



Operational Issues in Platform-Based Supply Chain Information Systems: A BERTopic-Based Analysis

Platform Tabanlı Tedarik Zinciri Bilgi Sistemlerinde Operasyonel Sorunlar: BERTopic Tabanlı Bir Analiz

Oğuz Emre Balkar^a Yunus Emre Özenoğlu^b Okan Alkan^{c*} Veysel Şahin^d

^a Asst. Prof., Erzincan Binali Yıldırım University, Vocational School, Department of Management and Organization, Erzincan/Türkiye, oguz.balkar@erzincan.edu.tr, ORCID: 0000-0001-5853-0926

^b PhD Student, Atatürk University, Institute of Social Sciences, Management Information Systems, Erzurum/Türkiye, yunusemre.ozenoğlu12@ogr.atauni.edu.tr, ORCID: 0000-0001-5051-4622

^c Lect.Dr., Erzincan Binali Yıldırım University, Vocational School, Department of Computer Technologies, Erzincan/Türkiye, oalkan@erzincan.edu.tr, ORCID: 0000-0002-6526-1645

^d PhD Student, Atatürk University, Institute of Social Sciences, Management Information Systems, Erzurum/Türkiye, veysel.sahin18@ogr.atauni.edu.tr, ORCID: 0000-0002-0827-9478

ARTICLE INFO

Article Type

Research Article

Keywords

Supply Chain Information Systems
Operational Issues
App Review Mining
Topic Modeling
BERTopic Analysis

Received: May, 01, 2026

Accepted: Sep, 17, 2026

ABSTRACT

This study aims to identify the operational issues encountered by business partners in platform-based supply chain information systems through mobile application reviews. In this context, 5 612 raw reviews of five business partner/seller applications available on the Google Play Store were collected, and 5 541 reviews remained analyzable after preprocessing. Since the initial BERTopic model was dominated by short and general satisfaction expressions, a problem-focused dataset consisting of 2 854 reviews was constructed. The problem-focused BERTopic analysis produced six meaningful topic clusters and one outlier/mixed problem pool. Model robustness was examined through topic coherence, topic diversity, stability checks, and intercoder reliability in manual topic labeling. The findings showed that the most critical issue areas are return, commission, and seller cost problems (K1), missing order and customer notifications (K2), and login, password, and access problems (K3). The study positions partner applications as operational components of platform-based supply chain information systems and contributes to the literature by transforming BERTopic outputs into a process improvement-focused decision support matrix.

MAKALE BİLGİSİ

Makale Türü

Araştırma Makalesi

Anahtar Kelimeler

Tedarik Zinciri Bilgi Sistemleri
Operasyonel Sorunlar
Uygulama Yorumu Madenciliği
Konu Modelleme
BERTopic Analizi

Geliş Tarihi : 01 Mayıs 2026

Kabul Tarihi: 17 Eylül 2026

ÖZ

Bu çalışma, platform tabanlı tedarik zinciri bilgi sistemlerinde iş ortaklarının karşılaştığı operasyonel sorunları mobil uygulama yorumları üzerinden belirlemeyi amaçlamaktadır. Bu kapsamda Google Play Store'da yer alan beş iş ortağı/satıcı uygulamasına ait 5 612 ham yorum toplanmış, ön işleme sonrasında 5 541 yorum analiz edilebilir nitelikte kalmıştır. İlk BERTopic modelinde kısa ve genel memnuniyet ifadelerinin konu kümelerini baskılaması nedeniyle, 2 854 yorumdan oluşan problem odaklı veri seti oluşturulmuştur. Problem odaklı BERTopic analizi altı anlamlı konu kümesi ve bir aykırı/karma problem havuzu üretmiştir. Modelin sağlamlığı; konu tutarlılığı, konu çeşitliliği, kararlılık kontrolleri ve manuel konu etiketleme sürecinde kodlayıcılar arası güvenilirlik ölçütü aracılığıyla incelenmiştir. Bulgular, en kritik sorun alanlarının iade, komisyon ve satıcı maliyeti (K1), sipariş ve müşteri bildirimlerinin gelmemesi (K2) ile giriş, şifre ve erişim sorunları (K3) olduğunu göstermektedir. Araştırma, iş ortağı uygulamalarını platform tabanlı tedarik zinciri bilgi sistemlerinin operasyonel bileşenleri olarak konumlandırmakta ve BERTopic çıktılarını süreç iyileştirme odaklı bir karar destek matrisine dönüştürerek yazına katkı sağlamaktadır.

* Corresponding Author / Sorumlu Yazar

Cite as/Atıf: Balkar, O.E., Özenoğlu, Y.E., Alkan, O. and Şahin, V. (2026). Operational Issues in Platform-Based Supply Chain Information Systems: A BERTopic-Based Analysis. *International Journal of Economics, Business and Politics*, 10(2), 572-601. <https://doi.org/10.29216/ueip.1942368>



This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license. / Bu makale, [Creative Commons Atıf \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) lisansının hüküm ve koşulları altında dağıtılan açık erişimli bir makaledir.

1. Introduction

The operations of businesses and the structures of the supply chains they are involved in have gradually changed with digital transformation, which is one of the significant developments of our time. In particular, digital platforms, where online commerce environments are located, have become places that increase transparency, coordination, and efficiency in supply chain management by algorithmically facilitating the exchange of information, products, and services (Zhao, 2025: 612–613).

These platforms also function as multi-party marketplaces. They provide process integration benefits for the disconnection problem, one of the complex problems in supply chain management (SCM), by reducing transaction costs among different market actors (Culotta et al., 2024: 449). The financial capacities of these environments, supported especially by the advancement of cloud computing (CC), the internet of things (IoT), and artificial intelligence (AI) technologies, enable firms to change their organizational and information technology (IT) structures and create innovative business models (Ivanov et al., 2022: 1–2). Thus, through integrated value creation, traditional supply chains are transformed into dynamic networks in which multifaceted flows of SCM-based activities take place and interact in real time (Johnson & Whang, 2002: 414).

Therefore, these online information systems (which traders can operate as business partners and/or vendors) include operational processes such as order fulfillment, inventory tracking, invoicing, payment processing, delivery, returns, and support. Businesses and/or individuals can reach a wide customer base through these information systems with the help of these integrated processes, that also avoid the high marketing, hardware, and financing costs associated with establishing their own independent IT infrastructure (Sorkun, 2019: 152).

These mobile applications and web systems, operating within the framework of SCM information system (IS) logic, can offer advantages in terms of registration processes, product listing processes, pricing autonomy, ease of delivery, accounts receivable collection, and support services for vendors (Kumar et al., 2021: 2316–2317). Nevertheless, the extent to which vendors and business partners effectively use these digital interfaces and benefit from SCM IS tools can vary depending on certain circumstances.

Operational problems in online SCM IS infrastructures can have significant negative consequences on platform performance, information quality, vendor satisfaction, trust in the system, and supply chain visibility. Weaknesses in operational integration and deficiencies in technical infrastructure in implemented business flows can lead to order delays, inaccurate inventory information, and inadequate response times in customer service (Urrea et al., 2024: 3808). In such a situation, disruptions in logistics processes, insufficient information flow, and inadequate support services can erode vendor trust in the platform and this leads users to abandon it.

However, technical disruptions such as system instabilities and data synchronization errors not only cause financial losses for vendors but also increase operational risks throughout the entire supply chain network. Therefore, system owners look at mobile application reviews to understand the platform experiences, expectations, and operational problems encountered by their partners and vendors. This data provides a valuable and rich source of user feedback (Yi et al., 2025: 1–2).

These reviews include evaluations made by consumers or business partners in uncontrolled environments after adopting products and services and they provide far more realistic and in-depth information than traditional survey methods in revealing the strengths and weaknesses of systems that can be improved. The large volume of textual data coming in daily from millions of users in app stores creates a vast pool of information that is difficult to analyze manually. However, when accurately analyzed, this data helps platform owners make strategic decisions regarding current market dynamics and user needs (Sagama et al., 2024: 2–3).

Topic modeling approaches gains prominence for extracting meaningful themes from large volumes of short, unstructured, and often colloquial textual data. BERTopic, a particularly advanced deep learning-based method, is considered a more meaningful and interpretable topic cluster for analyzing mobile application reviews than traditional topic modeling approaches, owing to its ability to capture contextual relationships ((Halim et al., 2026: 2239; Zhao et al., 2026: 1101–1102).

The BERTopic model employs Uniform Manifold Approximation and Projection (UMAP) to dramatically reduce the dimensionality of high-dimensional text embeddings. At the same time, this process preserves the underlying semantic relations among them. Additionally, the model features the use of Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) that detects density-based clusters in the embedding space reduced by UMAP. Also, class-based term frequency-inverse document frequency (c-TF-IDF) comes in handy when it is time to identify the most representative keywords and phrases of the topic clusters. Thanks to this synergy, BERTopic can produce topics that not only make semantic sense but are also very easy to interpret even when they are extracted from brief, chaotic and unstructured texts created by users.

This solid technical framework enables the pulling out of very consistent, highly understandable, and context-sensitive topic groups, even from noisy and complex texts (Putri et al., 2026). Therefore, complicated and chaotic complaints can be turned into meaningful thematic structures.

Yet a look at the literature indicates that a large part of topic modeling research based on mobile app reviews is simply about end-consumer reviews, general app usability experiences, or buyer complaints (Hanifah et al., 2025: 287). In these papers, the viewpoint from the seller and business partner side of operations is quite forgotten in e-commerce and digital platform research.

While most of the prior research has concentrated on the customers' side of the platform experience, the supplier-side operational issues in partner and seller applications have been comparatively neglected. This paper fills this gap by analyzing partner reviews and looking at them from a platform-based supply chain information systems perspective. By adopting this approach, this paper shifts the analytical focus from conventional end-user satisfaction to the operational experiences of sellers, restaurants, and business partners who utilize platform applications as daily supply chain interfaces.

One of the main theoretical implications of our research is that we conceptualize partner and seller apps as on the supply side operational components of platform-based supply chain information systems. Instead of seeing these apps simply as mobile software tools, we view them as digital conduits that business partners utilize for handling various platform-related operations such as order processing, returns, product information, notifications, access, payments, and other operational tasks. This perspective change allows us to associate our results more closely with the literature on information systems success, logistics service quality, operational risk management, and platform governance.

Methodologically, this study contributes by leveraging Google Play Store reviews as naturally occurring, user-generated data and applying BERTopic to detect operational problem clusters within short, unstructured review texts. In addition, the topic modeling outputs are strengthened through model validation indicators, manual topic labeling, intercoder reliability, and the transformation of topic clusters into an operational decision support matrix. Thus, the study does not limit BERTopic outputs to descriptive topic labels, instead it converts them into interpretable and actionable process improvement categories.

The managerial contribution of the study is that it provides platform managers with a prioritization framework for identifying which operational problems require urgent intervention. By linking topic clusters with operational process categories, the study shows whether issues such as

return and commission policies, missing order notifications, login and access failures, application errors, usability problems, or product upload difficulties should be treated as immediate, high-priority, or medium-priority improvement areas. Thus, the results present a pragmatic guide to upgrading partner experience, platform stability, and supply-side procedural regularity.

2. Theoretical Framework

2.1. Platform-Based Supply Chain Information Systems

In the contemporary business landscape, platform-based supply chain information systems serve as pivotal digital infrastructures that interconnect multiple supply chain actors, thereby enabling them to coordinate operational activities, share information, and manage transactions. These systems connect sellers, restaurants, logistics service providers, customers, payment systems, and platform support units within a shared digital space. Therefore, digital platforms both facilitate commercial exchanges and change coordination mechanisms, operational responsibilities, and information flows among the actors who depend on each other (Culotta et al., 2024: 449).

In platform-based supply chain structures, partner and seller applications are the supply-side operational interfaces of the platform. Partners use these applications to manage orders, change product info, check stock, process returns, keep track of payment and commission transactions, talk to customers, and get platform-related alerts. Therefore, these apps should not be thought of simply as mobile software tools. In fact, they are operational pieces of supply chain information systems with direct impact on process continuity, supply chain visibility, and partner experience.

