Understanding IT Product Purchasing Behavior of MSMEs Using Sequential Pattern Mining Approaches

Rendro Kasworo*, R.Rizal Isnanto, Budi Warsito
Doctoral Program of Information Systems-Postgraduate School,
Universitas Diponegoro, Semarang, Indonesia, 50241

***Abstract*—Sequential pattern mining is a crucial analytical method for understanding purchasing behavior and uncovering hidden patterns in transactional data. Unlike most prior studies that apply sequential pattern mining primarily in consumer-oriented retail settings or evaluate algorithms in isolation, this study investigates IT product purchasing behavior among Small and Medium Enterprises (SMEs) within a B2B digital transformation context through a direct comparative evaluation of three widely used algorithms: Apriori, PrefixSpan, and CloSpan. A series of controlled experiments was conducted on the same transactional datasets to assess algorithm performance in terms of accuracy, computational efficiency, and redundancy reduction. The results show that Apriori discovers exhaustive patterns at the cost of higher computational complexity, PrefixSpan achieves faster sequence extraction with balanced accuracy, and CloSpan effectively reduces redundancy by generating closed sequential patterns. Beyond pattern discovery, this study translates support, confidence, and lift metrics into actionable decision-support insights, highlighting how different algorithmic characteristics can be aligned with retention strategies, service bundling, and targeted interventions. These findings provide distinct methodological and practical contributions by positioning sequential pattern mining as a data-driven decision-support tool to accelerate digital transformation initiatives among SMEs in the IT product ecosystem.**

***Keywords*—Sequential pattern mining; Apriori; PrefixSpan; CloSpan; SMEs; digital transformation**

I. Introduction

Micro, Small, and Medium Enterprises (MSMEs) are important for economic growth, especially in developing countries. They not only contribute significantly to Gross Domestic Product (GDP) but also create more jobs and maintain economic stability [1], [2]. However, MSMEs face numerous challenges in adopting and utilizing Information Technology (IT) products. Various obstacles, including low digital literacy, limited capital, and a lack of solutions tailored to business needs, often hinder the digitalization process.

In recent years, sequential pattern mining (SPM) has proven to be an effective data mining technique for discovering relational patterns and transaction sequences in large-scale datasets [3],[4]. In contrast to association rule mining, which only identifies relationships between items without considering order, SPM is capable of uncovering temporal patterns in purchasing behavior [5]. This capability is highly relevant for MSMEs as it can reveal product adoption stages— for example, customers who initially purchase internet connectivity services and subsequently switch to digital productivity applications or content services. Such information can serve as a valuable reference for IT product providers in developing more personalized and adaptive marketing strategies.

Several algorithms have been developed to support the SPM process. Apriori is a fundamental algorithm widely used to identify frequent itemsets; however, its complex candidate generation process can lead to efficiency issues [6]. PrefixSpan utilizes a more efficient pattern-growth approach through prefix-based partition exploration without repeatedly generating candidates [7]. Meanwhile, CloSpan focuses on mining closed sequential patterns, producing more concise, non-redundant, and semantically meaningful results [8]. According to Arena et al., each algorithm presents its own advantages and limitations in terms of accuracy, computational complexity, and scalability. Thus, it is essential to carry out comparative evaluations in certain contexts [9].

Although SPM has been widely applied in retail and e-commerce domains, its application remains limited in the context of IT product adoption by MSMEs. Most previous research has focused on consumer (B2C) markets and has not adequately explored the business-to-business (B2B) context involving MSMEs. Furthermore, relatively few studies have compared the performance of several SPM algorithms simultaneously in analyzing the IT product purchasing behavior of MSMEs.

For instance, research by Aloysius and Binu demonstrated the application of PrefixSpan in the retail sector to organize products based on consumers' sequential shopping patterns, illustrating a typical application of SPM in the B2C domain [10]. Trivonanda et al. applied SPM in e-commerce recommendation systems in Indonesia and highlighted the dominant use of SPM in the consumer context [11]. Additionally, a literature review by Ezeife specifically mapped SPM-based e-commerce recommendation methods and emphasized algorithm performance in the B2C realm [12]. Similarly, Anwar and Uma examined the use of PrefixSpan in cross-domain recommendation systems, such as book recommendations, but remains within the consumer consumption paradigm [13].

In the telecommunications sector, studies examining service bundling recommendations do exists; however, they generally rely on other methods, such as constructive preference elicitation [14], instead of SPM. Other research in the telco field has applied SPM to customer churn prediction [15], with a focus on behavioral sequences leading to churn rather than sequences of IT product adoption. Moreover, many recent churn studies have shifted to explainable AI (XAI) approaches, further

*Corresponding author.

indicating that SPM studies in the context of product adoption remain relatively scarce [16].

Meanwhile, the broader MSME literature primarily discusses the adoption of digital technology. For example, Ballerini et al. examined factors influencing e-commerce adoption in manufacturing MSMEs [17], Kgakatsi et al. analyzed the impact of Big Data usage on the performance of MSMEs in South Africa [18], and Cooper et al. highlighted the role of artificial intelligence in new product development in MSMEs [19]. However, no prior research has specifically examined the sequence patterns of IT product purchases by MSMEs in a B2B context using a comparative SPM algorithm approach.

This study aims to identify sequential patterns of IT product purchases among MSMEs by applying and comparing the Apriori, PrefixSpan, and CloSpan algorithms, while evaluating their respective strengths and limitations in terms of accuracy, computational efficiency, and pattern redundancy reduction. By doing so, this research addresses an existing gap in the literature by providing empirical evidence on the comparative performance of SPM algorithms and offering actionable insights to support data-driven decision-making and digital transformation initiatives within the MSME IT ecosystem.

