Toward Enhanced Customer Transaction Insights: An Apriori Algorithm-based Analysis of Sales Patterns at University Industrial Corporation

Alex Alfredo Huaman Llanos[1], Lenin Quiñones Huatangari[2], Jeimis Royler Yalta Meza[3],
Alexander Huaman Monteza[4], Orestes Daniel Adrianzen Guerrero[5], John Smith Rodriguez Estacio[6]

Informatic and Language Center, National University of Jaen, Jaen, Peru[1]
Data Science Research Institute, National University of Jaen, Jaen, Peru[2]
Direction of Production Center of Goods and Services, National University of Jaen, Jaen, Peru[3]
Social Science Academic Department, National University of Jaen, Jaen, Peru[4]
University Industrial Corporation, National University of Jaen, Jaen, Peru[5, 6]

***Abstract*—The University Industrial Corporation (CIU) at the National University of Jaen offers a range of consumable products, encompassing nectar, water, coffee, chocolate, and chocoteja. However, its sales transactions function without a systematic analysis. To address this, the study gathered and analyzed sales data from March to November 2023, aiming to identify and delineate associations among frequently co-purchased products, revealing underlying interdependencies and associations. Employing text mining methodologies, this study preprocessed and analyzed 1542 sales records using the Apriori algorithm, culminating in the extraction of 17 association rules. Among these rules, three standout associations were uncovered: the purchase of chocolate, chocoteja and water suggests a purchase of nectar; chocolate, nectar and water acquisitions correlate with chocoteja purchases; lastly chocolate and nectar purchases are associated with chocoteja acquisitions. These findings provide insights to augment potential production adjustments within the CIU, enabling the leveraging of established associations to boost sales and revenue. Moreover, the identified rules serve as a cornerstone for decision-makers, actionable guidance for stakeholders, enabling the identification of co-purchased products, fostering informed production planning, fine-tuning marketing strategies for customer relationship management (CRM), and enhancing CIU's market competitiveness and profitability.**

***Keywords*—Apriori algorithm; association rules; Customer Relationship Management (CRM); decision making; text mining**

I. Introduction

Recent advancements in information technologies have propelled transformative opportunities in several public and private institutions, streamlining services and operational efficiencies for the populations. Consequently, this technological progression has empowered companies to acquire, store, analyze, and interpret vast volumes of data, marking a pivotal shift in the significance of dataset. Thus, the consequential impacts transcend the technological domain, influencing the restructuring of business strategies and marketing activities [1].

Simultaneously, the widespread accessibility of the Internet and the burgeoning domain of e-commerce have accumulated extensive repositories of customer transactional data within websites. Employing sophisticated data mining techniques has enabled the aggregation and analysis of these massive datasets within complex web structures [2], [3]. This revolutionizing data mining movement has significantly shaped the information landscape and society, facilitating the process of transforming large amounts of data into valuable insights and knowledge [4]. The fundamental objective of data mining remains centered on the discovery of high-frequency sets, fostering the extraction of implicit associations, subsequently aiding in informed decision-making. As observed in [5], there exists an intriguing correlation between seemingly unrelated products, such as diapers and beer, co-placement of which resulted in a mutual sales upsurge.

Moreover, text mining, a subset of data mining, involves the transformation of unstructured and semi-structured textual data from vast databases into digital formats for information extraction. Its functional applications span a wide range, including semantic text mining, word feature mining, association rule mining, text clustering analysis, and trend prediction [6].

On the other hand, data mining association algorithms serve as indispensable tools, widely embraced by major tech companies such as Amazon, Google, Netflix, Facebook, and Twitter. These algorithms effectively uncover intricate interrelations among variables within extensive datasets. By extracting interesting correlations, frequencies, and patterns from association rules [7], proving instrumental in predicting relevant elements in subsequent analysis [8].

The array of association rule algorithms spans the Apriori algorithm, Frequent Pattern (FP)-Tree, Equivalence Class Clustering, bottom-up Lattice Traversal (ECLAT) algorithm, and the gray association method. Notably, the Apriori algorithm developed by Agrawal and Srikant [9], stand as the classical paradigm for mining frequent itemset in the domain of association rules analysis [10]. This algorithm operates on a bottom-up search methodology, progressively assembling generated itemset to ensure the identification of frequent item subsets.

The University Industrial Corporation (CIU), an integral component of the Production Center of Goods and Services at the National University of Jaen (UNJ), plays a pivotal role in advancing academic, research, and production endeavors in alignment with the institution's overarching goals. As a part of its unwavering commitment to foster scientific inquiry, promote the UNJ brand, and deliver competitive, high-quality products, the CIU takes the lead in exploring new market opportunities and devising scalable, replicable business strategies.

This study endeavors to perform a comprehensive analysis of product purchases by employing association rules. To accomplish this objective, the research leveraged the renowned Apriori algorithm, which adeptly identifies frequent items within transactional databases. The research dataset, spanning sales from March to November 2023, meticulously focuses on key products offered by the CIU, including chocolate, chocoteja, water, nectar, and coffee.

The organization of this paper is as follows: Section I gives a brief background on data mining. Section II, presents the related research, highlighting key advancements and gaps in the field. Section III outlines the rigorous methodology employed in this study. The core findings, accompanied by in-depth analysis and contextual discussions are conducting in Section IV, and Section V draws the meaningful conclusions, and outlines avenues for future research.

II. State of the Art

The implementation of association rules through the Apriori algorithm has been used in different aspects of knowledge, showcasing its versatility and applicability in diverse fields such as marketing, health sciences, computer science, transportation, education, labor, and mining. In marketing, its utility extended to the dissection of purchase transaction-based product promotions [8], recommending e-commerce links [2], and characterizing e-customer behavior to discern purchase patterns among two customer groups [3]. Noteworthy studies have included analysis of purchased products [11], examination of consumer purchasing patterns [12], encompassing market basket analysis [13], and online shopping through RFMDR model [14]. Furthermore, it found practical application in minimum spanning tree-based shopping [15], and Internet marketing strategies in the cosmetics sales in Taiwan [16].

In the realm of health sciences, association rules' usage facilitated the sifting of potential analgesics from a pool of 311 cases treated with compounded drug prescriptions, extracting data on clinical symptoms and types of Chinese herbs [17]. Notably, it facilitated the development of an automatic diagnostic system for breast cancer detection [18], and the identification of distinguishing factors between dementia patients, and caregivers linked to long-term care services [19]. Furthermore, it revealed a set of frequent items aiding in fetal anomaly detection [20], and uncovered the primary combination of Chinese herbs for Alzheimer's disease treatment within acupuncture practices [21]. Research efforts extended into investigating acupuncture point combinations for the treatment of hemiparesis [22], exploring protein-gene interactions from omics data [23], and identifying links between Polycystic Ovary Syndrome (PCOS), and hormonal imbalances utilizing the DEODORANT model [24]. Associations between attributes characterizing the perfusion patterns in normal subjects illuminated the diagnosis of Alzheimer’s disease [25], unraveling pathogenesis linked to thyroid disease [26], analysis of comorbidity in residents afflicted by chronic diseases [27], and classification of anxiety in palliative patients [28]. Moreover, it delineated specific correlations between individuals with dementia and caregivers based on various dementia subtypes [29].

