UDC 004.891:004.65:339.138(045)
DOI:10.18372/1990-5548.87.20881

[1]Olena Arsirii,
[2]Dmitriy Ivanov

A MODEL AND METHOD OF TRANSACTIONAL-BEHAVIORAL DATA MINING FOR B2B CONTENT PERSONALIZATION

Odesa Polytechnic National University, Odesa, Ukraine
E-mails: [1]e.arsiriy@gmail.com ORCID 0000-0001-8130-9613
[2]ivanovdima9988@gmail.com ORCID 0009-0009-3958-5310

Abstract**—***The growth of financial significance and structural complexity of the B2B e-commerce market segment, alongside the necessity to increase its efficiency, has determined the relevance of developing a model and method for the intelligent analysis of B2B customer transactional and behavioral data for content personalization. This study analyzes data mining methods based on Apriori, FP-Growth, and Eclat algorithms and data structures, identifying ways to improve them for B2B-specific data analysis. A conceptual model for analyzing B2B customer commercial activity has been developed, incorporating the product of item quantity and individual B2B transaction price. Furthermore, the UP-Growth (Utility Pattern Growth) method has been developed, utilizing a weighted node utility calculation within the tree structure instead of the standard frequency counters used in FP-Growth. The paper provides examples of constructed association rules and sequential patterns, accompanied by explanations of their economic significance. The impact of the derived association rules and sequential patterns on the formation of personalized product, information, and recommendation content within B2B e-commerce systems is examined.*

Keywords—Data mining; association rules; Apriori; FP-Growth; Eclat; quality metrics; e-commerce; B2B systems; transactional and behavioral data.

I. INTRODUCTION

Modern B2B e-commerce systems operate in a highly competitive online environment, where the speed and accuracy of meeting wholesale customer needs depend on content quality. Current content quality standards are based on a personalized approach to its creation. The necessity for effective content personalization in B2B systems is further validated by the immense financial scale of this segment. The global B2B e-commerce market, valued at $11.54 trillion in 2024, significantly surpasses the B2C market ($6.55 trillion). Projections of its growth to $60.62 trillion by 2034 indicate that even a slight increase in customer interaction efficiency through personalization will yield a multi-trillion dollar economic impact [1] – [3].

At the same time, it has been established that traditional content personalization methods are designed for the B2C e-commerce segment and are based on customer data mining – specifically association analysis aimed at identifying hidden relationships between various elements (products and actions) in large B2C transaction databases [4]. Well-known association rule mining methods such as Apriori, FP-Growth, and Eclat are effective in B2C transactional databases containing many homogeneous transactions, as customers often make impulsive purchases driven by emotions and simple user experience [5].

Since B2B purchasing is rational, high-value, and driven by ROI (Return on Investment) requirements, there is an urgent need to analyze and adapt existing models and methods of association analysis to address B2B-specific challenges. These include the complexity of the Buying Center (where purchasing decisions are made by a group of people or departments within the buyer organization rather than a single individual), individual pricing conditions, data sparsity (infrequent but large orders), and the necessity for long-term partnerships [6]. Therefore, the adaptation and modernization of data mining methods that utilize transactional data to identify association rules and typical purchasing patterns, as well as behavioral data to understand purchasing scenarios and the needs of various roles in B2B e-commerce systems, is a highly relevant task. Solving this problem will enable the creation of relevant and rationally justified product, information, and recommendation content based on a personalized approach to meeting the needs of B2B e-commerce customers.

II. LITERATURE REVIEW

In article [7], a B2B e-commerce system is defined as a comprehensive online platform or a set of interconnected technological solutions designed to facilitate commercial transactions, information exchange, and the automation of business processes between two or more legal business entities. It is well-established that B2B (Business-to-Business) transactions involve the exchange of goods, services, or information between two or more business clients [8] – [10]. According to the transaction concept, the primary aspect is wholesale trade [8]. This business model allows retailers to purchase products at a lower price and resell them to consumers with a markup.

Another significant concept is marketing [9]. B2B marketing strategies focus on building relationships with other enterprises rather than individual consumers. This often involves a longer sales cycle and more complex decision-making processes. A business client (Customer) in a B2B context refers to other companies (enterprises) rather than individuals. Such customers possess distinct needs and purchasing behaviors compared to individual consumers [10]. B2B companies must understand the specific requirements and challenges of their business clients to provide value solutions. Collaboration (Collaborative) is a key element of B2B relationships. Enterprises often work closely together to achieve common goals. This collaborative approach fosters long-term partnerships and innovation [10].

Thus, the B2B transaction concept model "Wholesale, Marketing, Customer, Collaborative" significantly influences the formation of personalized content in B2B e-commerce systems, unlike B2C systems, where content is primarily generated through statistical analysis of purchase frequencies and general behavioral patterns of mass consumers. Detailed differences between these two e-commerce operational models across key indicators are presented in Table I. The comparative table was formulated by the authors based on an analysis of information provided in [11] – [14].