The subsequent literature review section discusses related studies on MSMEs, data-driven marketing, and SPM techniques. The methodology section outlines the research design, data collection, and preprocessing procedures, and the application of the Apriori, PrefixSpan, and CloSpan algorithms. The results and discussion section presents and analyzes the experimental findings and identifies sequential purchasing patterns. Finally, the conclusion section summarizes the key findings, highlights the contributions of the study, and outlines directions for future research.

II. Literature Review

A. *MSMEs and the Challenges of Adopting Information Technology*

Micro, Small, and Medium Enterprises (MSMEs) constitute a vital sector of the global economy, especially in developing countries. In Indonesia, MSMEs contribute approximately 61.7% of the national Gross Domestic Product (GDP) and provide employment for 97% of the workforce, making them the backbone of the national economy [20]. As a crucial economic pillar, MSMEs play an essential role not only as drivers of economic growth but also in maintaining social stability through job creation.

B. *Data-Driven Marketing and Recommendation Systems*

The development of digital technology has brought major changes, including the adoption of data-driven marketing approaches, in which strategic decisions are based on the analysis of customer behavior derived from historical transaction data and digital interactions. This approach enables companies to better understand consumer preferences and more accurately forecast future needs. In the context of MSMEs, data-driven marketing enhances the effectiveness of marketing strategies through personalized offerings, cross-selling, and product bundling tailored to customers' actual needs.

C. *Sequential Pattern Mining*

Sequential Pattern Mining (SPM) is a subfield of data mining that explores the discovery of sequential patterns within transaction data or temporal interactions [4], [21], [22]. Unlike static association rule mining, SPM enables researchers and practitioners to identify not only correlations among items that frequently appear together but also the order in which these items occur.

SPM has been widely applied across various fields, including retail, e-commerce, healthcare, and telecommunications. For instance, in retail environments, SPM helps analyze consumer shopping patterns, such as their tendency to purchase hardware before software or support services [23],[24]. In healthcare, SPM assists in monitoring medication usage and patient diagnoses, which are valuable for clinical decision support systems.

There are three main aspects commonly addressed in SPM research, namely:

1) Pattern accuracy: The extent to which an algorithm identifies patterns that accurately represent real-world behavior.

2) Computational efficiency: This aspect relates to execution time and resource requirements, especially when processing large-scale datasets.

3) Redundancy reduction: This refers to the ability to generate concise and meaningful patterns without duplication or irrelevant variations.

D. *Main Algorithms in Sequential Pattern Mining*

Three prominent algorithms discussed in the literature are Apriori, PrefixSpan, and CloSpan.

1) Apriori: Apriori is a classical algorithm originally introduced to discover frequent itemsets and later adapted for use in SPM. This algorithm employs a candidate generation-and-test approach that requires enumerating all possible candidate patterns [25]. While this approach is comprehensive, it suffers from high computational complexity, especially when applied to large datasets.

2) PrefixSpan (Prefix-Projected Sequential Pattern Mining): PrefixSpan introduces a pattern-growth paradigm by dividing the dataset into prefix-based projected databases. This method avoids explicit candidate exploration and improves processing efficiency. Numerous studies have shown that PrefixSpan outperforms Apriori in large-scale datasets.

3) CloSpan (Closed Sequential Pattern Mining): CloSpan is designed to address the problem of redundant results. The algorithm only generates closed patterns, which are patterns that do not have any superpatterns with the same support. This method produces more concise and meaningful results, facilitating interpretation by researchers and practitioners.

Despite widespread use, existing SPM approaches continue to face several challenges and limitations. Apriori-based methods suffer from high computational complexity due to exhaustive candidate generation, rendering them less efficient for large-scale transactional data. PrefixSpan improves efficiency by eliminating candidate generation; however, it may still produce a substantial number of redundant patterns, which can complicate interpretation. CloSpan addresses redundancy by mining closed sequential patterns but may overlook infrequent yet potentially meaningful sequences.

Moreover, most prior studies applying SPM focus predominantly on consumer-oriented (B2C) contexts, such as retail and e-commerce, while empirical investigations in MSME-oriented and B2B IT product adoption remain limited. Existing research also tends to evaluate algorithms in isolation rather than through direct comparative analysis using an identical dataset. These limitations highlight the need for a comprehensive comparative evaluation of SPM algorithms in the context of MSME IT product purchasing behavior, particularly to support data-driven decision-making and digital transformation initiatives**.**

III. METHODOLOGY

A. Research Design

This study employed a comparative experimental approach to assess the performance of three SPM algorithms: Apriori, PrefixSpan, and CloSpan. The research design focused on analyzing IT product purchasing patterns among MSMEs based on transaction data. Each algorithm was tested separately on the same dataset to ensure an unbiased comparison of results.

B. Data Collection

The data used in this study were collected from transaction records of IT product purchases made by MSMEs operating in the trade and services sectors. The data collection process involved two main stages.

First, primary data were collected from actual transactions recorded by MSMEs and IT service provider partners. These transaction records included purchase dates, product names, product categories (e.g., High Speed Internet, Metro Ethernet, ASTINet, Content, Music, Games, and Managed Capacity Network), and the sequence of purchases made by customers within a specific time period. These primary data became the foundation for building a sequence database for analysis using the SPM algorithm.

Second, secondary data were gathered from sales reports from authorized IT product distributors and B2B e-commerce databases. These data served as a benchmark dataset to support the analysis and ensure that the patterns identified were not limited to specific MSMEs but also reflected broader trends in the IT product distribution sector. Table I presents sample data from IT product sales reports.

