahyudi và Dyah Erny Herwindiati và Janson Hendryli}, tạp chí={Jurnal Ilmu Komputer dan Sistem Informasi} , năm={2023}, url={https://api.semanticscholar.org/CorpusID:260186039} } | Công nghệ hiện đang phát triển nhanh chóng, cho phép tất cả các tổ chức phát triển mạng lưới của họ với sự trợ giúp của nó và tạo ra các phương thức bán hàng hiện dựa vào công nghệ để hỗ trợ đưa ra các đánh giá đúng đắn. Khi có thể truy cập được dữ liệu giao dịch đã lưu, mọi doanh nghiệp sẽ có thể thực hiện chiến lược tiếp thị của mình để tối đa hóa các giao dịch của khách hàng. Sử dụng nó để lợi thế của bạn. Phân tích giỏ thị trường bằng cách sử dụng FP-Growth và thuật toán tiên nghiệm trong các giao dịch hỗ trợ lập kế hoạch chiến lược và cơ cấu sản phẩm kinh doanh. Thuật toán FP-Growth và thuật toán Apriori phối hợp tốt với nhau. Người ta có thể đánh giá hiệu quả của việc sử dụng thuật toán tiên nghiệm và thuật toán Tăng trưởng FP bằng cách áp dụng cả hai thuật toán này |
| 9 | @article{Ghous2023DeepLB, title={Phân tích giỏ thị trường dựa trên Deep Learning bằng cách sử dụng Quy tắc hiệp hội}, tác giả={Hamid Ghous và Mubasher Malik và Iqra Rehman}, tạp chí={Tạp chí khoa học máy tính và thông tin KIET}, năm={2023}, url={https://api.semanticscholar.org/CorpusID:264297776} } | Phân tích giỏ thị trường (MBA) là một kỹ thuật khai thác dữ liệu hỗ trợ các nhà bán lẻ xác định thói quen mua hàng của khách hàng đồng thời đưa ra các quyết định tiếp thị mới vì mong muốn của người mua thường xuyên thay đổi theo nhu cầu mở rộng; do đó, dữ liệu giao dịch ngày càng lớn. Có nhu cầu triển khai các phương pháp Deep Learning (DL) để xử lý dữ liệu đang phát triển nhanh chóng này. Trong nghiên cứu trước đây, nhiều tác giả đã tiến hành MBA áp dụng DL và luật kết hợp (AR) trên tập dữ liệu bán lẻ. AR xác định mối liên kết giữa các mặt hàng để tìm thứ tự khách hàng đặt các mặt hàng vào giỏ . AR chỉ được sử dụng để khai thác các mặt hàng được mua thường xuyên từ bộ dữ liệu bán lẻ. Có lỗ hổng trong việc phân loại các quy tắc này và dự đoán mục giỏ hàng tiếp theo bằng cách sử dụng DL trên tập dữ liệu giao dịch. Công trình này đề xuất một khung sử dụng AR làm lựa chọn tính năng trong khi áp dụng các phương pháp DL để phân loại và dự đoán. Các thử nghiệm được tiến hành trên hai bộ dữ liệu, InstaCart và dữ liệu thực tế từ Bites Bakers, hoạt động như một cửa hàng đang phát triển với ba chi nhánh và 2233 sản phẩm. AR được phân loại ở mức phân chia 80,20 và 70,30 bằng cách sử dụng CNNN, Bi-LSTM và CNN-BiLSTM. Kết quả xem xét mô phỏng ở cả hai phần tách cho thấy Bi-LSTM hoạt động với độ chính xác cao, khoảng 0,92 trên tập dữ liệu InstaCart. Ngược lại, CNN-BiLSTM hoạt động tốt nhất với độ chính xác khoảng 0,77 trên tập dữ liệu Bites Bakers. |
| 10 | @article{Fageeri2023MBAMB,  title={MBA: Market Basket Analysis Using Frequent Pattern Mining Techniques},  author={Sallam Osman Fageeri and Mohammad A. Kausar and Arockiasamy Soosaimanickam},  journal={International Journal on Recent and Innovation Trends in Computing and Communication},  year={2023},  url={https://api.semanticscholar.org/CorpusID:259884237}  } | Phân tích giỏ thị trường (MBA) này là một kỹ thuật khai thác dữ liệu sử dụng thuật toán khai thác mẫu thường xuyên để khám phá các mẫu xảy ra đồng thời giữa các mặt hàng thường được mua cùng nhau. Nó thường được sử dụng trong các doanh nghiệp bán lẻ và thương mại điện tử để tạo ra các quy tắc kết hợp mô tả mối quan hệ giữa các mặt hàng khác nhau và đưa ra đề xuất cho khách hàng dựa trên giao dịch mua trước đó của họ. MBA là một công cụ mạnh mẽ để xác định các mô hình xảy ra đồng thời và tạo ra những hiểu biết sâu sắc có thể cải thiện các chiến lược tiếp thị và bán hàng. Mặc dù nhiều công việc đã được thực hiện để xử lý chi phí tính toán cho việc khám phá các tập phổ biến, nhưng nó vẫn cần được khám phá và phát triển nhiều hơn. Trong bài báo này, chúng tôi giới thiệu một kỹ thuật cấu trúc dữ liệu dựa trên Bitwise hiệu quả để khai thác mẫu phổ biến trong cơ sở dữ liệu quy mô lớn. Thuật toán quét cơ sở dữ liệu gốc một lần, sử dụng các biểu diễn dữ liệu dựa trên Bitwise cũng như bố cục cơ sở dữ liệu theo chiều dọc, so với thuật toán Apriori và FP-Growth nổi tiếng. Kỹ thuật dựa trên bitwise cải thiện các vấn đề của việc truyền nhiều lần trên cơ sở dữ liệu gốc, do đó giảm thiểu thời gian thực hiện. Các thử nghiệm mở rộng đã được thực hiện để xác thực kỹ thuật của chúng tôi, kỹ thuật này vượt trội hơn Apriori, Éclat, FP-growth và H-mine về thời gian thực hiện Phân tích giỏ thị trường . |
| 11 | @article{Rahayu2023MarketBA, title={Phân tích giỏ thị trường bằng thuật toán tăng trưởng FP để thiết kế chiến lược tiếp thị bằng cách xác định mô hình mua hàng của người tiêu dùng}, tác giả={Silvia Anggun Rahayu}, tạp chí={Tạp chí khoa học dữ liệu ứng dụng}, Year={2023} , url={https://api.semanticscholar.org/CorpusID:257059297} } |  |
| 12 | @article{Hadi2023MarketBA, title={Phân tích giỏ thị trường bằng thuật toán Apriori để tìm chính sách tài chính hiệu quả kết hợp với lập trình R}, tác giả={Isnen Hadi và Al Ghozali và Arief Wibowo}, tạp chí={Giải mã: Jurnal Pendidikan Teknologi Informasi}, năm ={2023}, url={https://api.semanticscholar.org/CorpusID:265174547} } | Chính sách tài khóa thúc đẩy nền kinh tế của một quốc gia và là chính sách hiệu quả nhất để khôi phục nền kinh tế của một quốc gia. Khi suy thoái kinh tế xảy ra, chính sách tài khóa giúp một quốc gia tăng tổng cầu trên thị trường hàng hóa và dịch vụ. Nghiên cứu này đề xuất một chính sách tài khóa kết hợp có thể được thực hiện dựa trên dữ liệu lịch sử. Vì vậy, nghiên cứu này tập trung vào việc sử dụng các luật kết hợp để hỗ trợ những người ra quyết định (các cơ quan quản lý áp dụng các chính sách tài chính phù hợp trong thời đại VUCA (Biến động, Không chắc chắn, Phức tạp, Không rõ ràng) toàn cầu). Vì vậy, phương pháp nghiên cứu thực nghiệm đã được sử dụng trong nghiên cứu này để đưa ra kết quả tốt nhất. Quy tắc kết hợp. Nghiên cứu được thực hiện qua sáu giai đoạn để rút ra kết luận, cụ thể là xác định vấn đề trong mẫu nghiên cứu và tổng quan tài liệu liên quan đến thuật toán tiên nghiệm, thu thập dữ liệu, tiền xử lý dữ liệu, xác định tham số, kết quả nghiên cứu xây dựng thuật toán tiên nghiệm và kiến ​​thức trích xuất được hình thành từ các thuật toán tiên nghiệm. Dựa trên kết quả thử nghiệm sử dụng thuật toán tiên nghiệm, đã thu được 657 quy tắc với độ biến thiên tối thiểu từ hai đến sáu tập mục. Quy tắc 6 tạo ra giá trị trung bình của hiệu suất ngân sách là 90,65. Quy tắc 6 cũng nói rằng giá trị hiệu quả hoạt động của ngân sách có thể được tăng lên bằng cách phân bổ các nguồn tài trợ và đảm bảo rằng chi tiêu hoạt động hiệu quả nhất có thể. |
| 13 | @article{2023AssociaRM, title={Khai thác quy tắc hiệp hội để phân tích giỏ thị trường trên cửa hàng bán lẻ trực tuyến}, tác giả={}, tạp chí={Tạp chí quốc tế về kỹ thuật cơ khí}, năm={2023}, url={https://api .semanticscholar.org/CorpusID:260007060} } |  |
| 14 | @article{Zalmi2023PenerapanMB, title={Penerapan Phân tích giỏ hàng Dengan Metode Apriori Pada WFZ Book Store}, tác giả={Wahyuni ​​Fithratul Zalmi}, tạp chí={INFORMATIKA}, năm={2023}, url={https://api. ngữ nghĩascholar.org/CorpusID:256486464} } | Cửa hàng sách WFZ là Cửa hàng sách cung cấp nhiều loại sách khác nhau. Quá trình bán hàng được thực hiện ngoại tuyến trong đó người tiêu dùng phải đến cửa hàng để mua hàng. Quá trình ghi dữ liệu bằng sổ cái gây khó khăn cho WFZ Book Store trong việc lập báo cáo bán hàng vì phải thu thập dữ liệu giao dịch vào sổ cái, gây khó khăn trong việc biết tồn kho sách hiện có vì phải kiểm tra mọi giao dịch bán hàng trong sổ cái . WFZ Book Store yêu cầu một hệ thống thông tin có thể quản lý dữ liệu bán sách trực tuyến, khách hàng có thể dễ dàng xem dữ liệu sách, tồn kho và đặt chỗ trực tuyến. Với hệ thống thông tin Cửa hàng sách wfz có thể tạo điều kiện thuận lợi cho khách hàng trong việc quản lý dữ liệu sách và lưu trữ dữ liệu khách hàng đã thực hiện giao dịch, Cửa hàng sách WFZ không gặp khó khăn trong việc cung cấp dữ liệu giao dịch mua sách cho khách hàng để dữ liệu có thể được sử dụng cho đề xuất sách bằng cách tìm tập dữ liệu thường xuất hiện nhất trong tập dữ liệu. Kỹ thuật khai thác dữ liệu đã được sử dụng rộng rãi để khắc phục các vấn đề đang tồn tại, một trong số đó là ứng dụng thuật toán tiên nghiệm để tìm luật kết hợp được hình thành từ bộ dữ liệu giao dịch mua sách. Như vậy sẽ biết được mối liên hệ giữa tựa sách đã mua. Quy tắc kết hợp giữa các đầu sách được hình thành từ quá trình khai thác sau này có thể được Cửa hàng sách WFZ sử dụng để tăng số lượng sách được mua, ngoài ra nó có thể được sử dụng để mua sách từ quy tắc kết hợp của những cuốn sách được mua thường xuyên, nó có thể cũng được phát triển thành cơ sở tri thức cho hệ thống gợi ý mua sách. Đầu ra ở dạng phần mềm phân tích thuật toán tiên nghiệm. |