TABLE I. COMPARATIVE ANALYSIS OF B2B AND B2C E-COMMERCE SYSTEM OPERATIONAL MODELS BY KEY METRICS

B2B (Business-to-Business)	**B2C (Business-to-Consumer)**
Target Audience	
Smaller number of buyers. Consists of other companies, startups, enterprises, large corporations, government agencies, and non-profit organizations.	*Large number of buyers* – individual consumers.
Transaction Value	
High value, larger volumes, lower frequency.	*Low value*, smaller volumes, higher frequency
Purchase Motivation	
Rational: focused on business efficiency, cost reduction, profit maximization, and long-term goals.	*Emotional/Impulsive*: focused on satisfying personal needs and desires, based on price, brand, and trends.
Decision-Making Duration	
Longer and more complex: involves multiple stakeholders, negotiations, potential testing, and contracts.	*Shorter and simpler*: often a single-person decision, quick purchase.
Sales Cycle	
Long (from 1 to 6 months).	*Short* (from 1 to 6 days).
Customer Relationship Features	
Long-term partnerships: emphasis on trust-building, ongoing support, and customized solutions.	*Short-term*, transactional.
Marketing Features	
Focus on logic, value proposition, ROI, personalized offers, direct sales, and content marketing.	*Focus on emotions*, brand image, advertising, social media, promotions, and convenience.
Payment Methods	
Complex methods (credit lines, invoicing, bank transfers, deferred payments). Higher priority on security.	*Quick and convenient methods* (credit/debit cards, mobile wallets, electronic payments).
Product/Service Customization	
High degree of customization, specialized solutions.	*Standardized* "ready-to-use" products.

For instance, [11], [12] emphasize that the vast majority of B2C e-commerce platforms utilize static content generation methods based on fixed rules or universal recommendations for individual customers. Such an approach fails to account for purchasing patterns, the specific business characteristics of wholesale clients, seasonality, or interdependencies between products. This limits service efficiency and reduces the likelihood of repeat purchases [13], [14]. Addressing these issues requires a systematic analysis of existing models, methods, and technologies for association analysis of transactional and behavioral customer data to modify them for enhancing the personalization of product, information, and recommendation content within B2B e-commerce systems.

Research into established data analysis models and methods for Association Rule Mining indicates that they are primarily focused on identifying frequently recurring combinations of products or events within B2C e-commerce transactions, without considering their quantity or individual price [15] – [17]. This limitation is critical for B2B systems, as it fails to account for the actual value of wholesale transactions. Essentially, these methods are built upon the concept of frequent itemsets, which appear in sets and are linked to the notion of frequency – the primary determinant of value in B2C transactions. Conversely, for B2B transactions, the determining factors are cost and volume.

The fundamental toolkit for identifying such frequent patterns is based on three key approaches that differ in their search space traversal methods and data representation structures. Specifically, these include the well-known Apriori algorithm [15], which performs candidate generation followed by the pruning of infrequent itemsets; the FP-Growth algorithm, which constructs an FP-Tree for subsequent mining without candidate generation [16]; and the Eclat algorithm, which identifies frequent itemsets through the intersection of pre-constructed TID-lists (Transaction ID lists) [17]. Drawing on existing literature, we examine the advantages and disadvantages of these popular association analysis algorithms in greater detail (Table II).

TABLE II. COMPARATIVE ANALYSIS OF ASSOCIATION RULE MINING ALGORITHMS: ADVANTAGES AND DISADVANTAGES

Operational Principle	Advantages	Disadvantages
Apriori		
Candidate Generation	Simple, classical approach.	Slow performance on large datasets.
FP-Growth		
FP-Tree (Frequent Pattern Tree)	High speed, no candidate generation required.	More complex implementation.
Eclat		
TID-list Intersection (Vertical Data Format)	Simple and fast on small to medium datasets.	High memory consumption on large-scale data.

The method for mining frequent itemsets in transactional databases through iterative candidate generation and testing is based on the Apriori algorithm [15]. The fundamental idea of the algorithm relies on the Apriori Property – the principle that any subset of a frequent itemset must also be frequent. The algorithm operates bottom-up (from shorter to longer itemsets), using frequent sets identified at the previous step to generate candidates for the next one. During this process, candidate sets that are guaranteed not to be frequent are discarded (pruning). However, the efficiency of Apriori in B2B is limited due to high computational overhead when analyzing long transactions.

The FP-Growth (Frequent Pattern Growth) algorithm is more adaptive to large datasets, as it minimizes the number of database scans by constructing a compact data structure – the FP-Tree (Frequent Pattern Tree) [16]. Based on this structure, a recursive search for possible product combinations or behavioral patterns is performed. Despite the implementation complexity of FP-Growth, the authors believe that the capability to mine data without candidate generation is critical for complex B2B catalogs. Therefore, this specific algorithm is proposed for further enhancement to analyze the transactional and behavioral data of B2B customers.

The third approach is implemented in the Eclat (Equivalence CLAss Transformation) algorithm [17], which is based on a vertical data representation and transaction identifier (TID-sets) intersection operations, ensuring fast mining in sparse databases. The core idea of Eclat is to store a list of transaction identifiers (TID-lists) for each item where it appears and to find frequent itemsets by intersecting these lists.

Despite their technical excellence, all mentioned association mining methods, in their classical form, are oriented exclusively toward the binary presence of an element in the basket. This necessitates their modification to account for the quantity and value parameters of wholesale operations.

Thus, the conducted research into the specifics of B2B e-commerce systems and existing association mining methods has revealed a significant gap between the theoretical possibilities of pattern discovery in customer transactional and behavioral data and their practical effectiveness in creating personalized content within a B2B environment. This disparity determines the aim and objectives of this research.