Table I illustrates transaction records for IT products purchased by MSMEs and authorized distributors. Each transaction includes information on the customer, transaction ID, purchase time, product name, product category, quantity, price, and data sources. Data obtained from MSMEs represent actual transactions and serve as the primary data source, while data from distributors and e-commerce function as secondary sources for comparison. Overall, the table describes IT product consumption patterns in the trade and services sector, which are then analyzed to identify purchasing sequence patterns using the SPM approach.

C. Data Preprocessing

The data preprocessing stage aimed to ensure consistency and suitability of the collected data before applying the SPM algorithm. This stage involved removing duplicate records, correcting data entry errors, handling missing values, and normalizing product name formats to ensure uniformity. Products were categorized into seven main groups: High Speed Internet, Metro Ethernet, ASTINet, Content, Music, Games, and Managed Capacity Network.

Next, the data were sorted based on each customer's transaction time to obtain a valid purchase sequence. In the final stage, the data were converted into a sequence database consisting of the purchase sequence of each customer, which was subsequently processed using the Apriori, PrefixSpan, or CloSpan algorithms to identify meaningful purchase patterns.

TABLE I. SAMPLE DATA OF IT PRODUCT SALES REPORT

Customer	Transaction ID	Date and Time of Transaction	Product Name	Category	Amount	Price	Description
UMKM01	TRX001	2025-01-05 10:30:00	High Speed Internet Plan A	High Speed Internet	1	500,000	Recorded by the MSME
UMKM01	TRX002	2025-01-15 15:45:00	Metro Ethernet 10 Mbps	Metro Ethernet	1	2,500,000	Recorded by the MSME
UMKM02	TRX003	2025-02-02 09:20:00	Basic ASTINet	ASTINet	1	1,500,000	Recorded by the MSME
UMKM02	TRX004	2025-02-12 13:00:00	Premium Content Streaming	Content	2	1,000,000	Recorded by the MSME
UMKM03	TRX005	2025-03-07 11:10:00	Online Game Voucher	Game	3	300,000	Recorded by the MSME
UMKM04	TRX007	2025-03-18 14:00:00	Managed Capacity Network Silver	Managed Capacity Network	1	2,000,000	Recorded by the MSME
DIST01	TRX101	2025-01-06 09:00:00	High Speed Internet Plan B	High Speed Internet	1	750,000	Listed in the official distributor report
DIST02	TRX102	2025-01-25 16:30:00	Metro Ethernet 20 Mbps	Metro Ethernet	1	4,500,000	Listed in the official distributor report
DIST02	TRX103	2025-02-05 10:15:00	ASTINet Pro	ASTINet	1	2,800,000	Listed in the official distributor report
DIST03	TRX104	2025-02-15 12:30:00	Basic Content Streaming	Content	1	500,000	Listed in the official distributor report
DIST04	TRX105	2025-03-10 14:45:00	Online Game Voucher	Game	5	500,000	Listed in the official distributor report
DIST04	TRX106	2025-03-20 15:30:00	Premium Music Streaming	Music	2	240,000	Listed in the official distributor report

TABLE II. DATA PREPROCESSING

Customer	Transaction Sequence	Date and Time of Transaction	Product Name	Category
UMKM01	1	2025-01-05 10:30:00	High Speed Internet Plan A	High Speed Internet
UMKM01	2	2025-01-15 15:45:00	Metro Ethernet 10 Mbps	Metro Ethernet
UMKM02	1	2025-02-02 09:20:00	ASTINet Basic	ASTINet
UMKM02	2	2025-02-12 13:00:00	Premium Content Streaming	Content
UMKM03	1	2025-03-07 11:10:00	Online Game Voucher	Game
UMKM04	2	2025-03-18 14:00:00	Managed Capacity Network Silver	Managed Capacity Network
DIST01	1	2025-01-06 09:00:00	High Speed Internet Plan B	High Speed Internet
DIST02	1	2025-01-25 16:30:00	Metro Ethernet 20 Mbps	Metro Ethernet
DIST02	1	2025-02-05 10:15:00	ASTINet Pro	ASTINet
DIST03	2	2025-02-15 12:30:00	Basic Content Streaming	Content
DIST04	1	2025-03-10 14:45:00	Online Game Voucher	Game
DIST04	1	2025-03-20 15:30:00	Premium Music Streaming	Music

The preprocessing results presented in Table II show the sequence of IT product purchases for each customer after data cleaning and normalization. Each customer (MSME or distributor) is represented by a transaction ID sorted by date to identify sequential purchasing patterns. For example, UMKM01 initially purchased High Speed Internet Plan A and subsequently purchased Metro Ethernet services. Similarly, UMKM03 purchased an online game voucher, followed by a music streaming product.

Product categories were standardized into seven main groups to ensure consistent analysis. This format prepares the data for processing using the SPM algorithm to identify customers' most frequent purchase sequential patterns.

The pre-processed data were converted into a sequence database in the next stage. This process was accomplished by grouping all transactions by customer and sorting them by purchase time.

Each transaction in the sequence was represented as an element in a sequence, showing that each customer had a unique purchase sequence. For example, customer UMKM01 had the sequence [High Speed Internet → Metro Ethernet], while customer UMKM03 had the sequence [Game → Music].

Customers who only made one transaction were still categorized as a single sequence, for example, UMKM04 → [Manage Capacity Network].

TABLE III. EXAMPLE OF PURCHASE SEQUENCE REPRESENTATION

Customers	Purchase Sequence
UMKM01	[High Speed Internet → Metro Ethernet]
UMKM02	[ASTINet → Content]
UMKM02	[ASTINet → Content]
UMKM03	[Game → Music]
UMKM04	[Managed Capacity Network]
DIST01	[High Speed Internet]
DIST02	[Metro Ethernet → ASTINet]
DIST03	[Content]
DIST04	[Game → Music]

This representation allows for systematic analysis of the entire purchasing patterns of various customers. The sequence database served as the primary input for SPM algorithms, including AprioriAll, PrefixSpan, and CloSpan, to identify recurring patterns that describe customers' tendencies to purchase IT products sequentially. Table III shows an example of a sequence database representation.