| 15 | @article{VL2023EnhancingCE, title={Nâng cao trải nghiệm khách hàng và hiệu suất bán hàng trong cửa hàng bán lẻ bằng cách sử dụng khai thác quy tắc hiệp hội và phân tích giỏ thị trường}, tác giả={Helen Josephine VL và Dhivya Rajan}, tạp chí={2023 Hội nghị quốc tế lần thứ 14 về truyền thông máy tính và Công nghệ Mạng (ICCCNT)}, Year={2023}, pages={1-5}, url={https://api.semanticscholar.org/CorpusID:265406860} } | Hoạt động kinh doanh bán lẻ tăng trưởng đều đặn hàng năm và tuyển dụng rất nhiều người trên toàn cầu, đặc biệt là với sự phổ biến ngày càng tăng của mua sắm trực tuyến. Tính chất cạnh tranh của lĩnh vực có tốc độ phát triển nhanh này ngày càng thể hiện rõ ràng trong những năm gần đây. Khách hàng mong muốn kết hợp những lợi thế của thói quen mua hàng cũ với sự dễ dàng sử dụng công nghệ mới. Do đó, các nhà bán lẻ phải đảm bảo rằng chất lượng sản phẩm được duy trì khi đáp ứng nhu cầu và yêu cầu của khách hàng. Bài nghiên cứu này chứng minh giá trị tiềm năng của các kỹ thuật phân tích dữ liệu tiên tiến trong việc cải thiện trải nghiệm của khách hàng và hiệu suất bán hàng tại một cửa hàng bán lẻ. Các thuật toán Apriori, FP-Growth và Eclat được áp dụng trong dữ liệu giao dịch theo thời gian thực để khám phá các mối liên hệ và mẫu trong dữ liệu giao dịch. Các tham số hỗ trợ, độ tin cậy và tỷ lệ nâng được sử dụng và thuật toán apriori đưa ra một số bộ mục ứng cử viên có độ dài tăng dần và loại bỏ những bộ không cung cấp hỗ trợ theo ngưỡng yêu cầu. Chúng tôi xác định giá trị nâng cao hơn khi xem xét thịt, sữa và sữa chua đông lạnh. nếu khách hàng quyết định mua chung bất kỳ mặt hàng nào trong số này thì có khả năng khách hàng sẽ mua mặt hàng thứ 3 trong nhóm đó. Nghiên cứu đã đạt được Điểm tin cậy cao dành cho các Mặt hàng như Bánh mì bán thành phẩm và Sữa nên các sản phẩm này nên được bán cùng nhau, Tiếp theo là Thực phẩm đóng gói và bánh cuộn. Khi các nhà bán lẻ tiếp tục phải đối mặt với sự cạnh tranh ngày càng tăng và áp lực phải cải thiện hoạt động của mình, các kỹ thuật nói trên có thể cung cấp cho bạn một công cụ hữu ích để hiểu thói quen và thị hiếu mua hàng của người tiêu dùng, đồng thời sử dụng kiến ​​thức đó để đưa ra các quyết định dựa trên dữ liệu nhằm tối ưu hóa vị trí sản phẩm, nâng cao chất lượng khách hàng. sự hài lòng và thu hút doanh số bán hàng. |
| 16 | @article{Samboteng2022MARKETBA, title={PHÂN TÍCH GIỎ HÀNG THỊ TRƯỜNG CÁC MẪU HÀNH CHÍNH DỮ LIỆU MUA HÀNG CỦA NGƯỜI TIÊU DÙNG SỬ DỤNG CÔNG NGHỆ KHAI THÁC DỮ LIỆU}, tác giả={Lukman Samboteng và Rulinawaty Rulinawaty và M. Rachmat Kasmad và Mutmainnah Basit và Robbi Rahim}, tạp chí={Tạp chí của Khoa học Kỹ thuật Ứng dụng}, năm={2022}, url={https://api.semanticscholar.org/CorpusID:247142858} } | Thực phẩm là thành phần giúp con người trưởng thành, phát triển và đạt được thành tựu. Vì lý do này, chất lượng thực phẩm và các loại thực phẩm phải được xem xét sao cho an toàn khi tiêu dùng và được quản lý. Một số thực phẩm có nguồn gốc thực vật thường được cộng đồng chế biến và tiêu thụ, thậm chí là cần thiết nhất trong chế biến thực phẩm. Trong trường hợp này, nghiên cứu được thực hiện bằng cách sử dụng khai thác dữ liệu với thuật toán phân tích giỏ hàng thị trường để thu được thông tin rất có giá trị nhằm quyết định việc tồn kho loại nguyên liệu cần thiết. Phương pháp Phân tích dựa trên thị trường được sử dụng để phân tích tất cả dữ liệu và tạo mẫu cho từng dữ liệu. Một phương pháp Phân tích dựa trên thị trường được đề cập là quy tắc kết hợp với thuật toán tiên nghiệm. Thuật toán này tạo ra các giao dịch bán hàng có mối liên kết chặt chẽ giữa các mặt hàng trong giao dịch được sử dụng làm đề xuất bán hàng giúp người dùng (chủ sở hữu) nhận được đề xuất khi người dùng xem chi tiết về bộ mặt hàng đã mua. Từ kết quả của các thử nghiệm trong nghiên cứu này, người ta nhận thấy rằng độ hỗ trợ tối thiểu (minsup) và độ tin cậy tối thiểu (minconf) càng lớn thì càng mất ít thời gian để đưa ra khuyến nghị và càng đưa ra ít khuyến nghị hơn, nhưng các khuyến nghị được đưa ra đều đến từ giao dịch thường xuyên xuất hiện. |
| 17 | @inproceedings{Javier2022TemporalVP, title={Hồ sơ trực quan tạm thời về phân tích giỏ thị trường}, tác giả={Francisco Javier và Moreno Arboleda và Giovanni P{\'e}rez Ortega và Jaime Alberto và Guzm{\'a}n Luna}, năm ={2022}, url={https://api.semanticscholar.org/CorpusID:260515634} } | Phân tích giỏ thị trường cho phép các nhà phân tích hiểu được hành vi của khách hàng. Trong bài báo này, chúng tôi đề xuất một kỹ thuật mới để tạo hồ sơ khách hàng liên quan đến lịch sử mua sản phẩm của họ. Sau khi có được hồ sơ của khách hàng, chúng tôi trình bày một kỹ thuật trực quan để so sánh hồ sơ của khách hàng một cách thân thiện và tương tác theo thời gian. Để cho thấy tính khả thi của đề xuất của chúng tôi, chúng tôi đã thực hiện thử nghiệm với tập dữ liệu về hoạt động mua hàng của thị trường bán lẻ . Kết quả cho thấy đề xuất của chúng tôi có thể hữu ích cho việc lập kế hoạch tồn kho và xác định khách hàng có xu hướng mua hàng tương tự. |
| 18 | @article{Umar2022MarketBA, title={Phân tích giỏ thị trường Quy tắc của Hiệp hội Menggunakan dan Algoritma Apriori Pada Produk Penjualan Mitra Swalayan Salatiga}, tác giả={Elfira Umar và Daniel HF Manongga và Ade Iriani}, tạp chí={JURNAL MEDIA INFORMATIKA BUDIDARMA}, Year= {2022}, url={https://api.semanticscholar.org/CorpusID:253001514} } | Phân tích Rổ thị trường đang học cách quản lý các hiệp hội xử lý dữ liệu trong các lĩnh vực khác nhau. Mục đích chính của phân tích Giỏ hàng trong lĩnh vực bán hàng là truyền tải một thông điệp quan trọng đến công ty để công ty có thể tìm ra mô hình hành vi nhập hàng vào giỏ hàng của người tiêu dùng để đối tác đưa ra quyết định. Trong nghiên cứu này, Thuật toán Apriori được sử dụng để tính đến những thay đổi xảy ra trong dữ liệu. Nghiên cứu này thảo luận về các kỹ thuật khai thác dữ liệu trong việc phân tích những mặt hàng nào được người tiêu dùng thường mua cùng lúc nhất để họ có thể thay đổi vị trí các mặt hàng gần nhau nhằm tăng hiệu ứng mua hàng ngẫu hứng. Kết quả thu được là 5 quy tắc trong đó một trong các quy tắc đạt được giá trị tin cậy cao nhất khi mua thuốc lá, mặt hàng chiếm ưu thế được lấy đồng thời, đó là trứng bằng cách đạt được giá trị tin cậy có thể đáp ứng yêu cầu độ tin cậy cao nhất, cụ thể là 67%. |
| 19 | @article{Idris2022ComparisonOA, title={So sánh các thuật toán Apriori, Apriori-TID và FP-Tăng trưởng trong Phân tích giỏ thị trường tại các cửa hàng tạp hóa}, tác giả={Andi Ilhamsyah Idris và Eliyah Acantha M Sampetoding và Valian Yoga Pudya Ardhana và Irene Maritsa và Adrisumatri Sakri và Hidayatullah Ruslan và Esther Sanda Manapa}, tạp chí={The IJICS (Tạp chí Quốc tế về Tin học và Khoa học Máy tính)}, Year={2022}, url={https://api.semanticscholar.org/CorpusID:257018537} } | Phân tích giỏ thị trường là phân tích hành vi của người tiêu dùng cụ thể từ một nhóm / nhóm nhất định. Phân tích giỏ thị trường thường được sử dụng làm điểm khởi đầu để tìm kiếm kiến ​​thức từ giao dịch dữ liệu khi chúng ta không biết mình đang tìm kiếm mô hình cụ thể nào. Phân tích giỏ thị trường trong nghiên cứu này được áp dụng để tìm kiếm các mô hình mua hàng tạp hóa tại các cửa hàng tạp hóa và sau đó phân tích theo mùa. Nghiên cứu này nhằm mục đích so sánh các phương pháp Apriori, Apriori TID và FP-Growth trong việc xác định hành vi giao dịch của người tiêu dùng và tính toán số lượng giao dịch của người tiêu dùng trong một số mùa dựa trên dữ liệu thu được từ cơ sở dữ liệu Phân tích giỏ thị trường . Trong kết quả nghiên cứu này, được biết, FP-Growth có hiệu suất tốt nhất trong số hai thuật toán còn lại, nhưng sử dụng nhiều bộ nhớ hơn các thuật toán khác. Thuật toán Apriori-TID sử dụng bộ nhớ nhẹ hơn và nhanh hơn Thuật toán Apriori |