In the Information Technology (IT) context, a methodological approach has been proposed to advance requirements engineering within the enterprise software domain [30]. Also, significant contributions include outlier cleaning in network measurement data [31], processing voluminous datasets through membrane computing models [32], and developing mobile e-commerce recommender systems for online shopping [33]. Additionally, notable efforts have been directed towards bolstering the security of global cyberspace [34] and enhancing the water wave optimization algorithm [35].

In the transportation sector, specific methodologies have been tailored to scrutinize and process data derived from Iranian railroad accidents [36], and identifying risk factors associated with freight truck accidents [37]. Educational research has surfaced various funding initiatives. Among these are studies analyzing user behavior when requesting texts from the library loans [38], uncovering natural products and anti-migraine nutraceuticals from extensive classical medical literature collections [39], and rule extraction from the scoring records of 2002 computer science students at the Mongolian University of Science and Technology [40].

Within the domain of employment, through investigations were conducted to explore the correlations between employment status, and the employability indicators of maritime graduates [41]. Another study proposed a customer potential value matrix, designed to segment applicants based on their potential value and willingness to engage in purchases, thereby enhancing the scope of customer segmentation strategies [42]. Lastly, in the mining sector, an investigation scrutinized the link between structural deformation and gold mineralization, offering valuable insights into the intricate relationship [43].

This study is based on the pressing need to integrate innovative and technological approaches in the business environment. By using Apriori algorithms and association rules in CIU-Jaen, the complexity of large volumes of data is explored, revealing latent patterns and underlying relationships, that generate opportunities for improvement and facilitating agile decision-making. The exhaustive analysis of the association rules allows deepening the knowledge of the operational dynamics and the behavior of the customers, unveiling behavioral patterns and preferences, facilitating the design of solid strategies to boost growth and ensure sustainable development. The adoption of technological tools and data-based strategies within the business field is a key step to stimulate and promote innovation and adaptability in a competitive and ever-evolving environment.

III. Methodology

The research consists of four crucial steps and is visually represented in Fig. 1:

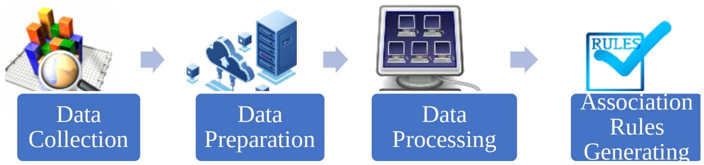
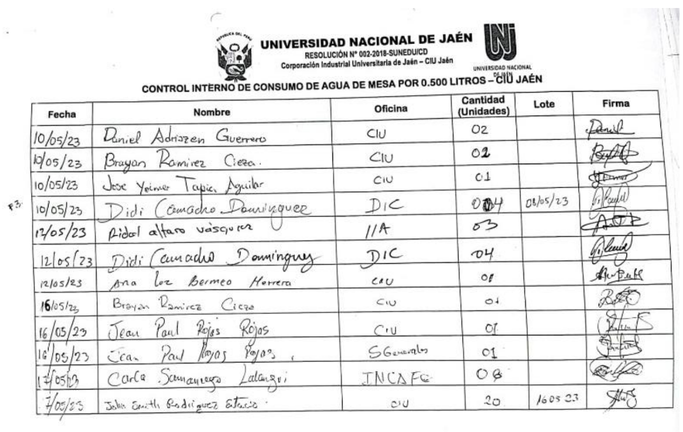


Fig. 1. Apriori algorithm schema.

A. Data Collection

The dataset was meticulously procured a comprehensive dataset from the sales register maintained by the University Industrial Corporation. The corporation employs a manual record-keeping register paper, which includes detailed information about products, such as chocolate, chocoteja, coffee, nectar, and water. The data collection efforts spanned form March to November 2023 as shows in Fig. 2.



UNIVERSIDAD NACIONAL DE JAÉN
RESOLUCIÓN N° 002-2018-SUNEDU/CD
Corporación Industrial Universitaria de Jaén – CIU Jaén

CONTROL INTERNO DE CONSUMO DE AGUA DE MESA POR 0.500 LITROS - CIU JAÉN

Fecha	Nombre	Oficina	Cantidad (Unidades)	Lote	Firma
10/05/23	Daniel Adriazen Guerrero	CIU	02		
10/05/23	Brayan Ramirez Cieza	CIU	02		
10/05/23	Jose Yeimer Tapia Aguilar	CIU	01		
10/05/23	Didi Camacho Dominguez	DIC	04	01/05/23	
12/05/23	Ridol alfaro vasquez	IIA	03		
12/05/23	Didi Camacho Dominguez	DIC	04		
12/05/23	Ana Luz Bermeo Herrera	CIU	01		
16/05/23	Brayan Ramirez Cieza	CIU	01		
16/05/23	Jean Paul Rojas Rojas	CIU	01		
16/05/23	Jean Paul Rojas Rojas	SGeneral	01		
17/05/23	Carla Samaniego Lalangui	INCAFE	08		
17/05/23	John Smith Rodriguez Alvarez	CIU	20	16/05/23	

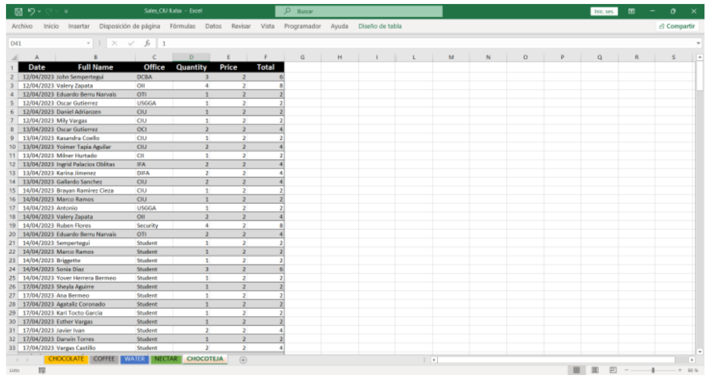
Fig. 2. CIU-Jaen data collection.

