# Anomaly Detection in Customer Purchasing Patterns: A CRISP-DM Approach

Ariana Nurlayla Syabandini*, Lucinda Laurent*, Rayhan Adhisworo Jardine Hidayat*,
Syahrul Apriansyah*, Wahyu Hidayat*, Timothy Orvin Edwardo*, Andi Pujo Rahadi*
*Faculty of Computer Science
Universitas Indonesia
Depok, Indonesia
Emails: ariana.nurlayla@ui.ac.id, lucinda.laurent@ui.ac.id, rayhan.adhisworo@ui.ac.id,
syahrul.apriansyah@ui.ac.id, wahyu.hidayat22@ui.ac.id, timothy.orvin@ui.ac.id, ap.rahadi@gmail.com

***Abstract*—Understanding anomalous purchasing patterns provides online retailers with valuable insights for inventory management, marketing, and risk mitigation. This research applies the CRISP-DM methodology to detect and analyze anomalies in customer purchasing behaviors using a UK-based online retail dataset spanning one year. We implemented a multi-stage approach integrating RFM-based customer segmentation, FP-Growth association rule mining, and hybrid anomaly detection combining Isolation Forest and Local Outlier Factor algorithms. Our analysis identified five distinct customer segments with unique purchasing behaviors and segment-specific anomalous patterns. Key findings revealed different anomaly drivers across segments: confidence/conviction metrics in casual buyers, leverage/lift in high-value customers, and support metrics in wholesale segments. Anomalous rules frequently involved seasonal products, emerging trends, or segment-specific preferences. The study demonstrates that combining segmentation with association rule mining enhances anomaly detection effectiveness by providing contextual baselines for normal behavior in each customer group. We also proposed a deployment framework for operationalizing these insights in retail environments. The integrated approach presented enables retailers to proactively identify unusual purchasing patterns, optimize inventory, develop targeted marketing strategies, and address potential operational issues before they impact business performance.**

***Index Terms*—Anomaly detection, Association rule mining, CRISP-DM, Customer segmentation, RFM analysis, FP-Growth, Isolation Forest, Local Outlier Factor, Retail analytics**

## I. Introduction

Understanding customer purchasing patterns is vital for online retailers competing in today's market. Transaction data analysis helps optimize inventory, develop marketing strategies, and identify issues including fraud and market shifts [1]. Detecting anomalous purchasing patterns (unusual product combinations, transaction spikes, or frequency changes) can reveal emerging trends, seasonal patterns, or unique customer segments [2]. Traditional statistical methods struggle with these complex patterns in large datasets, requiring advanced data mining approaches [3].

### A. CRISP-DM Methodology and Technical Approach

We adopt the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology with its six phases (business understanding, data understanding, data preparation, modeling, evaluation, and deployment) [4], ensuring our technical analyses align with practical retail needs. Our approach combines association rule mining using FP-Growth [5], which efficiently identifies frequent patterns with a compact tree structure outperforming Apriori methods, and anomaly detection techniques. For the latter, we employ Isolation Forest [6], which isolates anomalies through recursive partitioning, and Local Outlier Factor (LOF) [7], which compares local density patterns, making it suitable for varied retail purchasing behaviors.

### B. Research Questions and Objectives

This study addresses: (1) How to effectively detect anomalies in customer purchasing patterns; (2) How to determine whether anomalies stem from specific customer segments; and (3) Which frequent transaction patterns provide business optimization insights. We aim to segment customers based on purchasing behavior, discover significant association rules within segments, detect anomalies deviating from established patterns, and provide actionable business insights.

### C. Related Work

Previous research has explored individual aspects of retail data mining. Miguéis et al. [1] proposed lifestyle-based customer segmentation, while Han et al. [5] introduced FP-Growth for association rules. Liu et al. [6] developed Isolation Forest and Breunig et al. [7] introduced LOF for anomaly detection, but their combined application to retail data remains limited. While Kannan and Bhaskaran [8] applied association rules to e-commerce for recommendations and Blázquez-García et al. [9] reviewed time-series anomaly detection for retail, few studies integrate these techniques comprehensively.

Our research addresses this gap by combining customer segmentation, association rule mining, and anomaly detection within the CRISP-DM framework specifically for online retail transaction analysis.

