Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2026. Vol. 5. No. 2; e-ISSN: 2808-4519

FP-Growth for Data-Driven Purchase Pattern Analysis and Product Recommendations at Flanetqueen Store

Sopa Marwah[1]*, Nining Rahaningsih[2], Irfan Ali[3], Indra Wiguna Marthanu[4], Kaslani[5]

[1,3]Informatics Engineering, STMIK IKMI Cirebon
[2,4,5]Computerised Accounting, STMIK IKMI Cirebon
shofamrwh1@gmail.com[1*], niningr157@yahoo.co.id[2], irfanaali0.0@gmail.com[3], i.wiguna322@gmail.com[4], kaskaslani@gmail.com[5]

Abstract

The advancement of information technology has encouraged the use of data analytics to support data-driven business decision-making. This study aims to analyze purchasing patterns of hoodie products and provide product recommendations for customers at Flanetqueen Store using the FP-Growth (Frequent Pattern Growth) algorithm. The research applies the Knowledge Discovery in Database (KDD) framework, consisting of five stages: data selection, preprocessing, transformation, data mining, and interpretation/evaluation. The dataset comprises hoodie sales transactions recorded from January to December 2024. Data analysis was conducted using RapidMiner Studio version 10.3 with a minimum support of 0.2 and minimum confidence of 0.4. The analysis produced 26 itemsets and 11 association rules indicating product correlations. The strongest rule, Bloods → Champion, achieved a confidence of 0.414, revealing that customers who purchased Bloods hoodies were also likely to buy Champion hoodies. These findings were used to design cross-selling strategies and generate relevant product recommendations. The study demonstrates that FP-Growth effectively extracts frequent purchase patterns and contributes to the development of data-driven recommendation systems in the local fashion retail industry.

Keywords: *Data Mining, FP-Growth, Association Rule, Product Recommendation, Hoodie, RapidMiner, Flanetqueen*

1. Introduction

The rapid advancement of information technology has transformed how retail businesses collect and analyze transaction data. Every sales activity generates valuable information that reflects customer preferences, product relationships, and evolving market trends. When processed effectively, these data support data-driven decision-making and enable retailers to optimize product placement and promotion strategies [1], [2]. Data mining plays a crucial role in extracting hidden knowledge from large datasets and is widely used in retail for identifying co-purchase patterns through association rule mining [3].

Among various association mining algorithms, FP-Growth is highly efficient due to its ability to avoid candidate generation, a limitation found in traditional Apriori-based approaches. FP-Growth compresses transaction data into an FP-Tree structure, allowing fast extraction of frequent itemsets and strong association rules [3], [4]. Studies have shown that FP-Growth produces reliable results for market basket analysis, supporting cross-selling, product bundling, and customer behavior insights in the retail domain [1], [5]. Its efficiency makes it suitable for small and medium businesses that require powerful analytic tools with minimal computational cost.

However, despite the proven benefits of data mining, many small and medium-sized enterprises (SMEs) still face difficulties adopting analytics solutions due to limited digital infrastructure, unstructured data management, and low readiness for data-driven culture [6], [7]. Research indicates that SMEs struggle with integrating analytic capabilities into daily operations, reducing opportunities to improve sales performance and customer targeting [8], [9]. These challenges result in valuable retail transaction data being underutilized, particularly in small apparel shops where purchasing behavior patterns can significantly influence marketing outcomes.

Flanetqueen Store, a local retailer specializing in hoodie sales, experiences similar issues. Although the store collects transaction data continuously, the information has not been analyzed to understand customer buying patterns or product correlations. Such limitations hinder the development of more targeted promotional strategies. Research in apparel retail highlights that customer segmentation and association analysis can greatly enhance personalization and sales performance [10], [5]. Therefore, applying FP-Growth to analyze hoodie purchase patterns at Flanetqueen Store presents an opportunity to support evidence-based decision-making and improve the store's marketing effectiveness.