From this perspective, operational challenges encountered by partner application operators may yield consequences that extend far beyond the mere dissatisfaction of individual end-users. Problems with returns, disputes over the returns, late sending of notices, failures in user login, errors in the systems, and problems associated with product upload may disturb activities related to order fulfillment, reverse logistics, payment tracking, inventory accuracy, as well as coordination among the seller and platform. Accordingly, mobile application reviews generated by business partners constitute a critical and substantive repository of data for diagnosing supply-side operational inefficiencies within platform ecosystems.

2.2. Information Systems Success and Technology Acceptance

The information systems success model provides a robust theoretical framework for explaining operational challenges within partner and seller applications. DeLone and McLean's model shows how system quality, information quality, and service quality can lead to use, user satisfaction, and net benefits (DeLone & McLean, 2003: 19). Quality within platform-based supply chain information systems can be systematically evaluated across three core dimensions: system, information, and service quality. System quality includes technical attributes such as application stability, login reliability, access continuity, response time, and error-free transaction flows. Concurrently, information quality is assessed through the accuracy, timeliness, completeness, and visibility of operational data, including orders, inventory, and payments, while service quality pertains to support mechanisms, problem resolution, and platform responsiveness.

Additionally, the technology acceptance literature provides valuable insights into how business partners evaluate these systems (Venkatesh et al., 2003: 425). Davis (1989: 319) pointed out that perceived usefulness and perceived ease of use are probably the most essential factors in technology acceptance. When it comes to partner applications, perceived usefulness is mainly focused on whether the application can help operational tasks such as receiving orders, uploading products, tracking payments, and handling customer communication. Perceived ease of use, on the other hand, can be linked to interface clarity, application speed, navigation quality, and the level of simplicity in completing operational tasks.

Therefore, the recurring issues in partner applications are not only technical flaws. They can lower perceived usefulness, make the platform seem less easy to use, cause dissatisfaction, and even result in stopping the continued platform use (Bhattacharjee, 2001: 351; Wixom and Todd, 2005: 86). For instance, login and access issues are a sign of system quality problems, missing order and customer notifications point to information quality issues. Also, problems with usability or slowness are very close to perceived ease of use. This theoretical perspective helps understand the topic clusters which are derived from BERTopic as a sign of information system quality and technology acceptance issues.

2.3. Logistics Service Quality and Operational Risk

Logistics service quality as discussed in logistics literature serves as a major theoretical perspective for understanding operational issues in the supply chain information systems that are platform-based (Balkar, 2025: 77; Murfield et al., 2017: 264). Service quality in digital commerce interfaces is typically measured by factors like efficiency, system availability, fulfillment, reliability, recovery, and customer service (Collier & Bienstock, 2006: 260-261; Parasuraman et al., 2005: 213; Rabinovich & Bailey, 2004: 651; Wolfinbarger & Gilly, 2003: 184). While the focus of logistics service quality research has been on the customer side of things, the same set of parameters is also important for partners who are conducting their day-to-day business activities on platform-based solutions.

From the perspective of sellers and restaurants, logistics service quality is largely determined by efficient order handling, flawless return processing, effective delivery coordination, fluid payment transactions, and responsive platform support (Jain et al., 2021: 1; Rao et al., 2011: 167). Issues with things like making return decision, commission, shipping fees, product approval, stock updating, and notification failures might harm one of the main quality aspects of the platform from a partner point of view. Logically, partner applications cannot be considered separate from the delivery of logistics services. They are, in fact, the digital channels through which logistics and operational processes are not only experienced but also controlled (Vakulenko et al., 2019: 306).

Operational risk management literature also supports this view. Platform operations could raise risks associated with process interruptions, data delays, system failures, cost uncertainties, and coordination problems (Urrea et al., 2024: 3807). If these risks materialize within partner applications, they may be presented as unfulfilled orders, delayed responses to consumer inquiries, account inaccessibility, transaction failures, system crashes, or opaque return and commission processes. Therefore, BERTopic clusters from the partner reviews can be seen as operational risk cues that expose the shortcomings in platform governance, reverse logistics, cost sharing, information flow, and system reliability.

2.4. App Review Mining and BERTopic-Based Topic Modeling

Mobile app reviews represent a highly valuable form of user-generated feedback which not only helps point out software bugs but also uncover user needs and the aspects in which the product can be improved. Market research based on app reviews shows that massive review data have the potential to assist software developers and other stakeholders in grasping the views of users, ranking the features as per users' demands, and developing a high-quality product (Chen et al., 2014: 767; Guzman & Maalej, 2014: 153; Maalej et al., 2016: 48; Martin et al., 2017: 817).

Topic modeling methods can come in handy for extracting hidden themes from large volumes of short and unstructured review texts (Blei et al., 2003: 993). BERTopic is a topic modeling technique that, apart from the traditional topic modeling, uses sentence embeddings based on transformers, dimensionality reduction, density-based clustering, and class-based TF-IDF topic representation (Grootendorst, 2022: 1). Sentence embeddings facilitate the representation of semantically similar review texts in a more effective way while UMAP and HDBSCAN help find

meaningful clusters even in noisy and heterogeneous user-generated data (McInnes et al., 2017: 1; Reimers & Gurevych, 2019: 3982).

This research employs BERTopic for discovering explanatory topics and converting non-formatted partner reviews into problem clumps. Utilizing this methodological approach, it becomes evident that various categories of operational challenges are primarily concentrated around return and commission policies, notification quality, login and access reliability, application errors, system usability, product management, and multifaceted operational risks. Therefore, BERTopic is a medium through which user-generated textual data is linked to supply chain information systems analysis.

2.5. Theoretical Positioning of the Study

The study is in the intersection of several research areas including platform-based supply chain information systems, the success of information systems, technology acceptance, logistics service quality, operational risk management, and app review mining. The study's main theoretical proposition is that partner and seller applications must be thought of as the supply-side operational interfaces of platform-based supply chain information systems. With this view, seller, restaurant, and business partner comments are not mere application reviews but rather operational signals that indicate system quality, information quality, service quality, usability, platform governance, and supply chain process risks.

The BERTopic-discovered topical clusters serve as a method for organizing and connecting user-generated complaints to the theoretical constructs within these domains. Problems with returns, commission, and seller costs are primarily linked to platform governance, reverse logistics, cost sharing, and logistics service quality. Missing orders and customer notifications point to flaws in information quality and order management. Login, password, access difficulties, and application errors are indicators of lack of system quality and operational reliability. Usability and slowness issues are related to perceived ease of use and technology acceptance. Product upload, approval, and freezing issues are usually related to product and inventory information quality. Lastly, the outlier/mixed problem category represents diverse operational risks which are not possible to classify under a single theoretical category. Table 1 shows the theoretical linkage between the BERTopic-derived operational problem clusters and the corresponding theoretical constructs.

Table 1: Theoretical Mapping of BERTopic-Derived Operational Problem Clusters

Topic Cluster	Related Theoretical Structure	Rationale
K1: Return, Commission, and Seller Cost Problems	Platform governance, reverse logistics, cost sharing, logistics service quality	Return, commission, shipping cost, and seller loss problems are related to platform policies, reverse logistics processes, cost allocation, and perceived fairness in logistics service delivery.
K2: Missing Order and Customer Notifications	Information quality, order management, information systems success model	Missing order alerts and invisible customer messages indicate problems in the timeliness, completeness, and reliability of information flow.
K3: Login, Password, and Access Problems	System quality, access reliability, information systems success model	Login failures, password problems, and account access interruptions reflect weaknesses in system availability, access continuity, and technical reliability.
K4: Application Errors and Functional Failures	System quality, operational risk management	Application crashes, map errors, search failures, and transaction screen problems interrupt operational workflows and create process-level risks.
K5: Usability, Interface, and Application Slowness	Perceived ease of use, usability, technology acceptance model	Interface complexity, slowness, and poor navigation reduce perceived ease of use and may negatively affect continued system use.
K6: Product Upload, Approval, and Freezing Problems	Product/inventory information quality, process integration	Product upload, category, barcode, price update, and approval problems are related to the accuracy and traceability of product and inventory information.
K7: Outlier/Mixed Problem Pool	Heterogeneous operational risk pool	Mixed comments include multiple or ambiguous operational problems that cannot be assigned to a single theoretical category and therefore require separate qualitative review.

3. Methodology

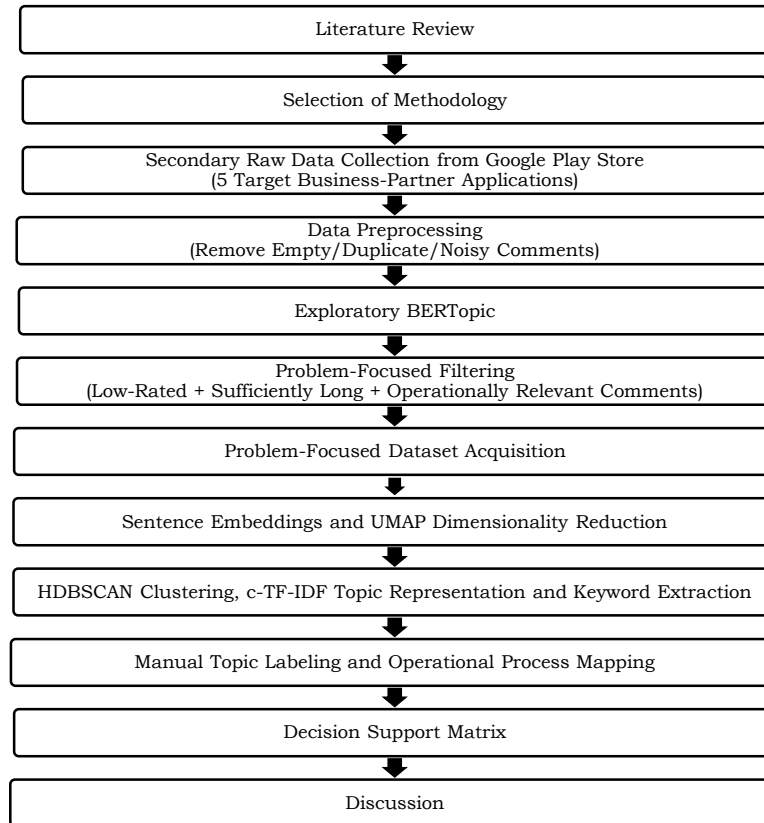
3.1. Research Design

The research presented here is an exploratory text mining study with the goal of uncovering the operational challenges of business partners using platform-based SCM ISs. The data is composed of mobile app reviews authored by users who function as sellers, restaurants, or business partners within digital platforms. The principal methodology of the research is first to extract significant sets of topics from the raw user reviews and then to analyze these sets against the supply chain and operational processes literature from a management information systems viewpoint.

BERTopic was selected for this study due to its efficacy in processing user feedback data. For instance, mobile application reviews for Supply Chain Management Information Systems (SCM IS) are typically characterized as short and unstructured texts. Consequently, this method is highly capable of extracting distinct topic clusters based on underlying semantic similarities. Besides measuring overall user satisfaction, it was also directed at problem identification of operational processes business partners issues such as ordering, inventory, returns, payments, notifications, technical access, support, and usability.

The research workflow is presented in Figure 1. As shown in the figure, Google Play Store reviews were first collected on an application basis. The collected data were then preprocessed to obtain an analyzable review dataset. In the next stage, an exploratory BERTopic model was applied to the entire preprocessed dataset. This preliminary model showed that short and general satisfaction expressions dominated the topic structure. Therefore, a problem-focused dataset was constructed in line with the objective of the study, and BERTopic was applied again to identify operational problem clusters. The resulting topic clusters were then manually labeled, mapped to operational process categories, and transformed into a decision support matrix.

