

ID	Cite	Abstract
1	<pre>@article{Saranyadevi2021ACG,   title={A Context-free Grammar based Association Rule Mining Technique for Network Dataset},   author={S. Saranyadevi and R. Murugeswari and Sankaranarayanan Bathrinath},   journal={Journal of Physics: Conference Series},   year={2021},   volume={1767},   url={https://api.semanticscholar .org/CorpusID:234030541} }</pre>	<p><b>A Context-free Grammar based Association Rule Mining Technique for Network Dataset</b></p> <p>Experimental outcomes prove that the proposed ARM – Grammar is effective than the traditional ARM approach and the techniques to mine required data by means of grammars.</p> <p>Among various data mining concepts like prediction, clustering, classification, association and outlier discovery, association is a useful technique to extract the interesting relations among data items effectively. Association technique is applied in a number of applications like marketing, education, chemical, bioinformatics, computational linguistics and etc. The important purpose of association is to provide useful information of buying preferences of customers in supermarket in order to increase the sales opportunity, which is called as market- basket analysis. Till now there are many algorithms were developed, but the usage of formal grammars in association rule mining (ARM) is a latest technique to mine required data by means of grammars. In this paper ARM is performed using Context –free Grammar (CFG) – (ARM – Grammar) and the experiments are conducted on MATLAB 2017 software using network dataset, KDDCUP’99. Experimental outcomes prove</p>

		that the proposed ARM – Grammar is effective than the traditional ARM approach.
2	<pre>@article{Saranyadevi2021ACG,   title={A Context-free Grammar based Association Rule Mining Technique for Network Dataset},   author={S. Saranyadevi and R. Murugeswari and Sankaranarayanan Bathrinath},   journal={Journal of Physics: Conference Series},   year={2021},   volume={1767},   url={https://api.semanticscholar.org/CorpusID:234030541} }</pre>	<p><b>Temporal Visual Profiling of Market Basket Analysis</b></p> <p>A novel technique to generate the profile of a customer with regard to his/her product purchase history is proposed and a visual technique is presented, to compare in a friendly and interactive way the profiles of customers over time.</p> <p>Market basket analysis allows analysts to understand the behaviour of customers. In this paper, we propose a novel technique to generate the profile of a customer with regard to his/her product purchase history. After obtaining the profile of a customer, we present a visual technique, to compare in a friendly and interactive way the profiles of customers over time. To show the expediency of our proposal, we performed experiments with a dataset of purchases of a retail market. The results showed that our proposal can be useful for stock planning and identifying customers with similar buying tendencies.</p>
3	<pre>@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},</pre>	<p><b>Market basket analysis using apriori algorithm to find consumer patterns in buying goods through transaction data (case study of Mizan computer retail stores)</b></p> <p>From the results of the analysis in this research, it can be concluded that for the combination of 2 items with the</p>

	<pre>url={https://api.semanticscholar .org/CorpusID:234119984} }</pre>	<p>highest confidence value 100% and the lift ratio value 3.39, a consumer will buy a Joystick and Laptop and for a combination of 3 items, he will also buy a keyboard.</p> <p>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.</p>
--	--	--

		<p>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.</p>
4	<pre>@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} }</pre>	<p><b>Market Basket Analysis for Books Sales Promotion using FP Growth Algorithm, Case Study : Gramedia Matraman Jakarta</b></p> <p>Gramedia does not yet have tools to analyze shopping cart patterns that aim to carry out product promotions appropriately, so tools are needed to analyze past sales data to promote what books should be promoted using the market basket analysis method or shopping basket analysis.</p> <p>For retail companies such as Gramedia stores, promotion and strategies to sell books are important, so tools are needed to analyze past sales data. Gramedia does not yet have tools to analyze shopping cart patterns that aim to carry out product promotions appropriately. To promote what books should be promoted using the market basket analysis method or shopping basket analysis. The algorithm used in</p>

		<p>the data mining process is Frequent Pattern Growth (FP Growth) because it is faster in processing large data. The data analyzed is historical data on book sales from January to March 2020 which is taken randomly (random sampling). The framework used in the data mining process is the Cross Industry Standard Process for Data Mining (CRISP-DM) and the tool used is the Rapid Miner using a market basket analysis framework. With a minimum support of 0.003 and a minimum confidence 0.3 using the FP-Growth algorithm to produce an item set of 7 rules to recommend product promotions. The algorithm results are also in accordance with the business understanding phase of CRISP-DM.</p>
5	<pre>@article{Ghassani2021MARKETBA,   title={MARKET BASKET ANALYSIS USING THE FP-GROWTH ALGORITHM TO DETERMINE CROSS-SELLING},   author={Fildzah Zia Ghassani and Asep Jamaludin and Agung Susilo Yuda Irawan},   journal={Jurnal Informatika Polinema},   year={2021},   url={https://api.semanticscholar.org/CorpusID:239700060} }</pre>	<p><b>Market basket analysis using the fp-growth algorithm to determine cross-selling</b></p> <p>Association rules are a method in data mining that functions to identify items that have a value that is likely to appear simultaneously with other items that can be applied using one of the marketing techniques.</p> <p>KAOICHEM Sinergi Mandiri Cooperative is a cooperative that provides various kinds of basic needs such as basic foodstuffs that can meet the needs of its members. The cooperative transaction data is only stored as a report. Association rules are a method in data mining that functions to identify items that</p>

		<p>have a value that is likely to appear simultaneously with other items. One implementation of the association method is Market Basket Analysis. The data used are transaction data for November 2019. Data mining is one of the processes or stages of the KDD method. The data mining process is carried out using the FP-Growth algorithm, which is one of the algorithms for calculating the sets that often appear from data. Researchers analyzed transaction data using the Rapid Miner Studio tools. In the data mining process using FP-Growth the researcher determines a minimum support value of 3% and a minimum confidence of 50%. The association process using these values produces 3 strong rules, namely if ades 350 ml, then fried / lontong with a support value of 0.030 and confidence 0.556 and if fried st, then fried / lontong with a support value of 0.048 and confidence 0.639, and if nasi uduk / bacang , then fried / rice cake with a support value of 0.031 and confidence 0.824. The results of the association rules can be applied using one of the marketing techniques, namely cross-selling to increase the sales of the cooperative.</p>
6	<p>@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},</p>	<p><b>MBA: Market Basket Analysis Using Frequent Pattern Mining Techniques</b></p> <p>This paper introduces an efficient Bitwise-Based data structure technique for mining frequent pattern in large-scale</p>

	<pre> year={2023}, url={https://api.semanticscholar .org/CorpusID:259884237} } </pre>	<p>databases, which outperform Apriori, Éclat, FP-growth, and H-mine in terms of execution time for Market Basket Analysis.</p> <p>This Market Basket Analysis (MBA) is a data mining technique that uses frequent pattern mining algorithms to discover patterns of co-occurrence among items that are frequently purchased together. It is commonly used in retail and e-commerce businesses to generate association rules that describe the relationships between different items, and to make recommendations to customers based on their previous purchases. MBA is a powerful tool for identifying patterns of co-occurrence and generating insights that can improve sales and marketing strategies. Although a numerous works has been carried out to handle the computational cost for discovering the frequent itemsets, but it still needs more exploration and developments. In this paper, we introduce an efficient Bitwise-Based data structure technique for mining frequent pattern in large-scale databases. The algorithm scans the original database once, using the Bitwise-Based data representations as well as vertical database layout, compared to the well-known Apriori and FP-Growth algorithm. Bitwise-Based technique enhance the problems of multiple passes over the original database, hence,</p>
--	---	---