III. MATHEMATICAL FOUNDATIONS OF TRANSACTIONAL-BEHAVIORAL DATA MINING

The study is based on the mathematical framework of Data Mining, specifically the concept of Association Rule Mining. Formally, the problem is defined over a set of transactions $D = \{T_1, T_2 \ldots T_m\}$, where each transaction T consists of a set of items (Stock Keeping Units, SKU), $I = \{i_1, i_2 \ldots i_n\}$. The process of discovering dependencies relies on calculating probabilistic metrics, such as *Support* and *Confidence*. However, the classical approach implemented in the Apriori and FP-Growth algorithms utilizes a binary model (the fact of an item's presence), which fails to account for the financial significance of B2B operations. This paper proposes an extension of the mathematical framework of Frequent Itemset Mining toward High-Utility Itemset Mining (HUIM). This enables a transition from simply counting the frequency *count*(i_k) to calculating a utility function $U(i_k)$, which, within the developed conceptual model, is defined as the product of the SKU quantity and its individual price.

IV. The Aim and Objectives of the Research

The aim of this research is to improve the model and method for intelligent analysis of customer transactional and behavioral data to enhance content personalization within B2B e-commerce systems. To achieve this aim, the following objectives have been identified:

- to develop a conceptual model for analyzing the commercial activity of B2B customers, incorporating the product of item quantity and individual B2B transaction price;
- to develop the UP-Growth (Utility Pattern Growth) method, utilizing a weighted node utility calculation within the tree structure instead of the frequency counters used in FP-Growth;
- to analyze the practical research results regarding the developed model and method, and to discuss ways of enhancing the personalization of product, information, and recommendation content in B2B e-commerce systems based on these findings.

V. The Research Materials and Methods

A. Development of a Conceptual Model for B2B Commercial Activity Analysis

Taking into account the fundamental concepts of association analysis in transactional and behavioral data of B2C platforms, such as *transaction*, *sequence*, *sequential pattern*, and *association rule*, as well as the *quality metrics* for their construction, we represent the structure of the B2B commercial activity conceptual model as a three-level hierarchy: "*Entities*, *Relationships*, *Metrics*".

The first level, Entities, contains data objects. It serves as the foundation of the conceptual model, transforming "raw" data into structured information. The entity level consists of sets of item positions I, where each i_k (Stock Keeping Unit, SKU) is the atomic unit of analysis; transactions T, which are logical groupings of products considering volume and individual price (rational context); and customer sequences S, which are chronological chains of transactions reflecting the customer lifecycle. It should be noted that in practice, a B2B entity is a single wholesale order placed by a buyer company, including multiple SKUs and specific financial terms. An SKU has a unique alphanumeric code used by the enterprise (seller, distributor, retailer) for internal identification, tracking, and inventory management [18]. Below are the formal definitions of the «*Entities*» within the conceptual model.

The set of item positions I is defined as

$$I = \{i_1, i_2 \ldots i_n\}, \tag{1}$$

where i_k is the identifier of an item (SKU) available for ordering in the B2B system.

A transaction T is a *logical unit of operation* occurring at a single point in time τ. It is a subset of I (i.e., $T \subseteq I$) and is defined as a set of m tuples:

$$T = \{t_1, t_2 \ldots t_m\}. \tag{2}$$

Each element t_k of the transaction T is a tuple describing a purchased or processed item, accounting for its quantity (Quantity) q and cost p:

$$t_k = i_k, q_k, p_k, \tag{3}$$

where $i_k \in I$ is the unique item identifier (SKU); q_k is the quantity of the item in the transaction; p_k is the individual unit price (Price) applied to the specific customer, reflecting personalized discounts and purchase volume.

A sequence s_k in B2B is a time-ordered activity log of a single customer (or a Buying Center), consisting of a set of transactions T that occurred at different points in time τ:

$$s_k = T_{\tau_1}, T_{\tau_2}, \ldots, T_{\tau_m}, \tag{4}$$

where $T_{\tau j}$ is a B2B transaction (2) that occurred at time τ_j; the sequence is strictly ordered by time, such that $\tau_1 < \tau_2 < \ldots < \tau_m$.

The second level, Logical Relationships, contains the so-called analytical vectors. By establishing these connections, data is transformed into knowledge about behavior. The second level is based on:

- association rule vectors, these describe horizontal relationships, representing item positions (i_k, q_k, p_k) (3) that are purchased together. Such vectors serve as the foundation for "product personalization" and the formation of product bundles;

- sequential pattern vectors, these describe vertical (temporal) relationships s_k (4), representing the sequence of transactions over time. These vectors are the basis for "predictive personalization", recommendation content, and the management of replenishment cycles.

Below are the formal definitions of the "Relationships" within the conceptual model.

An Association Rule (AR) in B2B reflects the simultaneous presence of products, services, or actions within a single transaction T (2). If I is the set of all elements (SKUs / services / actions), and A and B are two sets of elements such that $A \subset I$, $B \subset I$, and $A \cap B = \varnothing$, then an association rule is formulated as:

$$AR = A \to B, \tag{5}$$

which means – *if* a set of elements A is present in a wholesale order (Transaction), *then* there is a high probability that a set of elements B is also present in the same order.