| 20 | @article{Jirapatsil2022MarketBA,  title={Market Basket Analysis for Fresh Products location improvement: A case study of E-Commerce Business Warehouse},  author={Pattraporn Jirapatsil and Naragain Phumchusri},  journal={Proceedings of the 4th International Conference on Management Science and Industrial Engineering},  year={2022},  url={https://api.semanticscholar.org/CorpusID:250624756}  } | Phân tích giỏ thị trường (MBA) sử dụng kỹ thuật khai thác dữ liệu làm công cụ phân tích để hiểu mối liên hệ giữa nhiều mặt hàng. Nó là một công cụ hữu ích để trích xuất thông tin từ lượng lớn dữ liệu trong nhiều lĩnh vực công nghiệp, ví dụ: tạp hóa, siêu thị, nhà bán lẻ, nhà kho, phòng trưng bày di động, thư viện, vườn thú, v.v. Công ty nghiên cứu điển hình bán sản phẩm tươi sống trong kinh doanh E-commence và hiện tại vị trí sản phẩm trong kho không hiệu quả, gây ra sự chậm trễ trong quá trình lấy hàng. Vì vậy, mục tiêu của bài viết này là đề xuất phương pháp phân tích giỏ hàng thị trường để hiểu rõ hơn về các giao dịch lịch sử, một bộ dữ liệu ghi lại kết quả liên quan đến hoạt động mua bán của công ty nghiên cứu điển hình. Thuật toán Apiori được áp dụng cho các quy tắc kết hợp để phân tích 2366 dữ liệu giao dịch trong khoảng thời gian từ tháng 7 đến tháng 12 năm 2021. Kết quả phân tích dữ liệu sau đó được sử dụng để sắp xếp lại vị trí sản phẩm trong kho nhằm giảm khoảng cách lấy hàng trung bình trên mỗi đơn hàng. Kết quả cho thấy khoảng cách trung bình trên mỗi đơn hàng có thể giảm 54,4%. |
| 21 | @article{Ma2022NovelNR, title={Phương pháp đề xuất nhóm tiếp theo mới dựa trên thông tin giỏ hàng tuần tự}, tác giả={Li-Ching Ma}, tạp chí={Nghiên cứu thương mại điện tử}, năm={2022}, khối lượng={23} , pages={2399-2418}, url={https://api.semanticscholar.org/CorpusID:247573177} } | A market basket is a set of items included in a retail assortment that a customer buys on a shopping trip. The purpose of market basket analysis is to persuade a customer to spend more money through upselling or cross-selling. Most recommendation systems only suggest a single next-item or the top n items that a customer is most likely to buy. A company might succeed in convincing a customer to spend more money to increase sales revenue if a recommendation system can suggest the next or top n groups of items that customers are likely to buy according to the items in their basket. Based on the similarity upper approximation clustering, Borda majority count and PrefixSpan algorithm, this paper proposes a novel next-group recommendation approach according to sequential market basket information. Compared with the previous methods, the proposed approach can provide next-group instead of next-item recommendation, which may create more opportunities for customers to increase their spending. |
| 22 | @inproceedings{Putrevu2022MarketBA,  title={Market Basket Analysis Using Apriori Algorithm with Pruning Approach},  author={Umamaheswararao Putrevu and Chandrasekhar Pj},  year={2022},  url={https://api.semanticscholar.org/CorpusID:247625723}  } | This research paper is about Market Basket Analysis, an important component of Business Analytics in retail companies to determine the sales for different segments of customers to improve customer satisfaction and to increase profitability of the company. This is totally done by association rule mining in which it analyses the customer behaviour against the purchasing item from market. It analyses the customer purchasing pattern and generate frequent Itemset. After generation of frequent Itemset it is easy to find most popular Itemset and least priority Itemset from large transactional database instead of reading it manually. Generation of frequent Itemset will enhance the market strategy, placement of goods in an organized manner and many more. Market Basket Analysis helps in increase in sales of goods and for profitable business. In the present study Market Basket Analysis for a leading shopping mall is studied and analysed using frequent Itemset mining and decision tree techniques. The frequent Itemset are extracted from the market basket database using the efficient apriori algorithm and generated association rules to discover product associations and base for retailer‟s promotion strategy on them. Market basket analysis is one possible way to find out which items can be put together in super markets. Pruning of association rules resulted in best outcomes. |
| 23 | @article{Priyanto2022IMPLEMENTATIONOM,  title={IMPLEMENTATION OF MARKET BASKET ANALYSIS WITH APRIORI ALGORITHM IN MINIMARKET},  author={Abdul Hafiidh Priyanto and Amalia Beladinna Arifa},  journal={Jurnal Teknik Informatika (Jutif)},  year={2022},  url={https://api.semanticscholar.org/CorpusID:256128251}  } | The rapid growth of the retail business has an impact on increasing the economic growth of the community. The retail business has high profit potential in areas that have a large population such as Indonesia. A retail business that is popular among the public is a modern market retail business or convenience store. With the rapid growth, it gives a tendency between convenience stores to compete. By designing a marketing strategy is one of the efforts to win the competition in supermarkets. Management needs to understand the purchase behavior made by customers, this action is useful to find out the products that customers are popularly buying. Association algorithm is a form of algorithm in the field of data mining that serves to provide correlation between one item and another. there are several popular algorithms in applying association algorithms one of which is the a priori algorithm created by Agrawal and Srikant in 1994. To support the understanding of customer purchase patterns, it is necessary to implement market basket analysis that has the ability to recognize pattern patterns from transaction data in a convenience store. Performance in market basket analysis also needs to be tested to handle a lot of transaction data, considering that the recording of sales transaction data continues to run over time. The implementation carried out using flask is one of the implementations that is relevant to technological developments, this implementation results in a relatively short data speed with the factor that the magnitude of transaction data is middle to lower, which is 14,963 transaction data. |
| 24 | @article{MAlawadh2022ASO,  title={A Survey on Methods and Applications of Intelligent Market Basket Analysis Based on Association Rule},  author={Monerah M. Alawadh and Ahmed M. Barnawi},  journal={Journal on Big Data},  year={2022},  url={https://api.semanticscholar.org/CorpusID:248651754}  } |  |
| 25 | @article{Nuraeni2022OPTIMIZATIONOM,  title={OPTIMIZATION OF MARKET BASKET ANALYSIS USING CENTROID-BASED CLUSTERING ALGORITHM AND FP-GROWTH ALGORITHM},  author={Fitri Nuraeni and Dewi Tresnawati and Yoga Handoko Agustin and Gisna Fauzi},  journal={Jurnal Teknik Informatika (Jutif)},  year={2022},  url={https://api.semanticscholar.org/CorpusID:256127322}  } | The proliferation of the food and beverage sales business requires the creativity of business owners to offer their flagship products to every consumer, both new and subscribed consumers. A large number of menu choices makes the ordering process long because consumers are confused about which menu will be the best choice. the seller to be able to provide the right recommendations so that orders can take place faster. Shopping cart analysis is an activity that has often been done to find out the items found that are sold simultaneously. The FP-Growth association method is a faster algorithm for generating association rules, but the association process in large dataset sizes tends to add large items so that the accuracy value of association rules decreases. So that in this study, the grouping of datasets was carried out using a clustering model with a centroid-based algorithm, namely k-means, k-medoids, and fuzzy c-means. This research was conducted through dataset collection, dataset preparation, clustering modeling, evaluation of clustering models using DBI and silhouette index, association modeling, and evaluation of association models using lift ratio. The results of this study showed that the clustering model with the best DBI and silhouette index values ​​was at k=3 for k-means, k=2 for k-medoids, and k=7 for fuzzy c-means. The number of association rules is generated from the grouped data set using fuzzy c-means, but the highest average lift ratio is in the association rules generated from the grouping data set using k-means. From the association model using k-means and FP-Growth, 32 unique association rules were found with the 4 most frequently found items, namely cireng chili oil, regal milk coffee, banana cheese, and vietnam drip. |