		<p>minimizes the execution time. Extensive experiments have been carried out to validate our technique, which outperform Apriori, Éclat, FP-growth, and H-mine in terms of execution time for Market Basket Analysis.</p>
7	<pre>@article{Bagaskara2023AnalisisDI,   title={ Analisis dan Implementasi Market Basket Analysis (MBA) Menggunakan Algoritma Apriori dengan Dukungan Visualisasi Data},   author={Septembri Rio Bagaskara and Dwi Hosanna Bangkalang},   journal={Jurnal Sistem Komputer dan Informatika (JSON)},   year={2023},   url={https://api.semanticscholar.org/CorpusID:260034709} }</pre>	<p><b>Analisis dan Implementasi Market Basket Analysis (MBA) Menggunakan Algoritma Apriori dengan Dukungan Visualisasi Data</b></p> <p>The result of this research is an association rule on the historical data of MSME Culture Coffee customer purchases which can be used as a basis for the sales and marketing strategy of Culture Coffee MSMEs to increase business revenue.</p> <p>Culture Coffee MSME is one of the MSMEs engaged in the culinary field and is experiencing business competition. A marketing strategy is needed with the right decision-making process so that the business can survive and excel. UMKM Culture Coffee uses a point of sales application to accommodate the transaction process and record transactions. Historical customer data can be processed into a basis for decision making for marketing strategies that effectively increase sales. However, the transaction data has not been used optimally. There is a need to analyze historical customer data that can generate information to form marketing strategies. Market Basket</p>



		<p>Analysis (MBA) is one of the methods in data mining used in knowing products that tend to be purchased together by customers known as Association Rule. Association rules produce products in the form of packages or bundling which are used as marketing strategies. The marketing strategy obtained is supported by data visualization which contains information from the data. Apriori algorithm is used to generate association rules. The result of this research is an association rule on the historical data of MSME Culture Coffee customer purchases. Based on these rules, recommendations for selling menu packages to customers can be given. The purpose of this research is to find customer purchasing patterns which are used as the basis for decision making in determining menu sales. The results showed 2 product packages, namely, nuggets and french fries with sausages and french fries with a support and confidence value of 12.5% and 37.6% with 10.8% and 29% respectively. The results of this study can be used as a basis for the sales and marketing strategy of Culture Coffee MSMEs to increase business revenue.</p>
8	<p>@article{Nafi2023MarketBA,  title={Market Basket Analysis for Sales Transaction in Shopping Stores},  author={Mohd Noor Azam Nafi and Azni Sharlina Zakaria and Nur Izzati Mohamad Arif and Siti Nurhafizah Mohd Shafie and Nasuhar Ab. Aziz and Omar Kairan},  journal={International Journal of Academic</p>	<p><b>Market Basket Analysis for Sales Transaction in Shopping Stores</b></p> <p>The findings showed that there are 13 interesting rules of association revealed in this study and found that most</p>

	<p>Research in Business and Social Sciences},  year={2023},  url={https://api.semanticscholar  .org/CorpusID:257119151}  }</p>	<p>products that were purchased together are tissues, condiments, instant food, cooking oil, meat, biscuits, dry goods, beverages, and cleaning products.</p> <p>Market Basket Analysis (MBA) system is a widely used technique among marketers, especially for undirected data mining analysis. MBA is also known as product association analysis and the outcome of this analysis is called association rules. The outcome can be used to schedule marketing or advertising strategies and design catalogs for different shop layouts. Discovering the pattern from the customer's buying habits in the shopping stores was collected in their buying transaction. This study aims to compare the item purchased by the respondents between Store A and Store B and to find out the most potential products that customers have bought along with a specific category of products. Convenience non-probability sampling was involved with structured questionnaires of items in store was collected to analyze data. Association analysis was used by analyzing the result from support, confidence, and lift. The findings showed that there are 13 interesting rules of association revealed in this study. Moreover, the result also found that most products that were purchased together are tissues, condiments, instant food, cooking oil, meat, biscuits, dry goods, beverages, and</p>
--	--	---

		cleaning products.
9	<pre>@article{Gupta2023ECommerceMB,   title={E-Commerce Market Basket Analysis using Apriori Algorithm},   author={Khushi Gupta and Kashyapi Shah and Ameya A Kadam},   journal={INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT},   year={2023},   url={https://api.semanticscholar.org/CorpusID:263659317} }</pre>	<p><b>E-Commerce Market Basket Analysis using Apriori Algorithm</b></p> <p>The usage of the Apriori Algorithm is presented to implement market basket analysis to identify purchase patterns for items that are frequently bought together by customers to improve sales of multi-product stores by enhancing product placement based on consumers' shopping habit.</p> <p>This paper presents the usage of the Apriori Algorithm to implement market basket analysis to identify purchase patterns for items that are frequently bought together by customers. The results of this analysis are primarily used to improve sales of multi-product stores by enhancing product placement based on consumers' shopping habit. In this particular scenario, we have used the data from an online E-commerce store that caters to customers across the world, but primarily focused to the United Kingdom. Key Words: Market Basket, Apriori Algorithm, Association Rules, E-Commerce, Consumer Behaviour, Data Mining</p>
10	<pre>@article{Ghous2023DeepLB,   title={Deep Learning based Market Basket Analysis using Association Rules},   author={Hamid Ghous and Mubasher Malik and Iqra Rehman},   journal={KIET Journal of Computing and Information Sciences},   year={2023},</pre>	<p><b>Deep Learning based Market Basket Analysis using Association Rules</b></p> <p>This work proposes a framework using AR as a feature selection while applying DL methods for classification</p>

	<pre>url={https://api.semanticscholar .org/CorpusID:264297776} }</pre>	<p>and prediction on the transactional dataset, conducted on two datasets, InstaCart and real-life data from Bites Bakers.</p> <p>Market Basket Analysis (MBA) is a data mining technique assisting retailers in determining the customer's buying habits while making new marketing decisions as the buyer's desire frequently changes with expanding needs; therefore, transactional data is getting large every day. There is a demand to implement Deep Learning (DL) methods to manipulate this rapidly growing data. In previous research, many authors conducted MBA applying DL and association rules (AR) on retail datasets. AR identifies the association between items to find in which order the customer place items in the basket. AR is only used in mining frequently purchased items from retail datasets. There is a gap in classifying these rules and predicting the next basket item using DL on the transactional dataset. This work proposes a framework using AR as a feature selection while applying DL methods for classification and prediction. The experiments were conducted on two datasets, InstaCart and real-life data from Bites Bakers, which operates as a growing store with three branches and 2233 products. The AR classified at 80,20 and 70,30 splits using CNNN, Bi-LSTM, and CNN-BiLSTM. The results considering simulation at</p>
--	--	---

		both splits show that Bi-LSTM performs with high accuracy, around 0.92 on the InstaCart dataset. In contrast, CNN-BiLSTM performs best at an accuracy of around 0.77 on Bites Bakers dataset.
11	@article{Gupta2023ECommerceMB, title={E-Commerce Market Basket Analysis using Apriori Algorithm}, author={Khushi Gupta and Kashyapi Shah and Ameya A Kadam}, journal={INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT}, year={2023}, url={https://api.semanticscholar.org/CorpusID:263659317} }	<p><b>E-Commerce Market Basket Analysis using Apriori Algorithm</b></p> <p>The usage of the Apriori Algorithm is presented to implement market basket analysis to identify purchase patterns for items that are frequently bought together by customers to improve sales of multi-product stores by enhancing product placement based on consumers' shopping habit.</p> <p>This paper presents the usage of the Apriori Algorithm to implement market basket analysis to identify purchase patterns for items that are frequently bought together by customers. The results of this analysis are primarily used to improve sales of multi-product stores by enhancing product placement based on consumers' shopping habit. In this particular scenario, we have used the data from an online E- commerce store that caters to customers across the world, but primarily focused to the United Kingdom. Key Words: Market Basket, Apriori Algorithm, Association Rules, E-Commerce, Consumer Behaviour, Data Mining</p>
12	@article{Mooy2023ContextualMB, title={Contextual Market Basket Analysis	<b>Contextual Market Basket Analysis during Covid-19</b>