Figure 1: Research Workflow



3.2. Data Source and Sample

The data source consists of user reviews of mobile applications on Google Play Store that function as business partner, seller, and restaurant panels. Three criteria were considered in application selection. First, the application had to be used directly by sellers, restaurants, businesses, or business partners. Second, it had to include functions related to supply chain and operational processes such as orders, products, inventory, returns, payments, delivery, support, or platform management. Third, it had to contain user reviews on Google Play Store and be technically suitable for data extraction.

In line with these criteria, Trendyol Go Restaurant Panel, GetirişYerim, Yemeksepeti Partner, Hepsiburada İş Ortağım, and Trendyol Seller Panel were included in the study. Getir Franchise Panel was initially evaluated but excluded since it contained no reviews. Thus, the study focuses not on consumer applications, but on the supply-side applications used by business partners and sellers in platform ecosystems, as shown in Table 2.

Table 2: Application-Based Review Counts

Application	Raw Review Count	Review Count After Preprocessing
Trendyol Go Restaurant Panel	106	104
GetirişYerim	180	177
Yemeksepeti Partner	176	176
Hepsiburada İş Ortağım	1 255	1 234
Trendyol Seller Panel	3 895	3 850
Total	5 612	5 541

Source: Created by authors according to Google Play Store data.

First, 5 612 raw user comments were collected, as seen in Table 2. After preprocessing, 5 541 comments, which are fit for analysis, were left. Trendyol Seller Panel had the maximum number of comments, suggesting that it is most likely used intensively by the sellers and, therefore, it provides a source of rich user feedback. However, the number of reviews was not evenly distributed among the applications. Therefore, simply comparing the applications based on the absolute number of reviews would not be appropriate. Additionally, the application-level results were primarily analyzed using percentage distributions to mitigate the impact of unequal review volumes. Nevertheless, the overwhelming volume of reviews originating from the Trendyol Seller Panel may have disproportionately influenced the global BERTopic model compared to applications with smaller datasets. Consequently, the integrated model results should be interpreted as reflecting macro-level operational problem patterns within the overall dataset, rather than a perfectly balanced comparative analysis across all individual platforms.

3.3. Data Collection Process

Data were retrieved from the Google Play Store on March 30, 2026, using Python (v.3.12) programming language. The dataset includes user reviews from August 6, 2020 to March 30, 2026. The *google-play-scraper* (v.1.2.7) tool was utilized to gather review text, review date, rating, and app details. The usernames fields were almost available but those were left out in the analysis and findings.

The *pandas* (v.2.2.2) library helped in structuring the data, merging application-level datasets, and turning it into a data frame ready for analysis. To store the datasets in Excel files for the next analytical stages, the *openpyxl* (v.3.2.9) library was chosen. At first, the reviews were acquired separately for each application and later merged into one dataset. Thus, a data structure was formed that not only facilitated comparisons based on the application but also allowed the execution of a single topic modeling procedure.

Only publicly available reviews on Google Play Store were used in the data collection process. No personal data analysis was conducted to identify users directly. Reviews were evaluated only at an aggregated level for text mining and topic modeling purposes. In this respect, the study is based on secondary data analysis using publicly available user feedback from digital platforms.

3.4. Data Preprocessing

After data collection, reviews were prepared for topic modeling. First, empty and missing texts were removed. Duplicate records were then eliminated, unnecessary spaces were cleaned, and review texts were standardized. Reviews consisting only of symbols, meaningless character sequences, or single words that did not contain sufficient contextual information were excluded from the dataset. Very short evaluations that could not be associated with operational processes and did not provide meaningful content for topic modeling were also excluded.

Review texts were converted to lowercase, unnecessary spaces and formatting inconsistencies were removed, and the texts were structured for BERTopic analysis. Following preprocessing, 5 541 reviews remained from the initial 5 612 raw reviews. This dataset was used in the first stage for a general BERTopic model. However, the first model showed that short, general satisfaction expressions dominated the topic clusters. Expressions such as “good,” “super,” “thanks,” “great,” and “excellent,” which did not directly represent operational processes, formed dominant topics. Accordingly, a problem-focused sub-dataset was constructed in line with the purpose of the study.

3.5. Construction of the Problem-Focused Dataset

Since the primary objective of this study was to isolate operational challenges within business partner applications, an aggregate topic model encompassing the entire review corpus was deemed insufficient. The initial BERTopic model was applied to the entire preprocessed dataset only as an exploratory preliminary analysis. In that model, general satisfaction expressions such as “good,” “super,” “thanks,” “great,” and “excellent” formed dominant clusters and reduced the visibility of supply chain and operational problem areas. Consequently, the main analyses were conducted on a problem-focused dataset rather than on the general dataset.

To explain the construction of the problem-oriented dataset, the sequential filtering phases are first delineated. Crucially, dataset refinement was not restricted solely to numerical ratings. Instead, a multi-criteria evaluation framework was employed, wherein each review was jointly assessed based on its star rating, a minimum textual length threshold, the presence of operation-specific keywords, and its overall semantic relevance. This methodology was adopted because the primary objective of this study extends beyond tracking general user engagement; rather, it aims to isolate and diagnose the distinct operational challenges encountered by sellers, restaurants, and ecosystem partners within platform-based supply chain information systems.

Table 3: Filtering and Dataset Reduction Flow

Filtering Stage	Operation / Criterion	Number of Reviews Retained	Number of Reviews Excluded	Explanation
Raw dataset	Reviews collected from five business partner/seller applications on the Google Play Store	5 612	-	Initial review dataset collected before preprocessing
Preprocessing	Removal of empty, missing, duplicate, noisy, symbolic, or contextually meaningless reviews	5 541	71	Reviews that did not provide analyzable textual content were removed
Problem-focused filtering	Selection based on rating score, minimum word count, operational keyword presence, and operational relevance of review content	2 854	2 687	Reviews directly related to operational problems were retained
Final analytical dataset	Problem-focused dataset used in the main BERTopic analysis	2 854	-	Dataset used to identify operational problem clusters

Source: Created by authors.

According to Table 3, 71 reviews have been removed from the dataset during the preprocessing step. Consequently, 5 541 reviews were left for analysis. In the next step, 2 854 reviews were selected as the problem-focused dataset, whereas 2 687 reviews were eliminated because they did not satisfy the criteria for operational problem relevance. Therefore, the ultimate BERTopic analysis was performed on 51.51% of the preprocessed dataset.

Initially, in selecting one, two, and three-star reviews as low-rated reviews for the problem-focused dataset, only ratings were not the criterion of selection. Besides, they were included only if they contained five words and operationally relevant words. Likewise, it was not the case that four- and five-star reviews were excluded automatically only because they contained operational problem statements. Consequently, a predominantly positive review could still be included in the dataset if it explicitly referenced challenges related to order notifications, product uploads, payments, commissions, returns, user authentication, technical access, or application errors.

Table 4: Inclusion and Exclusion Criteria for the Problem-Focused Dataset

Criterion Type	Operational Rule	Rationale
Inclusion criterion 1	Reviews with one, two, or three-star ratings were considered low-rated reviews.	Low ratings were assumed to indicate dissatisfaction or a negative user experience.
Inclusion criterion 2	Reviews containing at least five words were retained.	Very short texts generally did not provide sufficient contextual information for topic modeling.
Inclusion criterion 3	Reviews containing operationally relevant keywords were included.	The study focused on operational issues such as orders, stock, returns, payments, commissions, delivery, support, login, and notifications.
Inclusion criterion 4	Four- and five-star reviews were included if they contained operational problems.	Some users reported operational problems despite assigning positive ratings.
Exclusion criterion 1	Reviews containing only general satisfaction expressions were excluded.	Expressions such as “good,” “thanks,” or “excellent” did not represent operational problems.
Exclusion criterion 2	One-word or contextually unclear comments were excluded.	These comments did not provide sufficient semantic content for BERTopic analysis.
Exclusion criterion 3	Reviews consisting only of symbols, meaningless character strings, or irrelevant expressions were excluded.	Such reviews were considered noisy and unsuitable for text mining.
Exclusion criterion 4	Short reviews that could not be associated with operational processes were excluded.	The study specifically aimed to identify supply chain and platform operation-related issues.

Source: Created by authors.

The operational keyword list contained the following Turkish words: *sipariş, stok, ürün, iade, ödeme, komisyon, kurye, teslimat, destek, panel, uygulama, hata, giriş, şifre, bildirim, onay, barkod, fiyat, kategori, müşteri, mesaj*.

These keywords relate to operational aspects such as order management, inventory and product management, return and payment processes, commission and cost issues, courier and delivery operations, support services, application errors, login and password problems, notification flow, approval processes, barcode and price updates, category management, customer communication, and message visibility.

To direct the model toward operational problems, Turkish stop words as well as general satisfaction words such as “good,” “super,” “thanks,” “great,” “excellent,” and “successful” were treated as noise-generating expressions during the problem-focused analysis. These expressions were considered non-discriminatory because they did not directly represent operational processes and had dominated the first model.

Reviews rated with four or five stars were not automatically excluded if they contained operational problems. A generally positive review, for example, would be one where some of the issues related to order notifications, product uploads, payments, commissions, returns, login, or technical errors were mentioned as the negatives. If a short review just expressed one's general satisfaction without mentioning any of the problems of operation, then it was excluded.

After this filtering process, we obtained a problem-focused dataset consisting of 2 854 reviews. Consequently, the topic modeling method was aimed at operational problems rather than general user satisfaction expressions, hence providing an analysis framework that is in line with the supply chain information systems and management information systems perspective of the research.

Operational keyword lists were devised by the researchers; thus, their composition could have affected the problem-centered dataset and the subsequent topic structure. To limit the influence, the filtering did not simply use keyword matching. Along with the rating scores and minimum word count, the semantic relevance of the review content was also determined. Besides, four and five-star reviews were not entirely rejected if they held operational problem statements. For that reason, the problem-centric dataset is best regarded as a purposive analytical subset aimed at uncovering operational issues rather than a full representation of all user opinions. This weakness is also highlighted in the limitations section of the article.

3.6. Robustness Check for Keyword-Driven Filtering Bias

To address potential concerns that the predefined operational keywords might have distorted the latent topic structure, three alternative dataset configurations were deployed to verify the robustness of the results. The first version of the model took the whole preprocessed dataset as input. The second version of the model utilized only low-rated reviews, such as one, two, and three-star comments. The third version of the model worked with the final problem-focused dataset approved by rating score, minimum word count, operational keyword presence, and semantic relevance criteria.

This comparative analysis was undertaken to investigate whether the problem-focused filtering strategy inadvertently altered the underlying topic structure, or conversely, if it successfully attenuated the prevalence of generic satisfaction expressions, thereby enhancing the visibility of core operational challenge domains. Topic -1 ratio was calculated by dividing the number of reviews assigned to Topic -1 by the total number of reviews in the relevant dataset. The robustness check results are presented in Table 5.

Table 5 shows that in the whole preprocessed dataset most of the data were short and general satisfaction expressions, whereas, a low-rated-only dataset resulted in a very diverse structure with a large proportion of Topic -1. This observation implies that filtering by ratings only does not enable the discovery of stable and understandable customer complaint clusters. On the other hand, the problem-focused dataset lessened the predominance of general satisfaction expressions and resulted in more operational problem areas that are interpretative with analysis. Hence, rather than randomly guiding the model towards unrelated topics, the problem-focused filtering strategy enhanced the identification of supply-side operational issues in platform-based supply chain information systems.

