ILKOM Jurnal Ilmiah Vol. 18, No. 2, August 2026, pp. 221-236
Accredited 2nd by KEMDIKTISAINTEK No. 10/C/C3/DT.05.00/2025; E-ISSN 2548-7779 | P-ISSN 2087-1716

Ensemble Association Analysis on Kalla Toyota's Employee Data Using Apriori, FP-Growth, and ECLAT

Ahmad Zaky [a,1,*]; Lilis Nur Hayati [a,2]; Herdianti Darwis [a,3]; Roesman Ridwan Raja [b,4]
[a] *Universitas Muslim Indonesia, Jl. Urip Sumoharjo No.km.5, Panaikang, Kec. Panakkukang, Kota Makassar, Sulawesi Selatan 90231, Indonesia*
[b] *Kyushu Institute of Technology, 1-1 Sensuicho, Tobata Ward, Kitakyushu, Fukuoka 804-8550, Japan*
[1] *13020210052@umi.ac.id;* [2] *lilis.nurhayati@umi.ac.id;* [3] *herdianti.darwis@umi.ac.id;* [4] *raja.roesman-ridwan757@mail.kyutech.jp*
** Corresponding author*

Article history: Received March 16, 2025; Revised April 22, 2025; Accepted April 22, 2026; Available online August 10, 2026.

Abstract

The effective management of employee data plays an important role in supporting organizational decision-making, particularly in today's data-driven business environment. This study examines the identification of association patterns within employee data at Kalla Toyota by applying a combined approach using the Apriori, ECLAT, and FP-Growth algorithms. The dataset includes information such as demographic characteristics, educational background, marital status, and job classification. Prior to analysis, the data were carefully preprocessed to improve consistency and ensure suitability for pattern discovery. Relevant variables were then selected using statistical measures, including Cramér's V, Kendall's Tau, and Chi-Square tests, to capture meaningful relationships among attributes. With a minimum support threshold set at 10%, the combined method produced 84 association rules considered significant. These patterns were further explored using visual tools such as network graphs and matrix plots to better understand the relationships between variables. The findings highlight notable connections among factors such as gender, generational groups, job roles, and marital status. These insights may assist the company in refining its human resource strategies, particularly in areas such as recruitment, employee development, and retention. This study shows that combining multiple association rule techniques can provide a more comprehensive understanding of employee data and support more informed decision-making.

Keywords: Association Rules; Apriori; FP-Growth; ECLAT; Frequent Itemsets.

Introduction

In today's competitive organizational landscape, effective management of employee data is crucial to improving operational efficiency and supporting strategic decision-making [1], [2]. Human Resource Analytics, driven by data mining techniques, has become an essential tool for uncovering hidden patterns that inform workforce planning, recruitment, and retention strategies.

Employee datasets typically contain diverse attributes such as demographic profiles, educational backgrounds, marital statuses, and job roles, offering valuable insights when analyzed appropriately [2]. However, the increasing complexity and volume of such data often exceed the capabilities of conventional analytical methods, necessitating the use of more advanced techniques such as association rule mining [4], [5].

Previous studies have demonstrated the effectiveness of Apriori, FP-Growth, and ECLAT algorithms in discovering frequent itemsets across various domains such as retail, healthcare, and social sciences [4], [5], [6], [7]. However, relatively few studies have explored the ensemble integration of these algorithms for human resource datasets, particularly in employee data analysis. Ensemble approaches have been widely applied in human resource analytics to enhance accuracy in classification and pattern exploration tasks, including studies on employee attrition [16].

To address this gap, this study proposes an ensemble framework that combines Apriori, FP-Growth, and ECLAT algorithms to uncover frequent patterns in Kalla Toyota's employee data. The contribution of this study is positioned in the systematic validation of pattern consistency across multiple association rule algorithms. Rather than focusing solely on methodological combination, the framework evaluates the degree of agreement among Apriori, FP-Growth,

and ECLAT to strengthen the interpretability and robustness of the extracted patterns in human resource analytics. Furthermore, identified patterns are visualized through network graphs and matrix plots to support practical interpretation, offering actionable insights for strategic workforce development.

This analytical approach aligns with the growing demand for data-informed strategies in human capital management, where decisions are increasingly expected to be supported by empirical insights. By adopting a systematic framework that integrates proven data mining methods, organizations can better align their workforce strategies with operational goals while navigating the dynamic challenges of today's labor environment.