## II. Business and Data Understanding

### A. Business Context and Objectives

The online retail industry operates in a highly competitive environment where understanding customer behavior is crucial for maintaining a competitive edge, particularly for wholesale-focused retailers selling souvenirs and gift items. This study addresses the detection and interpretation of anomalous purchasing patterns, which can manifest as unusual product combinations, unexpected transaction spikes, or sudden changes in purchasing frequency.

Our business objectives are:

1) **Strategic Marketing:** Develop targeted strategies based on customer segmentation for effective product bundling and cross-selling opportunities.
2) **Risk Mitigation:** Identify potential fraud or operational issues by distinguishing between normal variations and truly anomalous behavior.

Key stakeholders include retail management, marketing teams, and fraud prevention units, with expected benefits of reduced inventory costs, increased sales, and early issue detection.

### B. Dataset Description

The "Online Retail" dataset from the UCI Machine Learning Repository [10] contains 541,909 transactions from December 1, 2010, to December 9, 2011, representing a UK-based company that primarily sells souvenirs and gift items to wholesale buyers. The dataset comprises eight primary variables as described in Table I.

TABLE I
DESCRIPTION OF THE ONLINE RETAIL DATASET VARIABLES

| Variable | Description |
|---|---|
| InvoiceNo | A 6-digit unique identifier for each transaction. Invoices starting with 'C' indicate cancellations. |
| StockCode | A 5-digit product identifier. |
| Description | Product name/description. |
| Quantity | Number of units purchased per transaction. |
| InvoiceDate | Date and time of the transaction. |
| UnitPrice | Price per unit in British Pounds (£). |
| CustomerID | A 5-digit unique customer identifier. |
| Country | Customer's country of residence. |

### C. Exploratory Data Analysis

The exploratory data analysis (EDA) was conducted on the "Online Retail" dataset from the UCI Machine Learning Repository, comprising 541,909 transactions from December 1, 2010, to December 9, 2011. Key observations include 135,080 missing CustomerID entries (24.9%) and 1,454 missing Descriptions (0.3%), with 5,268 duplicate records. Quantity ranged from -80,995 to 80,995 (mean 9.55, std 218.08), and UnitPrice from -11,062.06 to 38,970 GBP, with a median Quantity of 3 and moderate UnitPrices dominating most transactions.

Negative Quantity values were associated with 'C'-prefixed InvoiceNos (cancellations) and 'A'-prefixed entries (internal adjustments, lacking CustomerIDs and occasionally showing negative UnitPrices). Temporal analysis indicated a peak of approximately 80,000 transactions in November 2011 and a low of 20,000 in February 2011, with no transactions recorded on Saturdays. Geographically, the UK accounted for 89.8% of transactions, followed by Germany (4.4%) and France (2.1%), while a right-skewed customer transaction distribution reflected a mix of retail and wholesale buyers. These insights into transaction patterns, cancellation behaviors, and data anomalies informed subsequent data cleaning, feature engineering, and anomaly detection efforts.

## III. Data Preparation

We constructed a structured pipeline to resolve data quality issues and engineer key features for anomaly detection.

### A. Data Cleaning

Out of 541,909 records, 133,626 (24.7%) had missing `CustomerID`. Rather than dropping them, we added a `CustomerType` attribute, generated synthetic IDs from `InvoiceNo`, and removed 1,454 records with missing `Description` (identified by `StockCode`).

We also found 5,268 duplicates, which we addressed by merging repeated products in the same invoice and removing fully identical rows.

We standardized formats by converting `CustomerID` to string, `InvoiceDate` to datetime, and fixing `UnitPrice` (decimal correction). Invalid entries such as negative values, cancellations ("C"/"A" invoices), and "Unspecified" countries (1,938 rows) were removed. Text fields were trimmed and capitalized consistently.

### B. Feature Engineering

We extracted features to improve model interpretability:

- **Product Categories:** Derived from `Description` using rule-based pattern matching (e.g., themes, bundles, location).
- **Transaction Value:**

$$\boxed{\text{TotalPrice} = \text{Quantity} \times \text{UnitPrice}} \tag{1}$$

- **Temporal Features:** Weekday and hour extracted to capture seasonality.
- **Customer Metrics:**:
  - *Recency:* Days since last transaction,
  - *Frequency:* Total transactions per customer.