D. Pattern Analysis

At this stage, the data processed into a sequence database were analyzed using an SPM approach to identify sequential patterns in customer transactions. Three main algorithms were employed: AprioriAll, PrefixSpan, and CloSpan.

1) AprioriAll was used to find correlations between products that frequently appear together or sequentially, by measuring support and confidence as indicators of pattern strength.

2) PrefixSpan was applied to extract frequently occurring purchase sequences by developing prefixes from existing sequences without explicitly generating candidate patterns.

3) CloSpan was chosen to generate closed sequential patterns that cannot be expanded further without lowering the support value, resulting in a more concise but informative representation of the pattern.

E. Pattern Evaluation

After obtaining purchasing patterns using the SPM algorithm, a pattern evaluation stage was conducted to assess the validity and strength of the correlations among items in the sequence. This evaluation was conducted using several standard measures, including:

1) Support: This metric measures how frequently a particular pattern appears in the entire transaction database. A high support value indicates that the pattern is popular and generally relevant.

2) Confidence: This metric indicates the probability that a subsequent item or sequence occurs after the initial pattern has occurred. Higher confidence values indicate stronger and more reliable associations.

3) Lift (optional): This metric measures how frequently a pattern occurs compared to its probability of occurring by chance. It helps identify patterns that are substantially significant rather than coincidental.

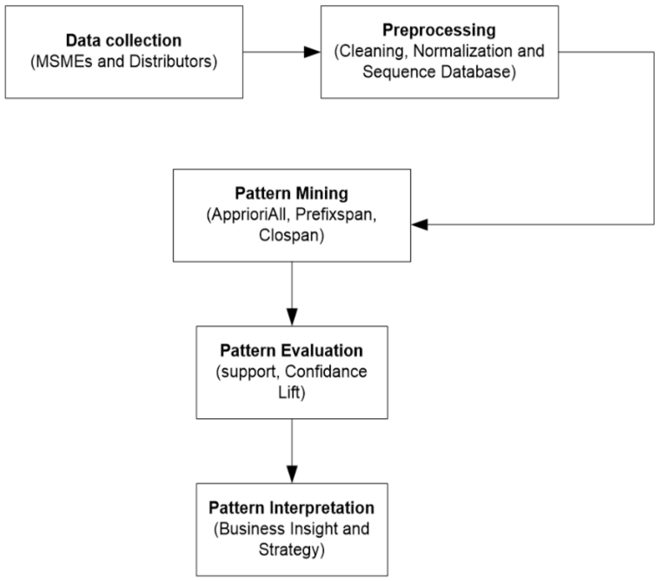

Fig. 1. Research methodology stages.

These metrics allowed researchers to filter meaningful patterns and eliminate those with low frequency or weak correlations. The evaluation process led to validated findings suitable for interpretation in the context of IT product purchasing behavior among MSMEs and distributors.

F. Pattern Interpretation

Pattern interpretation was performed after evaluating metrics, such as support, confidence, and lift, to ensure their validity and significance. At this stage, the identified sequential purchase patterns were related to the real-world context of MSME consumer behavior and IT product distribution practices. For example, the Game → Music pattern suggests that customers who purchase game-related services are likely to subsequently purchase digital music products. Similarly, the pattern Content → High Speed Internet indicates an increased demand for high-speed internet access following the purchase of digital content.

This interpretation is beneficial for analyzing purchasing behavior and serves as a foundation for formulating business strategies, such as product bundling, personalized service recommendations, and optimized IT service distribution. Thus, the identified patterns offer not only theoretical contributions but also practical value for decision-making and the development of more targeted marketing strategies. Fig. 1 illustrates the flowchart of the research stages.

IV. Results and Discussion

A. Results

This section presents the key findings from the SPM analysis of IT product transaction data that were preprocessed into a sequence database. The results highlight:

- Associations and purchasing sequence tendencies among categories/products.
- Consistent and recurring patterns observed for both MSME customers and distributors.
- Practical implications for service bundling strategies, recommendations, and service improvement.

The results are presented by algorithm (AprioriAll, PrefixSpan, and CloSpan) and customer segment (MSMEs vs. distributors), along with evaluation metrics (support, confidence, and lift), to assess the strength and significance of the identified patterns.

B. Test Simulation Scenario

The dataset used in this study consists of transactional records from 500 MSME customers (CUST0001–CUST0500). The data were generated to simulate realistic IT product purchasing behavior, with each customer having between one and six sequential transactions recorded over the period from January to March 2025. The dataset includes the following characteristics:

1) Number of transactions per customer: Randomly generated between one and six transactions.

2) Product categories: Seven main groups (High Speed Internet, Metro Ethernet, ASTINet, Content, Music, Game, and Manage Capacity Network), each comprising several product variants.

3) Transaction attributes: Customer ID, transaction ID, transaction date and time, product name, category, quantity, and price.

4) Time period: Transactions conducted between January and March 2025 with random time sequences representing the chronological order of purchases.

Table IV presents the top ten purchase sequence rules ranked by confidence for all customers during the January-March 2025 period. The inclusion of support and lift values enables comparison of pattern consistency and relevance for service recommendations and bundling strategies.