	<p>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}  }</p>	<p>Using the Artificial Neural Network Approach that is connected to Market Basket Analysis, it can analyze and compare purchasing patterns and can identify rules that were formed before and after covid-19; several rule changes were found due to changes in people's behavior patterns.</p> <p>One form of Data Mining application to analyze Market Basket Analysis. Market Basket Analysis helps identify buying patterns formed from concurrent transactions. One of the problems with Market Basket Analysis is that customer needs vary according to season and time of day, especially during this covid-19 season. For this purpose, by using the Artificial Neural Network (ANN) Approach that is connected to Market Basket Analysis, it can analyze and compare purchasing patterns and can identify rules that were formed before and after covid-19; several rule changes were found due to changes in people's behavior patterns.</p>
13	<p>@article{Purnomo2023PenerapanDM,  title={Penerapan Data Mining Dalam Menganalisis Pola Belanja Konsumen Menggunakan Market Basket Analysis},  author={Sari Asih Kencono Purnomo and Heny Ispur Pratiwi and Muhammad Ibnu Sa'ad},  journal={METIK JURNAL},  year={2023},  url={https://api.semanticscholar.org/CorpusID:266702253}  }</p>	<p><b>Penerapan Data Mining Dalam Menganalisis Pola Belanja Konsumen Menggunakan Market Basket Analysis</b></p> <p>Providing recommendations for layout changes aims to make it easier for consumers to shop, increase the possibility of impulse buying by consumers, and maximize product display,</p>

		<p>thereby reducing the accumulation of goods in the Purnama Store Warehouse.</p> <p>Currently, almost every activity is related to data. in the business sector, daily sales transaction data stored in the database system will always increase and accumulate. The existing data is only used as an archive by the shop owner so that it has an impact on sales strategies that are not implemented well, even though the existing data can be processed into information to determine the layout of goods so that it has an impact on increasing the occurrence of impulse buying, increasing or maintaining turnover, and minimizing product waste. accumulate until it expires which can be detrimental to the shop. The aim of this research is to find consumer shopping patterns using Marker Basket Analysis. This research method is called market basket analysis or also called association rules, which is a data mining technique for finding patterns that often appear simultaneously in transaction data, so that it can be used as a method for finding information about what kinds of goods are frequently used. purchased by consumers simultaneously. The results of this research, based on data analysis using the Rapidminer application, found 25 associative relationships or rules with a lift ratio value of more than 1, these rules become a reference in determining the layout of goods.</p>
--	--	--

		Providing recommendations for layout changes aims to make it easier for consumers to shop, increase the possibility of impulse buying by consumers, and maximize product display, thereby reducing the accumulation of goods in the Purnama Store Warehouse.
14	<pre>@article{Wahyudi2023MarketBA,   title={Market Basket Analysis dengan Perbandingan Metode Apriori dan FP-Growth Pada Data Transaksi XYZ},   author={Rizki Nofrian Wahyudi and Dyah Erny Herwindiati and Janson Hendryli},   journal={Jurnal Ilmu Komputer dan Sistem Informasi},   year={2023},   url={https://api.semanticscholar.org/CorpusID:260186039} }</pre>	<p><b>Market Basket Analysis dengan Perbandingan Metode Apriori dan FP-Growth Pada Data Transaksi XYZ</b></p> <p>Analysis of the market basket using the FP-Growth and a priori algorithm in transactions that aid in strategic planning and business product structuring.</p> <p>Technology is currently advancing quickly, allowing all organizations to grow their networks with its aid and create sales methods that now rely on technology to aid in making the proper judgments. When saved transaction data is accessible, every business will be able to implement its marketing strategy to maximize client transactions. use it to your advantage. Analysis of the market basket using the FP-Growth and a priori algorithm in transactions that aid in strategic planning and business product structuring. The FP-Growth algorithm and the Apriori algorithm work well together. One can evaluate the effectiveness of the employment of the a priori algorithm and the FP-Growth algorithm by applying both of them</p>
15	@article{Idris2022ComparisonOA,	<b>Comparison of Apriori,</b>



	<p>title={Comparison of Apriori, Apriori-TID and FP-Growth Algorithms in Market Basket Analysis at Grocery Stores},</p> <p>author={Andi Ilhamsyah Idris and Eliyah Acantha M Sampetoding and Valian Yoga Pudya Ardhana and Irene Maritsa and Adrisumatri Sakri and Hidayatullah Ruslan and Esther Sanda Manapa},</p> <p>journal={The IJICS (International Journal of Informatics and Computer Science)},</p> <p>year={2022},</p> <p>url={https://api.semanticscholar.org/CorpusID:257018537}</p>	<p><b>Apriori-TID and FP-Growth Algorithms in Market Basket Analysis at Grocery Stores</b></p> <p>In the results of this study, it is known that FP-Growth has the best performance among the other two algorithms, but uses more memory than other algorithms.</p> <p>Market Basket Analysis is an analysis of consumer behavior specifically from a certain group/group. Market Basket Analysis is generally used as a starting point for seeking knowledge from a data transaction when we do not know what specific pattern we are looking for. Market Basket Analysis in this study is applied to the search for patterns of purchasing groceries at grocery stores and then analyzed by season. This study aims to compare the Apriori, Apriori TID and FP-Growth methods in determining consumer transaction behavior and calculating the quantity of consumer transactions in several seasons based on data obtained from the Market Basket Analysis database. In the results of this study, it is known that FP-Growth has the best performance among the other two algorithms, but uses more memory than other algorithms. The Apriori-TID algorithm uses lighter and faster memory than the Apriori Algorithm</p>
16	<p>@article{Hadi2023MarketBA,</p> <p>title={Market Basket Analysis Using Apriori</p>	<p><b>Market Basket Analysis Using Apriori Algorithm to Find</b></p>