A Sequential Pattern (SP) in B2B reflects the time-ordered occurrence of transactions or behavioral events (3) that form a customer activity chain. If S is the set of all B2B customer sequences (3), and α and β are two sequences of transactions or events such that $\alpha = \langle T_{\tau_a} \rangle, \beta = \langle T_{\tau_b} \rangle$. A sequential pattern is formulated as:

$$\mathrm{SP} = \alpha \Rightarrow \beta, \tag{6}$$

which means – *if* at time τ_α a customer performed an event or a set of transactions α, *then* there is a high probability that at a subsequent time τ_β, $\tau_\alpha < \tau_\beta$ they will perform an event or a set of transactions β. The notation "$\Rightarrow$" is used to specifically emphasize the temporal sequence.

The third level, Evaluation and Validation, contains *quality metrics* whose values serve as tools for selecting the most significant rules or patterns that carry economic value, namely:

- *Support* (S) characterizes the prevalence and scale of a rule or pattern.
- *Confidence* (C) characterizes the reliability of the prediction provided by a rule or pattern.
- *Lift* (L) characterizes the degree of correlation between transactions or patterns, accounting for its non-triviality and intellectual novelty.
- *Leverage* (Le) characterizes the absolute benefit or weight of a rule or pattern for the business.
- *Conviction* (V) characterizes the stability of a rule or pattern against random fluctuations.

Below are the formal definitions of the «Metrics» within the conceptual model for constructing AR.

Support (S) of an *AR* indicates how frequently the set $\{A \cup B\}$ appears in the total database of B2B transactions D. Thus, for the *AR* A→B, the support S_{AB} (A→B)$ is defined as:

$$S_{AR}(A \to B) = S_{AR}(A \cup B) = \frac{\left|\{T \in D \mid A \cup B \subseteq T\}\right|}{|D|}, \tag{7}$$

where $\left|\{T \in D \mid A \cup B \subseteq T\}\right|$ is the total number of transactions containing both sets A and B, and $|D|$ is the total number of transactions in the B2B database.

Confidence (C) of the AR $A \to B$ indicates the probability that if a customer purchased set A, they will also purchase set B within the same transaction. This is a direct measure of the predictive value of set A for the sale of set B within a single wholesale order. Thus, for the AR $A \to B$, the confidence $C_{AR}(A \to B)$ is defined as:

$$C_{AR}(A \to B) = \frac{S_{AR}(A \cup B)}{S_{AR}(A)}. \tag{8}$$

For B2B systems, the quality evaluation (Support, Confidence) of an AR must account for not only the number of transactions $|D|$, but also the total financial value or volume of items A and B within these transactions; that is, it must be *weighted* according to definition (3).

Lift (L) indicates how much more frequently the set $\{A \cup B\}$ appears than would be expected if A and B were independent. Thus, for the AR A→B, Lift $L_{AR}(A \to B)$ is defined as:

$$L_{AR}(A \to B) = \frac{C_{AR}(A \to B)}{S_{AR}(A)} = \frac{S_{AR}(A \cup B)}{S_{AR}(A) S_{AR}(B)}. \tag{9}$$

If $L_{AR} > 1$, there is a positive correlation (rational sense) in the joint purchase; if $L_{AR} < 1$, there is a negative correlation (substitute goods). This is a key metric for B2B, as it filters out trivial rules that appear frequent simply due to the high overall popularity of one of the items.

Leverage (Le) measures the difference between the actual frequency of the co-occurrence of sets A and B and the frequency that would be expected if A and B were completely independent of each other. Thus, for the AR $A \to B$, Leverage $Le_{AR}(A \to B)$ is defined as:

$$Le_{AR}(A \to B) = S_{AR}(A \cup B) - S_{AR}(A) \cdot S_{AR}(B). \tag{10}$$

If Le_{AR} = 0, then elements A and B are independent, and their co-occurrence in orders is purely coincidental. If $Le_{AR} > 0$, there is a strong logical connection between the elements, meaning they appear together more frequently than expected. In B2B, this serves as a signal for creating product bundles or implementing joint logistics. If $Le_{AR} < 0$, the elements appear together less frequently than expected, which may indicate that the products are competitors or substitutes.

Conviction (V) indicates how frequently the AR (5) yields incorrect predictions by measuring the dependence of set A on set B. Thus, for the AR $A \rightarrow B$, Conviction $V_{AR}(A \rightarrow B)$ is defined as:

$$V_{AR}(A \rightarrow B) = \frac{1 - S_{AR}(B)}{1 - C_{AR}(A \rightarrow B)}. \quad (11)$$

A high V_{AR} value indicates a very strong rule, where the presence of set A significantly "forces" the appearance of set B. This metric is particularly useful for filtering rules when set A has very high overall popularity.

Without further detailed elaboration, we define the *Metrics* for constructing *SP* (6). These are applied to the sequence database S, where event β occurs after event α

Support of the *SP*:

$$S_{SP}(\alpha \Rightarrow \beta) = \frac{|\{s \in S \mid \alpha \text{ precedes in sequential } s\}|}{|S|}. \quad (12)$$

Confidence of the *SP*

$$C_{SP}(\alpha \Rightarrow \beta) = \frac{S_{SP}(\alpha \Rightarrow \beta)}{S_{SP}(\alpha)}. \quad (13)$$

Lift of the *SP*

$$L_{SP}(\alpha \Rightarrow \beta) = \frac{S_{SP}(\alpha \Rightarrow \beta)}{S_{SP}(\alpha) \cdot S_{SP}(\beta)}. \quad (14)$$

Leverage of the *SP*

$$Le_{SP}(\alpha \Rightarrow \beta) = S_{SP}(\alpha \Rightarrow \beta) - S_{SP}(\alpha) \cdot S_{SP}(\beta). \quad (15)$$

Conviction of the *SP*

$$V_{SP}(\alpha \Rightarrow \beta) = \frac{1 - S_{SP}(\beta)}{1 - C_{SP}(\alpha \Rightarrow \beta)}. \quad (16)$$

Thus, the proposed conceptual model of B2B commercial activity enables the formalization of complex, multi-factor interaction processes between the supplier and the customer. It integrates static association analysis for optimizing product content with dynamic sequential pattern analysis for proactive demand forecasting.