In Table IV, the pattern with the highest confidence shows a consistent transition tendency, even though it may not be the most frequent. The rules ASTINet → Content (confidence = 0.67; lift = 1.90) and High Speed Internet → Content (confidence = 0.66; lift = 1.87) indicate that the purchase of content services commonly follows the activation of connectivity services. The patterns Managed Capacity Network → Game (confidence = 0.62; lift = 1.78) and Content → Game (confidence = 0.58; lift = 1.67) suggest strong cross-selling potential within entertainment-related services. Although the pattern Content → Managed Capacity Network has merely 1.6% support, it exhibits a very high lift value (3.36), making it a worthwhile target for marketing campaigns. Regarding service upgrades, the rules Game → ASTINet (confidence = 0.57; lift = 2.86) and ASTINet → Metro Ethernet (confidence = 0.53; lift = 1.43) represent promising upgrade paths toward more reliable connectivity. Meanwhile, Metro Ethernet → Metro Ethernet stands out as a repeat or renewal pattern (confidence = 0.52; lift = 1.40), as does Music → Music (confidence = 0.50; lift = 2.23), indicating strong subscription retention.

Table V shows the top ten purchase sequence rules based on support, reflecting the most common pattern observed in the MSME segment from January to March 2025. The table includes the columns Antecedent, Consequent, Pattern, Length, Support Count, Support, Confidence, and Lift.

Based on Table V, the most prevalent patterns in the MSME segment are dominated by repeat or renewal behavior (X→X). The sequence Metro Ethernet → Metro Ethernet is the most frequent (support = 0.1960; confidence = 0.5241; lift = 1.4012), followed by Content → Content (support = 0.1740; confidence = 0.4971; lift =1.4204) and Game → Game (support = 0.1560; confidence = 0.4509; lift = 1.3031). Similarly, High Speed Internet → High Speed Internet (support = 0.1160; confidence = 0.3816; lift = 1.2552) and ASTINet → ASTINet (support = 0.0980; confidence = 0.4900; lift = 2.4500) also show strong retention. The presence of the trigram Metro Ethernet → Metro Ethernet → Metro Ethernet (two rules with support values of 0.0900 due to identical three-step patterns) indicates persistent subscription behavior. Other recurring patterns, such as Content → Content → Content (support = 0.0720) and Music → Music (support = 0.0720; confidence = 0.3214; lift = 1.4349), reinforce the dominance of retention-oriented patterns. Although the results focus on support, the high lift observed for certain services—such as Managed Capacity Network → Managed Capacity Network (support = 0.0760; confidence = 0.4471; lift = 2.6298)—shows strong repeat associations within more specific segments. These findings emphasize the importance of retention, renewal strategies, and loyalty initiatives for services with the highest repeat customer reach, while also identifying high-value services with high lift for more targeted offerings.

TABLE IV. TOP-10 SEQUENTIAL PURCHASE RULES BY CONFIDENCE: MSME SEGMENT

Antecedent	Consequent	Pattern	Length	Support Count	Support	Confidence	Lift
ASTINet → Content	Content	ASTINet → Content → Content	3	8	0.0160	0.6667	1.9048
High Speed Internet → Content	Content	High Speed Internet → Content → Content	3	19	0.0380	0.6552	1.8719
Managed Capacity Network → Game	Game	Managed Capacity Network → Game → Game	3	8	0.0160	0.6154	1.7786
Content → Game	Game	Content → Game → Game	3	15	0.0300	0.5769	1.6674
Content → Managed Capacity Network	Managed Capacity Network	Content → Managed Capacity Network → Managed Capacity Network	3	8	0.0160	0.5714	3.3613
Game → ASTINet	ASTINet	Game → ASTINet → ASTINet	3	8	0.0160	0.5714	2.8571
ASTINet → Metro Ethernet	Metro Ethernet	ASTINet → Metro Ethernet → Metro Ethernet	3	8	0.0160	0.5333	1.4260
Metro Ethernet	Metro Ethernet	Metro Ethernet → Metro Ethernet	2	98	0.1960	0.5241	1.4012
Music → Music	Music	Music → Music → Musik	3	18	0.0360	0.5000	2.2321
Managed Capacity Network → Content	Content	Managed Capacity Network → Content → Content	3	7	0.0140	0.5000	1.4286

TABLE V. TOP 10 RULES BY SUPPORT

Antecedent	Consequent	Pattern	Length	Support Count	Support	Confidence	Lift
Metro Ethernet	Metro Ethernet	Metro Ethernet → Metro Ethernet	2	98	0.1960	0.5241	1.4012
Content	Content	Content → Content	2	87	0.1740	0.4971	1.4204
Game	Game	Game → Game	2	78	0.1560	0.4509	1.3031
High Speed Internet	High Speed Internet	High Speed Internet → High Speed Internet	2	58	0.1160	0.3816	1.2552
ASTINet	ASTINet	ASTINet → ASTINet	2	49	0.0980	0.4900	2.4500
Metro Ethernet → Metro Ethernet	Metro Ethernet	Metro Ethernet → Metro Ethernet → Metro Ethernet	3	45	0.0900	0.4592	1.2278
Metro Ethernet	Metro Ethernet → Metro Ethernet	Metro Ethernet → Metro Ethernet → Metro Ethernet	3	45	0.0900	0.2406	1.2278
Managed Capacity Network	Managed Capacity Network	Managed Capacity Network → Managed Capacity Network	2	38	0.0760	0.4471	2.6298
Content → Content	Content	Content → Content → Content	3	36	0.0720	0.4138	1.1823
Music	Music	Music → Music	2	36	0.0720	0.3214	1.4349

Table VI shows the ten sequential purchase rules based on the highest lift values. Lift indicates how strongly the antecedent and consequent are linked beyond random chance; a lift value greater than one indicates a positive association, meaning the higher the value, the stronger the relative correlation. Unlike confidence, which reflects the certainty of a transition, or support, which reflects prevalence, lift emphasizes patterns with strategic value even if they are not the most common, such as certain repeat patterns or strong upgrade paths in customer subsegments. Therefore, Table VI complements Tables IV (confidence-based) and Table V (support-based).