B. Data Preparation

The acquired data was diligently organized within an Excel worksheet (.xlsx), delineating key attributes such as date, client details, product names, work office, quantity and pricing fields, which is depicted in Fig. 3. Subsequently, a structured flat file (.csv) was generated, employing comma-separated values for seamless data manipulation.

C. Data Processing

Leveraging the powerful capabilities of the R software ecosystem. Specifically, we harnessed the tidyverse, arules, plyr, and arulesviz libraries to transform raw data into meaningful insights, as shown in Fig. 4. The focus of the study was on extracting association rules, a critical step in uncovering hidden patterns and relationships within the dataset. To enhance interpretability, the visualization of the extracted rules was using the arulesviz library. The graphical representations employed three distinct layouts: Fruchterman-Reingold, Kamada-Kawai, and Reingold-Tilford graphs. These layouts elegantly depicted the interconnected nodes corresponding to products like chocolate, chocoteja, nectar, water, and coffee. Arrows extending from antecedent to consequent nodes illustrated the direction of association, emphasizing the sequential relationships. The size of circles positioned at the nexus of these arrows conveyed support values, reflecting the frequency of occurrence for each rule. Additionally, coloration was strategically applied to denote the associated lift values.



	Date	Full Name	Office	Quantity	Price	Total
2	12/04/2023	John Sempertegui	DCBA	3	2	6
3	12/04/2023	Valery Zapata	OII	4	2	8
4	12/04/2023	Eduardo Berru Narvais	OTI	1	2	2
5	12/04/2023	Oscar Gutierrez	USGGA	1	2	2
6	12/04/2023	Daniel Adrianzen	CIU	1	2	2
7	12/04/2023	Mily Vargas	CIU	1	2	2
8	13/04/2023	Oscar Gutierrez	OCI	2	2	4
9	13/04/2023	Kasandra Coello	CIU	1	2	2
10	13/04/2023	Yoimer Tapia Aguilar	CIU	2	2	4
11	13/04/2023	Milner Hurtado	CII	1	2	2
12	13/04/2023	Ingrid Palacios Oblitas	IFA	2	2	4
13	13/04/2023	Karina Jimenez	DIFA	2	2	4
14	13/04/2023	Gallardo Sanchez	CIU	2	2	4
15	14/04/2023	Brayan Ramirez Cieza	CIU	1	2	2
16	14/04/2023	Marco Ramos	CIU	1	2	2
17	14/04/2023	Antonio	USGGA	1	2	2
18	14/04/2023	Valery Zapata	OII	2	2	4
19	14/04/2023	Ruben Flores	Security	4	2	8
20	14/04/2023	Eduardo Berru Narvais	OTI	2	2	4
21	14/04/2023	Sempertegui	Student	1	2	2
22	14/04/2023	Marco Ramos	Student	1	2	2
23	14/04/2023	Briggette	Student	1	2	2
24	14/04/2023	Sonia Diaz	Student	3	2	6
25	14/04/2023	Yover Herrera Bermeo	Student	1	2	2
26	17/04/2023	Sheyla Aguirre	Student	1	2	2
27	17/04/2023	Ana Bermeo	Student	1	2	2
28	17/04/2023	Agataliz Coronado	Student	1	2	2
29	17/04/2023	Kari Tocto Garcia	Student	1	2	2
30	17/04/2023	Esther Vargas	Student	1	2	2
31	17/04/2023	Javier Ivan	Student	2	2	4
32	17/04/2023	Darwin Torres	Student	1	2	2
33	17/04/2023	Vargas Castillo	Student	2	2	4

Fig. 3. Data stored in excel.

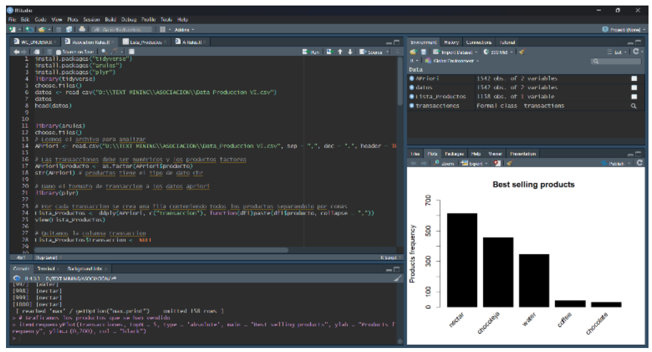


Fig. 4. Use R software for data processing.

D. Association Rules Generating

Three criteria were used to identify the relationship: support confidence, and lift.

- Support: It is the percentage of transactions in the database that contain both itemset A and B. The degree of support $A \Longrightarrow B$ in the rule $A \Longrightarrow B$, is the probability that a given set of itemset contain A and B, which is expressed by the value of probability P(AUB) [44]. A high degree of support indicates that the mining results are consistent and that the provided rules are effective association rules. On the other side, a low degree of support indicates that the data mining results appear only occasionally and the provided rules have little value for the research. Eq. (1) represents the definition of support of the association rule between A and B [45].

$$\text{Support } A \Longrightarrow B = P(A \cap B) = \frac{|A \cup B|}{|D|} \tag{1}$$

- Confidence: It is the percentage of transactions in database D with item set A that also contain item set B [5]. Confidence is calculated using the conditional probability and is expressed relative to the item set support [46] and is represented by Eq. (2):

$$\text{Confidence A} \Longrightarrow \text{B} = \frac{support\ (A \Longrightarrow B)}{support\ (A)} = \frac{P(\text{A} \cap \text{B})}{\text{P(A)}} \quad (2)$$

In Eq. (2), support $(A \Longrightarrow B)$ is the number of transactions containing the itemset A and B, support (A) is the number of transactions containing the itemset A [44].

- Lift: Used to measure the frequency of A and B together, if both sets of items are statistically independent of each other [47]. The calculation is as shown in Eq. (3):

$$\text{Lift A} \Longrightarrow \text{B} = \frac{confidence\ (A \Longrightarrow B)}{support\ (A)} = \frac{P(\text{A} \cap \text{B})}{\text{P(A)P(B)}} \quad (3)$$

The lift of the rule A $\Longrightarrow$ B shows how much the probability of B will increase, if A occurs [48]. There are three cases:

- When lift $(A \Longrightarrow B) > 1$, then there is a positive interdependence between the antecedent and consequent; so, the rule is considered valuable.
- When lift $(A \Longrightarrow B) < 1$, then there is a negative interdependence between the antecedent and consequent.
- When $(A \Longrightarrow B) = 1$, then A and B are independent and there is no correlation between them.

Therefore, the higher the measure of lift, the higher the interest of the generated rules will be. So, with the help of this measure, it will classify the rules that meet the minimum thresholds of support and confidence.