2. Literature Review

2.1. Data selection

Data selection is the initial stage of the KDD process that focuses on identifying and extracting relevant data from broader organizational data sources. Only attributes that contribute directly to the research objectives are retained, ensuring the dataset is meaningful and analytically valuable [1], [4]. In this study, hoodie sales transaction records collected between January and December 2024 were extracted from the Flanetqueen Store's internal transaction log. The data were stored in Microsoft Excel format, containing transaction IDs, item names, purchase timestamps, and product brand attributes. This stage aims to ensure that the dataset accurately reflects customer purchasing behavior and contains sufficient transactional diversity to generate strong association rules.

2.2. Pre-processing

Preprocessing is crucial for improving data quality by addressing inconsistencies, noise, and inaccuracies present in raw transaction records. According to prior research, effective preprocessing significantly enhances the accuracy of pattern discovery and minimizes model bias [3], [5]. In this study, preprocessing involved removing redundant and incomplete entries, correcting inconsistent product labeling, and eliminating irrelevant attributes that did not contribute to association rule mining. Duplicate transaction IDs were resolved, spelling variations of product names were standardized, and missing values were handled through deletion when they did not meet analytical requirements. This step ensures that the dataset is clean, consistent, and ready for transformation.

2.3. Transformation

Transformation converts the cleaned dataset into a suitable format for the data mining algorithm—particularly FP-Growth, which requires transaction-itemset input structures. Previous studies highlight that poor transformation can reduce algorithm efficiency and alter support-count accuracy [4], [5]. In this research, transformation involved aggregating item names into transaction-based item lists, renaming attributes to uniform identifiers, and assigning proper roles (such as "id" and "item") using RapidMiner operations. The process also included converting the dataset into a market basket structure where each transaction contained a set of hoodie brands purchased together. This restructuring enables FP-Growth to accurately identify frequent itemsets from the transaction data.

2.4. Data mining

Data mining is the core stage of the KDD framework where algorithms are applied to extract hidden patterns. FP-Growth was selected due to its superior computational efficiency compared to Apriori, especially for medium-sized datasets [3], [4]. FP-Growth constructs a compact FP-Tree to mine frequent itemsets without candidate generation, making it suitable for retail analysis where transactions may contain multiple items. In this study, the algorithm was implemented using RapidMiner Studio 10.3 with a minimum support threshold of 0.2 and a minimum confidence threshold of 0.4. The algorithm identified recurring product combinations and quantified the strength of associations among hoodie brands.

2.5. Interpretation

The interpretation stage focuses on analyzing and understanding the patterns discovered during data mining. Effective interpretation transforms data-driven patterns into actionable business insights, such as cross-selling opportunities and promotional strategies [1], [11]. In this research, association rules generated by FP-Growth were evaluated based on support, confidence, and lift values to determine their analytical significance. Rules with strong relationships—such as *Bloods → Champion*—indicate complementary product preferences among customers. The findings were then translated into practical recommendations for product bundling, inventory management, and marketing decision-making at Flanetqueen Store.

3. Research Methods

This study adopts a structured research methodology that follows the stages illustrated in the revised flowchart. The methodology begins with a general preparation phase, continues with a detailed KDD process consisting of five analytical steps, and concludes with the development of promotional and bundling strategies derived from discovered association rules. Each stage is implemented sequentially to ensure accuracy, consistency, and relevance of findings toward business decision-making at Flanetqueen Store.

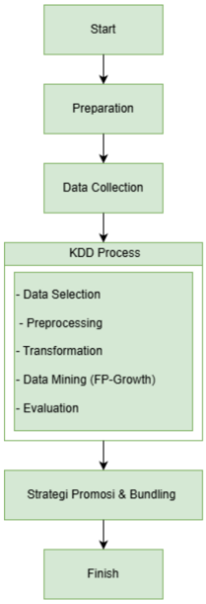


Fig. 1. Research Stages

3.1. Preparation

The preparation stage focuses on organizing and verifying the availability of essential research components. This includes preparing the research design, selecting the appropriate data mining technique (FP-Growth), determining supporting tools (Microsoft Excel and RapidMiner Studio 10.3), and outlining evaluation metrics such as support, confidence, and lift. The purpose of this stage is to ensure that all foundational elements required for the data analysis process are ready and aligned with the study's objectives.

3.2. Data Collection

Relevant hoodie sales transaction data from January to December 2024 were selected from the Flanetqueen Store database. This includes transaction ID, hoodie brand, and timestamp information. Only records contributing to the identification of purchasing patterns were included to ensure the analytical value and focus of the dataset.