## IV. Modeling and Evaluation

### A. Modeling Approach

To detect anomalies in online retail customer purchasing patterns, we implemented a multi-stage approach that integrates several data mining techniques:

1) **Customer Segmentation using K-Means**: Grouping customers based on purchasing behavior using RFM (Recency, Frequency, Monetary) metrics to provide a baseline context of normal behavior.
2) **Association Rule Mining with FP-Growth**: Identifying normal product purchasing patterns within each customer segment to form a basis for comparison.
3) **Anomaly Detection**: Identifying transactions and association patterns that significantly deviate from baseline patterns in their respective customer segments using a combination of Isolation Forest and Local Outlier Factor.

*B. Customer Segmentation with K-Means*

Customer segmentation was performed as a supporting step to achieve the main project objectives of anomaly detection and association rule mining. We used RFM metrics, which are the most relevant metrics for grouping customers based on purchasing behavior:

- **Recency**: Indicates how recently a customer has transacted (activity indicator)
- **Frequency**: Indicates how often a customer transacts (loyalty indicator)
- **Monetary**: Indicates how much money a customer spends (customer value indicator)

This segmentation helps us: (1) understand variations in normal shopping patterns per segment, (2) provide a baseline for anomaly detection, (3) generate insights for business strategies, and (4) provide a richer context for further analysis.

*1) K-Means Hyperparameter Tuning:* The determination of the optimal number of clusters was performed by evaluating the Silhouette Score for various values of $k$ from 2 to 10 (Figure 1).

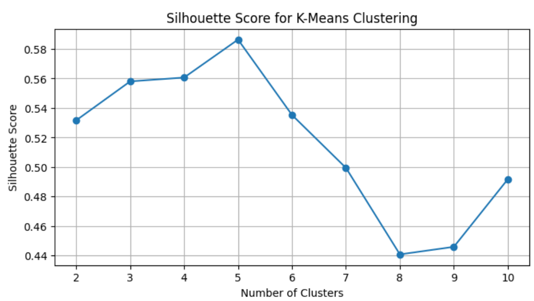


Fig. 1. Silhouette Score for K-Means with various cluster numbers

The highest Silhouette Score of 0.586 was achieved at $k = 5$, indicating that five clusters provide the best separation between clusters. This score is quite good (values $> 0.5$ are considered feasible), showing that the data within each cluster is sufficiently cohesive and well-separated from other clusters.

*2) Customer Cluster Interpretation:* The $k = 5$ clustering results are shown in a 3D plot of the original RFM dimensions (Figure 2). Minor random position adjustments (jitter) were applied to reveal overlapping customer data points. (Figure 2).

Based on the average RFM values, the five customer clusters can be characterized as follows:

- **Cluster 0 - "Casual Buyers"**: Customers with Recency of 48.7 days, Frequency of 3.2, and Monetary value of 1,339. These are still active customers but don't transact frequently and have moderate expenditure. Includes some guest customers with moderate recency and spending.
- **Cluster 1 - "Inactive Customers"**: Customers with Recency of 255.7 days, Frequency of 1.3, and Monetary value of 700. These are inactive customers who rarely transact and have low expenditure. Dominated by guest customers.

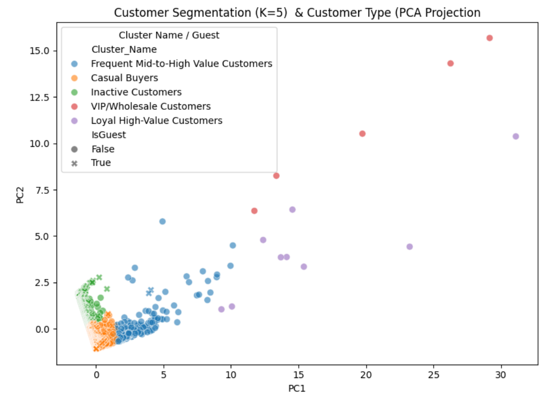


Fig. 2. Customer Cluster Visualization (K-Means, 5 Clusters)

- **Cluster 2 - "Loyal High-Value Customers"**: Customers with Recency of 6.9 days, Frequency of 113.2, and Monetary value of 62,209. These are very active customers who transact frequently and have high expenditure.
- **Cluster 3 - "VIP/Wholesale Customers"**: Customers with Recency of 7.2 days, Frequency of 40.4, and Monetary value of 205,560. These are very active customers with extreme expenditure (possibly wholesale customers).
- **Cluster 4 - "Frequent Mid-to-High Value Customers"**: Customers with Recency of 15.4 days, Frequency of 20.4, and Monetary value of 12,452. These are active customers who transact quite frequently and have relatively high expenditure. Includes rare guest outliers with high spending, with potential fraud indicated by unusual guest Monetary.