IV. Result Analysis and Discussions

The extracted association rules, meticulously presented in Table I, offer profound insights into the intricate relationships and patterns inherent in the consumer transactions within the dataset. Simultaneously, the graphical representation in Fig. 5, vividly illustrates the frequency distribution of the sold products. Upon discerning the bar chart, a discernible hierarchy in product popularity emerges. Foremost among the products is nectar, indisputably leading in sales volume. Subsequently, chocoteja claims the second position, closely followed by water in third. However, it is imperative to note that, coffee and chocolate, manifest a comparatively lower acceptance rate among consumers.

This analysis not only informs products positioning within the market, but also lays the foundation for strategic decision-making, offering a profound understanding of consumer preferences and market dynamics. The identification of less favored products underscores potential areas for marketing enhancement, contributing to an informed approach for maximizing sales and overall business efficacy.

The outcomes showed in Table I illustrated 17 association rules, each distinguished by a pivotal support attribute, signifying the frequency of rule occurrence within the dataset. It's important to state that, a higher threshold correlates with an augmented frequency of rule manifestation. For instance, the acquisition of chocoteja and nectar frequently coincides with chocoteja, and vice versa.

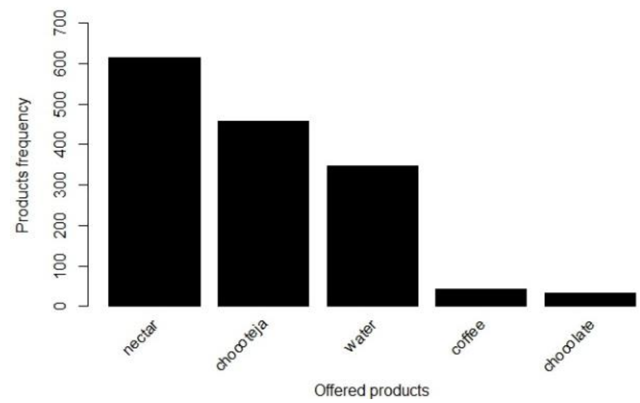


Fig. 5. Bar chart of best-selling products.

TABLE. I ASSOCIATION RULES OBTAINED

Nº	Association rules	Support	Confidence	Lift	Leverage
1	{chocolate} => {chocoteja}	0.0181	0.6563	1.6593	0.0072
2	{chocolate} => {nectar}	0.0173	0.6250	1.1787	0.0026
3	{coffee} => {chocoteja}	0.0225	0.6190	1.5652	0.0081
4	{coffee} => {nectar}	0.0268	0.7381	1.3920	0.0075
5	{chocolate, water} => {chocoteja}	0.0121	0.8235	2.0822	0.0063
6	{chocolate, chocoteja} => {water}	0.0121	0.6667	2.2184	0.0066
7	{chocolate, water} => {nectar}	0.0130	0.8824	1.6641	0.0052
8	{chocolate, nectar} => {water}	0.0130	0.7500	2.4957	0.0078
9	{chocolate, chocoteja} => {nectar}	0.0155	0.8571	1.6166	0.0059
10	{chocolate, nectar} => {chocoteja}	0.0155	0.9000	2.2755	0.0087
11	{coffee, water} => {nectar}	0.0199	0.9200	1.7351	0.0084
12	{coffee, nectar} => {water}	0.0199	0.7419	2.4689	0.0118
13	{chocoteja, water} => {nectar}	0.0484	0.6154	1.1606	0.0067
14	{chocolate, chocoteja, water} => {nectar}	0.0121	1.0000	1.8860	0.0057
15	{chocolate, nectar, water} => {chocoteja}	0.0121	0.9333	2.3598	0.0070
16	{chocolate, chocoteja, nectar} => {water}	0.0121	0.7778	2.5881	0.0074
17	{chocoteja, coffee, water} => {nectar}	0.0121	0.9333	1.7603	0.0052

Table I comprehensively details 17 rules scrutinized with 1542 purchase records involving five different products. Each entry includes a robust rule alongside its corresponding metrics, including support, confidence, lift and leverage. Although the vast dataset and modest support might diminish the impact on confidence verification, three strong rules, considering the aforementioned parameters, warrant attention:

- Rule 14: {chocolate, chocoteja, water} ⇒{nectar}
- Rule 15: {chocolate, nectar, water} ⇒{chocoteja}
- Rule 10: {chocolate, nectar} ⇒{chocoteja}

Examining rule 14, the support stands at 1.21%, confidence at 100%, lift at 1.886, and leverage at 0.00557. These results have a perfect confidence, indicating that whenever chocolate, chocoteja, and water are present, nectar is also present. The high lift value; suggest a strong association between the items, making it a significant rule.

It is the percentage of transactions in the database that contain both itemset A and B. The degree Rule 15, with a support of 1.21%, confidence of 93.33%, lift of 2.3598, and leverage of 0.007, indicating that when chocolate, nectar, and water are present, there is a strong likelihood of chocoteja being present. The lift of 2.3598 suggests a significant association.

Lastly, rule 10 boasting 1.55% support, 90% confidence, 0.0087 leverage, and 2.2755 lift. The high confidence and lift make this rule significant. It indicates a strong association between chocolate and nectar leading to the presence of chocoteja. These selected rules are not only supported by high confidence but also exhibit substantial lift, indicating that the items involved are more likely to be purchased together than if they were chosen randomly. Leverage provides additional insights into the strength of the association, and in these cases, it complements the high confidence and lift values.

The Fig. 6 displays a grouped matrix for 17 rules, where rows represent the RHS (Right Hand Side), columns represent the LHS (Left Hand Side) items, and cells convey information about the strength of association between these elements. Colors denote the lift of products, while size represents the support between them. This graphical representation enhances the comprehension of association relationships between LHS and RHS elements based on metrics such as support and lift. Such a visual approach aids in identifying patterns and interpreting association rules within the context of CIU product purchases.

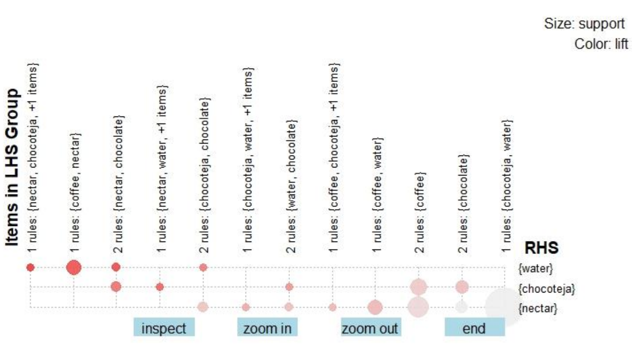


Fig. 6. Grouped matrix for 17 rules.