3.3. Data Mining Processing

The core methodological framework of this research is based on the Knowledge Discovery in Database (KDD) model. This model consists of five sequential steps that systematically transform raw data into meaningful patterns and actionable business insights.

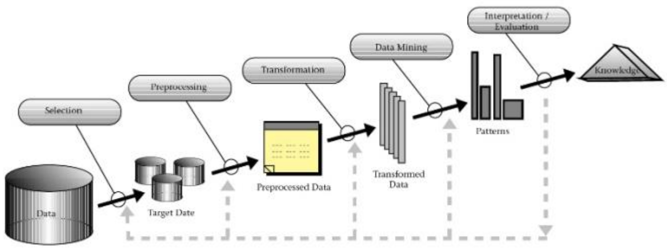


Fig. 2. Knowledge Discovery Precess in Databases (KDD)

3.4. Data Selection

Relevant hoodie sales transaction data from January to December 2024 were selected from the Flanetqueen Store database. This includes transaction ID, hoodie brand, and timestamp information. Only records contributing to the identification of purchasing patterns were included to ensure the analytical value and focus of the dataset.

Table 1: Sales Transaction Data

No	Transaction ID	Date	Product Name	Unit Price
1.	TRX-20240101-001	01/01/2024	Bloods	68500
2.	TRX-20240101-001	01/01/2024	Adidas	105000
3.	TRX-20240101-001	01/01/2024	Adidas	105000
4.	TRX-20240101-001	01/01/2024	3Second	78800

3.5. Pre-processing

To enhance data quality, preprocessing steps were applied to remove missing values, eliminate duplicate records, correct inconsistent item names, and discard irrelevant attributes. This ensures that the dataset is clean, consistent, and free from noise that could negatively impact analytical accuracy.

3.6. Transformation

The cleaned dataset was then transformed into a suitable input format for the FP-Growth algorithm. This included restructuring the data into a transaction-itemset format, renaming attributes for uniformity, and assigning appropriate roles (e.g., transaction ID as identifier). Transformation was performed using RapidMiner operators such as Aggregate, Rename, and Set Role.

3.7. Data Mining (FP-Growth)

The FP-Growth algorithm was applied to the transformed dataset to discover frequent itemsets and generate association rules. The algorithm was configured with:

a. Minimum support: 0.2
b. Minimum confidence: 0.4

FP-Growth was chosen due to its efficiency in avoiding candidate generation and its suitability for analyzing medium-sized retail datasets.

3.8. Evaluation

Each association rule generated by FP-Growth was evaluated using support, confidence, and lift metrics to determine its significance and reliability. Only rules that met or exceeded evaluation thresholds were considered for further analysis. This step ensures that the resulting insights are both statistically meaningful and practically valuable for decision-making.

4. Result and Discussion

4.1. Data Selection

The data that will be used in the data mining process is the transaction data recorded each month during the year 2024. The sales data include all hoodie products sold at Flanetqueen Store and are grouped into categories based on customer purchasing behavior.

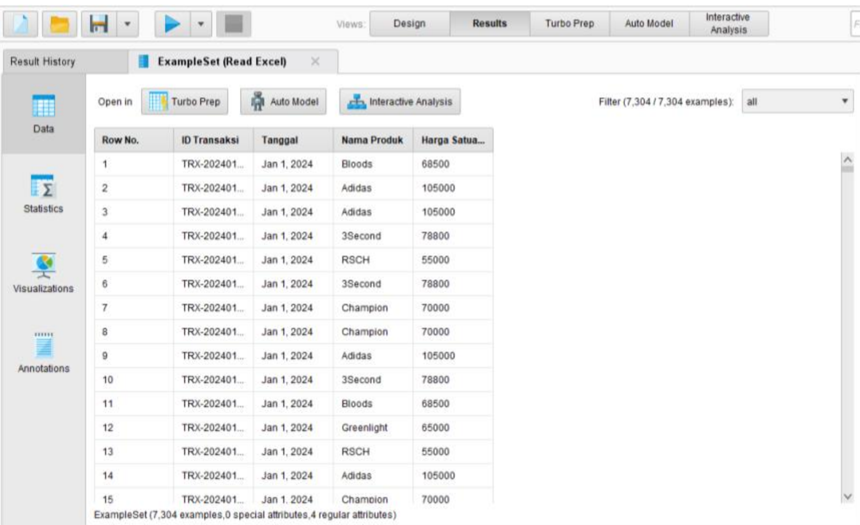


Fig. 3. Data Import Process

4.2. Pre-proccesing

At this stage, the sales transaction data does not have any missing values. Therefore, data cleaning is not necessary.