| 26 | @article{2022AnalysisOD,  title={Analysis of Data Mining at Supermarket X in Surabaya Using Market Basket Analysis to Determine Consumer Buying Patterns},  author={},  journal={Nusantara Science and Technology Proceedings},  year={2022},  url={https://api.semanticscholar.org/CorpusID:255697563}  } |  |
| 27 | @article{Karnila2022MARKETBA,  title={MARKET BASKET ANALYSIS ON TRANSACTION DATA USING THE APRIORI ALGORITHM},  author={Sri Karnila and Akbar Rizkyandi and Rio Kurniawan and Nurjoko Nurjoko},  journal={Jurnal TAM (Technology Acceptance Model)},  year={2022},  url={https://api.semanticscholar.org/CorpusID:252757734}  } | This research aims to get information about the relationship between sales patterns carried out by CV. Dian Abadi Jaya workshop by using APRIORI algorithms through transaction data sets carried out by customers. The subject of research is a record of shopping cart transactions made by customers, namely vehicle parts sales transactions and vehicle repair service transactions. The data collection techniques used are interviews and documentation. The criteria used in this research are a minimum of frequent itemset of 20 transactions with support criteria of 1,7%, confidence value of 40% and lift ratio value above 1. The results of the research have produced 9 sales pattern relationships with the highest confidence of 100%. The results that have been obtained are expected to help the CV. Dian Abadi Jaya workshop in making a decision for the next sale. |
| 28 | @article{Fitzsimon2022ASV,  title={A Shapley Value Index for Market Basket Analysis: Efficient Computation Using an Harsanyi Dividend Representation},  author={Jayden Fitzsimon and Shrikant Agrawal and K. K. Khade and Evan Shellshear and Jonathon Allport and Archie C. Chapman},  journal={IGTR},  year={2022},  volume={24},  pages={2250015:1-2250015:29},  url={https://api.semanticscholar.org/CorpusID:248804402}  } |  |
| 29 | @article{Trklmez2022UtilityfrequencyAU,  title={Utility-frequency analysis using customer market basket data in the retail sector},  author={G{\"o}kçe Baysal T{\"u}rk{\"o}lmez and Ipek Deveci Kocakoç},  journal={Int. J. Bus. Inf. Syst.},  year={2022},  volume={41},  pages={69-81},  url={https://api.semanticscholar.org/CorpusID:248383542}  } |  |
| 30 | @inproceedings{Nurmayanti2021MarketBA,  title={Market Basket Analysis with Apriori Algorithm and Frequent Pattern Growth (Fp-Growth) on Outdoor Product Sales Data},  author={Wiwit Pura Nurmayanti and Hanipar Mahyulis Sastriana and Abdul Rahim and Muhammad Munawir Gazali and Ristu Haiban Hirzi and Zuhut Ramdani and Muhammad Malthuf},  year={2021},  url={https://api.semanticscholar.org/CorpusID:233418197}  } | Indonesia is an equatorial country that has abundant natural wealth from the seabed to the top of the mountains, the beauty of the country of Indonesia also lies in the mountains that it has in various provinces, for example in the province of West Nusa Tenggara known for its beautiful mountain, namely Rinjani. The increase in outdoor activities has attracted many people to open outdoor shops in the West Nusa Tenggara region. Sales transaction data in outdoor stores can be processed into information that can be profitable for the store itself. Using a market basket analysis method to see the association (rules) between a number of sales attributes. The purpose of this study is to determine the pattern of relationships in the transactions that occur. The data used is the transaction data of outdoor goods. The analysis used is the Association Rules with the Apriori algorithm and the frequent pattern growth (FP-growth) algorithm. The results of this study are formed 10 rules in the Apriori algorithm and 4 rules in the FP-Growth algorithm. The relationship pattern or association rule that is formed is in the item "if a consumer buys a portable stove, it is possible that portable gas will also be purchased" at the strength level of the rules with a minimum support of 0.296 and confidence 0.774 at Apriori and 0.296 and 0.750 at FP-Growth. |
| 31 | @article{Qisman2021MarketBA,  title={Market basket analysis using apriori algorithm to find consumer patterns in buying goods through transaction data (case study of Mizan computer retail stores)},  author={M Qisman and Rudi Rosadi and Atje Setiawan Abdullah},  journal={Journal of Physics: Conference Series},  year={2021},  volume={1722},  url={https://api.semanticscholar.org/CorpusID:234119984}  } | Mizan Computer Shop is a shop that is engaged in the trading sector, especially in the field of selling computers and supporting accessories. Growing and increasing number of business actors in the computer sector, can makes the players challenged to be able to create unique differentiation and clear positioning. So, that consumers can differentiate from their competitors. Competitive and dynamic market conditions make every company should always observe competition in their business environment. Retail stores need to use all of available resources including data. Data processing is expected to be able to provide information that can be used to support marketing strategies. One of the data processing methods that are often used in marketing strategies is the use of data mining techniques i.e Market Basket Analysis using a priori algorithm. The application is designed using the waterfall method which starts from analyzing user needs, designing a process using UML which consists of: Use Case Diagrams, Activity Diagrams and Sequence Diagrams . This Market Basket Analysis application was built using the PHP programming language. From the results of the analysis in this research, it can be concluded that for the combination of 2 items with the highest confidence value 100% and the lift ratio value 3.39 i.e if a consumer buys a Laptop Charger, he will also buy a keyboard and for a combination of 3 items with the highest confidence value 100% and the lift ratio value 2.17 i.e if a consumer buys a Joystick and Laptop, he will also buy a mouse. |
| 32 | @article{Firmansyah2021MarketBA,  title={Market Basket Analysis for Books Sales Promotion using FP Growth Algorithm, Case Study : Gramedia Matraman Jakarta},  author={Firmansyah Firmansyah and Agus Yulianto},  journal={Journal of Information Technology Education},  year={2021},  volume={4},  pages={383-392},  url={https://api.semanticscholar.org/CorpusID:234151819}  } | Data Mining is the process of extracting information or something interesting from the data in the database so as to produce valuable information using techniques such as clustering, estimation, description, and others. Based on observations at AB Mart, there were 44 product items whose data was not revealed. This problem will be solved using data mining analysis. The purpose of this research is to apply market basket analysis to the sale of goods at AB Mart with the a priori algorithm. This research uses a clear structure of the framework, namely problem identification, literature study, data collection, calculation & analysis of association rules with a priori algorithm, forming association rules and making reports. The results of the sales transaction of AB Mart in August resulted in or generated relationships between shopping product items where the% purchase of Pepsodent was 115%, Frisian Flag 96%, Sugar 96%, Indomilk 93%, and Nasi Jempol 91%. The conclusion of this research is using Weka software with a priori algorithm which produces an association relationship between pepsodent goods and the number of transactions purchased |