*C. Association Rule Mining using FP-Growth*

After grouping customers, we implemented association rule mining to identify normal product purchasing patterns within each segment. The FP-Growth algorithm was chosen due to its advantages in handling large datasets compared to the Apriori algorithm.

*1) Justification for FP-Growth Selection:* FP-Growth was chosen based on dataset characteristics:

- The dataset contains 519,641 transaction rows covering 19,946 unique transactions, requiring an efficient algorithm.
- There are 3,921 unique products grouped into 1,522 categories, showing item diversity.
- FP-Growth has more efficient memory usage for large datasets with imbalanced distribution.

*2) Hyperparameter Tuning:* For each cluster, we performed grid search for min_support and min_confidence parameters, considering the number of rules, average lift, confidence, leverage, and conviction. The optimal parameter combinations for each cluster:

- **Cluster 0**: min_support = 0.03, min_confidence = 0.3, min_lift = 1.5
- **Cluster 1**: min_support = 0.03, min_confidence = 0.3, min_lift = 1.2

- **Cluster 2**: min_support = 0.05, min_confidence = 0.4, min_lift = 1.5
- **Cluster 3**: min_support = 0.03, min_confidence = 0.3, min_lift = 1.2
- **Cluster 4**: min_support = 0.03, min_confidence = 0.3, min_lift = 1.5

*3) Association Rule Mining Model Evaluation:* Table II shows the evaluation summary for each cluster.

TABLE II
ASSOCIATION RULE MINING MODEL EVALUATION BY CLUSTER

| Cluster | Avg. Support | Avg. Confidence | Avg. Lift | Max Lift | Avg. Conviction |
|---|---|---|---|---|---|
| Cluster 0 | 0.04 | 0.50 | 2.00 | 6.41 | 1.59 |
| Cluster 1 | 0.04 | 0.51 | 6.18 | 10.97 | 2.11 |
| Cluster 2 | 0.07 | 0.63 | 1.93 | 3.82 | 2.28 |
| Cluster 3 | 0.03 | 0.85 | 22.29 | 32.83 | 9.40 |
| Cluster 4 | 0.04 | 0.51 | 2.18 | 16.65 | 1.64 |

### D. Anomaly Detection Methodology

After establishing normal purchasing patterns for each customer cluster, we implemented a multi-level approach for anomaly detection by combining two main algorithms to detect anomalies in association rules:

1) **Isolation Forest**: Used to detect global anomalies by recursively isolating observations through random feature selection and threshold values.
2) **Local Outlier Factor (LOF)**: Used to detect local anomalies by comparing the local density of a data point with the local density of its neighbors.

For both anomaly detection algorithms, we performed hyperparameter tuning to ensure optimal performance by selecting the values that yielded the most stable predictions.

- **Isolation Forest**: The main parameter tuned was `contamination` (estimated proportion of anomalies in the dataset). We evaluated values of 0.01, 0.03, 0.05, 0.07, and 0.1, then selected the value that provided the best stability.
- **Local Outlier Factor**: The main parameter tuned was `n_neighbors` (number of neighbors to estimate local density). We evaluated values of 5, 10, 20, 30, and 50, considering the correlation between LOF scores for different values. For Cluster 3, we specifically used higher values of `n_neighbors` — 150, 200, 250, 300, and 350 — due to a large number of association rules with highly similar metric values. A larger neighborhood size was needed to obtain more accurate density estimations and better distinguish between normal and anomalous rules.