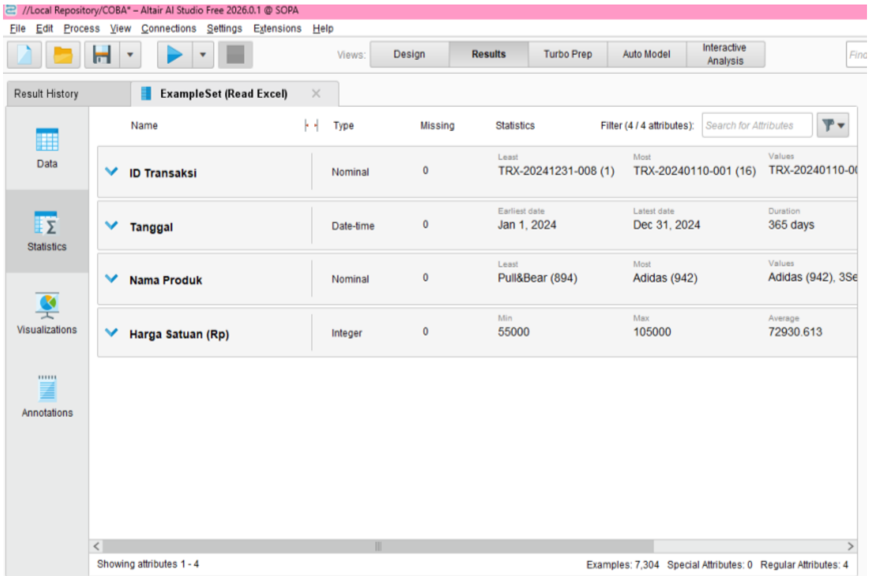


Fig 4. *Preprocessing*

4.3. Transformation

Data transformation aims to convert the dataset from individual transaction forms into a basket structure that includes all items in a single transaction. RapidMiner is used with the Group Aggregation operator to group product names based on transaction IDs. The transformation results in data with the following format:

Tabel 2. Data transformation results

Transaction ID	*Item Collection (Purchased Products)*
TRX-20240101-001	{Bloods, Adidas, 3Second, RSCH, Champion}
TRX-20240101-006	{3Second}

4.4. Data Mining (FP-Growth)

At this stage begins with data selection and then determining the class label, can be seen in Figure 5.

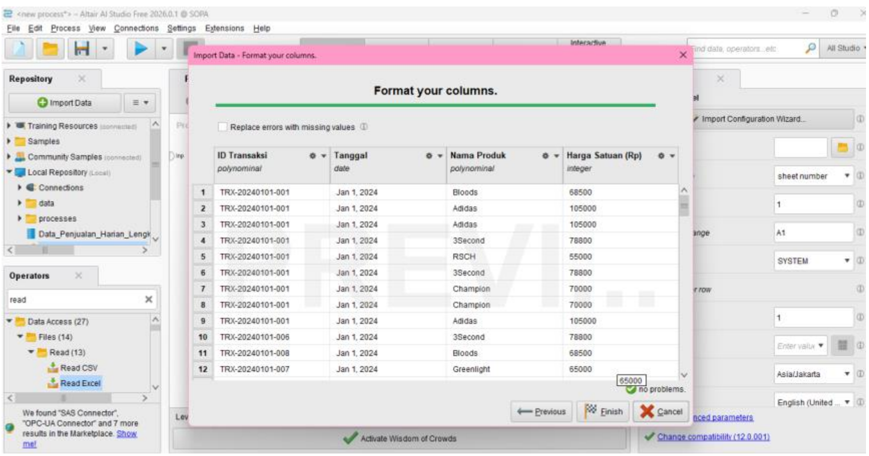


Fig 5. Class Label Determination

The model used in accordance with the FP-Growth algorithm consists of several key operators. The process begins with the Read Excel operator to import the transaction dataset, followed by the Aggregate operator to group items into transaction baskets. The Rename and Set Role operators are then used to adjust attribute names and assign the correct data roles for processing. The FP-Growth operator performs the frequent pattern mining, and the Create Association Rules operator generates the final association rules based on the defined support and confidence values.