<p>Algorithm to Find Effective Fiscal Policy Mix with R Programming},  author={Isnen Hadi and Al Ghozali and Arief Wibowo},  journal={Decode: Jurnal Pendidikan Teknologi Informasi},  year={2023},  url={https://api.semanticscholar.org/CorpusID:265174547}  }</p>	<p><b>Effective Fiscal Policy Mix with R Programming</b></p> <p>Fiscal policy drives a country's economy and is the most effective policy to restore a country's economy. When a recession occurs, the fiscal policy helps a country increase aggregate demand in the market for goods and services. This study proposes a fiscal policy mix that can be implemented based on historical data. So this research focuses on using association rules to assist decision-makers (regulators in adopting appropriate fiscal policies in the global VUCA (Volatility, Uncertainty, Complexity, Ambiguity) era. Therefore, an experimental research approach was used in this study to produce the best association rules. The research was carried out in six stages to obtain conclusions, namely problem identification in the research sample and literature review related to apriori algorithms, data collection, data pre-processing, parameter determination, research findings building a priori algorithms, and knowledge extraction formed from a priori algorithms. Based on the experimental results using the a priori algorithm, 657 rules were obtained with a minimum variation of two to six itemsets. Rule 6 produces an average value of the budget performance of 90.65. Rule 6 also says that the performance value of the budget can be increased by spreading out</p>
--	---

		funding sources and ensuring that operational spending is as efficient as possible.
17	<pre>@article{Zalmi2023PenerapanMB,   title={Penerapan Market Basket Analysis Dengan Metode Apriori Pada WFZ Book Store},   author={Wahyuni Fithratul Zalmi},   journal={INFORMATIKA},   year={2023},   url={https://api.semanticscholar .org/CorpusID:256486464} }</pre>	<p>Penerapan Market Basket Analysis Dengan Metode Apriori Pada WFZ Book Store</p> <p>The association rules between book titles formed from the mining process can later be used by the WFZ Book Store to increase the number of books purchased, besides that it can be used for the procurement of books from the association rules of frequently purchased books.</p> <p>WFZ Book Store is a Book Store that provides various types of books. The sales process is done offline where the costumer has to visit the store to make a purchase. The process of recording data using the ledger makes it difficult for WFZ Book Store to make sales reports because it has to collect transaction data in the ledger, it causes difficulties to know the available Book stock because it has to check every sales transaction in the ledger. WFZ Book Store requires an information system that can manage book sales data online agar costumers can easily view book data, stock and book online reservations. With the wfz Book Store information system can facilitate the customer in managing book data and store customer data that has made transactions also WFZ Book Store does not find it difficult to provide book purchase transaction data for the</p>

		<p>customer, so that the data can be used for book recommendations by finding the set of data that most often appear in a data set. Datamining techniques have been widely used to overcome existing problems, one of which is the application of a-priori algorithms to find association rules formed from book purchase transaction datasets. So it will be known the association between the title of the book purchased. The association rules between book titles formed from the mining process can later be used by the WFZ Book Store to increase the number of books purchased, besides that it can be used for the procurement of books from the association rules of frequently purchased books, it can also be developed into a knowledge base for the book purchase recommendation system. The Output is in the form of a priori algorithm analysis software.</p>
18	<p>@article{Samboteng2022MARKETBA, title={MARKET BASKET ANALYSIS OF ADMINISTRATIVE PATTERNS DATA OF CONSUMER PURCHASES USING DATA MINING TECHNOLOGY}, author={Lukman Samboteng and Rulinawaty Rulinawaty and M. Rachmat Kasmad and Mutmainnah Basit and Robbi Rahim}, journal={Journal of Applied Engineering Science}, year={2022}, url={https://api.semanticscholar.org/CorpusID:247142858} }</p>	<p><b>Market basket analysis of administrative patterns data of consumer purchases using data mining technology</b></p> <p>From the results of the trials in this study, it was found that the greater the minimum support and minimum confidence, the less time it takes to produce recommendations and the fewer recommendations are given, but the recommendations given come from transactions that often appear.</p> <p>Food is the ingredient that enables people to grow, develop,</p>

		<p>and achieve. For this reason, food quality and types of food must be considered so that they are safe for consumption and managed. Some plant-based foodstuffs are often processed and consumed by the community, even the most needed in food processing. In this case, the research was carried out using data mining with market basket analysis algorithms to obtain very valuable information to decide the inventory of the type of material needed. Market Based Analysis method is used to analyze all data and create patterns for each data. One method of Market Based Analysis in question is the association rule with a priori algorithm. This algorithm produces sales transactions with strong associations between items in the transaction which are used as sales recommendations that help users (owners) get recommendations when users see details of the itemset purchased. From the results of the trials in this study, it was found that the greater the minimum support (minsup) and minimum confidence (minconf), the less time it takes to produce recommendations and the fewer recommendations are given, but the recommendations given come from transactions that often appear.</p>
19	<p>@article{Jirapatsil2022MarketBA, title={Market Basket Analysis for Fresh Products location improvement: A case study of E-Commerce Business Warehouse}, author={Pattraporn Jirapatsil and Naragain</p>	<p><b>Market Basket Analysis for Fresh Products location improvement: A case study of E-Commerce Business Warehouse</b></p>

	Phumchusri}, journal={Proceedings of the 4th International Conference on Management Science and Industrial Engineering}, year={2022}, url={https://api.semanticscholar .org/CorpusID:250624756} }	<p>The goal of this paper is to propose a market basket analysis method to gain insights from historical transactions, a set of recording data result in connections with sales-purchase activities, of the case-study company.</p> <p>Market Basket Analysis (MBA) uses the data mining technique as an analysis tool to understand association among many items. It is a useful tool for extracting information from large amount of data in many industrial areas, e.g., grocery, supermarkets, retailers, warehouse, mobile showroom, libraries, zoos, etc. The case-study company sells fresh products in E-commerce business and currently has inefficient product location in warehouse, causing delays in picking process. Thus, the goal of this paper is to propose a market basket analysis method to gain insights from historical transactions, a set of recording data result in connections with sales-purchase activities, of the case-study company. Apriori algorithm is applied for association rules to analyze 2366 transactions data between July and December 2021. The results of data analysis are then utilized in rearranging products location in warehouse to reduce average picking distance per order. The results show that the average distance per order can be reduced by 54.4%.</p>
20	@article{Umar2022MarketBA, title={Market Basket Analysis Menggunakan	<b>Market Basket Analysis Menggunakan Association</b>

	<p>Association Rule dan Algoritma Apriori Pada Produk Penjualan Mitra Swalayan Salatiga},  author={Elfira Umar and Daniel H. F. Manongga and Ade Iriani},  journal={JURNAL MEDIA INFORMATIKA BUDIDARMA},  year={2022},  url={https://api.semanticscholar.org/CorpusID:253001514}  }</p>	<p><b>Rule dan Algoritma Apriori Pada Produk Penjualan Mitra Swalayan Salatiga</b></p> <p>Data mining techniques in analyzing what items are most often purchased at the same time by consumers are discussed so that they can change the placement of items that are close together to increase the impulse buying effect.</p> <p>Market Basket analysis is learning to manage associations in data processing in various fields. The main purpose of Market Basket analysis in the field of sales is to convey an important message to the company so that it can find out the behavior patterns of entering goods into the shopping basket by consumers so that partners can make a decision. In this study, the Apriori Algorithm is used to take into account changes that occur in the data. This study discusses data mining techniques in analyzing what items are most often purchased at the same time by consumers so that they can change the placement of items that are close together to increase the impulse buying effect. The results obtained are 5 rules where one of the rules obtains the highest confidence value when buying cigarettes, the dominant item is taken simultaneously, namely eggs by obtaining a confidence value that can meet the highest confidence requirements, namely 67%.</p>
21	@inproceedings{Putrevu2022MarketBA,	<b>Market Basket Analysis Using</b>

	<p>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}  }</p>	<p><b>Apriori Algorithm with Pruning Approach</b></p> <p>In the present study Market Basket Analysis for a leading shopping mall is studied and analysed using frequent Itemset mining and decision tree techniques.</p> <p>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</p>
--	---	--



		<p>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.</p>
22	<p>@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}  }</p>	<p><b>Optimization of market basket analysis using centroid-based clustering algorithm and fp-growth algorithm</b></p> <p>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-medoids, and k=7 for fuzzy c-me means, which means that the highest average lift ratio is in the association rules generated from the grouping data set using k-Means.</p> <p>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</p>