Method

As depicted in **Figure 1**, this study encompasses five primary stages that are systematically organized to address the research objectives.

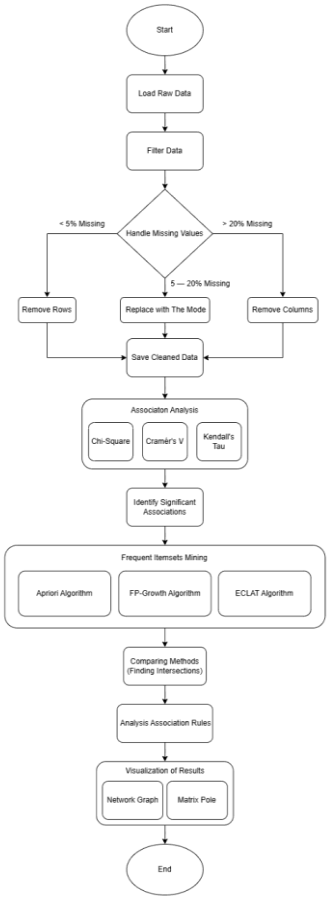

Figure 1. *Flowchart of the proposed methodology*

While the study adopts a five primary stages methodology at its core—comprising data preprocessing, association analysis, frequent itemset mining, algorithm comparison, and visualization—this flowchart further expands those stages into granular sub-steps for enhanced interpretability.

A. Data Pre-Processing

This stage focuses on transforming raw data from the company into a format suitable for subsequent analysis. Initially, the dataset comprised 82 columns and 1,000 entries, which presented challenges in terms of relevance and manageability. To address this, a filtering process was conducted to retain only the entries and columns deemed essential for the analysis. Irrelevant columns, such as email, phone number, and religion, were removed to streamline the dataset and eliminate unnecessary information. This meticulous curation not only reduced complexity but also ensured that the data was more focused and aligned with the objectives of the study, providing a solid foundation for further analytical processes.

Table 1 summarizes the dataset used in this study, detailing variables, their possible values, and types. The dataset includes demographic and employment-related attributes of employees at Kalla Toyota. For instance, the variable *EDUCATION* (Education) categorizes employees based on their highest educational attainment, ranging from elementary school *(SD)* to doctoral degree *(S3)*. *Jenis Kelamin* (Gender) distinguishes employees as 'P' for female and 'L' for male. *GENERATION* (Generation) groups employees into generational cohorts: Baby Boomer, Gen X, Millennial, and Gen Z, providing insight into generational differences. The variable *ROLE TYPE* (Role Type) indicates whether employees are in direct operational roles or supporting roles.

Furthermore, the *Status Kawin* (Marital Status) variable uses 'L' for single, 'K' for married, and 'C' for divorced, with numbers denoting the number of children. *Usia* (Age), a numerical variable, records the age range of employees from 19 to 62 years. Employment types are classified in the *Status* variable as 'MG' (internship), 'K' (contract), and 'T' (permanent). Lastly, *Golongan Darah* (Blood Type) identifies employees' blood types, including A, B, B+ AB, and O. This dataset captures diverse employee attributes, allowing for a comprehensive analysis of workforce patterns.

Missing data were addressed based on their proportion: columns with less than 5% missing data had the rows with missing values removed, columns with 5–20% missing values were imputed using the mode, and columns with more than 20% missing values, such as blood type were eliminated entirely. Deletion methods are appropriate when the proportion of missing data is small, while imputation techniques help maintain dataset integrity when a moderate amount of data is missing [8].

After addressing missing values and removing duplicate rows, the dataset was systematically refined, resulting in a reduced structure consisting of 21 columns and 977 entries. Among the retained columns were key attributes such as generation, highest education, and job role, which were deemed essential for subsequent analysis. This meticulous preprocessing step was crucial in ensuring that the dataset was both clean and well-structured, thereby minimizing noise and inconsistencies.By creating a reliable and organized foundation, this process aimed to enhance the validity and accuracy of the forthcoming analytical processes, allowing for more robust insights to be derived.