Table 5: Robustness Check of BERTopic Models Across Alternative Dataset Configurations

Model	Dataset	Number of Reviews	Number of Topics	Topic -1 Ratio (%)	Dominant Topics	Interpretation
Model 1	Entire preprocessed dataset	5 541	9 meaningful topics + 1 outlier topic	2.74	General satisfaction expressions such as “teşekkür,” “mükemmel,” “teşekkürler,” “süper,” “harika,” “güzel,” and “başarılı”	This model was used as an exploratory baseline. However, short and general satisfaction expressions dominated the topic structure and reduced the visibility of operational problem areas.
Model 2	Low-rated reviews only, 1–3 stars	2 931	11 meaningful topics + 1 outlier topic	79.02	General dissatisfaction, application opening/freezing problems, commission and cargo cost complaints, return-related seller losses, product upload issues, notification/access problems, and support-related complaints	This model reduced positive satisfaction noise; however, the very high Topic -1 ratio indicated that rating-based filtering alone produced a heterogeneous and weakly structured complaint pool. Therefore, low rating alone was not sufficient to obtain stable and interpretable operational topic clusters.
Model 3	Problem-focused dataset	2 854	6 meaningful topics + 1 outlier/mixed problem pool	18.64	Return/commission/cost problems; missing order and customer notifications; login/access problems; application errors; usability; product upload/approval problems	This model produced the most interpretable operational problem structure and was therefore used as the main analytical model.

Source: Created by authors.

3.7. BERTopic Topic Modeling Process

BERTopic topic modeling was applied to the problem-focused dataset. To generate semantic representations of short Turkish user reviews, the *sentence-transformers (v.5.5.1)/paraphrase-multilingual-mpnet-base-v2* embedding model was used. This model was selected because it can capture sentence-level semantic similarities in multilingual texts and is suitable for short Turkish user reviews.

In the BERTopic process, review texts were first transformed into embedding vectors. UMAP (v.0.5.12) was then used to reduce the high-dimensional vectors into a lower-dimensional space. The UMAP parameters were set as *n_neighbors*=15, *n_components*=5, *min_dist*=0.0, and *cosine* as the distance metric. HDBSCAN (v.0.8.43) was then employed to cluster semantically similar reviews. In the HDBSCAN stage, *min_cluster_size*=25, *min_samples*=5, and *euclidean* distance were used. Finally, the c-TF-IDF approach was applied to identify the distinctive words and phrases representing each topic cluster. In the vectorization process, the *n-gram* range was set to 1-2, *min_df* to 2, and *max_df* to 0.85. Thus, both single words and meaningful two-word expressions were included in topic representation.

The choice of model components and parameter values should be reported along with the methodological justification to shed light on the BERTopic setup and make it more reproducible. To ensure the validity of the robustness check, identical parameter configurations were maintained across all dataset variations. This approach guaranteed that any observed discrepancies were attributable solely to the underlying dataset structure rather than fluctuations in parameter settings.

Table 6: BERTopic Model Parameters and Methodological Rationale

Parameter / Component	Selected Value	Methodological Rationale
Embedding model	sentence-transformers/paraphrase-multilingual-mpnet-base-v2	This model was selected because it can capture sentence-level semantic similarities in multilingual short texts and is suitable for Turkish user reviews.
UMAP n_neighbors	15	This value was used to balance local semantic similarity and broader global structure in the review embeddings.
UMAP n_components	5	A five-dimensional reduced representation was preferred to provide a compact structure suitable for density-based clustering.
UMAP min_dist	0.0	This value was selected to allow semantically similar reviews to form tighter clusters.
UMAP metric	cosine	Cosine distance was used because it is commonly preferred for measuring semantic similarity between text embeddings.
HDBSCAN min_cluster_size	25	This value was selected to reduce very small and difficult-to-interpret clusters and to obtain more meaningful topic groups.
HDBSCAN min_samples	5	This value was used to provide a flexible clustering structure while still controlling noise in short and heterogeneous review texts.
HDBSCAN metric	euclidean	Euclidean distance was used in the reduced UMAP space for density-based clustering.
Vectorizer ngram_range	1-2	Both single words and meaningful two-word expressions were included in topic representation.
Vectorizer min_df	2	Very rare terms appearing in only one review were excluded to reduce noise.
Vectorizer max_df	0.85	Very frequent and overly general terms were limited to prevent them from dominating topic representations.
Topic representation	c-TF-IDF	Class-based TF-IDF was used to identify the most distinctive words and phrases representing each topic cluster.

Source: Created by authors.

This parameter configuration was adopted to obtain interpretable topic clusters from short, noisy, and unstructured Turkish review texts. In particular, the selected UMAP and HDBSCAN settings were intended to balance semantic compactness, cluster interpretability, and noise control. The vectorizer settings were used to capture both single-word and two-word operational expressions while reducing the influence of very rare or overly general terms.

Additional model validation indicators were calculated for the three dataset configurations to strengthen the methodological validity of the BERTopic results. These indicators included the number of meaningful topics, Topic -1 ratio, average topic size, topic coherence, topic diversity, and topic stability. Topic -1 ratio was calculated by dividing the number of reviews assigned to the outlier topic by the total number of reviews in the relevant dataset. Average topic size was calculated by dividing the number of non-outlier reviews by the number of meaningful topics. Topic coherence was calculated using the c_npmi metric based on topic keywords and tokenized review texts. Topic diversity was calculated as the ratio of unique topic words to the total number of topic words. Topic stability was calculated by comparing the similarity of topic keyword sets across repeated random-state configurations using average Jaccard similarity.

As shown in Table 7, the three BERTopic configurations produced different levels of interpretability. Model 1 had a low Topic -1 ratio and moderate stability; however, its topic structure

was dominated by short and general satisfaction expressions rather than operational problem areas. Model 2 focused only on low-rated reviews and produced a moderate topic stability value, but its very high Topic -1 ratio showed that rating-based filtering alone generated a heterogeneous and weakly structured complaint pool. Model 3, based on the problem-focused dataset, produced a more interpretable operational topic structure with six meaningful topics, a moderate Topic -1 ratio, and the highest topic diversity value. Therefore, the validation metrics support the use of the problem-focused dataset as the main analytical model.

Table 7: Model Validation Metrics Across BERTopic Configurations

Model	Dataset	Number of Meaningful Topics	Topic -1 Ratio (%)	Average Topic Size	Topic Coherence (c_npmi)	Topic Diversity	Topic Stability	Evaluation
Model 1	Entire preprocessed dataset	9	2.74	598.78	-0.499	0.732	0.664	Although the Topic -1 ratio was low, the topic structure was dominated by short and general satisfaction expressions, which reduced the visibility of operational issue areas.
Model 2	Low-rated reviews only, 1-3 stars	11	79.02	55.91	-0.285	0.791	0.662	The very high Topic -1 ratio indicates that rating-based filtering alone produced a heterogeneous complaint pool. Although this model produced acceptable topic diversity and moderate topic stability, the high outlier ratio shows that low-rating-based filtering alone did not provide a sufficiently interpretable operational topic structure.
Model 3	Problem-focused dataset	6	18.64	387.00	-0.612	0.950	0.628	This model produced the most interpretable operational problem structure by reducing general satisfaction noise and making supply-side operational issues more visible.

Source: Created by authors.

Note: Topic coherence was calculated using the c_npmi metric based on topic keywords and tokenized review texts. Topic diversity was calculated as the ratio of unique topic words to the total number of topic words. Topic stability was calculated by comparing the similarity of topic keyword sets across repeated random-state configurations using average Jaccard similarity.

This parameter structure was adopted to obtain meaningful and interpretable topic clusters from short and unstructured Turkish user reviews. The model produced six meaningful topic clusters and one outlier cluster. Topics were examined in terms of review count, average rating, negative review ratio, and representative reviews. Accordingly, each topic was evaluated not only as a label but also in terms of its effect on business partner experience.

3.8. Topic Labeling and Decision Support Matrix

Automatically generated BERTopic labels were not used directly as final findings. Instead, each topic cluster was systematically reviewed by considering topic keywords, representative reviews, average rating values, negative review ratios, and the contextual meaning of the review texts. Final topic labels were determined through a two-stage review process and aligned with the management information systems and supply chain management perspectives of the study. Thus, topic labels were based not only on lexical similarity, but also on the operational process and problem domain represented by the relevant comments.

Manual topic labeling was conducted through an independent dual-coding protocol to strengthen the reliability of the qualitative interpretation of BERTopic outputs. For this purpose, a representative subset of 170 reviews was selected from the problem-focused dataset. This subset included 20 representative reviews from each of the six meaningful topic clusters and 50 reviews from the outlier/mixed problem pool because of its heterogeneous structure. The selected reviews were independently coded at two levels: BERTopic-derived topic categories (K1-K7) and operational

process categories (S1–S6). Agreement between the two independent coding outputs was calculated using Cohen’s Kappa. Disagreements between the two coding outputs were reviewed and resolved through consensus before the final topic labels and operational process mappings were reported.

After the independent dual-coding process, disagreements were reviewed and resolved through consensus. The manual labeling process produced the following final topic labels: return, commission, and seller cost problems; missing order and customer notifications; login, password, and access problems; application errors and functional failures; usability, interface, and application slowness; product upload, approval, and freezing problems; and outlier/mixed problem pool. These topic headings were then associated with operational process categories such as payment and reverse logistics, information quality and order management, technical system performance, usability, stock/product management, and mixed operational risk.

User ratings were also included in the analysis to evaluate the effect of topic clusters on business partner experience. One- and two-star reviews were classified as negative, three-star reviews as neutral, and four- and five-star reviews as positive. The negative review ratio was calculated by dividing the number of one- and two-star reviews in each topic cluster by the total number of reviews in that cluster. Thus, each topic was evaluated not only by review count but also by the level of user dissatisfaction.

In BERTopic, Topic -1 represents reviews that cannot be assigned strongly to a specific density-based cluster during HDBSCAN clustering. Therefore, Topic -1 was not interpreted at the same level as the main topic clusters. It was reported separately as an outlier/mixed problem pool. This approach made it possible to retain operationally relevant comments in the analysis without distorting the interpretation of the main topic findings.

Since Topic -1 represented a relatively large outlier/mixed problem pool, an additional thematic sub-analysis was conducted for K7. In this process, all 532 reviews assigned to K7 were reviewed separately and coded according to their dominant operational problem theme. Each review was assigned to only one subcategory based on the most salient problem cue in the text. When a review contained multiple problem statements, the dominant operational issue was considered for coding. This procedure was used to avoid treating K7 merely as noise and to reveal the heterogeneous operational risk structure within the outlier/mixed problem pool.

In the final stage, topic clusters were transformed into a decision support matrix using a rule-based prioritization logic. For each topic, the matrix included review count, negative review ratio, average rating, operational process category, priority level, and managerial recommendation. Priority levels were determined by considering review count and negative review ratio together. These threshold values were defined by the researchers as an analytical prioritization rule to transform BERTopic-derived topic clusters into actionable managerial categories. Topics with at least 100 reviews and a negative review ratio of at least 70% were classified as “Urgent Intervention.” Topics with at least 50 reviews and a negative review ratio of at least 60% were classified as “High Priority.” Topics with at least 25 reviews and a negative review ratio of at least 50% were classified as “Medium Priority.” Topic -1, represented by K7, was classified as “Separate Review” because it reflected a heterogeneous operational risk pool rather than a homogeneous topic cluster. Consequently, the insights generated by BERTopic extended beyond mere descriptive topic clustering; instead, they were operationalized into a process-improvement-oriented decision support mechanism tailored for platform administrators and strategic decision-makers.

4. Findings

4.1. General Overview of the Problem-Focused Dataset

As a result of the filtering process described in the methodology section, 2 854 of the 5 541 preprocessed reviews were included in the problem-focused dataset. Table 8 presents the application-based distribution of this dataset.

Table 8: Application-Based Distribution of the Problem-Focused Dataset

Code	Application	Review Count After Preprocessing	Problem-Focused Review Count	Problem-Focused Review Ratio (%)
U1	Trendyol Go Restaurant Panel	104	46	44.23
U2	GetirİşYerim	177	101	57.06
U3	Yemeksepeti Partner	176	140	79.55
U4	Hepsiburada İş Ortağım	1 234	490	39.71
U5	Trendyol Seller Panel	3 850	2 077	53.95
Total		5 541	2 854	51.51

Source: Created by authors.

Note: U1-U5 denote the five business partner applications included in the study. U1 represents Trendyol Go Restaurant Panel, U2 represents GetirİşYerim, U3 represents Yemeksepeti Partner, U4 represents Hepsiburada İş Ortağım, and U5 represents Trendyol Seller Panel.