TABLE VI. TOP 10 RULES BY LIFT

Antecedent	Consequent	Pattern	Length	Support Count	Support	Confidence	Lift
Managed Capacity Network	Managed Capacity Network → ASTINet	Managed Capacity Network → Managed Capacity Network → ASTINet	3	3	0.0060	0.0353	3.5294
Content → Managed Capacity Network	Managed Capacity Network	Content → Managed Capacity Network → Managed Capacity Network	3	8	0.0160	0.5714	3.3613
ASTINet	ASTINet → Metro Ethernet	ASTINet → ASTINet → Metro Ethernet	3	10	0.0200	0.1000	3.3333
High Speed Internet → Managed Capacity Network	Managed Capacity Network	High Speed Internet → Managed Capacity Network → Managed Capacity Network	3	6	0.0120	0.5000	2.9412
Managed Capacity Network	Managed Capacity Network → Musik	Managed Capacity Network → Managed Capacity Network → Musik	3	5	0.0100	0.0588	2.9412
Game → ASTINet	ASTINet	Game → ASTINet → ASTINet	3	8	0.0160	0.5714	2.8571
Managed Capacity Network	Managed Capacity Network	Managed Capacity Network → Managed Capacity Network	2	38	0.0760	0.4471	2.6298
Musik → Managed Capacity Network	Managed Capacity Network	Musik → Managed Capacity Network → Managed Capacity Network	3	4	0.0080	0.4444	2.6144
Metro Ethernet → Managed Capacity Network	Managed Capacity Network	Metro Ethernet → Managed Capacity Network → Managed Capacity Network	3	6	0.0120	0.4286	2.5210
Managed Capacity Network	Managed Capacity Network → Content	Managed Capacity Network → Managed Capacity Network → Content	3	6	0.0120	0.0706	2.5210

Table VI also reveals the sequential purchase rules based on lift values, which measure the strength of the correlation between the antecedent and consequent beyond random chance. A lift value greater than one indicates a positive association, with higher values reflecting stronger relative correlations. In contrast to support, which emphasizes how common a pattern is, or confidence, which emphasizes the certainty of a transition, lift highlights a strong pattern even if it does not occur frequently. These findings suggest that high-lift but low-support patterns are suitable for targeted interventions. In contrast, high-lift patterns with higher support are relevant for larger-scale retention or bundling programs [26], [27].

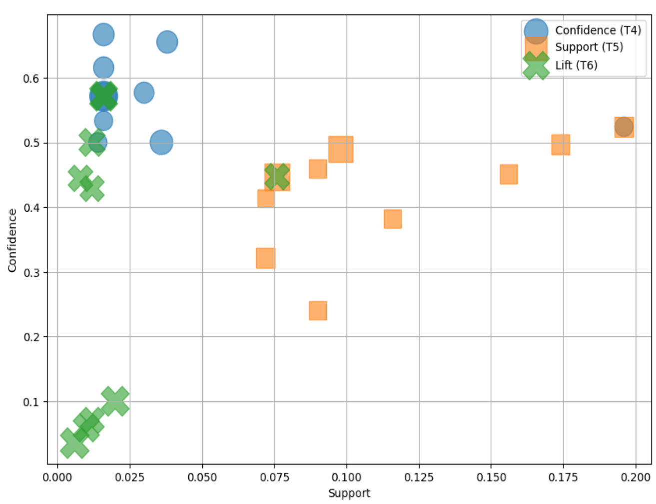

Fig. 2. Sequential comparison chart (confidence, support, and lift).

Furthermore, to provide a more comprehensive perspective, the analysis results in Tables IV (based on confidence), Table V (based on support), and Table VI (based on lift) are presented in the form of a comparison chart in Fig. 2. This visualization demonstrates how each metric highlights different patterns: rules with high confidence indicate consistency in transition behavior, rules with high support reflect the prevalence of purchasing behavior, and rules with high lift reveal strategic opportunities even when they occur infrequently. Therefore, combining these three perspectives offers a stronger foundation for developing marketing strategies, whether aimed at maintaining consistent customer behavior, increasing subscription retention, or implementing targeted strategic intervention value.

C. Discussion

The research results indicate that IT product purchasing behavior among MSMEs is strongly characterized by repetition and renewal patterns within the same service category. Frequent sequences such as Metro Ethernet → Metro Ethernet, Content → Content, Games → Games, and ASTINet → ASTINet indicate that once MSMEs adopt a particular IT service, they tend to continue using or renewing that service over time. This behavior reflects a preference for service stability and reliability, which is essential for business operations that rely on digital connectivity and continuous service availability.

Sequential patterns with high confidence levels further demonstrate a gradual process of digital adoption among MSMEs. Connectivity-related services, such as High-Speed Internet and ASTINet, often precede the adoption of content- or entertainment-based services. This sequence implies that MSMEs prioritize establishing a stable digital infrastructure before expanding into additional digital services. Such a progression aligns with the incremental nature of digital transformation, in which basic infrastructure is developed before higher-level digital capabilities are adopted.

Patterns with high support scores highlight the widespread prevalence of retention-oriented behavior among MSMEs. The dominance of repeat subscription patterns suggests that customer retention and service renewals are more influential than frequent product switching. This finding emphasizes the importance of long-term customer relationship management, where maintaining service quality and reliability is a key factor in maintaining MSME engagement with IT service providers.

Conversely, patterns with high upgrade values reveal strategic opportunities within more specific market segments. Certain services exhibit strong associations despite lower overall frequency, indicating a concentrated but highly engaged user group. These high upgrade patterns are highly relevant for targeted marketing strategies, customized service integrations, and differentiated loyalty programs. While high support patterns are suitable for broad retention strategies, high upgrade patterns offer insights for focused interventions aimed at specific MSME segments.