Tuning results for each cluster:

- **Cluster 0**: contamination = 0.07, n_neighbors = 20
- **Cluster 1**: contamination = 0.07, n_neighbors = 30
- **Cluster 2**: contamination = 0.07, n_neighbors = 20
- **Cluster 3**: contamination = 0.07, n_neighbors = 250
- **Cluster 4**: contamination = 0.1, n_neighbors = 30

### E. Anomaly Detection Results and Analysis

We identified several significant anomaly patterns in each customer cluster (where ratios denote the proportion of a product's frequency in anomalous rules relative to its frequency in normal rules).

*1) Cluster 0 - "Casual Buyers":*

- Key metrics determining anomalies were identified using SHAP values, with leverage and support emerging as the most influential.
- Detected 42 anomalies with Isolation Forest and 21 anomalies with LOF
- Categories frequently appearing in anomalies, such as feltcraft doll, spice tins, and jumbo vintage with very high occurrence ratios (appearing almost exclusively in anomalous rules)

*2) Cluster 1 - "Inactive Customers":*

- Key metrics determining anomalies were identified using SHAP values, with conviction and support emerging as the most influential.
- Detected 350 anomalies with Isolation Forest and 23 anomalies with LOF.
- Categories frequently appearing in anomalies, such as polkadot cup, gold tape, and roses regency with infinite, infinite, and 13.286 ratios, respectively

*3) Cluster 2 - "Loyal High-Value Customers":*

- Key metrics determining anomalies were identified using SHAP values, with support and leverage emerging as the most influential.
- Detected 69 anomalies with Isolation Forest and 26 anomalies with LOF
- Categories frequently appearing in anomalies, such as cream stripe, travel card, and babushka notebook with infinite, 6.594, and 2.098 ratios, respectively

*4) Cluster 3 - "VIP/Wholesale Customers":*

- Key metrics determining anomalies were identified using SHAP values, with leverage and support emerging as the most influential.
- Detected 290 anomalies with Isolation Forest and 1053 anomalies with LOF
- Categories frequently appearing in anomalies, such as card holder, wood board, and hook photo with very high occurrence ratios (appearing almost exclusively in anomalous rules)

*5) Cluster 4 - "Frequent Mid-to-High Value Customers":*

- Key metrics determining anomalies were identified using SHAP values, with leverage and support emerging as the most influential.
- Detected 42 anomalies with Isolation Forest and 27 anomalies with LOF
- Categories frequently appearing in anomalies, such as toy, piece polkadot, and regency teacup with very high occurrence ratios (appearing almost exclusively in anomalous rules)

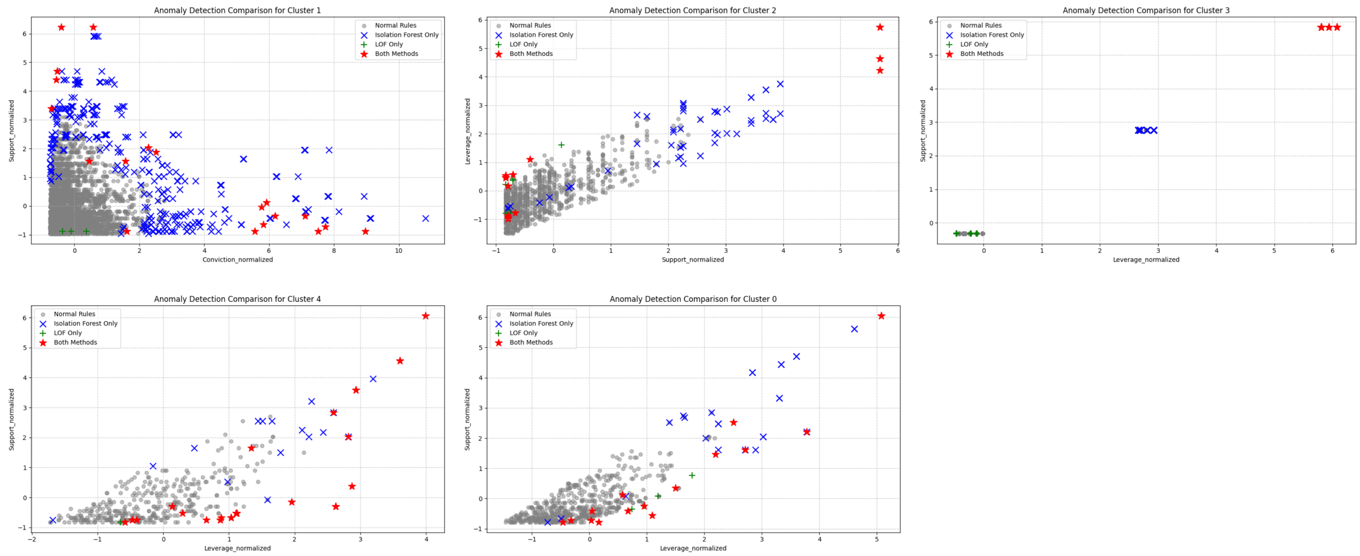


Fig. 3. Anomaly Detection Method Comparison Across All Customer Clusters.