# Giới thiệu

## Đặt vấn đề

Hành vi khách hàng là một lĩnh vực nghiên cứu đầy hấp dẫn, tập trung vào việc giải mã cách thức mà cá nhân, nhóm hay tổ chức lựa chọn, sử dụng và loại bỏ sản phẩm, dịch vụ, ý tưởng hay trải nghiệm để thỏa mãn nhu cầu và mong muốn của họ. Hiểu được hành vi khách hàng là chìa khóa giúp doanh nghiệp đưa ra những quyết định kinh doanh sáng suốt, từ đó phát triển sản phẩm phù hợp, tăng hiệu quả marketing, cải thiện trải nghiệm khách hàng và ultimately, gia tăng doanh thu và lợi nhuận. Nghiên cứu hành vi khách hàng cung cấp cho doanh nghiệp một bức tranh toàn cảnh về hành trình mua sắm của khách hàng, từ khi nhận thức nhu cầu cho đến khi sử dụng và đánh giá sản phẩm. Bằng cách áp dụng các phương pháp nghiên cứu như khảo sát, phỏng vấn, nhóm tập trung hay phân tích dữ liệu, doanh nghiệp có thể thu thập thông tin quý giá về sở thích, nhu cầu, động lực và hành vi của khách hàng. Nghiên cứu hành vi khách hàng mang lại nhiều lợi ích cho doanh nghiệp như phát triển sản phẩm và dịch vụ phù hợp, tăng hiệu quả marketing, cải thiện trải nghiệm khách hàng và tăng doanh thu và lợi nhuận. Nói tóm lại, hành vi khách hàng là một lĩnh vực nghiên cứu đầy tiềm năng và mang lại lợi ích to lớn cho doanh nghiệp. Doanh nghiệp nên đầu tư vào việc nghiên cứu hành vi khách hàng để đưa ra những quyết định kinh doanh sáng suốt, gia tăng lợi thế cạnh tranh và gặt hái thành công trong thị trường đầy biến động hiện nay.