B. Development of the Utility Pattern Growth (UP-Growth) Method

The mathematical foundation of the conceptual model (levels "Entities, Relationships, Metrics"), as demonstrated through expressions (1) – (16), enables the transition from classical frequent itemset analysis to the discovery of high-utility patterns. The Utility Pattern Growth (UP-Growth) method, unlike the basic Frequent Pattern Growth (FP-Growth) method [16], operates not with the number of item occurrences, but with the weighted utility of the tree nodes.

At the core of the proposed enhancement lies the replacement of the unit increment of the support counter with the calculation of Transaction Utility. According to the developed model (3), the utility of each element in a tree node is defined as the sum of the products of the quantity q_k and the individual price p_k across all associated transactions

The process of discovering utility patterns using the UP-Growth method consists of four key stages:

1) Calculation of SKU Potential Utility (Transaction-Weighted Utilization):

- instead of a simple support count, for each item position i_k, the tota*l utility* of all transactions in which it is present is calculated as the product $q_k \cdot p_k$, specifically: $U_k = q_k \cdot p_k$;
- items whose total utility is lower than the predefined *min_utility* threshold are discarded.

2) Construction of the UP-Tree (Utility Pattern Tree):

- items within transactions are sorted in descending order of their potential *utility*;
- when adding a transaction to the tree, nodes are not merely incremented (count + 1) but instead accumulate the utility value of the specific order;
- the compact structure of the UP-Tree preserves information regarding the financial significance of each prefix path

3) Generation of Conditional Potential Utility Bases, for each SKU, a conditional base is formed based on the paths in the tree; however, instead of frequency counters, the minimum potential utility values of the paths are used.

4) Identification of High-Utility Itemsets (HUI), through recursive traversal of the UP-Tree, itemsets are identified whose total utility exceeds the specified *min_utility* threshold,

$$U_k = q_k \cdot p_k \geq min\_utility.$$

Table III presents the input data for a test example of transactions, accounting for their utility, and the results of the first two stages of the UP-Growth method. Figure 1 illustrates the structure of the constructed Utility Pattern Tree.

TABLE III. INPUT DATA FOR THE TRANSACTIONAL TEST CASE ACCOUNTING FOR UTILITY AND RESULTS OF THE FIRST TWO STAGES OF THE UP-GROWTH METHOD

Input Data Accounting for Utility $q_k{\times}p_k$			**Calculation of Potential Utility** SKU **(TU)** **min_utility= $130**	
TID	Transaction Content (SKU, Quantity, Price)	Utility (U)	SKU	**TU≥$130**
T1	{A(1, $10), B(1, $20), D(1, $5), E(1, $5)}	$40	**A:**	T1($40) + T3($50) + T5($60) = **$150**
T2	{B(2, $20), C(1, $10), D(1, $5)}	$55	**B:**	T1($40) + T2($55) + T3($50) + T4($35) + T5($60) = **$240**
T3	{A(1, $10), B(1, $20), C(1, $10), E(2, $5)}	$50	**C:**	T2($55) + T3($50) + T5($60) = **$165**
T4	{B(1, $20), E(3, $5)}	$35	**D:**	T1($40) + T2($55) + T5($60) = **$155**
T5	{A(2, $10), B(1, $20), C(1, $10), D(2, $5)}	$60	**E:**	T1($40) + T3($50) + T4($35) = **$125**
Sorting SKUs by Descending Utility: B ($240) > C ($165) > D ($155) > A ($150), creating **UP-Growth**				
T1	**B, D, A,** Creating nodes: B(40) → D(40) →A(40)			
T2	**B, C, D**. Incrementing B (40 + 55 = 95) Adding a new branch: C(55) →D(55).			
T3	**B, C, A.** Incrementing B (95 + 50 = 145). Incrementing C (55 + 50 = 105). Adding a new branch C (105)→A(50)			
T4	**B.** Incrementing B (145 + 35 = 180)			
T5	**B, C, D, A.** Incrementing B (180 + 60 = 240). Incrementing C (105 + 60 = 165). Incrementing D (55 + 60 = 115). Adding a new branch D (115) → A(60)			

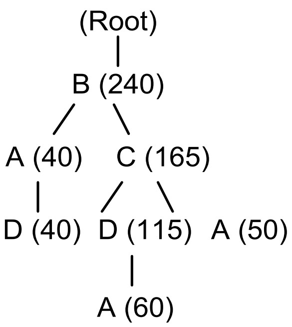

Fig. 1. UP-Tree Structure for the Test Case

At Step 3, for each item, we identify all paths in the constructed UP-Tree (Fig. 1) that contain this item and form a Conditional Utility Pattern Base. The analysis proceeds from the least valuable item to the most valuable (bottom-up through the priority list): $A{\to}D{\to}C{\to}B$ considering the threshold $min\_utility$ = 130. Unfortunately, the limited scope of this article does not allow for a detailed demonstration of the conditional UP-Tree construction process for each element in the priority list; however, it is precisely at this stage that economically inefficient associations are filtered out prior to their final generation, significantly optimizing B2B system resources when generating personalized content.