		<p>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.</p>
23	@article{Priyanto2022IMPLEMENTATIONOM, title={IMPLEMENTATION OF MARKET BASKET ANALYSIS WITH APRIORI ALGORITHM IN MINIMARKET},	<b>Implementation of market basket analysis with apriori algorithm in minimarket</b>

	<p>author={ Abdul Hafiidh Priyanto and Amalia Beladinna Arifa },  journal={ Jurnal Teknik Informatika (Jutif) },  year={ 2022 },  url={ https://api.semanticscholar.org/CorpusID:256128251 }  }</p>	<p>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.</p> <p><b>Abstract</b></p> <p>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</p>
--	---	---

		<p>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.</p>
24	<p>@article{FenesiaYunika2024MARKETBA,  title={MARKET BASKET ANALYSIS MENGGUNAKAN ALGORITMA HASH BASED PADA DATA PENJUALAN DI O2 SWALAYAN},  author={Fenesia Yunika Fenesia Yunika},  journal={Jurnal Informatika Kaputama (JIK)},  year={2024},  url={https://api.semanticscholar.org/CorpusID:266803185}  }</p>	<p><b>Market basket analysis menggunakan algoritma hash based pada data penjualan di o2 swalayan</b></p> <p>The algorithm used is the Hash Based and Association Rule Algorithm because it can overcome the determination of frequent itemsets from k-itemset candidates with a large number and speed up information on consumer spending patterns.</p> <p>O2 Supermarkets have stock of goods available in the warehouse. During this time there is a buildup of goods in the warehouse because there is no system to analyze transaction data so that losses occur. Therefore, a system is needed to manage these goods in order to reduce stocks of goods that are not selling well and make stocks of goods that sell well in warehouses through consumer spending patterns. To analyze data on consumer spending</p>

		<p>patterns, the Market Basket Analysis Method can be used. This method can be used to analyze consumer spending patterns. The algorithm used is the Hash Based and Association Rule Algorithm because it can overcome the determination of frequent itemsets from k-itemset candidates with a large number and speed up information on consumer spending patterns. With existing data and using this algorithm it can produce a lot of information. The results of this study get 6 Rules with a minimum Confidence of 50% and minimum Support <math>\geq 4</math>. From consumer spending patterns that can be used as a strategy to support sales, knowing which items are selling best, placing goods so that they are more efficient in picking up the goods desired by consumers, not only shopping patterns but also more controlled stock of goods.</p>
25	<p>@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}  }</p>	<p><b>Market Basket Analysis with Apriori Algorithm and Frequent Pattern Growth (Fp-Growth) on Outdoor Product Sales Data</b></p> <p>Using a market basket analysis method to see the association (rules) between a number of sales attributes, the pattern of relationships in the transactions that occur in outdoor stores in Indonesia is determined.</p> <p>Indonesia is an equatorial country that has abundant natural wealth from the seabed to the top of the mountains, the</p>

		<p>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.</p>
26	<p>@article{Karnila2022MARKETBA, title={MARKET BASKET ANALYSIS ON TRANSACTION DATA USING THE APRIORI ALGORITHM}, author={Sri Karnila and Akbar Rizkyandi and</p>	<p><b>Market basket analysis on transaction data using the apriori algorithm</b></p> <p>The research has produced 9</p>

	<p>Rio Kurniawan and Nurjoko Nurjoko},  journal={Jurnal TAM (Technology Acceptance Model)},  year={2022},  url={https://api.semanticscholar.org/CorpusID:252757734}  }</p>	<p>sales pattern relationships with the highest confidence of 100%.</p> <p>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.</p>
27	<p>@article{Lewis2021PenerapanDM,  title={Penerapan Data Mining Menggunakan Task Market Basket Analysis Pada Transaksi Penjualan Barang di Ab Mart dengan Algoritma Apriori},  author={Angie Lewis and Muhammad Zarlis and Zakarias Situmorang},  journal={JURNAL MEDIA INFORMATIKA BUDIDARMA},  year={2021},  url={https://api.semanticscholar.org/CorpusID:235542001}  }</p>	<p><b>Penerapan Data Mining Menggunakan Task Market Basket Analysis Pada Transaksi Penjualan Barang di Ab Mart dengan Algoritma Apriori</b></p> <p>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.</p>

		<p>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 &amp; 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</p>
28	<p>@article{iekli2021MarketBA,  title={Market Basket Analysis of Basket Data with Demographics: A Case Study in E-Retailing},  author={Ural G{"o}kay iekli and İnan Kabasakal},</p>	<p><b>Market Basket Analysis of Basket Data with Demographics: A Case Study in E-Retailing</b></p> <p>This study extends the</p>



	<pre> journal={ Alphanumeric Journal }, year={ 2021 }, url={ https://api.semanticscholar .org/CorpusID:238038517 } </pre>	<p>conventional basket analysis by incorporating demographic variables along with purchase transactions, and provides an example for the extraction of segment-specific rules that relate productlevel purchase decisions with gender, location, and age group.</p> <p>Businesses overcome with a high degree of competition that necessitates customer-focused strategies in most industries. In a digitalized business environment, the implementation of such strategies often requires the analysis of customer data. Market basket analysis is a well-known method in marketing that examines basket data to discover useful information about customers' purchase intentions. The analysis has been a playground for data mining researchers that aim to overcome with its practical challenges. Our study extends the conventional basket analysis by incorporating demographic variables along with purchase transactions. With such modification, we provide an example for the extraction of segment-specific rules that relate productlevel purchase decisions with gender, location, and age group. For this purpose, we present a case study on monthly basket data obtained from an e-retailer in Turkey. Our findings demonstrate association rules that might guide marketing practitioners who need to discover segment-</p>
--	---	---

		specific purchase patterns to designate personalized promotions.
29	<pre>@article{Lestari2021IMPLEMENTASIDM,   title={IMPLEMENTASI DATA MINING UNTUK MENENTUKAN POLA PENJUALAN DENGAN MARKET BASKET ANALYSIS},   author={Novia Lestari and Refika Fitria Gunawan},   journal={Insearch: Information System Research Journal},   year={2021},   url={https://api.semanticscholar .org/CorpusID:254918712} }</pre>	<p><b>Implementasi data mining untuk menentukan pola penjualan dengan market basket analysis</b></p> <p>By utilizing the sales transaction data of consumers supported by the Market Basket Analysis method, it is possible to determine the right marketing tactics for business actors so as to increase sales.</p> <p>The increasingly fierce competition in the business world lately has forced business actors to have the right strategy to increase their product sales, one of which is to determine sales patterns. However, along with the amount of sales transaction data that is carried out every day, it makes it difficult for business actors to analyze existing sales patterns. So we need a method that can help determine sales patterns from even large transaction data, one of which uses the Market Basket Analysis method which uses customer data that has been stored in the database to find new information in it. By utilizing the sales transaction data of consumers supported by the Market Basket Analysis method, it is possible to determine the right marketing tactics for business actors so as to increase sales</p>
30	<pre>@article{Rana2021ASA,   title={A Seasonal and Multilevel Association Based Approach for Market Basket Analysis in</pre>	<p><b>A Seasonal and Multilevel Association Based Approach for Market Basket Analysis in</b></p>