In the Fig. 7 shown below, there is a visual representation of the association rules extracted through the Apriori algorithm. Specifically, the use of the "htmlwidget" engine enables an interactive experience, allowing users to dynamically explore and analyze the generated rules. For instance, by clicking on a node or edge, users can access detailed information about the rule, including metrics such as support, confidence, and lift. This visualization proves valuable in comprehending purchase patterns and relationships among products, leading to meaningful applications in marketing strategies and business decision-making.

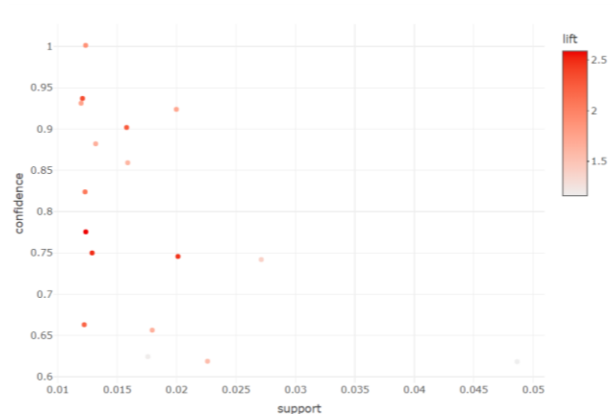


Fig. 7. Scatter plot for 17 association rules.

Analysis of Association Rules

Fig. 8, 9 and 10 depict the association rules, particularly highlighting the rule "chocolate, chocoteja, water ⟹ nectar", showcasing a modest support but a notably high confidence value. Similar observations are made for the rules "chocolate, nectar, water ⟹ chocoteja" and "chocolate, nectar ⟹ chocoteja". The graphical and Table II representations affirm the consistency and interrelations of these rules, with variations attributed to the arrangement elements.

The application of the Fruchterman-Reingold (FR) algorithm, grounded in particle physics principles, is evident in Fig 8. Guided by the notions that connected nodes should be proximate, and others should be proximate, and others should maintain a suitable distance [49]. The FR method considers a graph as a collection of vertices connected by edges with two types of forces acting on the vertices. The graph is generated with 500 iterations and an initial temperature parameter of 4.69.

For the creation of two-dimensional graphs, the Kamada-Kawai (KK) method that conceptualizes a graph as a system of spring, was employed [50]. Fig. 9 illustrates the KK algorithm with 1100 iterations and a constant vertex attraction parameter set to 22. In the KK algorithm, the placement of nodes ensures that their visual distance corresponds proportionally to their plotted distance.

TABLE. II STRONG ASSOCIATION RULES OBTAINED

Nº	Association rules	Support	Confidence	Lift	Leverage
1	{chocolate, chocoteja, water} => {nectar}	0.0121	1.0000	1.8860	0.0057
2	{chocolate, nectar, water} => {chocoteja}	0.0121	0.9333	2.3598	0.0070
3	{chocolate, nectar} => {chocoteja}	0.0155	0.9000	2.2755	0.0087

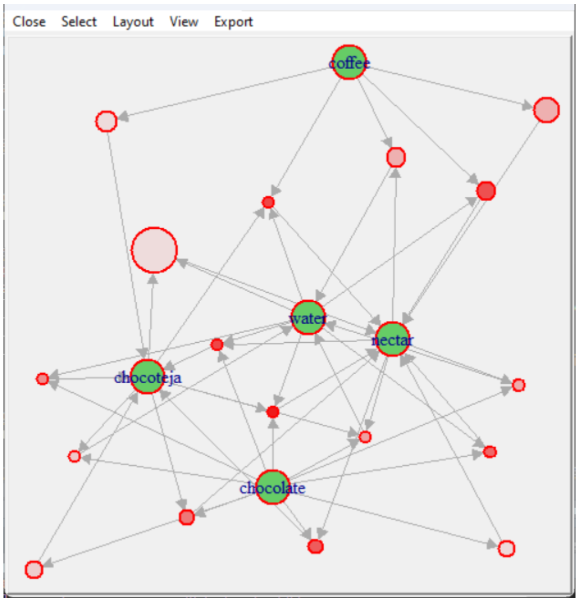


Fig. 8. Graph of Fruchterman-Reingold.

Finally, the Reingold-Tilford algorithm, developed by John Reingold and Jhon Tilford, is instrumental in arranging tree structures to optimize readability. This algorithm assigns coordinates to each tree node, aligning sibling nodes horizontally and positioning child nodes below parent nodes, fostering a visually comprehensible representation of hierarchical data [51]. Fig. 10 demonstrates the application of this algorithm, configured with 1 as the root vertex.

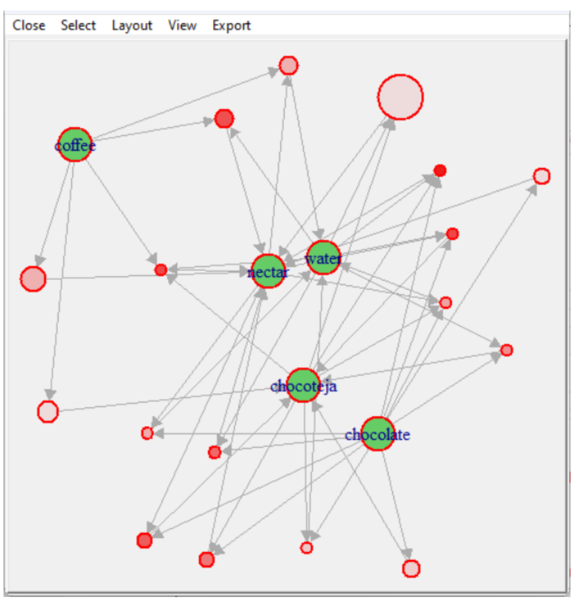


Fig. 9. Graph of Kamada-Kawai.

In Fig. 11, the 17 rules obtained are presented using a visual approach, which provides a graphic and comprehensible representation of the relationships among products. This is essential for interpreting purchase patterns and making strategic decisions in the business and marketing domains. The ability to interactively explore the rules will enable users to gather detailed information about the discovered associations, thereby contributing to a deeper understanding of consumer behavior. In summary, the graphical representations affirm the robustness of the association rules, offering insights into their relationships and highlighting the applicability of diverse algorithms for visualizing these intricate patterns with the dataset. These findings lay a foundation for nuanced interpretations and strategic decision-making.