B. Association Analysis

Association Association analysis aims to uncover relationships or patterns among variables in a dataset. This analytical approach is frequently utilized to explore the strength and nature of associations between variables, particularly in fields such as medical research and social sciences, where understanding variable interdependencies is crucial [2], [3]. In this study, the objective is to investigate the correlations and dependencies between various employee attributes using association rule mining techniques.

These methods are pivotal in revealing how different characteristics interact, guiding informed decision-making processes and subsequent analyses. Three widely recognized statistical methods were applied: Cramér's V, Kendall's Tau, and the Chi-Square test. These methods offer robust tools for evaluating the strength and significance of associations between categorical variables.

The Chi-Square test is a statistical method used to evaluate the association between two categorical variables. It compares the observed frequencies O in a contingency table with the expected frequencies E derived under the assumption of independence. This method is widely applied in various disciplines to test hypotheses about categorical data, including goodness-of-fit, independence, and homogeneity [11]. Equation (1) provides the formula for calculating the Chi-Square statistic.

$$x^2 = \sum \frac{(O - E)^2}{E} \tag{1}$$

In the Chi-Square formula, the summation symbol Σ indicates the process of adding all contributions from each category in the dataset. These contributions are based on the differences between the observed O and expected E frequencies. The observed frequency O reflects the actual distribution of values recorded, while the expected frequency E is calculated under the assumption of independence between variables. The summation provides an aggregate measure of how much the observed data deviates from what is expected. This aggregate is expressed as the Chi-Square statistic (x^2), which quantifies the overall strength and significance of the relationships within the data. The process is crucial for understanding patterns and dependencies, and the result forms the basis for determining degrees of freedom

(df). In the context of a contingency table with r rows and c columns, the degrees of freedom are calculated using the formula shown in (2).

$$df = (r - 1)(c - 1) \quad (2)$$

After performing the necessary calculations to obtain the Chi-Square statistic and identifying the degrees of freedom, the next critical step involves computing the p-value. This value is instrumental in assessing the statistical significance of the observed relationship between the variables under investigation. The Chi-Square statistic serves as a measure of the extent to which the observed frequencies deviate from the expected frequencies, assuming that the variables are independent and follow the null hypothesis.

The p-value provides a probabilistic framework to evaluate the likelihood of obtaining a Chi-Square statistic that is as extreme as, or more extreme than, the one calculated from the data, given that the null hypothesis holds true. In this context, the null hypothesis posits that no association exists between the variables, implying their independence.

A smaller p-value signifies stronger evidence against the null hypothesis, thus suggesting the presence of a statistically meaningful association between the variables. This interpretation is particularly important in determining whether the observed patterns in the data arise purely from random chance or reflect a genuine relationship that warrants further exploration. As such, the p-value serves as a cornerstone of hypothesis testing, providing a basis for informed decision-making in empirical research. It enables researchers to assess results with a quantified level of confidence, thereby strengthening the credibility of analytical outcomes and supporting transparent evaluation in scientific inquiry.

Table 2 demonstrates the Chi-Square test results for several pairs of categorical variables in the dataset. Each pair exhibits statistically significant associations, as indicated by their low p-values (all below 0.05). For instance, the relationship between *STATUS* and *LAYER* (Chi-Square: 122.888, p-value: 2.96×10^{-15}) suggests notable differences in employee statuses across organizational layers. Similarly, *EDUCATION* and *LAYER* (Chi-Square: 834.104, p-value: 2.22×10^{-102}) reflects how educational qualifications influence organizational placement. Significant associations are also observed between *LAYER* and *GENERATION* (Chi-Square: 651.131, p-value: 6.247×10^{-114}), and *EDUCATION* and *STATUS* (Chi-Square: 124.762, p-value: 2.36×10^{-16}), highlighting potential patterns related to generational trends and the impact of education on job roles. Overall, these findings reveal meaningful dependencies among the variables, providing a foundation for further analysis and interpretation.

The interpretation of the p-value is as follows:

P-value < 0.05: Indicates a statistically significant association between the variables, leading to the rejection of the null hypothesis.