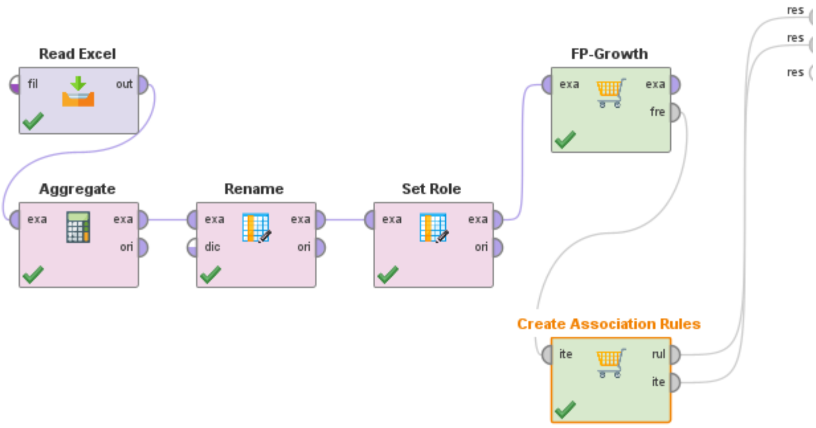


Fig 6. Fp- Growth Model

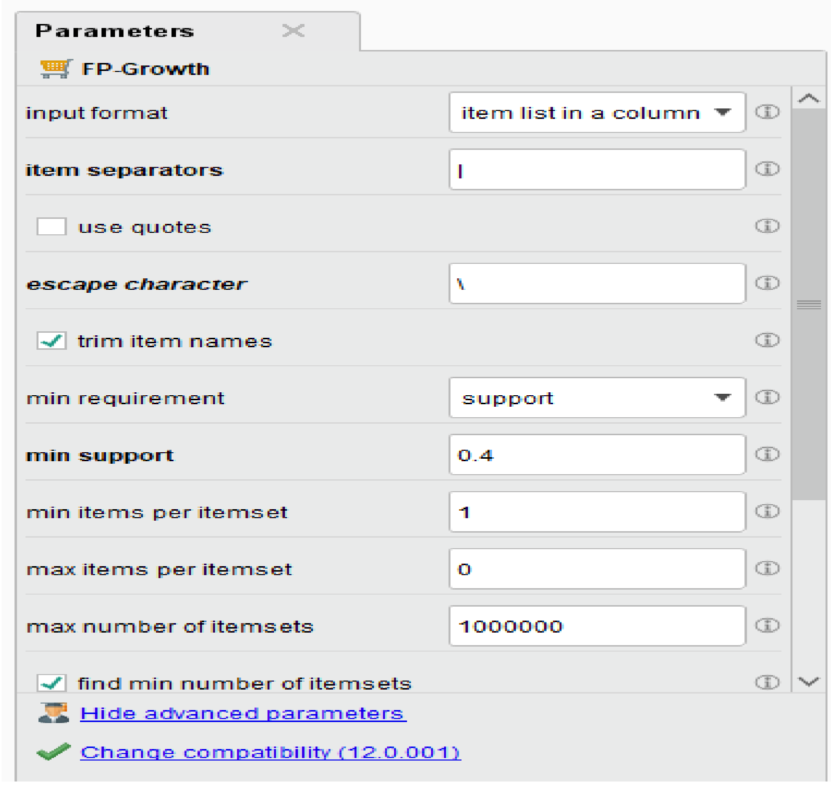


Fig 7. Division of Training Data and Testing Data

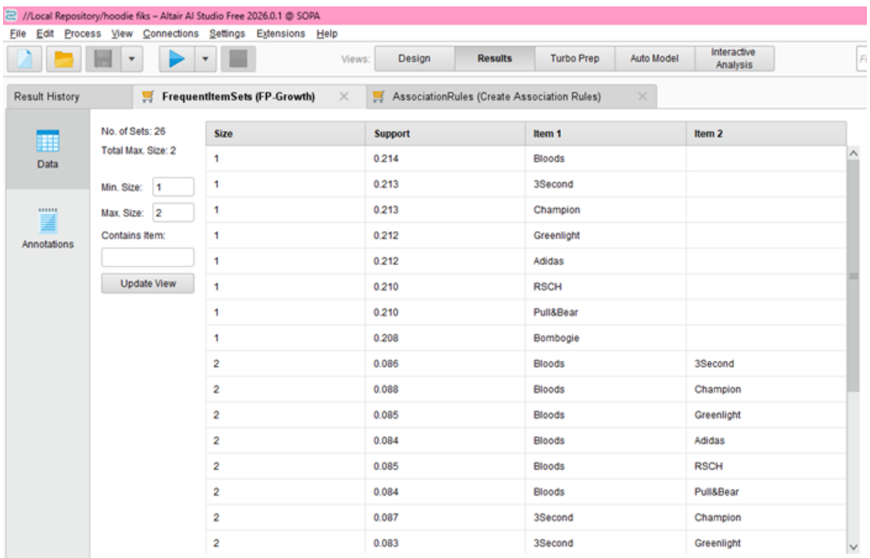
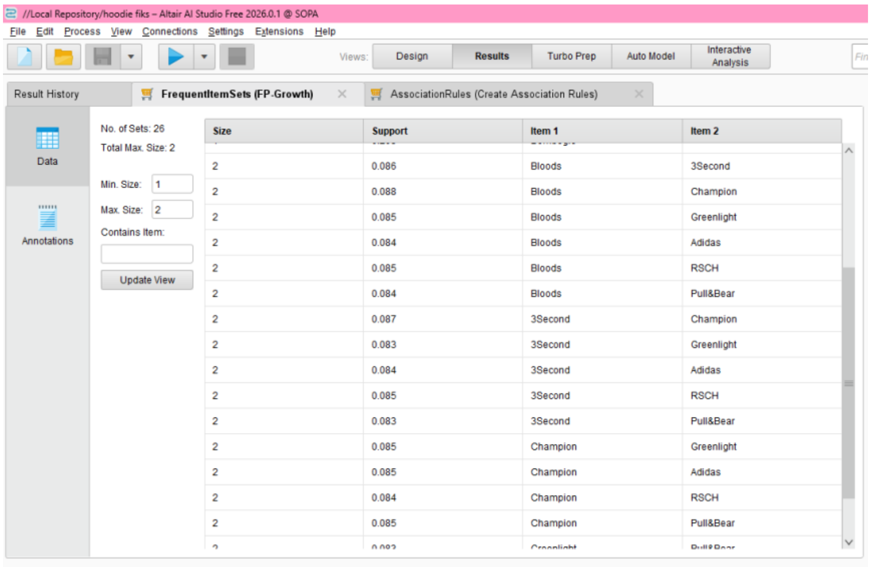


Fig 8. Prediction Results Data from 1 itemset



Size	Support	Item 1	Item 2
2	0.086	Bloods	3Second
2	0.088	Bloods	Champion
2	0.085	Bloods	Greenlight
2	0.084	Bloods	Adidas
2	0.085	Bloods	RSCH
2	0.084	Bloods	Pull&Bear
2	0.087	3Second	Champion
2	0.083	3Second	Greenlight
2	0.084	3Second	Adidas
2	0.085	3Second	RSCH
2	0.083	3Second	Pull&Bear
2	0.085	Champion	Greenlight
2	0.085	Champion	Adidas
2	0.084	Champion	RSCH
2	0.085	Champion	Pull&Bear

Fig 9. Prediction Results Data from 2 itemset

After completing the implementation in RapidMiner using the FP-Growth algorithm, the results can be seen in the visualization below. This pattern indicates a tendency for customers to buy more than one product of a brand with similar characteristics, especially local streetwear brands that are popular among consumers.