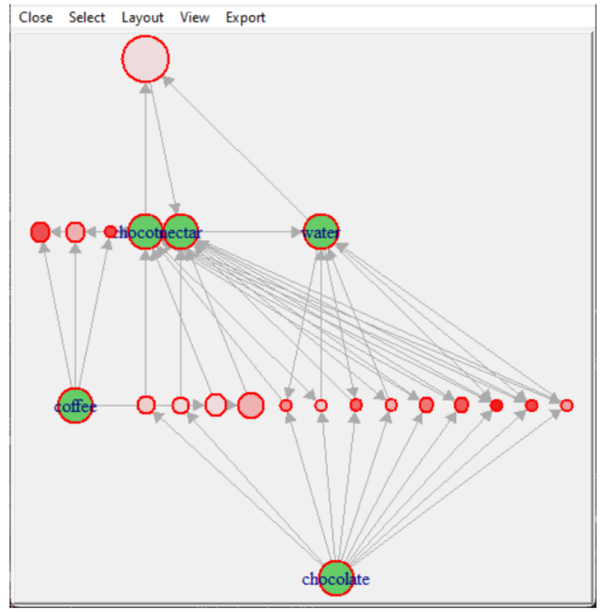


Fig. 10. Graph of Reingold-Tilford.

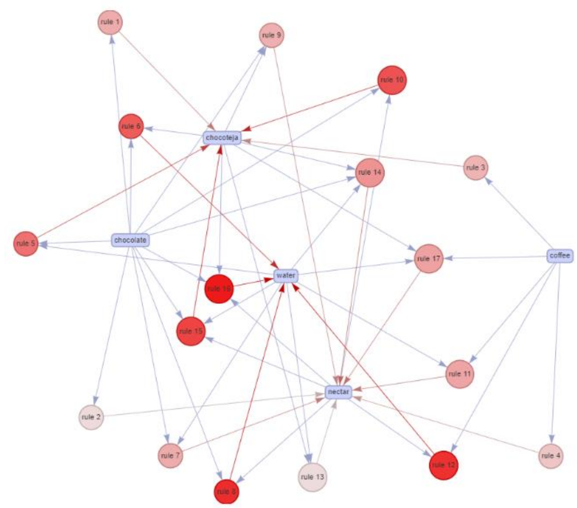


Fig. 11. Graph HTML widget.

In the context of the study, a pivotal challenge confronting marketer lies in augmenting transaction volumes among customers engaged in limited product acquisitions-typically one, two, or three items, a prevalent scenario at CIU-Jaen. Tackling this challenge necessitates the deployment of strategic initiatives, such as flash, strategically amalgamating product displays and promotional information to stimulate impulsive purchases. The subsequent exposition of sales patterns concerning nectar, water, coffee, chocolate and chocoteja, coupled with the discernment of robust association rules, furnishes a valuable foundation for crafting and executing such targeted strategies.

Existing research underscores the affirmative response of customers to well-crafted sales campaigns, particularly in Jaen, a commercial province, where such endeavors hold promising potential for customers attraction and retention. This case study, rooted in the CIU-UNJ market, injects a distinctive dimension into the scholarly discourse by acknowledging and integrating the unique economic and market dynamics inherent to the province of Jaen. Notably, transactions involving three or four products dominated the observe timeframe, a noteworthy insight considering potential marketing budget constraints encountered by institutions within the Jaen market. Despite these constrains, the analysis equips senior management and marketing leaders with the tools to proffer nuanced, tailor-made offering to their customers. Expanding on this point, the identification of frequently co-purchased product combinations empowers companies not only to optimize the physical placement of products within the store but also to devise targeted marketing strategies and product bundles capable of propelling sales.

Nevertheless, it is imperative to acknowledge the constraints imposed by the dataset size and the modest transaction volume averages. Transactions predominantly featuring one or two products limit the scope of the analysis. To enhance the robustness of future studies, it is recommended to use a larger database and the extension of the study duration. This methodological refinement aligns with the pursuit of more objective outcomes, facilitating longitudinal analyses that enable the evaluation of the enduring impacts of strategic marketing decisions on transactional volume and value.

V. Conclusion

This study represents a significant advancement in understanding customer transaction patterns through the application of the Apriori algorithm at CIU-Jaen. By uncovering valuable insights into product associations, the findings inform targeted marketing strategies, discount planning, and direct marketing campaigns, while seamlessly integrating with customer loyalty programs. This holistic approach, embedded within a comprehensive customer relationship management system, enhances the precision and efficacy of marketing initiatives, thereby offering a nuanced understanding of consumer behavior and fostering sustained business growth.

In a broader context, this research offers a nuanced understanding of consumer behavior in the unique market landscape of CIU-Jaen. The association rule identified, based on real transactional data, transcends conventional marketing paradigms, serving as a valuable guide for navigating product placements, promotional efforts, and customer relationship management strategies. As industries grapple with evolving consumer preferences, this study furnishes timely and insightful contributions as a valuable toolset available for strategic decision-making in the domain of marketing and business development.

Looking ahead, future research endeavors could build upon our findings by exploring additional data sources, such as customer feedback and demographic information, to further refine our understanding of consumer behavior. Additionally, investigating the scalability and adaptability of the Apriori algorithm in diverse market contexts could offer valuable insights for practitioners seeking to leverage data-driven approaches in their marketing strategies. Ultimately, by continuously refining our methodologies and insights, we can better anticipate and respond to the evolving dynamics of consumer preferences, driving innovation and growth in the field of marketing and business development.

Acknowledgment

This study was adopted by data reporting sales provided by the University Industrial Corporation. Besides, our sincere gratitude extends to the National University of Jaen for their invaluable support and collaboration throughout the research. Also, we express profound appreciation to every professional whose contributions enriched the depth and breadth of this study.

References

[1] E. Constantinides y S. J. Fountain, «Web 2.0: Conceptual foundations and marketing issues», J Direct Data Digit Mark Pract, vol. 9, n.o 3, pp. 231-244, ene. 2008, doi: 10.1057/palgrave.dddmp.4350098.

[2] L. Zheng, «Research on E-Commerce Potential Client Mining Applied to Apriori Association Rule Algorithm», 2020 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), pp. 667-670, ene. 2020, doi: 10.1109/ICITBS49701.2020.00146.

[3] G. Suchacka y G. Chodak, «Using association rules to assess purchase probability in online stores», Inf Syst E-Bus Manage, vol. 15, n.o 3, pp. 751-780, ago. 2017, doi: 10.1007/s10257-016-0329-4.

[4] R. Sumithra y S. Paul, «Using distributed apriori association rule and classical apriori mining algorithms for grid based knowledge discovery», In 2010 Second International conference on Computing, Communication and Networking Technologies, pp. 1-5, jul. 2010, doi: 10.1109/ICCCNT.2010.5591577.