The distribution of the problem-focused dataset based on the application also reveals an unbalanced structure. Trendyol Seller Panel alone was responsible for 2 077 out of the 2 854 problem-oriented reviews, while the other applications combined contributed fewer review volumes. Therefore, we interpreted the application-based findings by looking at not only the review numbers but also the proportional distributions. By doing this, we minimized the risk of giving too much importance to the applications with larger review volumes. Nevertheless, the overwhelming presence of Trendyol Seller Panel in the dataset must be kept in mind when looking at the overall topic model results.

Table 8 lays out the breakdown of application-based data and application codes. Trendyol Go Restaurant Panel is referred to as U1, GetirİşYerim as U2, Yemeksepeti Partner as U3, Hepsiburada İş Ortağım as U4, and Trendyol Seller Panel as U5. A problem-focused dataset makes up 51.51% of all preprocessed reviews. This number suggests that a large percentage of comments on business partner and seller applications highlight operational issues. The highest proportion of problem-focused reviews was found in the U3-coded Yemeksepeti Partner application (79.55%), and the second was U2-coded GetirİşYerim (57.06%) and third was U5-coded Trendyol Seller Panel (53.95%). Even though the problem-focused review ratio for U4-coded Hepsiburada İş Ortağım was lower (39.71%), due to the large number of reviews, this application also became a very important data source for analysis.

4.2. Results of BERTopic Topic Model

The BERTopic analysis conducted on the problem-focused dataset produced six meaningful topic clusters and one outlier/mixed problem pool. Table 9 presents the final topic names and topic codes. Return, commission, and seller cost problems are coded as K1; missing order and customer notifications as K2; login, password, and access problems as K3; application errors and functional failures as K4; usability, interface, and application slowness as K5; product upload, approval, and freezing problems as K6; and the outlier/mixed problem pool as K7.

Table 9: Problem-Focused BERTopic Topic Model Results

Code	BERTopic Topic Number	Topic Name	Review Count	Ratio in Dataset (%)	Average Rating	Negative Review Ratio (%)
K1	0	Return, Commission, and Seller Cost Problems	964	33.78	1.68	82.16
K2	1	Missing Order and Customer Notifications	589	20.64	1.87	76.57
K3	2	Login, Password, and Access Problems	557	19.52	1.86	75.58
K4	3	Application Errors and Functional Failures	103	3.61	2.16	67.96
K5	4	Usability, Interface, and Application Slowness	77	2.70	2.79	51.95
K6	5	Product Upload, Approval, and Freezing Problems	32	1.12	1.66	81.25
K7	-1	Outlier/Mixed Problem Pool	532	18.64	1.81	77.26

Source: Created by authors.

Note: K1-K7 represent BERTopic-derived operational problem clusters. K1-K6 refer to six meaningful topic clusters, whereas K7 represents Topic -1, namely the outlier/mixed problem pool generated by HDBSCAN.

Representative anonymized review excerpts were examined for each topic to improve the interpretability and credibility of the BERTopic-derived topic clusters. These excerpts were not used to identify users or applications; instead, they were paraphrased and anonymized to illustrate the dominant operational problem patterns represented by each topic cluster. Table 10 presents two representative anonymized review examples for each topic code.

Table 10: Representative Anonymized Review Excerpts for BERTopic-Derived Topic Clusters

Topic Code	Topic Name	Representative Anonymized Review Excerpts	Analytical Comment
K1	Return, Commission, and Seller Cost Problems	Example 1: "The return is accepted automatically, the shipping cost is charged to the seller, and commission deductions continue even when the seller experiences a loss." Example 2: "Commission and cargo costs are too high; after cancellations or returns, the financial burden still remains on the seller."	These examples show that K1 is mainly associated with seller-side cost pressure, return-related losses, commission deductions, and perceived unfairness in platform policies.
K2	Missing Order and Customer Notifications	Example 1: "Order notifications do not arrive on time, and the seller notices the order only after it is delayed." Example 2: "Customer messages are not visible in the panel, so the seller cannot respond to customers quickly."	These examples indicate information quality problems in order visibility, customer communication, and notification infrastructure.
K3	Login, Password, and Access Problems	Example 1: "The seller cannot log in to the account even though the password is correct, and the system repeatedly asks for verification." Example 2: "The application logs the user out frequently, making it difficult to access the seller panel during order processing."	These examples reflect system quality and access reliability problems that may interrupt daily operational continuity.
K4	Application Errors and Functional Failures	Example 1: "The application closes suddenly when the order screen is opened, preventing the seller from completing the transaction." Example 2: "Map, search, or transaction screens give errors and the process cannot be completed properly."	These examples show that K4 is related to functional failures, screen errors, transaction interruptions, and technical instability.
K5	Usability, Interface, and Application Slowness	Example 1: "The application is very slow, and screen transitions take too long during routine tasks." Example 2: "The interface is confusing, and reaching basic functions requires too many steps."	These examples indicate usability problems that may increase task completion time and reduce perceived ease of use.
K6	Product Upload, Approval, and Freezing Problems	Example 1: "Product upload and approval take too long; category or barcode information is rejected even when entered correctly." Example 2: "Price and stock updates do not appear correctly, and the product listing process freezes during updates."	These examples show that K6 is associated with product information quality, catalog management, stock visibility, and listing process problems.
K7	Outlier/Mixed Problem Pool	Example 1: "The review combines return costs, panel access problems, support delays, and notification problems in the same comment." Example 2: "The user expresses general dissatisfaction with several overlapping problems that cannot be assigned to a single dominant issue."	These examples indicate that K7 includes heterogeneous and overlapping operational problems that require separate qualitative review rather than direct interpretation as a homogeneous topic cluster.

Source: Created by authors.

Note: K1–K7 represent BERTopic-derived operational problem clusters. The review excerpts were paraphrased and anonymized to avoid disclosing usernames, store names, personal information, or offensive expressions.

Table 10 shows that each cluster of BERTopic-derived topics is grounded by representative anonymized review excerpts. Such examples help to explain the semantic as well as the functional basis of the topic labels. Financial and policy-related issues are what K1 has in common. K2 is about notification and information flow problems. K3 and K4 focus on system quality and technical reliability issues. K5 conveys usability and interface problems. K6 deals with product, stock, and catalog management. Lastly, K7 comprises assorted comments featuring several overlapping operational problems.

The six primary topics account for 2 322 reviews, whereas 532 reviews were placed in the outlier/mixed problem group labeled as K7. K7 is equivalent to Topic-1 in the BERTopic framework and refers to reviews that could not be grouped into a single topic cluster. Hence, K7 was not evaluated at the same depth as the major topic clusters. On the other hand, since K7 constituted 18.64% of the problem-oriented dataset, it was not simply discarded as an unaccounted residual

category. Rather, separate thematic sub-analysis was carried out to explore the diverse problem structure within this pool.

As shown in Table 9, the major source of problems appears to be K1. This segment comprises 964 customer reviews or 33.78% of the entire set of problem-focused discussions. It has the lowest average rating (1.68) and a significantly high percentage of negative reviews (82.16%). This implies that problems with returns, commissions, shipping costs, and seller losses are not simply monetary troubles but extended to issues of reverse logistics, cost sharing, platform governance, and transparency of the process. K2 is the second biggest section, with 589 reviews, making up 20.64% of the whole. Its average score of 1.87 and percentage of negative comments of 76.57% reveal a high degree of displeasure with the notification system and information dissemination during order handling. K3 is next with 557 customer feedback, 19.52% share, 1.86 average rating, 75.58% negative review ratio, bringing into focus the major problems of accessing, setting passwords, and logging. Even though K4, K5, and K6 represent smaller segments, their outcomes still reveal some significant operational vulnerabilities, particularly K6, which shows an extremely high negative review ratio of 81.25%.

Two coders who were independent of each other were asked to code a representative sample of 170 reviews to determine the reliability of manual topic classification and operational process mapping. This sample was composed of 20 reviews from each of the six significant topic clusters and 50 reviews from the outlier/mixed problem pool. For both topic-level tagging and operational process mapping, Cohen's Kappa was derived. The findings are summarized in Table 11.

Table 11: Intercooder Reliability Results for Manual Topic Labeling and Operational Process Mapping

Coding Level	Number of Reviews	Number of Coding Outputs	Reliability Measure	Value	Interpretation
Topic labeling, K1-K7	170	2	Cohen's Kappa	0.704	Substantial Agreement
Operational process mapping, S1-S6	170	2	Cohen's Kappa	0.693	Substantial Agreement

Source: Created by authors.

Note: K1-K7 represent BERTopic-derived operational problem clusters, whereas S1-S6 represent operational process categories used in the process mapping stage.

As shown in Table 11, Cohen's Kappa was 0.704 for topic-level labeling and 0.693 for operational process mapping. These values indicate a substantial level of agreement between the two coding outputs. After the reliability calculation, disagreements were reviewed and resolved through consensus. Therefore, the final topic labels and operational process categories were determined through both independent coding and consensus-based validation.

4.3. Sub-analysis of the Outlier/Mixed Problem Pool (K7)

Since K7 represented a relatively large share of the problem-focused dataset, a separate thematic sub-analysis was conducted for this outlier/mixed problem pool. The aim of this sub-analysis was not to reinterpret K7 as a homogeneous BERTopic cluster, but to identify the dominant operational problem patterns within this heterogeneous category. For this purpose, all 532 reviews assigned to K7 were reviewed and coded according to their dominant problem theme. The results are presented in Table 12.

As shown in Table 12, K7 should not be interpreted as random noise only. Although it was generated as Topic -1 by the HDBSCAN clustering process, the thematic sub-analysis shows that this pool contains several operationally meaningful problem patterns. The largest subcategories within K7 are platform governance, return, commission, and payment policy problems (23.87%), technical update, freezing, language, and performance problems (22.37%), and product, menu,

stock, and catalog management problems (20.11%). These findings indicate that K7 represents a heterogeneous operational risk pool rather than a meaningless residual category.

Table 12: Thematic Sub-analysis of the Outlier/Mixed Problem Pool (K7)

K7 Subcategory	Subcategory Name	Review Count	Ratio within K7 (%)	Representative Anonymized Review Excerpt	Analytical Comment
K7a	Platform governance, return, commission, and payment policy	127	23.87	"The seller reports high commission rates, delayed payments, return-related costs, and unfair cost allocation."	This subcategory indicates that a considerable part of K7 is related to platform governance, cost sharing, return policies, payment timing, and perceived fairness.
K7b	Product, menu, stock, and catalog management	107	20.11	"The seller states that product upload, menu editing, stock updates, price changes, or catalog approval processes are difficult or unstable."	These reviews point to operational problems in product information management, catalog control, stock visibility, and sales preparation processes.
K7c	Notification, customer communication, and order visibility	64	12.03	"The review states that order notifications or customer messages are not received reliably, causing missed orders or delayed responses."	This subcategory reflects information quality problems related to order visibility, notification reliability, and customer communication.
K7d	Account, panel access, and authentication	39	7.33	"The user reports repeated login problems, password prompts, account access issues, or difficulties using the seller panel."	These comments indicate access reliability problems and uncertainty in account, authentication, and panel-use processes.
K7e	Delivery, courier, and order timing	14	2.63	"The review describes late order visibility, courier coordination problems, delivery timing issues, or inaccurate delivery estimates."	This subcategory shows that some K7 reviews are related to delivery coordination, courier processes, and timing problems in order fulfillment.
K7f	Support and customer service	18	3.38	"The seller states that support channels are unresponsive, provide irrelevant answers, or do not resolve operational problems."	These reviews point to service quality weaknesses in support, problem resolution, and platform responsiveness.
K7g	Technical update, freezing, language, and performance problems	119	22.37	"The user reports that after updates the application freezes, gives errors, changes language settings, or becomes difficult to use."	This subcategory indicates technical instability, update-related disruption, language/interface problems, and general performance issues.
K7h	Unclassifiable or highly mixed expressions	44	8.27	"The review contains general dissatisfaction or multiple overlapping complaints that cannot be assigned to a single dominant problem theme."	These comments remain heterogeneous and require further qualitative review if a deeper diagnostic analysis is needed.
Total		532	100.00		

Source: Created by authors.