At Step 4, we search for utility patterns where the total weight of the nodes in the tree exceeds 130. For the input data of the test example, such a pattern is $\{B, C\}$ with a tree utility equal to 165.

In conclusion, it should be noted that for the personalization of B2B content, the use of the enhanced UP-Growth method solves the "cheap popular items" problem. This enables the B2B e-commerce system to recommend products that generate the primary profit rather than those that are merely frequently purchased, as is the case when using the FP-Growth method.

VI. RESULTS OF THE RESEARCH AND THEIR DISCUSSION

The practical validation of the developed conceptual model for B2B commercial activity analysis and the UP-Growth utility pattern mining method was conducted using transactional and behavioral data from the B2B e-commerce system "Baza Obuvi" [14] (bazaobuvi.com.ua). The specificity of the research object lies in the fact that the atomic unit of analysis (SKU) is a shoe carton (size assortment), which contains a fixed set of pairs of the same article in various sizes (e.g., from 36 to 41). Furthermore, each transaction includes financial parameters specific to B2B trade (individual price per carton, number of cartons, and seasonal supply conditions). Below are examples of the constructed association rules and sequential patterns, along with explanations of their underlying economic significance.

`**Association Rule**` `(Itemset):{Women's Sneakers Art. 403, q=5 cartons, p=4600 UAH/carton; Men's Sneakers Art. 510, q=5 cartons, p=5200`

UAH/carton}→ {Insole Set "Comfort", q=100 pc., p=15 UAH/pc}

Economic Significance, the constructed rule demonstrates a high Leverage (10) level, as the sale of the cross-sell item is tied to a high-value primary purchase, where the total transaction utility is: U(T) = (5 ·4600) + (5 ·5200) + (100 ·15) = 50500 UAH.

Association Rule (Product + Service):{Footwear Order q > 1000 pairs, U(T) > 500000 UAH} →{Granting a Line of Credit, term 60 days}

Economic Significance, the constructed rule links purchase volume with financial content (crediting), serving as an example of payment terms personalization based on rational justification (volume q and price p).

Sequential Pattern (Seasonality): ⟨Winter Boots, q=20 cartons, p=6500 UAH/carton⟩ $\xrightarrow{\Delta t=2mos}$ ⟨Salt Protection Products, q=200 pc., p=45 UAH/pc⟩

Economic Significance, the constructed pattern is aimed at forecasting the demand for care products based on a client's prior significant investment in the winter assortment. A time lag of 2 months is applied between transactions $T_{\tau 1}$ and $T_{\tau 2}$

Sequential Pattern (Forecasting): ⟨ (Request for individual price for the "Spring 2026" collection), (Receipt of technical documentation)⟩ $\xrightarrow{\Delta t\leq 14 days}$ ⟨Wholesale order U(T)≥150000UAH⟩

Economic Significance, the constructed pattern links behavioral events (price requests) with a transactional outcome. If events α (request for price and documentation) occur, the probability of a high-value transaction β taking place within 14 days is maximized.

In conclusion, we consider the authors' discussion proposals for enhancing content personalization using the B2B e-commerce system "Baza Obuvi" as a case study. Based on the results of the data mining performed on transactional and behavioral data, three key directions for content personalization have been identified: product, recommendation, and information-based.

For product personalization (Wholesale Bundles), the use of the Leverage metric Le_{AR} (10) allows the "Baza Obuvi" system module to automatically generate content in the form of "Wholesale Bundles." For instance, when a large batch (more than 10 cartons) of "Mermaid" women's footwear is ordered, the system offers the user a personalized bundle with related accessories (insoles, care products), where the price p_k is adjusted depending on the volume q_k. This effectively transforms a standard product page into an individual commercial offer

To generate recommendation content, sequential forecasting patterns are utilized. Applying sequential pattern analysis based on time lags (e.g., 2–3 months between "Winter" and "Spring" season purchases) enables the implementation of the *Restock Prediction* analytical function. A typical scenario for the "Baza Obuvi" system is as follows: if a client triggers the event «Request for individual price for a new leather footwear collection» the system generates recommendation content in the form of technical documentation and certificates of conformity based on the sequential pattern. This accelerates the decision-making cycle for the primary wholesale order.

Rationally justified information content is created using metrics such as Lift L_{AR}, which compares a client's basket with the patterns of clients within their specific cluster. If competitors in a certain region have already begun purchasing «Rubber Galoshes» and the current client has not, the system automatically sends an analytical "missed opportunity" report, stimulating the purchase. Furthermore, detecting behavioral anomalies, such as a sharp decrease in the purchase volume of top brands like "Baas" or "Bona", allows the system to proactively modify the content on the client's homepage, offering individual cooperation terms to prevent Churn.

Thus, the transition from frequency analysis to utility analysis via the developed conceptual model and the UP-Growth method allows a B2B system to transform from a simple product catalog into an intelligent assistant. This ensures the delivery of proactive and rationally justified content that directly impacts the profitability of both the supplier "Baza Obuvi" and its wholesale clients.