	<p>Retail Supermarket},  author={S. Rana and Mohammad Nazrul Islam Mondal},  journal={European Journal of Information Technologies and Computer Science},  year={2021},  url={https://api.semanticscholar.org/CorpusID:244090920}  }</p>	<p><b>Retail Supermarket</b></p> <p>The main contribution is to discover the hidden seasonal itemsets and extract the seasonal associations among products in additionally with the traditional strong regular rules in transactional database that shows the superiority for making season based merchandising decisions.</p> <p>Market Basket Analysis is an observational data mining methodology to investigate the consumer buying behavior patterns in retail Supermarket. It analyzes customer baskets and explores the relationship among products that helps retailers to design store layouts, make various strategic plans and other merchandising decisions that have a big impact on retail marketing and sales. Frequent itemsets mining is the first step for market basket analysis. The association rules mining uncovers the relationship among products by looking what products the customers frequently purchase together. In retail marketing, the transactional database consists of many itemsets that are frequent only in a particular season however not taken into consideration as frequent in general. In some cases, association rules mining at lower data level with uniform support doesn't reflect any significant pattern however there is valuable information hiding behind it. To overcome</p>
--	--	---

		those problems, we propose a methodology for mining seasonally frequent patterns and association rules with multilevel data environments. Our main contribution is to discover the hidden seasonal itemsets and extract the seasonal associations among products in additionally with the traditional strong regular rules in transactional database that shows the superiority for making season based merchandising decisions. The dataset has been generated from the transaction slips in large supermarket of Bangladesh that discover 442 more seasonal patterns as well as 1032 seasonal association rules in additionally with the regular rules for 0.1% minimum support and 50% minimum confidence.
--	--	---

## 1. Giới thiệu

### 1.1 Khái niệm

Giỏ thị trường là một nhóm các sản phẩm được thiết kế để theo dõi hiệu suất của một phân khúc thị trường cụ thể. Kinh tế giỏ thị trường tập trung vào Chỉ số giá tiêu dùng (CPI), theo dõi các mặt hàng tiêu dùng khác nhau và sử dụng mức giá của chúng để đưa ra ước tính về lạm phát

Phân tích giỏ thị trường (MBA), hay còn được gọi là phân tích giỏ hàng, là một kỹ thuật khai thác dữ liệu sử dụng các thuật toán khai thác mẫu thường xuyên để khám phá các mô hình đồng xuất hiện 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 liên kết mô tả mối quan hệ giữa các mặt hàng khác nhau và đưa ra đề xuất cho khách hàng dựa trên các giao dịch mua trước đó của họ [6]. Bằng cách này, phân tích giỏ thị trường cung cấp thông tin quý báu về hành vi mua hàng của khách hàng và giúp doanh nghiệp hiểu rõ hơn về nhu cầu và sở thích của khách hàng.

---

## 1.2 Ứng dụng

Phân tích giỏ hàng có thể được sử dụng để đạt được nhiều mục tiêu trong lĩnh vực thương mại điện tử và quản lý bán lẻ. Một số ứng dụng chính bao gồm:

- ❖ Hiểu sâu hơn về hành vi mua hàng của khách hàng [2]: Phân tích giỏ hàng giúp doanh nghiệp nhận biết các mối quan hệ giữa các sản phẩm, từ đó hiểu được sở thích và nhu cầu của khách hàng.
- ❖ Tối ưu hóa trải nghiệm mua sắm: Dựa trên thông tin từ phân tích giỏ hàng, doanh nghiệp có thể cải thiện việc bố trí sản phẩm trên trang web hoặc trong cửa hàng để tối ưu hóa trải nghiệm mua sắm của khách hàng.
- ❖ Tăng cường chiến lược giá cả: Phân tích giỏ hàng giúp xác định các mối quan hệ giữa sản phẩm và giá cả, từ đó doanh nghiệp có thể phát triển chiến lược giá cả hợp lý để kích thích mua sắm và tăng doanh số bán hàng.

## 1.3 Hiện nay đang được áp dụng như nào?

Phân tích giỏ hàng đã trở thành một công cụ quan trọng trong nhiều ngành công nghiệp, đặc biệt là trong thương mại điện tử và bán lẻ. Các doanh nghiệp sử dụng các công cụ phân tích dữ liệu và thuật toán máy học để tự động phân tích hàng tỷ giao dịch mua sắm hàng ngày và tạo ra thông tin chi tiết và chính xác về hành vi mua hàng của khách hàng. Điều này giúp họ đưa ra các quyết định kinh doanh thông minh về việc quản lý tồn kho, marketing, và chiến lược bán hàng.

## 1.4 Ví dụ

Theo thống kê nghiên cứu dữ liệu hành vi mua hàng của Wal-Mart: “Có 60% đàn ông đi siêu thị vào tối thứ 6, nếu họ mua bỉm cho trẻ em thì sẽ mua cả bia”. Điều này được lý giải như sau:

- ❖ Người mua bỉm là các ông bố mới kết hôn và có con.
- ❖ Trước khi có con: hành động cuối tuần của anh ta là đi nhậu nhẹt với bạn bè.
- ❖ Hiện tại: Anh ta có con nhỏ nên không đi được nữa, như vậy tối cuối tuần đi mua đồ cho con là cơ hội để anh ta đi ra khỏi nhà. Và tại siêu thị, anh ta gợi nhớ lại các bữa tiệc tùng, và sinh ra hành vi mua bia về nhà uống.

Dựa trên thông tin này, họ quyết định tái cấu trúc cửa hàng của mình để đặt các sản phẩm này gần nhau, tạo ra hiệu ứng mua sắm kết hợp, từ đó tăng doanh số bán hàng.

## 1.5 Phân tích giỏ hàng hoạt động như thế nào?

Phân tích giỏ hàng thực hiện bằng cách sử dụng các thuật toán và kỹ thuật phân tích dữ liệu để xác định các mẫu mua hàng phổ biến. Các bước cơ bản bao gồm:

- 
- ❖ Thu thập dữ liệu: Dữ liệu về các giao dịch mua sắm được thu thập từ các hệ thống bán lẻ hoặc thương mại điện tử.
  - ❖ Tiền xử lý dữ liệu: Dữ liệu được làm sạch và chuẩn hóa để loại bỏ dữ liệu không hợp lệ và phát hiện ra các mẫu mua hàng.
  - ❖ Áp dụng thuật toán: Các thuật toán phân tích dữ liệu như Apriori [3] hoặc FP-Growth [4] được áp dụng để tìm ra các luật kết hợp giữa các sản phẩm.
  - ❖ Hiển thị kết quả: Kết quả của phân tích được hiển thị dưới dạng các luật kết hợp, liệt kê các sản phẩm thường được mua cùng nhau và độ tin cậy của mỗi luật.

## **2. Nghiên cứu liên quan**

Trong 30 tài liệu tìm được đề cập đến rất nhiều phương pháp phân tích giỏ thị trường nhưng nhận thấy 2 phương pháp được sử dụng nhiều nhất là dùng thuật toán Apriori và FP-Growth

### **2.1 Thuật toán Apriori**

Thuật toán Apriori là một thuật toán trong lĩnh vực học máy và khai phá dữ liệu, được sử dụng để khám phá các quy luật kết hợp giữa các mặt hàng trong các tập dữ liệu chứa các giao dịch. Thuật toán này được phát triển để xử lý bài toán kỹ thuật ghép mặt hàng phổ biến (frequent itemset mining) và kỹ thuật tìm kiếm luật kết hợp (association rule learning).