The combined use of confidence, support, and lift metrics provides a comprehensive perspective on MSME purchasing behavior. Confidence captures consistent switching tendencies, support reflects dominant usage patterns, and lift highlights strong associations beyond random occurrence. Integrating these complementary metrics enables IT service providers to design balanced strategies that combine mass retention initiatives with targeted offerings, thereby enhancing data-driven decision-making.

These results collectively demonstrate that SPM is an effective approach for uncovering dominant and nuanced behavioral patterns in MSME IT product adoption. By translating these patterns into actionable insights, IT service providers can systematically and data-drivenly improve retention strategies, optimize service offerings, and support MSMEs' ongoing digital transformation.

From both methodological and practical perspectives, this study contributes to the SPM literature by providing empirical evidence on the comparative behavior of Apriori, PrefixSpan, and CloSpan when applied to the same B2B transactional dataset. Unlike prior studies that primarily emphasize pattern discovery in consumer-oriented contexts, this work clarifies the distinct analytical roles of support, confidence, and lift metrics in decision-oriented interpretation. Furthermore, the findings can be operationalized into actionable guidelines for IT product providers, where high-support patterns inform large-scale retention and renewal programs, high-confidence patterns support service upgrade recommendations, and high-lift, but lower-support patterns enable targeted interventions for specific MSME segments. This integrated interpretation strengthens data-driven decision-making and supports the digital transformation of MSMEs.

V. Conclusion

This study demonstrates that sequential pattern mining is an effective approach for identifying IT product purchasing behavior among MSMEs. The comparative analysis of the Apriori, PrefixSpan, and CloSpan algorithms reveals distinct algorithmic strengths: Apriori effectively identifies high-support patterns, PrefixSpan provides efficient sequence extraction, and CloSpan generates concise closed patterns by reducing redundancy. The empirical results further indicate that confidence values capture consistent transition behavior, support values reflect the prevalence of repeat and renewal purchases, and lift values highlight patterns with strong strategic significance.

Beyond methodological contributions, the findings reveal key characteristics of MSME digital adoption behavior. The dominance of repeat and renewal patterns underscores the importance of service continuity, while high-confidence and high-lift sequences suggest staged adoption pathways and targeted opportunities for service bundling, retention strategies, and differentiated offerings. These insights provide practical value for IT service providers in designing data-driven marketing strategies aligned with MSME digital transformation initiatives.

From a theoretical perspective, this study contributes to sequential pattern mining research by clarifying the distinct analytical roles of support, confidence, and lift through a direct comparative evaluation of Apriori, PrefixSpan, and CloSpan within a B2B MSME context. From a practical perspective, the findings offer actionable guidance for IT product providers, whereby high-support patterns inform large-scale retention strategies, high-confidence patterns support service upgrade pathways, and high-lift patterns enable targeted interventions for specific MSME segments.

Despite these contributions, this study is limited to transactional data collected within a specific observation period and service domain. Future research may extend this work by incorporating longer time horizons, cross-regional datasets, and advanced machine learning or hybrid modeling approaches to enhance predictive capabilities. The integration of non-transactional factors, such as organizational characteristics, digital readiness, and socio-demographic attributes, may further enrich behavioral interpretation. In addition, the development of analytics dashboards or interactive recommendation systems based on the proposed framework represents a promising direction for practical implementation and sustained digital transformation among MSMEs.

References

[1] Hasbiah, Teguh Suharyanto, Virda Febriani Putri Irwan, Fika Faridotul, and Chasanah, “Optimizing the Role of MSMEs in Regional Economic Growth: Challenges and Opportunities,” J. Econ. Educ. Entrep. Stud., vol. 5, no. 3, pp. 530–542, 2024, doi: 10.62794/je3s.v5i3.4310.

[2] T. T. H. Tambunan, “Micro and Small Enterprises’ Export Competencies and Cooperation in Indonesia,” J. Dev. Econ., vol. 9, no. 1, pp. 84–106, 2024, doi: 10.20473/jde.v9i1.53667.

[3] S. R. Wakchaure and R. G. Vishwakarma, “SPM: Study of Different Techniques of Sequential Pattern Mining,” Int. J. Intell. Syst. Appl. Eng., vol. 11, no. 5s SE-Research Article, pp. 442–457, Apr. 2023, [Online]. Available: https://www.ijisae.org/index.php/IJISAE/article/view/2806

[4] N. Mazumdar and P. K. D. Sarma, “Sequential pattern mining algorithms and their applications: a technical review,” Int. J. Data Sci. Anal., 2024, doi: 10.1007/s41060-024-00659-x.

[5] H. M. Huynh, L. T. T. Nguyen, N. N. Pham, Z. K. Oplatková, U. Yun, and B. Vo, “An efficient method for mining sequential patterns with indices,” Knowledge-Based Syst., vol. 239, p. 107946, 2022, doi: https://doi.org/10.1016/j.knosys.2021.107946.

[6] I. Riadi, H. Herman, F. Fitriah, S. Suprihatin, A. Muis, and M. Yunus, “Implementation of association rule using apriori algorithm and frequent pattern growth for inventory control,” J. Infotel, vol. 15, no. 4, pp. 369–378, 2023, doi: 10.20895/infotel.v15i4.980.

[7] D. S. Maylawati, H. Aulawi, and M. A. Ramdhani, “The concept of sequential pattern mining for text,” IOP Conf. Ser. Mater. Sci. Eng., vol. 434, no. 1, 2018, doi: 10.1088/1757-899X/434/1/012042.