## Giới thiệu

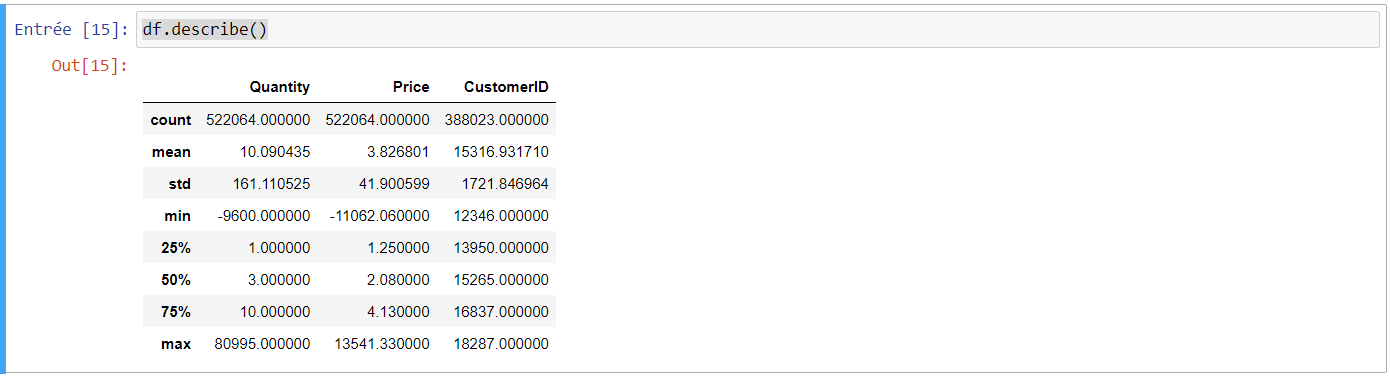
Ngành thương mại điện tử đang phát triển mạnh mẽ trên toàn cầu. Trước thách thức và cơ hội từ sự bùng nổ này, nghiên cứu được tập trung vào phân tích giỏ mua hàng của một cửa hàng E-commerce hướng tới việc phục vụ khách hàng toàn cầu, nhưng chủ yếu tập trung vào thị trường nội địa Vương quốc Anh. Kỹ thuật phân tích giỏ hàng thị trường (Market Basket Analysis) là kỹ thuật khai thác dữ liệu nhằm khám phá các mô hình mua sắm trong thị trường bán lẻ. Nó hoạt động bằng cách phân tích các "giỏ hàng" (bộ sưu tập các sản phẩm được mua cùng nhau) để xác định mối liên hệ tiềm ẩn giữa các mặt hàng. Vì vậy kỹ thuật được áp dụng để hoàn thành mục tiêu chính của nghiên cứu là khám phá những mẫu mua sắm tiềm ẩn trong dữ liệu giao dịch, từ đó hiểu rõ hơn về hành vi của khách hàng và đưa ra các chiến lược kinh doanh có hiệu quả cao. Ngoài ra để đạt được mục tiêu này, nghiên cứu sử dụng hai thuật toán phổ biến trong việc khai phá dữ liệu là Apriori và FP-Growth. Apriori được sử dụng để xác định những nhóm sản phẩm thường được mua cùng nhau, trong khi FP-Growth sử dụng cấu trúc cây FP để biểu diễn dữ liệu giao dịch, vượt trội hơn Apriori trong việc phát hiện những mẫu mua sắm phức tạp.

Nghiên cứu phân tích giỏ hàng mang lại lợi ích to lớn. Việc hiểu rõ hành vi khách hàng giúp xác định sản phẩm thường đi kèm, theo dõi xu hướng mua sắm theo thời gian và khu vực. Nhờ đó, chiến lược marketing được cá nhân hóa hiệu quả hơn, với khả năng đề xuất sản phẩm phù hợp và nhắm mục tiêu quảng cáo chính xác. Hơn nữa, nghiên cứu này còn hỗ trợ quản lý hiệu quả cho cửa hàng E-commerce và là nguồn thông tin quý giá để phát triển sản phẩm mới. Bằng cách xác định chính xác nhu cầu thị trường, cửa hàng có thể đưa ra các sản phẩm mới phù hợp với mong muốn và yêu cầu của khách hàng.

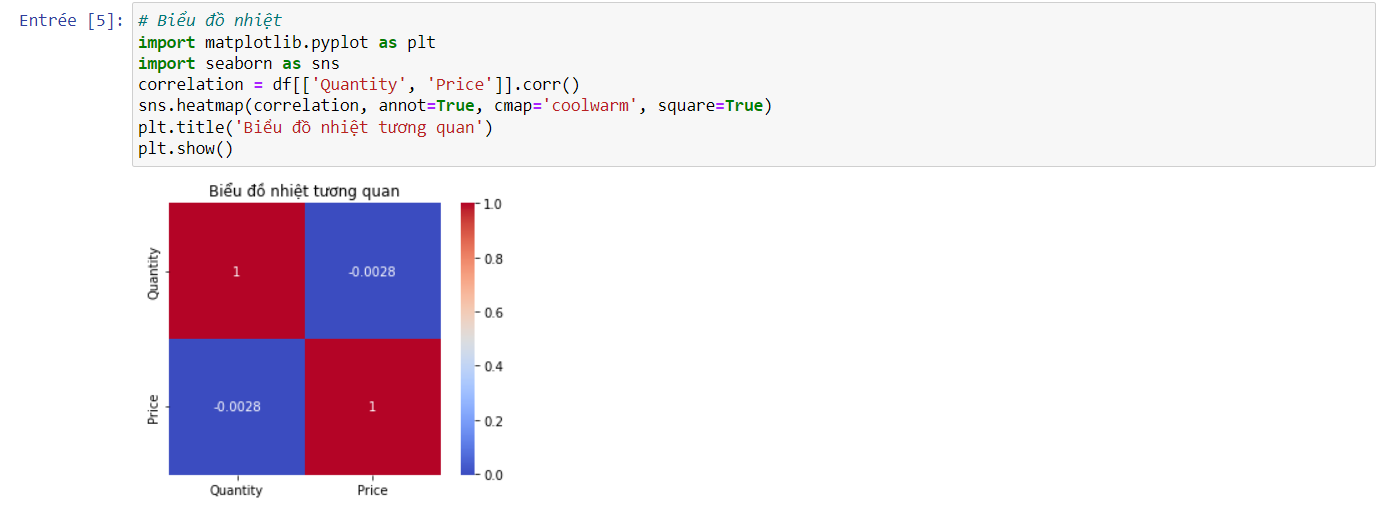
Tóm lại, phân tích giỏ mua hàng với Apriori và FP-Growth không chỉ là một công cụ hiệu quả để hiểu rõ hơn hành vi của khách hàng mà còn là cơ sở để xây dựng những chiến lược kinh doanh đầy hiệu quả cho cửa hàng E-commerce. Mặc dù nghiên cứu này tập trung vào thị trường Vương quốc Anh, những phương pháp và kết quả có thể áp dụng cho nhiều thị trường khác với những điều chỉnh và thích ứng phù hợp.