#### **2.1.1 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

#### **2.1.2 Các nghiên cứu**

Phân tích giỏ hàng thị trường bằng thuật toán apriori để tìm ra mô hình mua hàng của người tiêu dùng thông qua dữ liệu giao dịch (nghiên cứu trường hợp cửa hàng bán lẻ máy tính Mizan). Từ kết quả phân tích trong nghiên cứu này, có thể kết luận rằng đối với sự kết hợp của 2 mặt hàng có giá trị tin cậy cao nhất 100% và giá trị tỷ lệ nâng 3,39 thì người tiêu dùng sẽ mua Joystick và Laptop và đối với sự kết hợp của 3 mặt hàng, anh ấy cũng sẽ mua một bàn phím [3]

Phân tích và triển khai Phân tích giỏ thị trường (MBA) Thuật toán áp dụng cho các dữ liệu trực quan hóa. Kết quả của nghiên cứu này là một quy tắc kết hợp về dữ liệu lịch sử mua hàng của khách hàng Cà phê Văn hóa MSME, dữ liệu này có thể được sử dụng làm cơ sở cho chiến lược bán hàng và tiếp thị của các MSME Cà phê Văn hóa nhằm tăng doanh thu kinh doanh.[7]

---

Phân tích giỏ thị trường thương mại điện tử bằng thuật toán Apriori. Việc sử dụng Thuật toán Apriori được trình bày để thực hiện phân tích giỏ hàng thị trường nhằm xác định mô hình mua hàng cho các mặt hàng thường được khách hàng mua cùng nhau nhằm cải thiện doanh số bán hàng của các cửa hàng đa sản phẩm bằng cách tăng cường vị trí sản phẩm dựa trên thói quen mua sắm của người tiêu dùng.[9]

### 2.1.3 Ứng dụng vào phân tích giỏ thị trường

Thuật toán Apriori được ứng dụng rộng rãi trong phân tích giỏ hàng mua sắm để tìm ra các mẫu mua hàng phổ biến và xác định các mối quan hệ giữa các mặt hàng. Các ứng dụng cụ thể bao gồm:

- ❖ Marketing và Quảng cáo: Phân tích giỏ hàng có thể giúp các nhà bán lẻ hiểu được hành vi mua hàng của khách hàng, từ đó tối ưu hóa chiến lược quảng cáo và tiếp thị.[30]
- ❖ Quản lý tồn kho: Hiểu được các mối quan hệ giữa các mặt hàng có thể giúp các doanh nghiệp quản lý tồn kho hiệu quả hơn.
- ❖ Đề xuất sản phẩm: Dựa trên các quy luật kết hợp, hệ thống có thể đề xuất sản phẩm phù hợp với khách hàng dựa trên lịch sử mua hàng của họ.
- ❖ Thiết kế kệ hàng: Các quy luật kết hợp có thể được sử dụng để sắp xếp các mặt hàng trên kệ hàng một cách hiệu quả để tăng doanh số bán hàng.

## 2.2 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



---

### 2.2.1 Khái niệm

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

### 2.2.2 Các nghiên cứu

Phân tích giỏ thị trường với Phương pháp cải tiến Apriori và FP-Growth Pada Data Transaksi XYZ. 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.[14]

So sánh các thuật toán tăng trưởng Apriori, Apriori-TID và FP trong phân tích giỏ hàng tại các cửa hàng tạp hóa. 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.[15]

Tối ưu hóa phân tích giỏ hàng thị trường bằng thuật toán phân cụm dựa trên centroid và thuật toán FP-Growth. Kết quả của nghiên cứu này cho thấy mô hình phân cụm có giá trị chỉ số DBI và hình bóng tốt nhất là  $k=3$  đối với k-mean, k-medoids và  $k=7$  đối với mờ c-me nghĩa, có nghĩa là mức tăng trung bình cao nhất tỷ lệ nằm trong quy tắc kết hợp được tạo từ tập dữ liệu nhóm bằng k-Means.[22]

Phân tích giỏ thị trường với thuật toán Apriori và tăng trưởng mô hình thường xuyên (FP-Growth) trên dữ liệu bán sản phẩm ngoài trời. Bằng cách sử dụng phương pháp phân tích giỏ hàng thị trường để xem mối liên kết (quy tắc) giữa một số thuộc tính bán hàng, mô hình mối quan hệ trong giao dịch xảy ra tại các cửa hàng ngoài trời ở Indonesia được xác định.[25]

### 2.2.3 Ứng dụng vào phân tích giỏ thị trường:

Phân tích mẫu mua hàng phổ biến: FP-Growth có thể được sử dụng để phát hiện các mẫu mua hàng phổ biến trong dữ liệu giỏ hàng mua sắm, giúp các doanh nghiệp hiểu được xu hướng mua hàng của khách hàng.

Đề xuất sản phẩm: Dựa trên các mẫu phổ biến, FP-Growth có thể giúp hệ thống đề xuất các sản phẩm phù hợp với từng khách hàng dựa trên lịch sử mua hàng của họ.



---

Quản lý tồn kho: Hiểu được các mối quan hệ giữa các mặt hàng có thể giúp các doanh nghiệp quản lý tồn kho hiệu quả hơn bằng cách dự đoán nhu cầu sản phẩm.

### 3. Thực nghiệm (15 trang)

#### 3.1 Tiền xử lý dữ liệu

Dữ liệu hiện tại đang có là dữ liệu thô. Tiến hành sử dụng jupyter notebook để xử lý dữ liệu.

- Tiến hành tải và hiển thị dữ liệu: Đây là bước đầu tiên trong tiến trình tiền xử lý dữ liệu. Việc nhìn qua dữ liệu (có thể là 5 dòng đầu) giúp hình dung được cơ bản cấu trúc của dữ liệu, thông tin về dữ liệu

```
In [2]: import pandas as pd
f = pd.ExcelFile('D:/Assignment-1_Data.xlsx')
df = pd.read_excel(f)
```

```
In [3]: #In 5 dòng đầu df
df.head()
```

Out[3]:

	BillNo	Itemname	Quantity	Date	Price	CustomerID	Country
0	536365	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom
1	536365	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
2	536365	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom
3	536365	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
4	536365	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom

- Loại bỏ dữ liệu có số lượng âm hoặc bằng 0 (Ko có giá trị phân tích) và những dữ liệu không phải sản phẩm.

```
In [4]: #Loại bỏ dữ liệu có số lượng âm hoặc bằng 0 (Ko có giá trị phân tích)
df=df.loc[df['Quantity']>0]
```

```
In [5]: #Loại bỏ dữ liệu có giá âm hoặc bằng 0
df=df.loc[df['Price']>0]
```

```
In [6]: #Loại bỏ những dữ liệu không phải sản phẩm
df=df.loc[(df['Itemname']!='POSTAGE')&(df['Itemname']!='DOTCOM POSTAGE')&
(df['Itemname']!='Adjust bad debt')&(df['Itemname']!='Manual')]
```

- Điền các dữ liệu còn trống

---

```
In [4]: #Liệt kê các giá trị trống
df.isnull().sum()
```

```
Out[4]: BillNo          0
Itemname        1455
Quantity        0
Date            0
Price           0
CustomerID     134041
Country         0
dtype: int64
```

```
In [8]: #Điền dữ liệu null
df=df.fillna('-')
#Check lại còn giá trị null hay không:
df.isnull().sum()
```

```
Out[8]: BillNo          0
Itemname          0
Quantity          0
Date              0
Price             0
CustomerID        0
Country           0
dtype: int64
```