VII. CONCLUSION

This article addresses the relevant scientific and applied task of improving data mining for transactional and behavioral data to enhance content personalization efficiency within B2B e-commerce systems. The main results of the research are as follows:

1) A conceptual model for analyzing the commercial activity of B2B clients has been developed. Unlike existing approaches, it is based on a three-level structure "Entities, Relationships, Metrics". This model accounts not only for the occurrence of an event but also for the rational context of the transaction, the product of purchase volume q and individual price p, which is critical for the wholesale segment.

2) An enhanced UP-Growth (Utility Pattern Growth) method is proposed. It is based on a transition from the frequency counters of the classical

FP-Growth algorithm to the calculation of transaction-weighted potential utility for tree nodes. This shift allows for the reprioritization of association rule generation, focusing the personalization system on economically significant patterns that yield maximum profit (High-Utility Itemsets).

3) Validation of the developed solutions was conducted using the "Baza Obuvi" system. Practical results demonstrated that the use of weighted association rules transforms the B2B platform from a passive product catalog into an active decision support system. Proposed personalization scenarios (e.g., Wholesale Bundles, "missed opportunity" analytical reports) ensure increased client loyalty and optimization of the supplier's operational processes.

4) The use of sequential pattern analysis for implementing predictive personalization functions, such as *Restock Prediction*, has been justified. It was established that accounting for the time lag between transactions and the dynamics of behavioral indicators allows for the automatic adaptation of informational and recommendation content to the individual procurement cycle of the business client.

In summary, the research establishes the necessary theoretical foundation for transforming B2B e-commerce systems from passive catalogs into intelligent decision support systems, thereby increasing client loyalty and the overall economic efficiency of the platform.

REFERENCES

[1] Global Ecommerce Statistics: Trends to Guide Your Store in 2025 – Available from: https://www.shopify.com/enterprise/blog/global-ecommerce-statistics – [Accessed: January 2026].

[2] Insider Intelligence – Available from: https://www.emarketer.com/insights/mobile-commerce- shopping-trends-stats/ [Accessed: January 2026].

[3] T. S. Panasiuk, & I. R. Zubenko, Perspektyvy rozvytku B2C elektronnoi komertsii v sviti [Prospects for the development of B2C e-commerce in the world], 2024. https://doi.org/10.5281/zenodo.14603373

[4] O. O. Arsirii, D. V. Ivanov, & O. Yu. Babilunha, Analiz modelei, metodiv ta tekhnolohii stvorennia personalizovanoho kontentu v systemakh B2B elektronnoi komertsii [Analysis of models, methods and technologies for creating personalized content in B2B e-commerce systems]. Informatyka. Kultura. Tekhnika [*Informatics. Culture. Technology*], vol. 2, pp. 273–278. 2025. https://ict.op.edu.ua/index.php/journal/uk/issue/view/3

[5] Y.-H. Hu, Y.-L. Chen, & K. Tang, "Mining sequential patterns in the B2B environment," *Journal of Information Science*, 35(6), 677–694, 2009. https://doi.org/10.1177/0165551509103600 (Original work published 2009)

[6] D. Kahneman, (2011), "Thinking, Fast and Slow," *Farrar, Straus and Giroux*, August 2014 Statistical Papers vol. 55(3), Page 915, https://doi.org/10.1007/s00362-013-0533-y

[7] E. Turban, J. Outland, D. King, et al., (2018) Electronic Commerce 2018: A Managerial and Social Networks Perspective. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-58715-8

[8] A. Gunasekaran, et al., "E-Commerce and Its Impact on Operations Management. International," *Journal of Production Economics*, 75, 185–197, 2002. https://doi.org/10.1016/S0925-5273(01)00191-8

[9] Fred Wiersema, The B2B Agenda: The current state of B2B marketing and a look ahead, Industrial Marketing Management, Volume 42, Issue 4, 2013, pp. 470–488, https://doi.org/10.1016/j.indmarman.2013.02.015. (https://www.sciencedirect.com/science/article/pii/S0019850113000345)

[10] G. L. Lilien, "The b2b Knowledge Gap. International," *Journal of Research in Marketing*, 75, 196–210, 2016. https://doi.org/10.1016/j.ijresmar.2016.01.003

[11] N. M. Shveda, & O. I. Krauze, Elektronna komertsiia: suchasnyi stan ta stratehii rozvytku, ["E-commerce: current state and development strategies"] Mizhnarodnyi naukovyi zhurnal "Internauka". Seriia: Ekonomichni nauky [*International Scientific Journal "Internauka". Series: Economic Sciences*], Issue 2 (82), 2024. URL: https://doi.org/10.25313/2520-2294-2024-2-9639

[12] Y.-H. Hu, Y.-L. Chen, & K. Tang, "Mining sequential patterns in the B2B environment," *Journal of Information Science*, 35(6), 677–694, 2009. https://doi.org/10.1177/0165551509103600 (Original work published 2009)

[13] Systema elektronnoi komertsii B2B "Baza vzuttia" [B2B e-commerce system "Baza Obuvi"]. Available at: https://bazaobuvi.com.ua/ua/ [Accessed: January 2026]

[14] E. Cramer-Flood, "Global Ecommerce Forecast & Growth Projections," Available from: https://www.insiderintelligence.com/content/globalecommerce-forecast-2022 – [Accessed: January 2026]