### *F. Anomaly Cause Interpretation*

The analysis of anomaly characteristics reveals that specific customer segments drive certain anomalies. Categories such as cream stripe and babushka notebook predominantly appear in anomalous rules for Cluster 2 (Loyal High-Value Customers), indicating distinct purchasing patterns of this segment.

### *G. Business Insights*

To support actionable insights based on the clustering results, we defined a dictionary `cluster_strategies` that maps each customer cluster to a specific set of marketing tactics. This mapping enables the system to generate targeted business actions based on the cluster assignment. The strategies for each segment are summarized as follows:

- **Cluster 0 – Casual Buyers**: Focused on `cross_selling`, affordable `bundling`, and time-limited `promotions`. Fallback: personalized recommendations.
- **Cluster 1 – Inactive Customers**: Uses `reactivation` offers, `cross_selling`, and `discounts`. Fallback: general promotions.
- **Cluster 2 – Loyal High-Value Customers**: Targets customers with `upselling`, `personalization`, and `loyalty_programs`. Fallbacks include cross-selling and exclusive promotions.
- **Cluster 3 – VIP/Wholesale Customers**: Includes `wholesale` deals, `cross_selling`, and `special_services` such as dedicated account management. Fallback: exclusive bulk promotions.
- **Cluster 4 – Frequent Mid-to-High Value Customers**: Combines `cross_selling`, `upselling`, and `seasonal_promotions`. Fallback: personalized offers.

To connect these segment-level strategies with the underlying anomalies in customer behavior, we analyzed the dominant metrics behind each anomalous association rule. Each rule was associated with a primary metric (e.g., `support`, `leverage`, `conviction`) that reflects the nature of the pattern being flagged. These key metrics were then mapped to suitable marketing strategies, forming a rule-to-strategy pipeline.

- **Support**-driven anomalies were linked to strategies such as *bundling*, *promotions*, or *wholesale*, indicating frequent but potentially underutilized item combinations.
- **Leverage**-based anomalies suggested *cross-selling* or *wholesale* tactics, as they reflected associations with significant added value.
- **Conviction**-related anomalies were associated with *personalization*, *reactivation*, or *upselling*, highlighting rules with high directional confidence.

When a direct match between metric and strategy was not available, fallback strategies were used, guided by the behavioral profile of each cluster. This rule-based recommendation framework allows for interpretable, data-driven decision support tailored to both the customer segment and the specific nature of the detected anomaly.

Examples of strategy recommendations include cross-selling of `candles` and `t-lights` for *Casual Buyers* based on high leverage; reactivation offers for *Inactive Customers* driven by conviction-based rules involving `lunch boxes` and `picnic baskets`; cross-selling of `sign` and `tin` items for *Loyal High-Value Customers* informed by high support; wholesale deals for *VIP Customers* suggested by perfect-confidence associations among `doormat` variants; and complementary offers on `mugs` and `tins` for *Frequent Mid-to-High Value Customers* based on leverage-driven patterns.

### H. Model Limitations

Although our approach provides valuable insights, several limitations should be noted:

- **Imbalanced Clusters**: The distribution of customers across clusters is imbalanced (Cluster 0 dominates with 58% of transactions), which can affect statistics and rule mining.
- **Temporal Constraints**: The dataset only covers a one-year period, limiting the ability to analyze longer seasonal patterns.
- **Computational Limitations**: For clusters with very large numbers of transactions, we had to perform random sampling of 1,000 transactions to keep the mining process feasible.
- **Anomaly Detection Parameter Limitations**: The selection of contamination and n_neighbors parameters significantly influences detection results, but there is no standard method to evaluate the accuracy of these parameters in the context of unsupervised learning.
- **Product Categorization**: Product category extraction was performed with a rule-based approach, which may be less optimal compared to more sophisticated machine learning methods.