Note: K7 corresponds to Topic -1 in the BERTopic model. K7a–K7h represent thematic subcategories identified through the separate qualitative sub-analysis of the outlier/mixed problem pool.

This additional analysis also supports the decision not to include K7 in the main topic prioritization at the same level as K1–K6. While K1–K6 represent more homogeneous BERTopic-derived clusters, K7 contains multiple overlapping and ambiguous problem areas. Therefore, it should be evaluated as a separate qualitative review category for platform managers. In practical terms, the K7 sub-analysis suggests that platform administrators should periodically review outlier comments because they may contain early signals of emerging operational problems that are not yet strong enough to form a distinct topic cluster.

4.4. Mapping BERTopic-Derived Operational Problem Clusters to Operational Process Categories

In this phase of the study, the BERTopic-derived operational problem clusters (K1–K7) were systematically mapped onto predefined operational process categories (S1–S6). This cross-referencing was executed to enhance the interpretability of the topic modeling outcomes through the complementary lenses of management information systems (MIS) and supply chain management (SCM). K1 was associated with payment, commission, and reverse logistics processes; K2 with notification and information quality in order management; K3 and K4 with technical system performance; K5 with usability; K6 with stock/product management and technical performance; and K7 with the outlier/mixed problem pool. This mapping is presented in Table 13.

Table 13: Mapping BERTopic-Derived Operational Problem Clusters to Operational Process Categories

Topic Code	Topic Name	Process Code	Operational Process Category
K1	Return, Commission, and Seller Cost Problems	S1	Payment, Commission, and Reverse Logistics
K2	Missing Order and Customer Notifications	S2	Notification and Information Quality / Order Management
K3	Login, Password, and Access Problems	S3	Technical System Performance
K4	Application Errors and Functional Failures	S3	Technical System Performance
K5	Usability, Interface, and Application Slowness	S4	Usability
K6	Product Upload, Approval, and Freezing Problems	S5	Stock/Product Management and Technical Performance
K7	Outlier/Mixed Problem Pool	S6	Outlier/Mixed Problem Pool

Source: Created by authors.

Note: K1–K7 represent BERTopic-derived operational problem clusters, whereas S1–S6 represent operational process categories. K7 and S6 refer to the outlier/mixed problem pool that requires separate qualitative review.

In line with this matching in Table 13, the raw issue tags generated by the model have been placed within an interpretable framework consistent with the logistics and information systems literature. The categories K1 (cost-focused issues in payment, commission, and reverse logistics processes), K2 (information quality problems related to orders and customer notifications), K3 (login, password, and access problems), K4 (application errors and functional failures), K5 (interface, usability, and slowness problems), and K6 (product loading, approval, and freezing issues) are clearly defined. K7 represents a mixed and anomalous pool of problems that cannot be clearly assigned to a specific issue.

4.5. Application-Based Distribution of BERTopic-Derived Operational Problem Clusters

Table 14 summarizes the distribution of BERTopic-derived operational problem clusters by application. Here, U1–U5 are the five business partner applications, and K1–K7 are the BERTopic-derived operational problem clusters. This coding scheme was adopted to enhance visual readability and facilitate application-based comparative analysis.

Table 14: Application-Based Distribution of BERTopic-Derived Operational Problem Clusters (%)

Topic Code	U1	U2	U3	U4	U5
K1	26.09	9.90	22.86	20.00	39.09
K2	21.74	34.65	24.29	24.29	18.83
K3	21.74	19.80	14.29	26.94	18.05
K4	4.35	5.94	3.57	5.51	3.03
K5	8.70	4.95	2.14	4.69	2.02
K6	2.17	0.99	0.00	0.20	1.40
K7	15.22	23.76	32.86	18.37	17.57

Source: Created by authors.

Note: U1–U5 denote the five business partner applications included in the study. K1–K7 denote BERTopic-derived operational problem clusters. Percentages show the distribution of each operational problem cluster within the relevant application.

Distributions of reports through application-based channels reveal that most of the GetirİşYerim (U2) users are pointing out K2 type issues (34.65%), which implies order process and notification flow related problems. For Hepsiburada İş Ortağım (U4), K3 comes as the leading issue category with a frequency of 26.94%, which means that technical access and account management are the areas requiring immediate attention. Trendyol Go (U1) has K1 (26.09%) and K3 (21.74%) at almost the same level, which signifies that financial and technical matters are equally important for the restaurants. The Trendyol Seller Panel (U5) is mainly characterized by K1 (39.09%), which reflects sellers' worries about commissions, shipping costs, and return-related losses. Further, the high K7 percentage in Yemeksepeti Partner (U3) (32.86%) indicates a combination of problems that need a deep manual analysis.

4.6. Application-Based Distribution of Operational Process Categories

Table 15 provides a process-level perspective by showing the application-based distribution of operational process categories. In this table, U1–U5 represent the five business partner applications, while S1–S6 represent the operational process categories derived from the mapping of BERTopic topic clusters. This structure makes it possible to evaluate each application not only according to topic clusters, but also according to broader managerial process categories.

Table 15: Application-Based Distribution of Operational Process Categories (%)

Application Code	S1	S2	S3	S4	S5	S6
U1	26.09	21.74	26.09	8.70	2.17	15.22
U2	9.90	34.65	25.74	4.95	0.99	23.76
U3	22.86	24.29	17.86	2.14	0.00	32.86
U4	20.00	24.29	32.45	4.69	0.20	18.37
U5	39.09	18.83	21.09	2.02	1.40	17.57

Source: Created by authors.

Note: U1–U5 denote the five business partner applications included in the study, whereas S1–S6 denote operational process categories. S1 refers to payment, commission, and reverse logistics; S2 to notification and information quality/order management; S3 to technical system performance; S4 to usability; S5 to stock/product management and technical performance; and S6 to the outlier/mixed problem pool.

As illustrated in Table 15, the primary operational challenges vary significantly across the analyzed applications. For example, Trendyol Seller Panel (U5) mainly suffers from S1 (39.09%), which relates to financial policy and reverse logistics issues. On the other hand, Hepsiburada İş Ortağım (U4) is mainly affected by S3 (32.45%), which is all about system stability. Meanwhile, the major blockage in GetirİşYerim (U2) is S2 (34.65%) that covers order and notifications problems. S1 and S3 have the same weight in Trendyol Go (U1) (26.09%). A high S6 percentage in Yemeksepeti Partner (U3) (32.86%) is a hint of a compound problem structure.

4.7. Decision Support Matrix

The BERTopic outputs were transformed into a decision support matrix to make the analytical results more functional and to provide platform administrators with a concrete roadmap. Also, the matrix was designed to support managerial prioritization by considering both how frequently each topic was discussed and the level of user dissatisfaction reflected in negative review ratios and average ratings. This approach enabled the identification of operational areas where improvement actions should be treated as urgent, high-priority, medium-priority, or subject to separate qualitative reviews.

In this study, priority levels were produced through a rule-based prioritization logic. This logic was based on two indicators: review count and negative review ratio. Review count was used to represent the visibility and frequency of each operational problem cluster, while negative review ratio was used to represent the intensity of dissatisfaction associated with that cluster. The threshold values used in the decision support matrix were defined by the researchers as an analytical prioritization rule. Consequently, these thresholds should not be construed as universal

benchmarks, but rather as study-specific decision rules designed to translate BERTopic-derived topic clusters into actionable managerial categories.

The prioritization logic was structured as follows: topics with at least 100 reviews and a negative review ratio of at least 70% were classified as “*Urgent Intervention*”; topics with at least 50 reviews and a negative review ratio of at least 60% were classified as “*High Priority*”; topics with at least 25 reviews and a negative review ratio of at least 50% were classified as “*Medium Priority*”; and Topic -1, represented by K7, was classified as “*Separate Review*” because it reflected a heterogeneous operational risk pool rather than a homogeneous topic cluster. Based on this approach, Table 16 presents the priority level, rule-based prioritization logic, operational effects and MIS/SCM implications, and managerial recommendations for each topic code.

Table 16: Decision Support Matrix for BERTopic-Derived Operational Problem Clusters

Topic Code	Short Topic Name	Priority Level	Rule-Based Prioritization Logic	Operational Effects and MIS/SCM Implications	Managerial Recommendation
K1	Return / commission / cost	Urgent Intervention	Review count ≥ 100 and negative review ratio $\geq 70\%$	Directly affects seller cost, reverse logistics burden, platform trust, cost sharing, and process transparency.	Return decision mechanisms, commissions, and shipping costs should be redesigned to reduce seller losses and improve perceived fairness.
K2	Order / notification	Urgent Intervention	Review count ≥ 100 and negative review ratio $\geq 70\%$	Creates the risk of missed orders, delayed responses to customer questions, service delays, revenue loss, and reduced supply chain visibility.	Order and customer notification infrastructure should be strengthened through real-time alerts and notification-permission control mechanisms.
K3	Login / access	Urgent Intervention	Review count ≥ 100 and negative review ratio $\geq 70\%$	Access interruptions threaten order management, system availability, account continuity, and the continuity of seller operations.	Login, password verification, session continuity, and account access processes should be improved as a priority.
K4	Application errors	High Priority	Review count ≥ 50 and negative review ratio $\geq 60\%$	Interrupts workflows and creates operational failures in order, map, search, and transaction screens.	Error logs and map, search, order-screen, and transaction-flow failures should be monitored regularly.
K5	Interface / slowness	Medium Priority	Review count ≥ 25 and negative review ratio $\geq 50\%$	Affects user experience, task completion time, perceived ease of use, and the ability of partners to use the application efficiently.	Interface, speed, screen flow, and mobile usability improvements should be redesigned based on user tasks.
K6	Product upload / approval	Medium Priority	Review count ≥ 25 and negative review ratio $\geq 50\%$	Disrupts product listing, stock/price updates, catalog management, data accuracy, and sales-preparation processes.	Product upload, approval, category, barcode, and price-update processes should be made more stable and traceable.
K7	Mixed / outlier operational risks	Separate Review	Topic -1 / Separate qualitative review	Represents a heterogeneous operational risk pool including governance and cost issues, technical update problems, product/catalog management problems, notification and communication issues, account access problems, delivery-related concerns, support weaknesses, and highly mixed expressions.	K7 should be periodically reviewed through qualitative analysis instead of being directly prioritized with the main BERTopic clusters. Outlier comments should be monitored as possible early signals of emerging operational risks.

Source: Created by authors.

Note: K1–K7 represent BERTopic-derived operational problem clusters. Priority levels were determined through a researcher-defined rule-based prioritization logic based on review count and negative review ratio. K7 was classified as “*Separate Review*” because it represents Topic -1, namely a heterogeneous outlier/mixed problem pool rather than a homogeneous topic cluster.

As shown in Table 16, the classification of topics K1, K2, and K3 under the 'Urgent Intervention' category is attributable to their fulfillment of both rule-based prioritization thresholds: a review volume of ≥ 100 and a negative review ratio of $\geq 70\%$. The problem of return, commission, and seller cost, K1 was identified as a main problem that requires immediate attention because it affects seller cost burden, reverse logistics processes, platform trust, cost sharing, and process

transparency. Missed customer and order notifications that may lead to missing orders, delayed responses to customer questions, service delays, revenue losses and reduced supply chain visibility, that is why K2 was considered as an urgent task to be done. Likewise, K3 needs immediate action because it raises issues of login, password, and access that can disrupt system availability, account continuity, order management, and the continuity of seller operations.

K4 was characterized as a "*High Priority*" problem area since it satisfied the rule-based criteria of review count ≥ 50 and negative review ratio $\geq 60\%$. Application errors and functional failures disrupt operational workflows and generate process-level risks in order, map, search, and transaction screens. Therefore, K4 should be kept under surveillance periodically via error logs, application performance indicators, and transaction-flow controls.