[15] Rakesh Agrawal, Tomasz Imieliński, and Arun Swami, 1993. Mining association rules between sets of items in large databases. In Proceedings of the 1993 ACM SIGMOD international conference on Management of data (SIGMOD '93). Association for Computing Machinery, New York, NY, USA, 207–216. https://doi.org/10.1145/170035.170072

[16] J. Han, J. Pei, Y. Yin, et al., "Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach," *Data Mining and*

Knowledge Discovery, 8, 53–87, 2004. https://doi.org/10.1023/B:DAMI.0000005258.31418.83

[17] M. J. Zaki, "Scalable algorithms for association mining," *in IEEE Transactions on Knowledge and Data Engineering*, vol. 12, no. 3, pp. 372–390, May-June 2000, https://doi.org/10.1109/69.846291

[18] Grow your B2B sales with 6 smart B2B product bundling strategies: Available from: https://turis.app/b2b-ecommerce/grow-b2b-sales-6-b2b-product-bundle-strategies/ – [Accessed: January 2026]

[19] Grow your B2B sales with 6 smart B2B product bundling strategies: Available from: https://turis.app/b2b-ecommerce/grow-b2b-sales-6-b2b-product-bundle-strategies/ – [Accessed: January 2026].

[20] Investopedia: SKU: What It Is and How It Works By Andrew Bloomenthal Updated June 15, 2025, Reviewed by Samantha Silberstein, Fact checked by Ariel Courage: Available from: https://www.investopedia.com/terms/s/stock-keeping-unit-sku.asp – [Accessed: January 2026]

[21] O. O. Arsirii, O. M. Liubomska, O. V. Rudenko, & D. V. Ivanov, "Hybrid recommendation system for supporting UI/UX designers," *Informatics. Culture. Technology,* 2024, vol. 1, pp. 29–35. https://doi.org/10.15276/ict.01.2024.03 ISSN 2522-1523 (Online)

Received: November 28, 2025
Accepted: December 23, 2025
Published: February 22, 2026

Arsirii Olena. ORCID 0000-0001-8130-9613. Doctor of Science (Eng), Professor, Head of the Department of Information Systems. Odesa Polytechnic National University, Odessa, Ukraine.
Education: Odessa Polytechnic Institute, Odessa, Ukraine, (1981).
Research interests: Information technology; artificial intelligence; decision support systems; machine learning; neural networks.
Publications: more than 150 papers.
E-mail: e.arsiriy@gmail.com

Ivanov Dmitriy. ORCID 0009-0009-3958-5310 Postgraduate Student.
Department of Information Systems, Odesa Polytechnic National University, Odessa, Ukraine.
Education: Odesa Polytechnic National University, Odessa, Ukraine, (2016).
Research interests: Data mining, e-commerce, B2B systems.
Publications: 6.
E-mail: ivanovdima9988@gmail.com

О. О. Арсірій, Д. В. Іванов. Модель та метод інтелектуального аналізу транзакційно-поведінкових даних клієнтів для персоналізації B2B-контенту

Зростання фінансової значущості та структурної складності сегмента ринку B2B-систем електронної комерції, а також необхідність підвищення його ефективності зумовило актуальність розробки моделі та метода інтелектуального аналізу транзакційних та поведінкових даних B2B-клієнтів з метою персоналізації контенту систем електронної комерції. Проаналізовано методи інтелектуального аналізу які базуються на алгоритмах та структурах даних Apriori, FP-Growth та Eclat. Визначено шляхи їх удосконалення щодо аналізу транзакційних і поведінкових даних B2B-клієнтів. Розроблено концептуальну модель для аналізу комерційної активності B2B-клієнтів з урахуванням добутку кількості товару та індивідуальної ціни B2B-транзакції. Розроблено метод розростання ціннісних патернів UP-Growth (Utility Pattern Growth) з використанням розрахунку зваженої цінності вузлів дерева замість частотних лічильників в FP-Growth. Показано приклади побудованих асоціативних правил та послідовних шаблонів з поясненнями отриманого економічного сенсу. Розглянуто вплив отриманих асоціативних правил та послідовних шаблонів на формування персоналізованого товарного, інформаційного та рекомендаційного контенту системи B2B електронної комерції.

Ключові слова: інтелектуальний аналіз даних; асоціативні правила; Apriori; FP-Growth; Eclat; метрики якості; електронна комерція; B2B-системи; транзакційні та поведінкові данні.

Арсірій Олена Олександрівна. ORCID 0000-0001-8130-9613. Доктор технічних наук. Професор.
Завідувачка кафедри інформаційних систем, Національний університет «Одеська Політехніка», Одеса, Україна.
Освіта: Одеський політехнічний інститут, Одеса, Україна, (1981).
Напрямок наукової діяльності: Інформаційні технології; штучний інтелект; системи підтримки рішень; машинне навчання; нейронні мережі.
Кількість публікацій: більше 180 наукових робіт.
E-mail: aprodeus@gmail.com

Іванов Дмитро В'ячеславович. ORCID 0009-0009-3958-5310. Аспірант.
Кафедра інформаційних систем, Національний університет «Одеська Політехніка», Одеса, Україна.
Освіта: Одеський Національний політехнічний університет, Одеса, Україна, (2016).
Напрямок наукової діяльності: Інтелектуальний аналіз даних, електронна комерція, B2B-системи.
Кількість публікацій: 6.
E-mail: ivanovdima9988@gmail.com