P-value > 0.05: Suggests that the observed association could be due to random chance, and the null hypothesis cannot be rejected

Extensions such as partial and average conditional Kendall's Tau have been introduced to account for conditional dependencies, providing a more flexible tool for analyzing complex relationships in structured data [12]. This method is particularly effective when the variables are ranked or have ordinal data. The formula for calculating Kendall's Tau is given by the following equation (3).

$$t = \frac{(C - D)}{\sqrt{(C + D + T_x)(C + D + T_y)}} \quad (3)$$

Kendall's Tau is calculated based on several key components: C, which represents the number of concordant pairs where the rank order of both variables aligns; D, which indicates the number of discordant pairs where the rank order of the variables does not align; Tx, referring to the number of tied ranks in the first variable; and Ty, referring to the number of tied ranks in the second variable. The value of Kendall's Tau ranges from -1 to 1, where t = 1 signifies a perfect positive association, indicating that all pairs are concordant; t = -1 indicates a perfect negative association, where all pairs are discordant; and t = 0 reflects no association between the variables.

To assess the statistical significance of Kendall's Tau, a p-value is calculated. The p-value represents the probability of observing a correlation as extreme as or more extreme than the one found in the data, assuming no correlation exists, as show in (4).

$$(H_0 + t = 0) \quad (4)$$

The interpretation of the p-value provides insight into the statistical significance of a correlation. A p-value less than 0.05 indicates a significant correlation, leading to the rejection of the null hypothesis H_0. Conversely, a p-value equal to or greater than 0.05 suggests the absence of a statistically significant correlation, meaning that the null hypothesis H_0 cannot be rejected. This threshold is widely used to determine whether observed patterns in the data are likely due to chance or reflect true underlying relationships.

Table 3 displays the results of Kendall's Tau test, which evaluates the strength and direction of monotonic relationships between pairs of variables. For concordant pairs, *STATUS* and *LAYER* (t = 0.280, p-value = 4.613×10^{-21}) and *EDUCATION* and *LAYER* (t = 0.250, p-value = 2.574×10^{-18}) indicate moderate positive associations, suggesting aligned trends between these variables. Similarly, *LAYER* and *GENERATION* (t = 0.158, p-value = 8.862×10^{-8}) and *GENERATIONI* and *STATUS* (t = 0.360, p-value = 4.061×10^{-33}) show weaker yet significant positive correlations.

Conversely, discordant pairs, such as *EDUCATION* and *STATUS* (t = -0.209, p-value = 5.355×10^{-13}) and *EDUCATION* and *GENERATION* (t = -0.040, p-value = 1.674×10^{-1}), exhibit negative associations. The strength of these relationships varies, with the latter being weaker. Overall, the findings highlight both concordant and discordant relationships, offering insight into the alignment or divergence between key variables in the dataset.

Cramér's V is a widely used statistical measure designed to evaluate the strength of association between two categorical variables within the context of a contingency table. This metric is derived from the chi-square statistic and provides a standardized value that facilitates the interpretation of relationships between the variables.

The value of Cramér's V ranges from 0 to 1, where a value of 0 signifies the complete absence of association, indicating that the variables are entirely independent. Conversely, a value of 1 denotes a perfect association, reflecting a deterministic relationship where one variable can be fully predicted from the other. Intermediate values indicate varying degrees of association, with higher values corresponding to stronger relationships. The bounded nature of this metric makes it especially useful in comparing the relative strength of multiple variable pairs within a dataset, allowing analysts to prioritize the most relevant relationships for further analysis.

Due to its bounded nature, Cramér's V is particularly advantageous in comparing associations across different datasets or studies, regardless of the size of the contingency table or the scale of the data. This makes it a valuable tool in fields such as social sciences, marketing, and biomedical research, where understanding categorical relationships is critical for drawing meaningful conclusions from data. It is particularly useful for contingency tables with more than two categories for each variable [6], as show in (5).

$$v = \sqrt{\frac{x^2}{n(k-1)}} \quad (5)$$

Cramér's V is calculated using the Chi-Square statistic x^2, the total number of observations n, and the smaller of the number of rows (r) or columns (c) in the contingency table k. This measure provides a standardized value to assess the strength of association between categorical variables. The interpretation of Cramér's V is based on its value: a range between 0.0 and 0.1 indicates little to no association, values from 0.1 to 0.3 suggest a weak association, values between 0.3 and 0.5 imply a moderate association, and values greater than 0.5 represent a strong association. This framework allows researchers to quantify the relationship's intensity, aiding in meaningful data interpretation.