K5 and K6 were labeled "*Medium Priority*" problem areas, as they fulfilled the criteria of having review count ≥ 25 and negative review ratio $\geq 50\%$. K5 is mostly concerned with the issues related to interface, speed, navigation, usability, and perceived ease of use. Even though the number of its reviews is less than that of K1, K2, and K3, it can potentially increase the time for task completion and decrease the efficiency of everyday partner operations. K6 is linked with issues concerning product upload, approval, category, barcode, stock, price-update, catalog management, and data accuracy. These problems may cause a breakdown in product listing, sales preparation, inventory visibility, and platform-based product information management.

K7 was considered a "*Separate Review*" instead of a priority of urgent, high or medium as it is Topic -1 i.e. a heterogeneous outlier/mixed problem pool rather than a homogeneous BERTopic-derived topic cluster. The separate K7 sub-analysis revealed that this pool consists of issues related to governance and costs, technical update problems, product/catalog management problems, notification and communication issues, account access problems, delivery-related concerns, support weaknesses, and highly mixed expressions. Hence, K7 should be periodically reviewed via qualitative analysis and kept under observation as a possible source of early signs for emerging operational risks.

Overall, the decision support matrix takes a set of clustered topics on operational problems derived from BERTopic and turns them into managerial priorities that can be acted upon. The matrix employs a straightforward, yet highly systematic prioritization method based on a rule that combines review volume and negative review ratios. This approach provides a transparent framework for determining which operational problems require immediate intervention, which should be monitored at a high or medium priority level, and which warrant further analysis through a separate qualitative review. Hence, the matrix both enhances the descriptive capability of topic modeling and offers a handy decision support instrument for platform-based supply chain information systems.

5. Discussion

5.1. Interpretation of Key Operational Problem Areas

In this research, the authors investigated platform-based supply chain information systems' operational issues through the analysis of Google Play Store mobile business partner and seller applications reviews and the BERTopic topic modeling technique. The problem-oriented BERTopic analysis identified six significant problem-related topic groups and one outlier/mixed problem group. The results indicate that business partner applications issues are not solely related to technical application errors but also span various operational processes such as payment, commission, return, order notification, access, system performance, usability, product management, and platform governance.

The general results reveal that partner apps should not be treated as mere mobile front-ends anymore. Rather, it is more appropriate to regard them as crucial parts of platform-based supply

chain information systems. Such apps direct how business partners handle order processing, costing, returns, customer communication, technical access, product data, stock updating, and platform's daily operation. Hence, the operational challenges identified through application reviews serve as vital supply-side indicators of information systems quality, logistics service quality, technology acceptance, platform governance, and operational risk.

K1, return, commission, and seller cost problems, is obviously the most significant issue area in the study. In theory, this cluster is deeply connected with platform governance, reverse logistics, cost sharing, seller trust, and transparency. Return decisions, commission deductions, shipping cost allocations, and seller loss complaints suggest that platform policies are not only perceived as financial rules but also as governance mechanisms that shape fairness, trust, and operational sustainability. From the perspective of logistics service quality, return and cost-related problems can degrade the perceived reliability of the platform. From the perspective of operational risk, unclear cost allocation and return procedures may lead to financial ambiguity for sellers and a decrease in their trust in the platform ecosystem.

The component K2, which deals with missing orders and customer notifications, is tightly linked with the quality of information, order management, notification infrastructure, and supply chain visibility. In supply chains that use platform-based information systems, it is very important that order fulfillment and customer communication rely on information that flows in a timely, accurate, complete, and highly visible manner. If the system fails to generate order alerts or customer messages remain invisible, it could lead to missed orders, delayed response times, delivery interruptions, and eventually loss of revenue. Therefore, K2 is the tip of the iceberg as far as the notification side of the problem goes. Additionally, the entire supply chain might get impacted because of an overall lack of good quality information, which in turn might hamper supply chain visibility and coordination among the sellers, customers, and the platform systems.

K3 and K4 issues such as login/access problems and application errors are connected to the quality of system, accessibility, technical reliability and operational risk. Directly affecting the continuity of seller operations are login failures, password problems, session interruptions, application crashes, map errors, search failures, and transaction-screen problems. In terms of information systems success model, these issues show lack of system availability, reliability, response continuity, and error-free transaction flow. Also, in terms of operational risk, access and system failures can disrupt order management, product updates, client responses, and financial tracking. Therefore, K3 and K4 cannot be seen as just minor technical faults; they are system-level risks which may disrupt the continuity of supply-side platform operations.

K5, usability, interface, and application slowness prevalent among apps, are theoretically connected to the main components of usability, perceived ease of use, and technology acceptance. These aspects have lesser number of reviews as compared to K1, K2, and K3, however, they are very crucial as using these applications, the business partners perform their operational tasks on a daily basis. Slow screen transitions, confusing interface structures, and excessive navigation steps may increase the time taken to complete a task and, at the same time, reduce the perceived ease of use. Relating these problems to technology acceptance, it is quite possible that users' willingness to continue the use of the application in an effective manner may be negatively impacted. Hence, the key idea of K5 is that partner experience is not just about whether the system runs technically, but also whether the operational tasks can be done easily, quickly, and intuitively.

K6, the troubles related to product uploads, approval, and freezing, are linked to product and stock management, data accuracy, catalog management, and sales preparation. Issues with product upload, wrong category selections, barcode glitches, approval time, inaccurate stock updates, and price update problems are examples of how seller's skills to prepare products for sale can be directly impacted. Considering the perspective of information quality, such problems can be perceived as the lack of the rightness, wholeness, traceability, and the timeliness of product and

inventory data. From a supply chain process perspective, suboptimal catalog and inventory management can lead not only to diminished product visibility but also to protracted sales preparation workflows, and it also culminates in widespread operational inefficiencies. As a result, K6 proves that product information management is an essential part of platform-based supply chain information systems.

K7, the outlier/mixed problem pool, should be considered as a mix of different operational risks rather than just an instance of statistical noise. Although cluster K7 lacked internal homogeneity within the primary BERTopic architecture, a dedicated thematic sub-analysis revealed that it encapsulated substantive operational problem patterns, despite these issues being heavily intermingled. These mainly focused on platform governance and cost-related issues, technical update and performance problems, product/catalog management difficulties, notification and communication issues, account access problems, delivery-related concerns, support weaknesses, and highly mixed expressions. Consequently, Topic -1 generated during application review mining may still encapsulate weak yet critical signals of diffuse or nascent operational risks. Therefore, platform managers should not disregard outlier comments completely but rather, they should consider them regularly as a secondary diagnostic tool.

All things combined, these results reveal that identical platform features might cause different operational hurdles for the marketplace, rapid delivery, and restaurant partner systems. Although a few come up with quite severe cost and return issues, others uncover problems of notification, technical access, usability, product management, or some mixed patterns of problems. This is a clear indication that platform-based supply chain information systems need to be measured not only with general satisfaction scores but also with process-specific operational problem signals.

5.2. Theoretical Implications

Diverging from the established paradigm of consumer-centric review analysis, this study examines mobile application data through the lens of supply-side business partners, thereby offering a novel perspective on operational experiences within digital platform ecosystems. While extant research on digital platforms extensively documents customer satisfaction, user reviews, and service quality, the operational challenges embedded within partner-facing applications, such as those utilized by vendors, restaurants, and business affiliates, remain under-explored. Through the identification of partner apps as operational interfaces of platform-based supply chain information systems, the paper contributes to the literature by broadening the customer-focused perspective.

These results also establish a link between mining of app reviews and the success of information systems, quality of logistics services, technology acceptance, platform governance, and management of operational risk. K1 explains how seller trust and perceived fairness are influenced by platform governance, reverse logistics, and cost sharing. K2 reveals how information quality and notification reliability impact order management and supply chain visibility. K3 and K4 illustrate that system quality and technical reliability are considered both software issues and matters of continuous operation. K5 brings technology acceptance debates a step further by revealing the influence of usability and perceived ease of use on business partner experience. K6 points out the critical role of product and inventory information quality in platform-based supply chain operations. Lastly, K7 offers a new perspective on Topic -1 outputs by demonstrating that outlier comments may contain diverse but meaningful operational risk signals.

At this point, the study makes a theoretical connection between user-generated app reviews and supply-side operational analysis. It illustrates that partner reviews are a valuable tool to comprehend not only user satisfaction, but also to detect process-level weaknesses in platform-based supply chain information systems.

5.3. Managerial Implications

From a managerial perspective, this study provides platform decision-makers with a pragmatic framework by converting BERTopic-derived operational problem clusters into a rule-based decision support matrix. This tool operationalizes empirical findings to facilitate continuous process improvement and inform strategic platform governance. The matrix uses a researcher-defined threshold value based on the review count and the ratio of negative reviews to classify operational problem areas as urgent intervention, high priority, medium priority, or separate review. In this way, the study not only delivers descriptive topic modeling outcomes but also provides a clear prioritization mechanism that can help managerial decision-making in platform-based supply chain information systems.

According to the findings, cost transparency, consistent notification of order status, and regular technical access are the critical issues that need to be addressed right away. Mechanisms around returns, commissions, and shipping costs should be overhauled to both minimize sellers' losses and enhance fairness perception. Order and customer notification systems should be reinforced by using real-time alerts, giving control over notification permissions, and making the visibility of messages better. Besides that, login, password check, session continuity, and account access procedures need to be enhanced as interruptions in access can directly disrupt seller operations.

In planning, application errors, usability problems, and product management issues should be regarded as system health indicators. Error logs, map functions, search screens, order screens as well as transaction flows should be periodically analyzed to minimize operation disruptions. Design of interface speed, screen flow, and mobile usability enhancements should reflect the daily work of business partners. Product upload, approval, category, barcode, price update, and stock management processes should be made more reliable, have better traceability, and increased transparency

Additionally, the K7 sub-study indicates that platform managers should periodically evaluate outlier comments. This is because such peripheral feedback may signal nascent operational risks that have not yet crystallized sufficiently to form a distinct, homogeneous topic cluster. Therefore, Topic -1 results should not be discarded totally without consideration. They can be seen as additional diagnostics for finding broken, emerging, or cross-process operational issues.

6. Conclusion

This study contributes to related literature and practice in three main ways. First, it positions business partner and seller applications as supply-side operational components of platform-based supply chain information systems. This perspective shifts the focus from customer-side platform experiences to the operational experiences of sellers, restaurants, and business partners.

Additionally, the study provides a methodological contribution by using problem-focused app review mining and BERTopic-based topic modeling to identify operational problem areas in short and unstructured user-generated texts. The study does not limit topic modeling outputs to descriptive topic labels; instead, it links BERTopic-derived clusters with operational process categories, model validation indicators, representative anonymized review excerpts, K7 sub-analysis, and a rule-based decision support matrix.

Further, the research presents managerial insights by converting user-generated review information into actual improvement focus areas. The decision-support matrix assists platform administrators in distinguishing between operational anomalies requiring immediate remediation, issues designated for high- or medium-priority monitoring, and those reserved for subsequent qualitative analysis.

Also, this research has some limitations. The first major limitation of this study is that it relied solely on reviews on Google Play Store. Reviews on App Store, social media data, complaint platform data, and qualitative interview evidence were not taken into consideration. Therefore, the dataset only covers one public app review source and will miss other possible channels through which business partners may express their operational problems.

Secondly, the review volumes were not evenly distributed across different applications. Trendyol Seller Panel was the major contributor in terms of the number of reviews in both datasets (the preprocessed one and the one focused on problems). Although the study of application-based results was done using percentage distributions and not raw frequencies, the strong presence of this application might have affected the entire BERTopic thematic structure. Therefore, the results should not be understood as comparisons with equal weights among all platforms.

Additionally, the formulation of a problem-focused dataset involved elimination of short, general, and non-operational comments. Even though the filtering method was a must to highlight the areas of operational problems, it could have left out some user comments that were not directly linked to operational processes or were not textually rich. Hence, the problem-focused dataset is a deliberate analytical subset and should not be seen as a comprehensive presentation of all user comments.