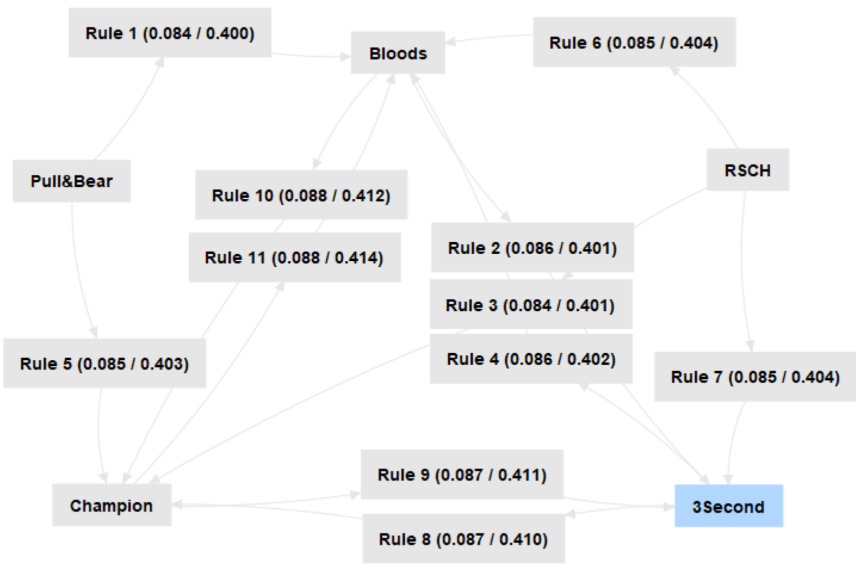


Fig 10. Visualization Results

The association rules generated from the FP-Growth algorithm are visualized in a network structure, where each node represents a hoodie brand and each connecting arrow indicates a rule formed between antecedent and consequent items. The numbers shown in parentheses represent the support and confidence values of each rule. Higher confidence values indicate a stronger likelihood that the consequent item appears when the antecedent item is purchased. The visualization shows that Bloods, Champion, and 3Second act as central nodes because they are frequently found in multiple rules. Rules 10 and 11 demonstrate that Bloods strongly influences the presence of Champion, with confidence values up to 0.414, making it one of the strongest association rules in the model. Similarly, Rules 8 and 9 indicate strong relationships leading toward 3Second, which also appears as a highly connected item in the network. Brands such as Pull&Bear and RSCH appear as supporting nodes showing weaker but relevant associations toward Champion and 3Second. The directional arrows indicate the direction of influence in purchasing patterns. Overall, the network structure visually confirms that Bloods, Champion, and 3Second are the most influential hoodie brands in customer purchasing behavior and form the core of the frequent itemsets found in this study.

AssociationRules

```
Association Rules
[Pull&Bear] --> [Bloods] (confidence: 0.400)
[Bloods] --> [3Second] (confidence: 0.401)
[RSCH] --> [Champion] (confidence: 0.401)
[3Second] --> [Bloods] (confidence: 0.402)
[Pull&Bear] --> [Champion] (confidence: 0.403)
[RSCH] --> [Bloods] (confidence: 0.404)
[RSCH] --> [3Second] (confidence: 0.404)
[3Second] --> [Champion] (confidence: 0.410)
[Champion] --> [3Second] (confidence: 0.411)
[Bloods] --> [Champion] (confidence: 0.412)
[Champion] --> [Bloods] (confidence: 0.414)
```

Fig 11. Avg Result

From the results of trials using the FP-Growth algorithm, the prediction performance produced 11 association rules that revealed patterns of repeat purchases among the streetwear brands studied. The confidence values ranged from 0.400 to 0.414, which means that around 40% to 41.4% of customers who purchased a product were likely to purchase other related products.