- Chuyển cột “Date” sang kiểu “datetime”. Sau đó tạo thêm các cột “Year” và “Month” từ cột “Date”

```
In [9]: # Chuyển cột 'Date' sang kiểu datetime
df['Date'] = pd.to_datetime(df['Date'])

# Tạo cột 'Year' từ cột 'Date'
df['Year'] = df['Date'].dt.year

# Tạo cột 'Month' từ cột 'Date'
df['Month'] = df['Date'].dt.month
```

```
In [10]: #Tạo cột Total Price
df['Total price']=df.Quantity*df.Price
```

- Tính doanh thu theo từng tháng. Nhận thấy năm 2010 chỉ có dữ liệu của tháng 12 nên xóa dữ liệu năm 2010

```
In [12]: #Tính total price trong mỗi tháng
df.groupby(['Year', 'Month'])['Total price'].sum()
```

```
Out[12]: Year  Month
2010  12      778386.780
2011   1      648311.120
        2      490058.230
        3      659979.660
        4      507366.971
        5      721789.800
        6      710158.020
        7      642528.481
        8      701411.420
        9      981408.102
       10     1072317.070
       11     1421055.630
       12      606953.650
Name: Total price, dtype: float64
```

```
In [14]: df = df.drop(df[df['Year'] == 2010].index)
```

## 3.2 Các phương pháp cơ sở

Luật kết hợp là một hướng quan trọng trong khai phá dữ liệu, giúp ta tìm được các mối liên hệ giữa các mục dữ liệu/thuộc tính (items) của dữ liệu, tìm các luật kết hợp “quý hiếm” và mang nhiều thông tin từ cơ sở dữ liệu tác nghiệp là một trong những hướng tiếp cận chính của lĩnh vực khai phá dữ liệu.

### 3.2.1 Các hướng tiếp cận trong khai phá luật kết hợp

- ❖ Luật kết hợp nhị phân
  - Các items chỉ được quan tâm là có hay không xuất hiện trong cơ sở dữ liệu giao tác chứ không quan tâm về mức độ hay tần xuất xuất hiện
  - Thuật giải Apriori
- ❖ Luật kết hợp có thuộc tính số và thuộc tính hạng mục
  - Dùng các phương pháp rời rạc hóa chuyển về dạng nhị phân để có thể áp dụng các thuật giải đã có
- ❖ Luật kết hợp tiếp cận theo hướng tập thô
  - Tìm kiếm luật kết hợp dựa trên lý thuyết tập thô
- ❖ Luật kết hợp nhiều mức
  - 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.
- ❖ Luật kết hợp mờ
  - Với những khó khăn gặp phải khi rời rạc hóa các thuộc tính số, luật kết hợp mờ khắc phục hạn chế đó và chuyển luật kết hợp về một dạng gần gũi hơn
- ❖ Luật kết hợp với thuộc tính được đánh trọng số
  - Các thuộc tính được đánh trọng số theo mức độ xác định nào đó

- Nhờ vậy, thu được những luật “hiếm” (tức là có độ hỗ trợ thấp nhưng mang nhiều ý nghĩa)
- ❖ Luật kết hợp song song
- Nhu cầu song song hóa và xử lý phân tán là cần thiết vì kích thước dữ liệu ngày càng lớn

### 3.2.2 Quy tắc kết hợp

- ❖ Cho  $I = \{I_1, I_2, \dots, I_n\}$  là một tập các mục (mặt hàng,...)
- ❖ Cho D là một tập các giao dịch mà mỗi giao dịch T là một tập các mục,  $T \subseteq I$
- ❖ Mỗi giao dịch có một mã định danh riêng gọi là TID
- ❖ Cho A là một tập các mục (mặt hàng). Một giao dịch T được gọi là chứa A khi và chỉ khi  $A \subseteq T$
- ❖ Một luật kết hợp được diễn đạt dưới hình thức  $A \Rightarrow B$ , với  $A \subset I, B \subset I, A \cap B = \emptyset$
- ❖ Ý nghĩa: Khi xuất hiện A thì B cũng xuất hiện (với xác suất nào đó)

VD: Mô tả cơ sở dữ liệu tác vụ A, C, D, T, W là các mục:  $T_i = 1, 2, 3, 4, 5, 6$  là các tác vụ

Tập các tác vụ	Các mục dữ liệu
1	A C T W
2	C D W
3	A C T W
4	A C D W
5	A C D T W
6	C D T

Mỗi giá trị của mục dữ liệu (Item) thể hiện thuộc tính xuất hiện (nhận giá trị 1) hay không xuất hiện (nhận giá trị 0) trong tác vụ

TID	A	C	D	T	W
$T_1$	1	1	0	1	1
$T_2$	0	1	1	0	1
$T_3$	1	1	0	1	1
$T_4$	1	1	1	0	1
$T_5$	1	1	1	1	1
$T_6$	0	1	1	1	0

### 3.2.3 Các thông số quan trọng của luật kết hợp

#### 3.2.3.1 Độ hỗ trợ (Support)

Độ hỗ trợ (Support) của  $X \Rightarrow Y$  là tỷ lệ phần trăm các bản ghi  $X \cup Y$  với tổng số các giao dịch có trong cơ sở dữ liệu

---


$$Support(X \Rightarrow Y) = \frac{|X \cup Y|}{n}$$

Support cung cấp thông tin về sự phổ biến của một quy luật kết hợp trong tập dữ liệu. Một support cao cho thấy rằng luật kết hợp xuất hiện thường xuyên trong dữ liệu, trong khi support thấp có thể chỉ ra rằng luật kết hợp ít phổ biến.

### 3.2.3.2 Độ tin cậy (Confidence)

Độ tin cậy (Confidence) của  $X \Rightarrow Y$  là tỷ lệ phần trăm của số giao dịch có chứa  $X \cup Y$  với số giao dịch có chứa  $X$

$$Confidence(X \Rightarrow Y) = \frac{Support(X \cup Y)}{Support(X)}$$

### 3.2.3.3 Lift

Chỉ số Lift đo lường mức độ mà một mục trong quy luật kết hợp (ví dụ: "nếu X xảy ra, thì Y cũng xảy ra") thực sự tăng cường khả năng xảy ra của mục khác so với trường hợp nếu chúng độc lập với nhau

$$Lift(X \Rightarrow Y) = \frac{Support(X \cup Y)}{Support(X) \times Support(Y)}$$

Nếu  $Lift > 1$ , điều này cho biết rằng sự xuất hiện của mục X có xu hướng làm tăng khả năng xuất hiện của mục Y, trong khi  $Lift < 1$  thì có nghĩa là sự xuất hiện của mục X có xu hướng làm giảm khả năng xuất hiện của mục Y. Trong khi đó,  $Lift = 1$  cho biết không có mối quan hệ giữa các mục.

### 3.2.3.4 Conviction

Chỉ số Conviction được sử dụng để đo độ tin cậy của một luật kết hợp. Conviction đo lường mức độ mà một quy luật kết hợp được xem là không phụ thuộc vào sự xuất hiện hay không xuất hiện của mục thứ hai trong quy luật đó, giả định rằng quy luật đã được xác định.

$$Conviction(X \Rightarrow Y) = \frac{1 - Support(Y)}{1 - Confidence(X \Rightarrow Y)}$$

Conviction càng cao thì mức độ tin cậy của quy luật càng thấp, và ngược lại. Nếu  $Conviction = 1$ , điều này cho thấy rằng sự xuất hiện của X và Y là hoàn toàn độc lập và không có mối quan hệ giữa chúng.