Table 4 presents the results of Cramér's V analysis, which measures the strength of associations between pairs of nominal variables. Most relationships exhibit weak associations, such as *STATUS* and *LAYER* (Cramér's V = 0.251) and *EDUCATION* and *LAYER* (Cramér's V = 0.278). A moderate association is observed between *LAYER* and *GENERATION* (Cramér's V = 0.471), indicating a stronger connection compared to the weakly associated pairs.

Notably, the relationship between *EDUCATION* and *GENERATION* shows a strong association (Cramér's V = 0.608), reflecting a significant dependency between educational background and generational grouping. Meanwhile, *EDUCATION* and *STATUS* (Cramér's V = 0.253) and *GENERATION* and *STATUS* (Cramér's V = 0.283) maintain weak associations. These findings offer an overview of how strongly key categorical variables are interconnected,

emphasizing certain notable relationships for deeper exploration. Cramér's V is a normalized measure that quantifies the strength of association between two categorical variables, independent of their scale or distribution [7].

C. Frequent Itemsets Mining

Frequent itemsets mining (FIM) identifies co-occurring combinations of variables (itemsets) within a dataset, foundational in association rule mining and widely applied in fields like market basket analysis and employee data analysis [29]. It evaluates itemsets based on support, the frequency of transactions containing them, to uncover meaningful patterns and relationships [8].

This study employs Apriori, FP-Growth, and ECLAT algorithms to analyze employee attributes, minimum support threshold of 0.1 (10%) was consistently applied among all three algorithms to identify frequent itemsets. A minimum support threshold of 10% was adopted, aligning with recommendations in previous studies suggesting thresholds between 5–20% as optimal for medium-sized datasets [15], [16]. Apriori leverages a join and prune approach to iteratively discover itemsets, emphasizing simplicity and effectiveness [9]. FP-Growth optimizes performance by using an FP-tree structure, reducing memory usage and processing time, particularly for large datasets [10]. ECLAT, with its depth-first search and intersection-based approach, excels in processing sparse datasets efficiently [11].

Apriori is more popular algorithm used to find all frequent item sets [12], [14]. The Apriori algorithm has been successfully applied in various domains to uncover meaningful patterns from transactional data [30]. **Table 5** lists the frequent itemsets identified using the Apriori algorithm, along with their corresponding support values. The support metric represents the proportion of transactions or records in the dataset that contain a specific itemset. For example, the combination *'LAYER STAFF'* and *'GENDER MALE'* has the highest support (0.608), indicating it is the most frequently occurring pair in the dataset. Similarly, the itemset *'GENERATION MILLENNIAL'* and *'LAYER STAFF'* exhibits a support value of 0.494, while *'STATUS PERMANENT'* and *'GENDER MALE'* appear with a support of 0.483.

Towards the lower end, itemsets like *'STATUS CONTRACT,' 'GENERATION GEN Z,' 'LAYER STAFF'* have a lower support value of 0.101, reflecting less frequent occurrences. These insights help in identifying prominent patterns and associations in the dataset, which may guide further analysis or decision-making. Let C_k denote the set of candidate itemsets of length k, where each itemset consists of k items. From this candidate set, we derive the set of frequent itemsets of length k, denoted by L_k, which includes all itemsets $X \in C_k$ whose support value meets or exceeds the predefined minimum threshold. The formula show in (6) and (7).

$$L_k = \{X \in C_k \mid Support(X) \geq Minimum\ Support\} \tag{6}$$

$$Support(X) = \frac{Total\ number\ of\ transactions\ containing\ X}{Number\ of\ transactions} \tag{7}$$

Ck is the set of candidates itemsets of length k

FP-Growth uses the FP-Tree technique which is famous for the divide and conquer methods and does not generate itemsets candidate generation [13], [18]. Formula for Building Frequent Itemsets, as show in (8).

$$Building\ Frequent = \{(X, Support) \mid X\ shares\ a\ prefix\ with\ an\ FP-tree\ path\} \tag{8}$$

To apply the FP-Growth algorithm to item data, the first step is to identify frequent itemsets by constructing an FP-Tree. After the pattern is established, the next step is to determine the Conditional Pattern Base, followed by the creation of the Conditional FP-Tree.

Table 6 highlights the frequent itemsets generated using the FP-Growth algorithm, along with their support values. The top-ranked itemset, *'GENERATION MILLENNIAL'* and *'LAYER STAFF'*, has highest support value of 0.494, signifying it as the most common combination in the dataset. Following this, *'LAYER STAFF,' 'ROLE TYPE DIRECT,'* and *'JENIS KELAMIN L'* exhibit a support value of 0.388, indicating a relatively strong presence as well.