5. Conclusion

This study demonstrates the effectiveness of the FP-Growth algorithm in identifying meaningful purchasing patterns within hoodie sales transactions at Flanetqueen Store. The implementation of the complete KDD process—covering data selection, preprocessing, transformation, mining, and evaluation—successfully produced a structured and reliable dataset that enabled the extraction of frequent itemsets and relevant association rules. The most significant contribution of this research lies in revealing strong product relationships that were not previously recognized by the store. The highest-confidence rule, Bloods → Champion, confirms that customers frequently purchase these products together, highlighting a consistent behavioral pattern. Additional rules involving 3Second, RSCH, and Pull&Bear further support the understanding of cross-brand associations that can be strategically leveraged. The novelty of this study is its application of FP-Growth to a localized fashion retail context, demonstrating that lightweight yet powerful data mining techniques can generate actionable insights even for small and medium-sized businesses. The findings provide direct operational value by supporting data-driven decision-making in product bundling, promotional design, and recommendation strategies. Through the identified associations, the store can optimize cross-selling opportunities and enhance customer engagement. Overall, the results confirm that FP-Growth is an efficient and practical solution for analyzing retail transaction data and can be expanded in future work through the integration of predictive models or real-time analytics to further support intelligent retail management.

Acknowledgement

The author's gratitude goes to the rector of the STMIK IKMI Cirebon campus who has lightened the financial burden for the author. Not to forget the author would like to thank the supervisors Mrs. Dra. Nining Rahaningsih, M.Si and Mr. Irfan Ali, M.Kom who patiently guided me until this research was completed.

References

[1] A. Sulistyohati, O. Opitasari, and S. Anisah, "Penerapan data mining menggunakan algoritma FP-Growth untuk pengenalan pola pembelian produk grosir," *Indones. J. Multidiscip. Expert.*, vol. 1, no. 2, pp. 34–41, 2023, doi: 10.31004/ijme.v1i2.20.

[2] L. S. Nasution, W. R. Maya, J. Halim, and M. M, "Data Mining Untuk Menganalisa Pola Pembelian Perak Dengan Menggunakan Algoritma Fp-Growth Pada Toko Emas Dan Perak Adi Saputra Tanjung," *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 3, no. 2, p. 96, 2020, doi: 10.53513/jsk.v3i2.2039.

[3] N. T. Saptadi, P. Chyan, and J. M. Leda, "Analysis of supermarket product purchase transactions with the association data mining method," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 7, no. 3, pp. 618–627, 2023, doi: 10.29207/resti.v7i3.4844.

[4] N. Zhu, "Research on customer relationship segmentation of apparel retail industry through data mining," *HighTech Innov. J.*, vol. 4, no. 2, 2023, doi: 10.28991/HIJ-2023-04-02-05.

[5] M. Namsraidorj, B. Ivanov, B. Sereeter, and E. Bukhsuren, "Increasing retail sales revenue using data mining techniques," *J. Inst. Math. Digit. Technol.*, vol. 4, no. 1, pp. 68–77, 2022, doi: 10.5564/jimdt.v4i1.2663.

[6] V. Shankar, "Big data and analytics in retailing," *NIM Mark. Intell. Rev.*, vol. 11, no. 1, pp. 36–40, 2019, doi: 10.2478/nimmir-2019-0006.

[7] R. Babalghaith, A. Aljarallah, and et al., "Factors affecting big data analytics adoption in small and medium enterprises," *Inf. Syst. Front.*, vol. 26, pp. 2165–2187, 2024, doi: 10.1007/s10796-024-10538-2.

[8] S. H. Bhatti, A. Ahmed, A. Ferraris, and et al., "Big data analytics capabilities and MSME innovation and performance: A double mediation model of digital platform and network capabilities," *Ann. Oper. Res.*, vol. 350, pp. 729–752, 2022, doi: 10.1007/s10479-022-05002-w.

[9] I. Alsolbi, F. H. Shavaki, R. Agarwal, and et al., "Big data optimisation and management in supply chain management: A systematic literature review," *Artif. Intell. Rev.*, vol. 56, no. Suppl 1, pp. 253–284, 2023, doi: 10.1007/s10462-023-10505-4.

[10] Z. Cheng, X. Zhu, and S. Gong, "Low-Resolution Face Recognition," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, pp. 605–621. doi: 10.1007/978-3-030-20893-6_38.

[11] P. Roaida Yanti, Q. Nurkhalisa Maradjabessy, and I. Promasanti Rachmadewi, "Determining the retail sales strategies using association rule mining," *Int. J. Adv. Appl. Sci.*, vol. 13, no. 3, pp. 530–538, 2024, doi: 10.11591/ijaas.v13.i3.pp530-538.