Further, the K7 / Topic -1 cluster appeared to have the highest ratio and was composed of various different elements. Even if K7 was looked at through a different expressive theme sub-analysis, the fact that it was mixed proves that it might be necessary to conduct further qualitative exploration. Thus, the K7 outcomes should be regarded as diverse operational risk signals rather than a uniform topic cluster.

Lastly, the subject of labeling was partially dependent on the interpretation of the researcher. Manual topic labeling and the mapping of operational procedures were assisted by intercoder reliability analysis to enhance the trustworthiness of this activity. However, the identification of BERTopic-generated clusters, operational categories, and selection of representative excerpts for the review still call for the use of analytical judgment.

The data sources in future research can be broadened by mixing Google Play Store reviews with App Store reviews, social media data, complaint platforms, and qualitative partner interviews. A multi-source design like this might lead to a fuller, richer understanding of the experience of business partners in platform-based supply chain information systems. Also, longitudinal analyses might be used to determine if operational problem areas evolve after platform updates, policy changes, pricing revisions, return-policy modifications, or interface redesigns. This would provide a way to assess whether platform interventions diminish recurring operational problems or lead to new areas of dissatisfaction.

Future research might tackle the problem of data imbalance at the platform level by implementing balanced sampling methods, creating platform-specific BERTopic models, doing stratified comparisons, or using weighted analytical methods. Developing platform-specific models may be beneficial to analyze if the operational issues of marketplace, restaurant partner, and rapid delivery platforms differ from one another. Moreover, upcoming studies can take a closer look at the K7 / Topic -1 pool via manual qualitative coding, interviews, or a combination of methods. Since it is possible that K7 holds faint signs of broken or new operational risks, additional examination of the extreme comments could reveal initial hints for platform governance, system design, and process improvement. Eventually, new research can focus on partner satisfaction, operational continuity, platform governance, and supply chain performance with mixed-method or longitudinal research designs. This approach will explain how operational issues in partners' applications affect overall platform performance and supply chain results.

Statement of Support and Appreciation: This research did not receive any external support.

Declaration of Researcher's Contribution Rate: The contribution rates of the authors in the preparation of this research are equal.

Conflict Declaration: The authors declare that they have no conflict of interest.

Research and Publication Ethics Statement: All rules specified in the "Higher Education Institutions Scientific Directive" were followed at every stage of this research. None of the actions specified under the heading "Actions Contrary to Scientific Research and Publication Ethics" of the directive have been carried out. During the writing process of this study, citations were made in accordance with ethical rules, and a bibliography was created. The work has been checked for plagiarism.

Ethics Committee Permission: This study does not require ethics committee approval since it does not involve humans or animals.

References

- Balkar, O. E. (2025). E-Ticaret Lojistiği Hizmet Kalitesi ile Müşteri Memnuniyeti Arasındaki İlişkide Dijital Okuryazarlığın Aracılık Etkisi. In *14. Ulusal Lojistik ve Tedarik Zinciri Kongresi* (pp. 77–78). Ekin Yayınevi.
- Bhattacharjee, A. (2001). Understanding Information Systems Continuance: An Expectation-Confirmation Model. *MIS Quarterly*, 25(3), 351–370. <https://doi.org/10.2307/3250921>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Chen, N., Lin, J., Hoi, S. C. H., Xiao, X., & Zhang, B. (2014). AR-miner: mining informative reviews for developers from mobile app marketplace. *Proceedings of the 36th International Conference on Software Engineering*, 1, 767–778. <https://doi.org/10.1145/2568225.2568263>
- Collier, J. E., & Bienstock, C. C. (2006). Measuring Service Quality in E-Retailing. *Journal of Service Research*, 8(3), 260–275. <https://doi.org/10.1177/1094670505278867>
- Culotta, C., Blome, C., & Henke, M. (2024). Theories of digital platforms for supply chain management: a systematic literature review. *International Journal of Physical Distribution & Logistics Management*, 54(5), 449–475. <https://doi.org/10.1108/IJPDLM-01-2023-0016>
- Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>
- DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. *Journal of Management Information Systems*, 19(4), 9–30. <https://doi.org/10.1080/07421222.2003.11045748>
- Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. <https://arxiv.org/abs/2203.05794>
- Gunasekaran, A., & Ngai, E. W. T. (2004). Information systems in supply chain integration and management. *European Journal of Operational Research*, 159(2), 269–295. <https://doi.org/10.1016/j.ejor.2003.08.016>
- Guzman, E., & Maalej, W. (2014). How Do Users Like This Feature? A Fine Grained Sentiment Analysis of App Reviews. *2014 IEEE 22nd International Requirements Engineering Conference (RE)*, 153–162. <https://doi.org/10.1109/RE.2014.6912257>
- Halim, D., Budi, I., Mubina, B. F., Budi Santoso, A., & Kresna Putra, P. (2026). Comparative Topic Modelling of Mobile Banking User Reviews Using LDA and BERTopic: A Case Study of wondr by BNI. *The Indonesian Journal of Computer Science*, 15(2), 2236–2246. <https://doi.org/10.33022/ijcs.v15i2.5118>

- Hanifah, A., Saputra, M., & Fa'rifah, R. Y. (2025). Exploring User Sentiments and Adoption Barriers in E-Wallet Services in Indonesia using BERTopic on Play Store Reviews. *2025 4th International Conference on Electronics Representation and Algorithm (ICERA)*, 286–291. <https://doi.org/10.1109/ICERA66156.2025.11086632>
- Ivanov, D., Dolgui, A., & Sokolov, B. (2022). Cloud supply chain: Integrating Industry 4.0 and digital platforms in the “Supply Chain-as-a-Service.” *Transportation Research Part E: Logistics and Transportation Review*, 160, 102676. <https://doi.org/10.1016/j.tre.2022.102676>
- Jain, N. K., Gajjar, H., & Shah, B. J. (2021). Electronic logistics service quality and repurchase intention in e-tailing: Catalytic role of shopping satisfaction, payment options, gender and returning experience. *Journal of Retailing and Consumer Services*, 59, 102360. <https://doi.org/10.1016/j.jretconser.2020.102360>
- Johnson, M. E., & Whang, S. (2002). E-Business and Supply Chain Management: An Overview and Framework. *Production and Operations Management*, 11(4), 413–423. <https://doi.org/10.1111/j.1937-5956.2002.tb00469.x>
- Kumar, A., Sikdar, P., & Saha, R. (2021). Seller experience assessment in online marketplace: a scale development study. *Benchmarking: An International Journal*, 28(7), 2315–2342. <https://doi.org/10.1108/BIJ-06-2020-0305>
- Li, L., Xu, S. X., Ning, Y., Liu, Y., & Yang, S. (2023). How should companies deploy their digital supply chain platforms to gain competitive advantages? An asset orchestration perspective. *Information & Management*, 60(6), 103842. <https://doi.org/10.1016/j.im.2023.103842>
- Li, S., & Lin, B. (2006). Accessing information sharing and information quality in supply chain management. *Decision Support Systems*, 42(3), 1641–1656. <https://doi.org/10.1016/j.dss.2006.02.011>
- Maalej, W., Nayebi, M., Johann, T., & Ruhe, G. (2016). Toward Data-Driven Requirements Engineering. *IEEE Software*, 33(1), 48–54. <https://doi.org/10.1109/MS.2015.153>
- Martin, W., Sarro, F., Jia, Y., Zhang, Y., & Harman, M. (2017). A Survey of App Store Analysis for Software Engineering. *IEEE Transactions on Software Engineering*, 43(9), 817–847. <https://doi.org/10.1109/TSE.2016.2630689>
- McInnes, L., Healy, J., & Astels, S. (2017). hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software*, 2(11), 205. <https://doi.org/10.21105/joss.00205>
- Murfield, M., Boone, C. A., Rutner, P., & Thomas, R. (2017). Investigating logistics service quality in omni-channel retailing. *International Journal of Physical Distribution & Logistics Management*, 47(4), 263–296. <https://doi.org/10.1108/IJPDLM-06-2016-0161>
- Parasuraman, A., Zeithaml, V. A., & Malhotra, A. (2005). E-S-QUAL. *Journal of Service Research*, 7(3), 213–233. <https://doi.org/10.1177/1094670504271156>
- Putri, N. A., Pravitasari, A. A., & Winarni, S. (2026). Topic-based sentiment analysis of Jamsostek mobile application reviews using the BERTopic method. *Commun. Math. Biol. Neurosci.*, 2026, Article ID 2. <https://doi.org/10.28919/cmbn/9682>
- Rabinovich, E., & Bailey, J. P. (2004). Physical distribution service quality in Internet retailing: service pricing, transaction attributes, and firm attributes. *Journal of Operations Management*, 21(6), 651–672. <https://doi.org/10.1016/j.jom.2003.11.004>
- Rao, S., Goldsby, T. J., Griffis, S. E., & Iyengar, D. (2011). Electronic Logistics Service Quality (e-LSQ): Its Impact on the Customer's Purchase Satisfaction and Retention. *Journal of Business Logistics*, 32(2), 167–179. <https://doi.org/10.1111/j.2158-1592.2011.01014.x>

- Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 3980–3990. <https://doi.org/10.18653/v1/D19-1410>
- Sagama, Y., Alamsyah, A., & Tricahyono, D. (2024). Generating Mobile App Improvement Recommendations Through User Review Analysis Using BERTopic and GPT. *2024 International Conference on Digital Business and Technology Management (ICONDBTM)*, 1–6. <https://doi.org/10.1109/ICONDBTM64135.2024.11122739>
- Sorkun, M. F. (2019). Factors That Impact the Logistics Service Performance in Online Marketplaces. *Journal of Yasar University*, 14(54), 152–167.
- Urrea, N. T., Vishkaei, B. M., & De Giovanni, P. (2024). Operational Risk Management in E-Commerce: A Platform Perspective. *IEEE Transactions on Engineering Management*, 71, 3807–3819. <https://doi.org/10.1109/TEM.2024.3358717>
- Vakulenko, Y., Shams, P., Hellström, D., & Hjort, K. (2019). Service innovation in e-commerce last mile delivery: Mapping the e-customer journey. *Journal of Business Research*, 101, 461–468. <https://doi.org/10.1016/j.jbusres.2019.01.016>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward A Unified View. *MIS Quarterly*, 27(3), 425–478. <https://doi.org/10.2307/30036540>
- Wixom, B. H., & Todd, P. A. (2005). A Theoretical Integration of User Satisfaction and Technology Acceptance. *Information Systems Research*, 16(1), 85–102. <https://doi.org/10.1287/isre.1050.0042>
- Wolfenbarger, M., & Gilly, M. C. (2003). eTailQ: dimensionalizing, measuring and predictingetail quality. *Journal of Retailing*, 79(3), 183–198. [https://doi.org/10.1016/S0022-4359\(03\)00034-4](https://doi.org/10.1016/S0022-4359(03)00034-4)
- Yi, J., Oh, Y. K., & Kim, J.-M. (2025). Unveiling the drivers of satisfaction in mobile trading: Contextual mining of retail investor experience through BERTopic and generative AI. *Journal of Retailing and Consumer Services*, 82, 104066. <https://doi.org/10.1016/j.jretconser.2024.104066>
- Zhao, J., Zhang, Z., Feng, J., & Li, A. (2026). Prioritizing App User Reviews Based on Multidimensional Factors: An Integrated BERTopic–SnowNLP Framework for Supporting Developer Decisions. *IEEE Transactions on Computational Social Systems*, 13(1), 1100–1109. <https://doi.org/10.1109/TCSS.2025.3627971>
- Zhao, Y. (2025). Platform Economy and Global Supply Chain Integration: A Technology-Driven Approach to Business Model Innovation. In *Proceedings of 2024 6th International Conference on Economic Management and Cultural Industry (ICEMCI 2024)* (pp. 612–620). Atlantis Press. https://doi.org/10.2991/978-94-6463-642-0_65