## V. Deployment

We successfully deployed our anomaly detection system as a production-ready Streamlit web application, operationalizing the complete CRISP-DM pipeline for real-world retail use. This implementation demonstrates practical business value beyond theoretical research.

### A. Production Implementation

The deployment consists of an interactive web application built with Streamlit, providing retail professionals with immediate access to customer segmentation, association rule mining, and anomaly detection capabilities. The system handles 541,909+ transaction records with optimized performance through advanced caching and model persistence.

Our technical stack includes:

- **Frontend**: Streamlit with custom CSS for professional presentation
- **ML Pipeline**: Scikit-learn implementation of K-Means, Isolation Forest, and LOF
- **Data Processing**: Pandas/NumPy with memory optimization for large datasets
- **Visualization**: Interactive Plotly dashboards with real-time updates
- **Performance**: Streamlit caching and Joblib model persistence reducing load times from minutes to seconds

### B. Interactive CRISP-DM Interface

Each CRISP-DM phase is implemented as an interactive module: (1) Business Understanding with objectives and success metrics, (2) Data Understanding featuring comprehensive EDA with 15+ visualizations, (3) Data Preparation showing cleaning pipelines and quality metrics, (4) Modeling workspace with real-time RFM clustering and configurable association rule mining, (5) Evaluation analytics with dual-algorithm anomaly detection, and (6) Deployment dashboard with operational monitoring.

## VI. Conclusion and Future Work

This study applied CRISP-DM methodology to detect anomalies in customer purchasing patterns within online retail transactions through an integrated approach combining segmentation, association rule mining, and anomaly detection techniques.

### A. Key Findings

Our research revealed:

- Five distinct customer segments with unique purchasing behaviors, from inactive to high-value wholesale buyers
- Segment-specific purchasing patterns with varying association strengths (average lift 1.93-22.29)
- While leverage and support were the primary anomaly-driving metrics across most segments, the Inactive Customers segment was characterized by higher influence of conviction and support.
- More consistent purchasing behavior in VIP customers versus unexpected strong associations in Inactive customers

### B. Limitations and Future Directions

Despite valuable insights, limitations included imbalanced cluster distribution, limited temporal scope, computational constraints, and challenges in unsupervised parameter optimization. Future research could:

- Develop balanced sampling techniques preserving pattern integrity
- Explore deep learning approaches for more sophisticated anomaly detection
- Integrate temporal modeling for better seasonal pattern identification
- Investigate hybrid supervised/unsupervised approaches
- Extend the methodology to cross-channel retail environments

## References


[1] V. L. Miguéis, D. Van den Poel, A. S. Camanho, and J. F. e. Cunha, "Customer data mining for lifestyle segmentation," *Expert Systems with Applications*, vol. 39, no. 10, pp. 9359–9366, 2012.

[2] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM computing surveys (CSUR)*, vol. 41, no. 3, pp. 1–58, 2009.

[3] J. Han, M. Kamber, and J. Pei, *Data mining: concepts and techniques*, 3rd ed. Morgan Kaufmann, 2011.

[4] R. Wirth and J. Hipp, "Crisp-dm: Towards a standard process model for data mining," *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, vol. 1, pp. 29–39, 2000.

[5] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 2000, pp. 1–12.

[6] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *2008 Eighth IEEE International Conference on Data Mining*. IEEE, 2008, pp. 413–422.

[7] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, “Lof: identifying density-based local outliers,” *ACM SIGMOD Record*, vol. 29, no. 2, pp. 93–104, 2000.
[8] S. Kannan and R. Bhaskaran, “Association rule mining and its applications in e-commerce business,” *International Journal of Computer Applications*, vol. 178, no. 1, pp. 12–14, 2019.
[9] A. Blázquez-García, A. Conde, U. Mori, and J. A. Lozano, “A review on outlier/anomaly detection in time series data,” *ACM Computing Surveys (CSUR)*, vol. 54, no. 3, pp. 1–33, 2021.
[10] D. Chen, “Online Retail,” UCI Machine Learning Repository, 2015, DOI: https://doi.org/10.24432/C5BW33.

‘