[5] J. Han, M. Kamber, y J. Pei, Data Mining: Concepts and Techniques, Third Edition. Amsterdam, Netherlands: Elsevier, 2012. [En línea]. Disponible en: https://linkinghub.elsevier.com/retrieve/pii/C20090618195

[6] J. Li, J. Wang, N. Xu, y Z. Zhou, «Analysis of Safety Risk Factors for Metro Construction Based on Text Mining Method», Tunnel Construction, vol. 37, n.o 2, pp. 160-166, feb. 2017, doi: 10.3973/j.issn.1672-741X.2017.02.006.

[7] R. Feldman y H. Hirsh, «Finding Associations in Collections of Text», J. Wiley, pp. 223-240, 1997.

[8] N. Riyadi, M. F. Mulki, y R. Susanto, «Analysis of Customers Purchase Patterns of E-Commerce Transactions Using Apriori Algorithm and Sales Forecasting Analysis with Weighted Moving Average (WMA) Methods», Scientific Research Journal, vol. VII, n.o VII, jul. 2019, doi: 10.31364/SCIRJ/v7.i7.2019.P0719670.

[9] R. Agrawal y R. Srikant, «Fast Algorithms for Mining Association Rules», in Proceedings of the International Conference on Very Large Data Bases, vol. 1, pp. 487-499, sep. 1994.

[10] J. Hong, R. Tamakloe, y D. Park, «Application of association rules mining algorithm for hazardous materials transportation crashes on expressway», Accident Analysis & Prevention, vol. 142, p. 105497, jul. 2020, doi: 10.1016/j.aap.2020.105497.

[11] M. Djukanovic, S. Rogic, L. Novicevic, V. Popovic-Bugarin, y M. Jovanovic, «Application of Apriori Algorithm for CRM Improvement - Case Study from Montenegro», 2022 8th International Conference on Computer Technology Applications, pp. 48-56, may 2022, doi: 10.1145/3543712.3543733.

[12] A. R. Efrat, R. Gernowo, y Farikhin, «Consumer purchase patterns based on market basket analysis using apriori algorithms», J. Phys.:

Conf. Ser., vol. 1524, n.o 1, p. 012109, abr. 2020, doi: 10.1088/1742-6596/1524/1/012109.

[13] Y. A. Ünvan, «Market basket analysis with association rules», Communications in Statistics - Theory and Methods, vol. 50, n.o 7, pp. 1615-1628, abr. 2021, doi: 10.1080/03610926.2020.1716255.

[14] W.-Y. Chiang, «To mine association rules of customer values via a data mining procedure with improved model: An empirical case study», Expert Systems with Applications, vol. 38, n.o 3, pp. 1716-1722, mar. 2011, doi: 10.1016/j.eswa.2010.07.097.

[15] M. A. Valle, G. A. Ruz, y R. Morrás, «Market basket analysis: Complementing association rules with minimum spanning trees», Expert Systems with Applications, vol. 97, pp. 146-162, may 2018, doi: 10.1016/j.eswa.2017.12.028.

[16] S. Liao, Y. Chen, y H. Hsieh, «Mining customer knowledge for direct selling and marketing», Expert Systems with Applications, vol. 38, n.o 5, pp. 6059-6069, may 2011, doi: 10.1016/j.eswa.2010.11.007.

[17] W. Lai et al., «An Apriori Algorithm-Based Association Analysis of Analgesic Drugs in Chinese Medicine Prescriptions RecordedfFrom Patients with Rheumatoid Arthritis Pain», Front. Pain Res., vol. 3, p. 937259, jul. 2022, doi: 10.3389/fpain.2022.937259.

[18] M. Karabatak y M. C. Ince, «An expert system for detection of breast cancer based on association rules and neural network», Expert Systems with Applications, vol. 36, n.o 2, pp. 3465-3469, mar. 2009, doi: 10.1016/j.eswa.2008.02.064.

[19] Y.-J. Chen, K.-M. Jhang, W.-F. Wang, G.-C. Lin, S.-W. Yen, y H.-H. Wu, «Applying Apriori algorithm to explore long-term care services usage status—Variables based on the combination of patients with dementia and their caregivers», Front. Psychol., vol. 13, p. 1022860, dic. 2022, doi: 10.3389/fpsyg.2022.1022860.

[20] M. Chen y Z. Yin, «Classification of Cardiotocography Based on the Apriori Algorithm and Multi-Model Ensemble Classifier», Front. Cell Dev. Biol., vol. 10, p. 888859, may 2022, doi: 10.3389/fcell.2022.888859.

[21] Y.-C. Wang, C.-C. Wu, A. P.-H. Huang, P.-C. Hsieh, y W.-M. Kung, «Combination of Acupoints for Alzheimer's Disease: An Association Rule Analysis», Front. Neurosci., vol. 16, p. 872392, jun. 2022, doi: 10.3389/fnins.2022.872392.

[22] Y.-F. Wang et al., «Combinations of scalp acupuncture location for the treatment of post-stroke hemiparesis: A systematic review and Apriori algorithm-based association rule analysis», Front. Neurosci., vol. 16, p. 956854, ago. 2022, doi: 10.3389/fnins.2022.956854.

[23] L. Ding et al., «Delayed Comparison and Apriori Algorithm (DCAA): A Tool for Discovering Protein–Protein Interactions From Time-Series Phosphoproteomic Data», Front. Mol. Biosci., vol. 7, p. 606570, dic. 2020, doi: 10.3389/fmolb.2020.606570.

[24] S. Pradeepa, K. Geetha, K. Kannan, y K. R. Manjula, «DEODORANT: a novel approach for early detection and prevention of polycystic ovary syndrome using association rule in hypergraph with the dominating set property», J Ambient Intell Human Comput, vol. 14, n.o 5, pp. 5421-5437, may 2023, doi: 10.1007/s12652-020-01990-4.

[25] R. Chaves, J. M. Górriz, J. Ramírez, I. A. Illán, D. Salas-Gonzalez, y M. Gómez-Río, «Efficient mining of association rules for the early diagnosis of Alzheimer's disease», Phys. Med. Biol., vol. 56, n.o 18, pp. 6047-6063, sep. 2011, doi: 10.1088/0031-9155/56/18/017.

[26] F. Liu y X. Zhang, «Hypertension and Obesity: Risk Factors for Thyroid Disease», Front. Endocrinol., vol. 13, p. 939367, jul. 2022, doi: 10.3389/fendo.2022.939367.

[27] Z. Yu, Y. Chen, Q. Xia, Q. Qu, y T. Dai, «Identification of status quo and association rules for chronic comorbidity among Chinese middle-aged and older adults rural residents», Front. Public Health, vol. 11, p. 1186248, jun. 2023, doi: 10.3389/fpubh.2023.1186248.