# Miêu tả dữ liệu

## Tổng quan:

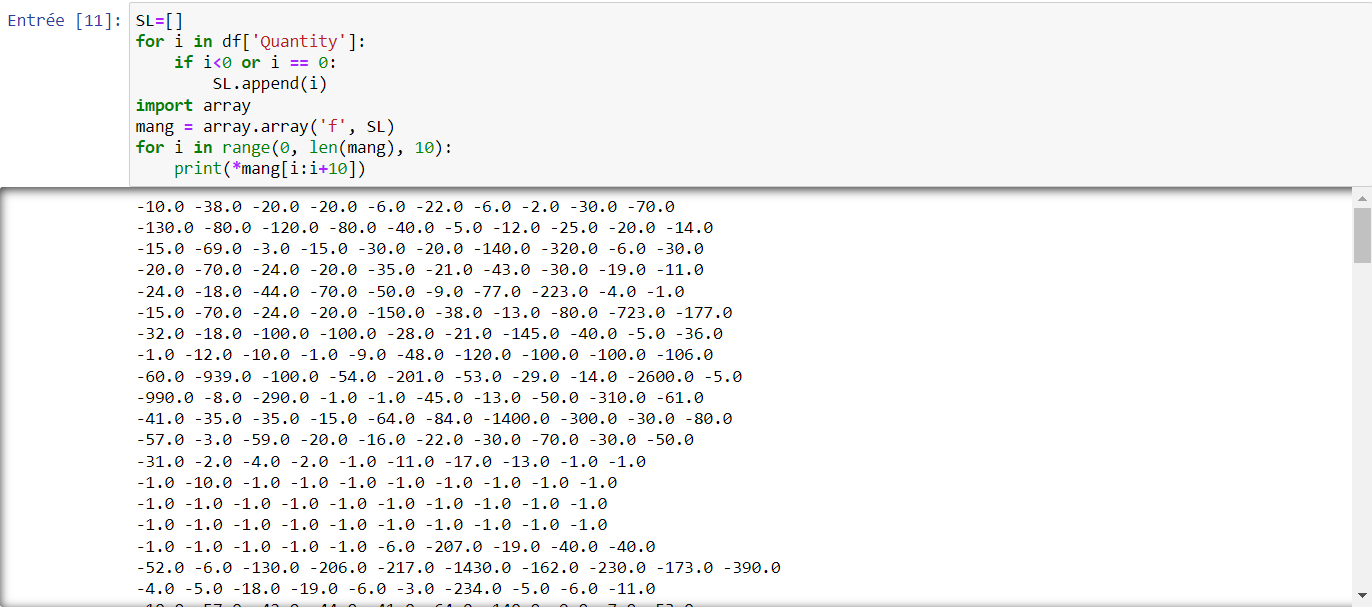
* Nguồn dữ liệu: Dữ liệu được cung cấp bởi giảng viên bộ môn
* Kết cấu dữ liệu:
* Dữ liệu chứa 7 thuộc tính và 522063 bản ghi
* 7 thuộc tính gồm:
* BillNo: mã hóa đơn của khách hàng
* Item name: danh mục sản phẩm khách hàng đã mua
* Quantity: số lượng đã mua
* Date: thời gian mua
* Price: giá sản phẩm
* CustomerID: mã khách hàng
* Country: quốc gia
* Miêu tả dữ liệu

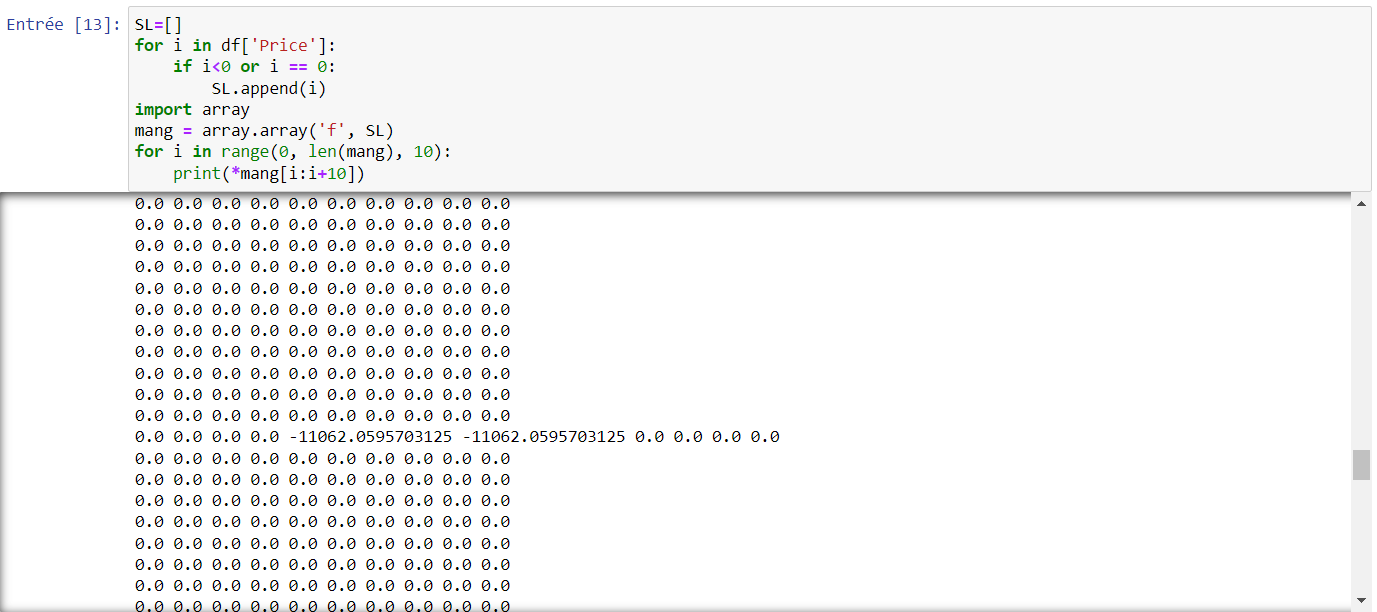
## Phân tích đặc điểm dữ liệu:

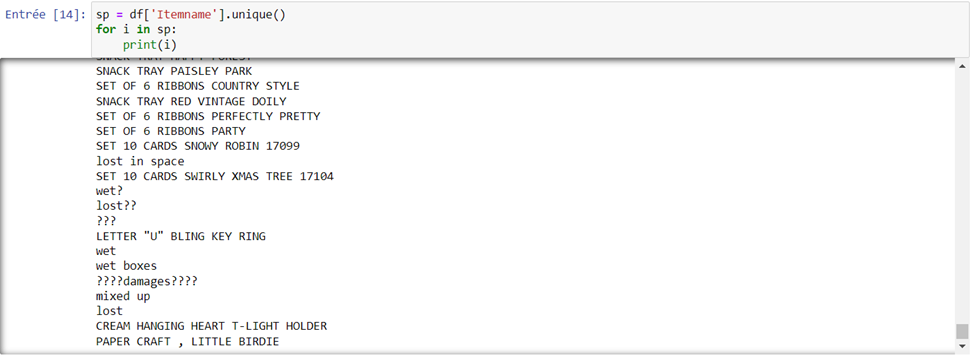
* Độ tương quan giữa các giá trị: có thể thấy dữ liệu có độ tin tưởng cao

## Các vấn đề của dữ liệu:

* Các dữ liệu trống: gồm 1455 bản ghi tại danh mục hàng đã mua và mã khách hàng thiếu 134041 bản ghi.



* Tại thuộc tính Quantity (số lượng mua): có sự sai sót dữ liệu khi số lượng nhỏ hơn hoặc bằng 0
* Tại thuộc tính Price (giá bán): cũng có sự sai sót dữ liệu khi giá bán nhỏ hơn 0



* Tại thuộc tính Item name (danh mục sản phẩm khách hàng đã mua): có sự sai sót dữ liệu khi có những bản ghi không đúng tên sản phẩm hoặc mất dữ liệu.

# Phương pháp thực hiện

## THUẬT TOÁN APRIORI

* Khái niệm: Là 1 thuật toán khai thác dữ liệu để tìm các tập hạng mục phổ biến và các luật kết hợp trong 1 cơ sở dữ liệu. Thuật toán này sử dụng kiến thức đã có từ trước(prior) về các thuộc tính của tập hạng mục phổ biến để giảm không gian tìm kiếm.
* Quy tắc:

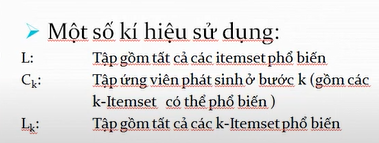
Gán nhãn cho các hạng mục: vd: A- quả táo, B- bánh mì, C- quả chuối, D- trứng, E - ếch => tập các hạng mục tập { ABC},{...}

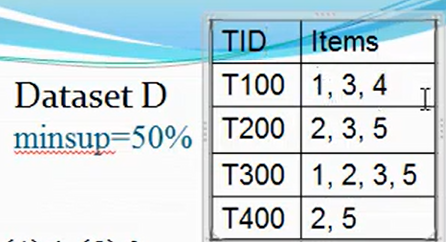
=> mẫu phổ biến : vd{B,E} xuất hiện nhiều trong các giao dịch mua

tập K đi kèm : vd: {A,B,C} danh sách cặp sản phẩm đi kèm như { A,B},{A,C}; danh sách 3 sản phẩm đi kèm {A,B,C}, {B,C,E}

=> Tập phổ biến đk: phổ biến tối thiếu là 40% hoặc xuất hiện 5 lần

* Gồm 2 giai đoạn:

1. Tìm tập L gồm tất cả các Itemset phổ biến ( có độ hộ trợ không nhỏ hơn min supp)
2. Phát sinh các luật từ phổ biến I

Ví dụ:

### Cách chạy của thuật toán:

#### Chạy duyệt Database: C1: {1}: 2; {2}: 3; {3}: 3; {4}: 1, {5}: 3

sau khi đếm xong, duyệt dựa trên minsup=50%=2

=> loại {4} do 1<2

L1: {1}: 2; {2}: 3; {3}: 3;, {5}: 3

Phát sinh C2 bằng cách ghét đôi lần lượt không trùng tập với nhau:

C2: {1,2}, {1,3}, {1,5}, {2,3}, {2,5}, {3,5}

Duyệt database lần 2: C2:  {1,2}: 1, {1,3}: 2, {1,5}: 1, {2,3}: 2, {2,5}: 3, {3,5}: 2

Sau khi đếm xong, duyệt tiếp ms=2

=> L2: {1,3}: 2, {2,3}: 2, {2,5}: 3, {3,5}: 2

Tương tự như vậy:

C3: {1,2,3}:1 , {1,3,5}: 1, { 2,3,5}: 2

=> L3: { 2,3,5}: 2

=> (Tính chất của APRIORI : **mọi tập con khác rỗng của 1 Itemset phổ biến cũng phải là 1 Itemset phổ biến)**

Ý nghĩa của  APRIORI: dựa vào tính chất của  APRIORI để giảm kích cỡ của tập ứng viên

## THUẬT TOÁN FP - GROWTH

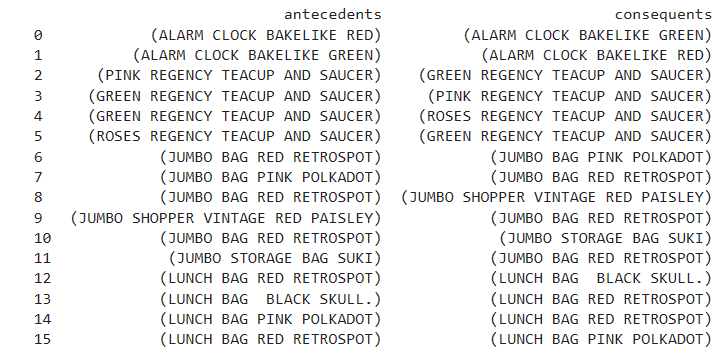
Thuật toán Apriori có một thiếu sót lớn. Sử dụng Apriori yêu cầu quét nhiều lần cơ sở dữ liệu để kiểm tra số lượng hỗ trợ của từng mục và các tập mục. Khi cơ sở dữ liệu lớn, điều này sẽ tiêu tốn một lượng đáng kể I / O đĩa và sức mạnh tính toán. Do đó, thuật toán FP-Growth được tạo ra để khắc phục sự thiếu hụt này. Nó chỉ quét cơ sở dữ liệu hai lần và sử dụng cấu trúc cây (FP-tree) để lưu trữ tất cả thông tin. Gốc đại diện cho null, mỗi nút đại diện cho một mục, trong khi sự liên kết của các nút là các tập phổ có thứ tự được duy trì trong khi hình thành cây. FP-tree ngắn gọn và được sử dụng để tạo trực tiếp các tập phổ biến lớn. Khi cây FP đã được xây dựng, nó sử dụng phương pháp phân chia và chinh phục đệ quy để khai thác các tập phổ biến.

FP-growth là phiên bản cải tiến của Thuật toán Apriori được sử dụng rộng rãi để khai thác mẫu thường xuyên (Khai thác theo quy tắc của Hiệp hội AKA). Nó được sử dụng như một quá trình phân tích để tìm ra các mẫu hoặc liên kết thường xuyên từ các tập dữ liệu. Ví dụ: dữ liệu giao dịch của cửa hàng tạp hóa có thể có mô hình thường xuyên là mọi người thường mua khoai tây chiên và bia cùng nhau. Thuật toán Apriori tạo ra các mẫu thường xuyên bằng cách tạo các tập phổ biến và khám phá tập phổ biến nhất qua một ngưỡng “số lượng hỗ trợ tối thiểu”. Nó làm giảm đáng kể kích thước của tập phổ biến trong cơ sở dữ liệu bằng một nguyên tắc đơn giản:

**Nếu một tập phổ biến là thường xuyên, thì tất cả các tập con của nó cũng phải thường xuyên.**

* Thuật toán FP- Growth dùng cấu trúc cây (FP - Tree) để nén CSDL chứa các item phổ biến
* Sau khi xây dựng được FP - Tree đó rồi, thì chia nhỏ Database lớn ra để khai thác các itemset.

# Phân tích, so sánh các kết quả:

Chúng tôi đã sử dụng 2 phương pháp là Apriori và FP-Growth để áp dụng với dữ liệu được cung cấp, kết quả thu được là như trên và như nhau.

Chúng tôi đã tìm được 15 nhóm sản phẩm phổ biến nhất dựa trên danh sách hóa đơn của khách hàng đã cung cấp với những hệ số như trên, cần chú ý tới chỉ số tin cậy là trên 60% và thước đo trên 86%.

Do kết quả là như nhau vì vậy có thể đánh giá 2 phương pháp trên đã áp dụng hiệu quả trên bộ dữ liệu là như nhau, sau đây là đánh giá so sánh về tính tối ưu giữa 2 phương pháp:

Dưới đây là một so sánh về ưu nhược điểm giữa hai đoạn mã:

Đối với phương pháp Apriori:

* Phương pháp:
* Sử dụng defaultdict(list) để nhóm các mục theo giá trị của cột “BillNo”
* Tạo một từ điển mới để lưu trữ dữ liệu nhóm
* Sử dụng zip để lặp qua dữ liệu, điều này phù hợp khi DataFrame ta tạo mới ở trên “df1” có cùng kích thước với cột “BillNo” và “Itemname”.
* Tận dụng các cấu trúc dữ liệu python cơ bản như danh sách và từ điển
* Ưu điểm:
* Dễ hiểu và dễ triển khai
* Phù hợp với các bộ dữ liệu nhỏ và đơn giản
* Không cần thư viên bên ngoài
* Nhược điểm:
* Khả năng mở rộng và xử lý dữ liệu bị giới hạn
* Không cung cấp các thuật toán tối ưu hóa cho việc tìm mẫu phổ biến và quy tắc kết hợp

Đối với phương pháp FP – growth:

* Phương pháp:
* Sử dụng groupby để nhóm các mục theo giá trị của cột “BillNo”
* Sử dụng mlxtend để thực hiện phân tích mẫu phổ biến và quy tắc kết hợp sử dụng FP – growth
* Có thể sử dụng cho các bộ dữ liệu lớn với hiệu suất cao hơn so với các cách tiếp cận cơ bản.
* Ưu điểm:
* Cung cấp một phương pháp hiệu quả để tìm kiếm mẫu phổ biến và quy tăvs kết hợp trong các bộ dữ liệu lớn.
* Sử dụng thuật toán FP – growth làm nền tảng, giúp tăng hiệu suất tính toán
* Nhược điểm:
* Yêu cầu sử dụng thư viện bên ngoài (mlxtend), đòi hỏi cài đặt thêm
* Có thể khó hiểu hơn, đặc biệt đối với những người mới bắt đầu hoặc không quen với các thuật toán khai thác dữ liệu.

Kết luận: Nếu bạn làm việc với các bộ dữ liệu nhỏ và đơn giản, hoặc chỉ cần nhóm các mục theo một cột cụ thể, thì phương pháp Apriori có thể là lựa chọn phụ hợp. Tuy nhiên, nếu bạn làm việc với các bộ dữ liệu lớn và cần tìm kiếm mẫu phổ biến và quy tắc kết một cách hiệu quả, thì FP – growth là lựa chọn tốt hơn.