Other notable patterns include *'EDUCATION BACHELOR DEGREE'* and *'GENERATION MILLENNIAL'* with a support of 0.348, as well as *'GENERATION MILLENNIAL'* and *'ROLE TYPE SUPPORTING'*, which have a support value of 0.316. Lower in the table, itemsets such as *'STATUS KAWIN L'* and *'LAYER STAFF'* have a support value of 0.309, while combinations like *'STATUS T,' 'ROLE TYPE SUPPORTING,' 'EDUCATION BACHELOR DEGREE'*

show a support of 0.255. The smallest frequent itemset reported is *'STATUS KAWIN L,' 'GENERATION GEN Z'*, with a support of 0.207.

These patterns provide insights into recurring relationships among attributes, enabling a deeper understanding of the dataset's structure and helping inform strategic decisions or further analysis, completing the formation of association rules through the FP-Growth algorithm [21].

Various algorithms have been introduced to process frequent itemsets. ECLAT based algorithms are one of the prominent algorithm used for frequent itemsets mining [22]. ECLAT algorithm is a depth first search-based algorithm, represents itemsets using the Transaction ID (TID) [7]. Formula frequent itemsets for ECLAT, as show in (9) and (10).

$$TID(X \cap Y) = TID(X) \cap TID(Y) \quad (9)$$

$$Support(X \cap Y) =| TID(X \cap Y) \quad (10)$$

In the context of association rule mining, $TID(X)$ represents the set of transaction IDs that include the itemset X. This allows for identifying all transactions in which the specified itemset occurs. Similarly, $X \cap Y$ denotes the intersection or combination of itemsets X and Y, referring to the set of transactions that contain both itemsets simultaneously. These concepts are fundamental in determining the relationships and patterns among itemsets within a dataset.

Table 7 presents the frequent itemsets identified using the ECLAT algorithm, showcasing patterns with varying support levels among employee attributes at Kalla Toyota. The highest support value is 0.478, corresponds to the itemset *'JENIS KELAMIN L'* and *'GENERATION MILLENNIAL'*, indicating that male employees from the millennial generation are the most prominent group in the dataset. Similarly, combinations such as *'GENERATION MILLENNIAL'* and *'LAYER STAFF'* with support 0.400, *'LAYER STAFF'* and *'EDUCATION BACHELOR DEGREE'* with support 0.380 suggest strong associations between millennial employees, staff-level positions, and bachelor's degree education.

Toward the bottom of the table, lower support values such as 0.242 for *'JENIS KELAMIN L', 'LAYER STAFF',* and *'KATEGORI JABATAN AFTERSALES'* highlight less frequent but specific patterns involving male employees working in aftersales roles. These findings provide critical insights into the demographic and professional structure of the workforce, allowing HR teams to design tailored policies for different employee segments. For example, the dominance of millennial males in staff positions could guide training programs or career development plans to address their specific needs and aspirations.

D. Algorithm Comparison Metrics

In this study, the comparative evaluation of the algorithms focuses on analyzing the intersection of frequent itemsets generated by the Apriori, FP-Growth, and ECLAT algorithms. This evaluation adopts a metric-driven approach designed to assess the consistency, robustness, and reliability of each algorithm in uncovering significant patterns embedded within the dataset. By examining the overlapping frequent itemsets, the analysis emphasizes the shared insights derived from these distinct methodologies. This process not only highlights the areas of convergence in the outputs but also contributes to a broader understanding of the relative effectiveness of these algorithms in the context of association rule mining, showcasing their ability to reveal meaningful patterns.

The primary objective of this comparison is to quantify the degree to which the three algorithms align in their outputs under identical parameter settings. This alignment is assessed by configuring all algorithms with a uniform minimum support threshold of 0.1, as depicted in **Figure 2**. The choice of this specific threshold ensures that the comparisons are conducted on an equitable basis, facilitating a straightforward evaluation of the frequent itemsets generated by each method. By standardizing the parameterization, the study guarantees a fair comparative framework, enabling a more precise examination of the commonalities and potential differences in the results produced by Apriori, FP-Growth, and ECLAT.