[28] O. Haas et al., «Predicting Anxiety in Routine Palliative Care Using Bayesian-Inspired Association Rule Mining», Front. Digit. Health, vol. 3, p. 724049, ago. 2021, doi: 10.3389/fdgth.2021.724049.

[29] K.-M. Jhang, M.-C. Chang, T.-Y. Lo, C.-W. Lin, W.-F. Wang, y H.-H. Wu, «Using The Apriori Algorithm To Classify The Care Needs Of Patients With Different Types Of Dementia», PPA, vol. 13, pp. 1899-1912, nov. 2019, doi: 10.2147/PPA.S223816.

[30] A. Soni, A. Saxena, y P. Bajaj, «A Methodological Approach for Mining the User Requirements Using Apriori Algorithm», Journal of Cases on Information Technology, vol. 22, n.o 4, pp. 1-30, oct. 2020, doi: 10.4018/JCIT.2020100101.

[31] H. Kuang et al., «An Association Rules-Based Method for Outliers Cleaning of Measurement Data in the Distribution Network», Front. Energy Res., vol. 9, p. 730058, oct. 2021, doi: 10.3389/fenrg.2021.730058.

[32] X. Liu, Y. Zhao, y M. Sun, «An Improved Apriori Algorithm Based on an Evolution-Communication Tissue-Like P System with Promoters and Inhibitors», Discrete Dynamics in Nature and Society, vol. 2017, pp. 1-11, 2017, doi: 10.1155/2017/6978146.

[33] Y. Guo, M. Wang, y X. Li, «Application of an improved Apriori algorithm in a mobile e-commerce recommendation system», IMDS, vol. 117, n.o 2, pp. 287-303, mar. 2017, doi: 10.1108/IMDS-03-2016-0094.

[34] Z. Li, X. Li, R. Tang, y L. Zhang, «Apriori Algorithm for the Data Mining of Global Cyberspace Security Issues for Human Participatory Based on Association Rules», Front. Psychol., vol. 11, p. 582480, feb. 2021, doi: 10.3389/fpsyg.2020.582480.

[35] Q. He et al., «Association Rule Mining through Combining Hybrid Water Wave Optimization Algorithm with Levy Flight», Mathematics, vol. 11, n.o 5, p. 1195, feb. 2023, doi: 10.3390/math11051195.

[36] A. Mirabadi y S. Sharifian, «Application of association rules in Iranian Railways (RAI) accident data analysis», Safety Science, vol. 48, n.o 10, pp. 1427-1435, dic. 2010, doi: 10.1016/j.ssci.2010.06.006.

[37] J. Hong, R. Tamakloe, y D. Park, «Application of association rules mining algorithm for hazardous materials transportation crashes on expressway», Accident Analysis & Prevention, vol. 142, p. 105497, jul. 2020, doi: 10.1016/j.aap.2020.105497.

[38] X. Zhang y J. Zhang, «Analysis and research on library user behavior based on apriori algorithm», Measurement: Sensors, vol. 27, p. 100802, jun. 2023, doi: 10.1016/j.measen.2023.100802.

[39] C. S. Zhang et al., «Natural products for migraine: Data-mining analyses of Chinese Medicine classical literature», Front. Pharmacol., vol. 13, p. 995559, oct. 2022, doi: 10.3389/fphar.2022.995559.

[40] P. Wang, L. Shi, J. Bai, y Y. Zhao, «Mining Association Rules Based on Apriori Algorithm and Application», 2009 International Forum on Computer Science-Technology and Applications, pp. 141-143, 2009, doi: 10.1109/IFCSTA.2009.41.

[41] F. Peng, Y. Sun, Z. Chen, y J. Gao, «Association Rule Mining Of Maritime Employability Demands Based On Apriori Algorithm», ICIC Express Letters, Part B: Applications, vol. 14, n.o 6, pp. 633-639, 2023, doi: 10.24507/icicelb.14.06.633.

[42] J.-B. Lin, T.-H. Liang, y Y.-G. Lee, «Mining important association rules on different customer potential value segments for life insurance database», 2012 IEEE International Conference on Granular Computing, pp. 283-288, ago. 2012, doi: 10.1109/GrC.2012.6468569.

[43] X. Mao, M. Tang, H. Deng, J. Chen, Z. Liu, y J. Wang, «Using association rules analysis to determine favorable mineralization sites in the Jiaojia gold belt, Jiaodong Peninsula, East China», Front. Earth Sci., vol. 11, p. 1174017, may 2023, doi: 10.3389/feart.2023.1174017.

[44] D. J. Prajapati, S. Garg, y N. C. Chauhan, «Interesting association rule mining with consistent and inconsistent rule detection from big sales data in distributed environment», Future Computing and Informatics Journal, vol. 2, n.o 1, pp. 19-30, jun. 2017, doi: 10.1016/j.fcij.2017.04.003.

[45] A. Valdivia et al., «What do people think about this monument? Understanding negative reviews via deep learning, clustering and descriptive rules», J Ambient Intell Human Comput, vol. 11, n.o 1, pp. 39-52, ene. 2020, doi: 10.1007/s12652-018-1150-3.

[46] R. Agrawal, T. Imielinski, y A. Swami, «Mining Association Rules between Sets of Items in Large Databases», In Proceedings of the 1993 ACM SIGMOD international conference on Management of data, pp. 207-216, jun. 1993, doi: https://doi.org/10.1145/170035.170072.

[47] T. Brijs, K. Vanhoof, y G. Wets, «Defining Interestigness for Association Rules», International Journal «Information Theories & Applications», vol. 10, n.o 4, pp. 370-375, 2003.

[48] S. Lee, Y. Cha, S. Han, y C. Hyun, «Application of Association Rule Mining and Social Network Analysis for Understanding Causality of Construction Defects», Sustainability, vol. 11, n.o 3, p. 618, ene. 2019, doi: 10.3390/su11030618.

[49] T. M. J. Fruchterman y E. M. Reingold, «Graph drawing by force-directed placement», Softw: Pract. Exper., vol. 21, n.o 11, pp. 1129-1164, nov. 1991, doi: 10.1002/spe.4380211102.

[50] T. Kamada y S. Kawai, «An algorithm for drawing general undirected graphs», Information Processing Letters, vol. 31, n.o 1, pp. 7-15, 1989, doi: 10.1016/0020-0190(89)90102-6.

[51] E. M. Reingold y J. S. Tilford, «Tidier Drawings of Trees», IIEEE Trans. Software Eng., vol. SE-7, n.o 2, pp. 223-228, mar. 1981, doi: 10.1109/TSE.1981.234519.