[8] M. Ceci and P. F. Lanotte, “Closed sequential pattern mining for sitemap generation,” World Wide Web, vol. 24, no. 1, pp. 175–203, 2021, doi: 10.1007/s11280-020-00839-2.

[9] S. Arena, E. Florian, F. Sgarbossa, E. Sølvsberg, and I. Zennaro, “A conceptual framework for machine learning algorithm selection for predictive maintenance,” Eng. Appl. Artif. Intell., vol. 133, no. April, 2024, doi: 10.1016/j.engappai.2024.108340.

[10] G. Aloysius and D. Binu, “An approach to products placement in supermarkets using PrefixSpan algorithm,” J. King Saud Univ. - Comput. Inf. Sci., vol. 25, no. 1, pp. 77–87, 2013, doi: 10.1016/j.jksuci.2012.07.001.

[11] R. Trivonanda, R. Mahendra, I. Budi, and R. A. Hidayat, “Sequential Pattern Mining for e-Commerce Recommender System,” in 2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS), 2020, pp. 393–398. doi: 10.1109/ICACSIS51025.2020.9263192.

[12] C. I. Ezeife and H. Karlapalepu, “A Survey of Sequential Pattern Based E-Commerce Recommendation Systems,” Algorithms, vol. 16, no. 10, pp. 1–28, 2023, doi: 10.3390/a16100467.

[13] T. Anwar and V. Uma, “CD-SPM: Cross-domain book recommendation using sequential pattern mining and rule mining,” J. King Saud Univ. - Comput. Inf. Sci., vol. 34, no. 3, pp. 793–800, 2022, doi: 10.1016/j.jksuci.2019.01.012.

[14] P. Dragone, G. Pellegrini, M. Vescovi, K. Tentori, and A. Passerini, “No more ready-made deals: Constructive recommendation for telco service bundling,” RecSys 2018 - 12th ACM Conf. Recomm. Syst., pp. 163–171, 2018, doi: 10.1145/3240323.3240348.

[15] T. P. Pimentel and R. R. Goldschmidt, “Sequential sentiment pattern mining to predict churn in CRM systems: A case study with telecom data,” ACM Int. Conf. Proceeding Ser., 2019, doi: 10.1145/3330204.3330220.

[16] D. Asif, M. S. Arif, and A. Mukheimer, “A data-driven approach with explainable artificial intelligence for customer churn prediction in the telecommunications industry,” Results Eng., vol. 26, no. March, p. 104629, 2025, doi: 10.1016/j.rineng.2025.104629.

[17] J. Ballerini, D. Herhausen, and A. Ferraris, “How commitment and platform adoption drive the e-commerce performance of SMEs: A mixed-method inquiry into e-commerce affordances,” Int. J. Inf. Manage., vol. 72, no. April, p. 102649, 2023, doi: 10.1016/j.ijinfomgt.2023.102649.

[18] M. Kgakatsi, O. P. Galeboe, K. K. Molelekwa, and B. A. Thango, “The Impact of Big Data on SME Performance: A Systematic Review,” 2024. doi: 10.3390/businesses4040038.

[19] R. G. Cooper, “SMEs’ use of AI for new product development: Adoption rates by application and readiness-to-adopt,” Ind. Mark. Manag., vol. 126, no. November 2024, pp. 159–167, 2025, doi: 10.1016/j.indmarman.2025.01.016.

[20] M. B. Legowo, B. Indiarto, and D. Prayitno, “A Conceptualization of The Sme Digitalization Model To Support The ‘SME-Go Digital’ Program in Indonesia,” Indones. J. Bus. Entrep., vol. 8, no. 3, pp. 421–429, 2022, doi: 10.17358/ijbe.8.3.421.

[21] S. Verstege, Y. Zhang, P. Wierenga, L. Paquette, and J. Diederen, “Using Sequential Pattern Mining to Understand How Students Use Guidance While Doing Scientific Calculations,” Technol. Knowl. Learn., vol. 29, no. 2, pp. 897–920, 2024, doi: 10.1007/s10758-023-09677-3.

[22] E. Doko, L. A. Bexheti, M. Hamiti, and B. P. Etemi, “Sequential pattern mining model to identify the most important or difficult learning topics via mobile technologies,” Int. J. Interact. Mob. Technol., vol. 12, no. 4, pp. 109–122, 2018, doi: 10.3991/ijim.v12i4.9223.

[23] S. M. Darshan, N. Nithinraj, and B. Jaisachin, “Integrating Data Mining and Predictive Modeling Techniques for Enhanced Retail Optimization,” Int. J. Comput. Sci. Inf. Secur., p. 22, 2024, [Online]. Available: https://google.academia.edu/JournalofComputerScience

[24] S. Saini, S. Saumya, and J. P. Singh, “Sequential purchase recommendation system for E-commerce sites,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 10244 LNCS, pp. 366–375, 2017, doi: 10.1007/978-3-319-59105-6_31.

[25] M. H. Santoso, “Application of Association Rule Method Using Apriori Algorithm to Find Sales Patterns Case Study of Indomaret Tanjung Anom,” Brill. Res. Artif. Intell., vol. 1, no. 2, pp. 54–66, 2021, doi: 10.47709/brilliance.v1i2.1228.

[26] G. Huang, W. Gan, and P. S. Yu, “TaSPM: Targeted Sequential Pattern Mining,” ACM Trans. Knowl. Discov. Data, vol. 18, no. 5, 2024, doi: 10.1145/3639827.

[27] J. A. Diaz-Garcia, M. D. Ruiz, and M. J. Martin-Bautista, “A survey on the use of association rules mining techniques in textual social media,” Artif. Intell. Rev., vol. 56, no. 2, pp. 1175–1200, 2023, doi: 10.1007/s10462-022-10196-3.