The results from **Figure 2** emphasize the shared and distinct patterns identified, providing critical insights into the strengths and limitations of each algorithm. Evaluating the intersection of frequent itemsets provides deeper insights into the patterns consistently identified among algorithms, aiding in reducing redundancy and ensuring a more accurate interpretation of the discovered patterns [23].

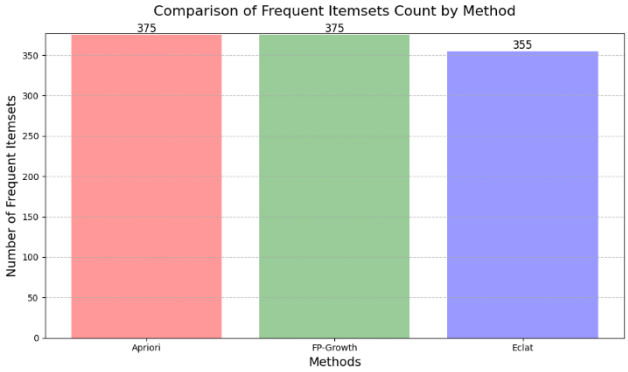

Figure 2. *Number of frequent itemsets on Aprori, FP-Growth, and ECLAT*

To achieve this objective, the frequent itemsets generated by each algorithm were systematically examined through a structured cross-comparison process. This procedure was intended to analyze variations and similarities in the outputs produced by Apriori, FP-Growth, and ECLAT when applied under identical parameter settings. By evaluating these results in parallel, the analysis is able to distinguish patterns that are algorithm-specific from those that consistently appear across methods. Such consistency is particularly important for assessing the reliability of frequent itemsets, as it indicates that the identified patterns are not merely artifacts of a single algorithmic mechanism. Through this comparative examination, overlapping itemsets are emphasized as stronger candidates for meaningful interpretation within the context of employee data analysis.

As depicted in **Figure 3**, the intersection of frequent itemsets illustrates patterns that are jointly identified by all three algorithms. This overlap represents a critical analytical outcome, as it reflects stability in pattern discovery despite inherent differences in search strategies, data representations, and traversal mechanisms employed by each method. Itemsets that persist across multiple algorithms suggest a higher level of structural relevance within the dataset and reduce the likelihood of spurious associations. From an analytical perspective, these intersected patterns can be regarded as more dependable indicators of underlying workforce characteristics. Their repeated emergence strengthens confidence in the observed associations and supports their use as key findings in subsequent interpretation.

The results further highlight the methodological advantages of adopting an ensemble-based approach for association rule mining. By combining multiple algorithms, the analysis leverages complementary strengths while mitigating the limitations inherent in individual techniques. This integration improves the overall robustness of pattern detection and enhances the interpretability of the findings. Moreover, ensemble strategies provide a balanced analytical framework that is well-suited for complex organizational datasets, where relationships among variables are often multidimensional. Consequently, the use of ensemble methods not only reinforces the validity of the discovered patterns but also demonstrates their practical relevance for supporting evidence-based decision-making and strategic planning in human resource management.

This reinforces the growing importance of ensemble methods in real-world applications, where complexity and data heterogeneity demand more robust and adaptive analytical frameworks.

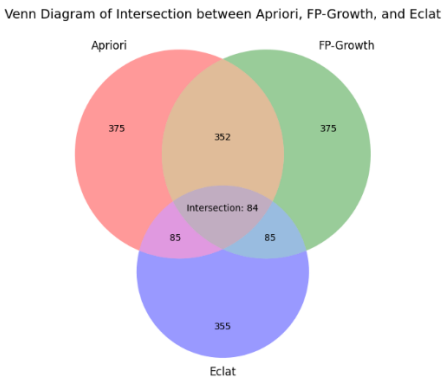

Figure 3. *Intersection from the methods*

E. Frequency Analysis with Visualization

Frequency analysis is a critical step in understanding the patterns of itemsets occurrences within a dataset. The technology of data visualization can extract and analyze huge amounts of data, extract useful data and present them in the form of images or graphics [24]. This study utilizes two primary visualization methods, namely the network graph and the matrix pole, to offer deeper and more intuitive insights into the relationships between itemsets and their frequency distribution.

In **Figure 4**, a network graph is employed to visually represent the relationships between itemsets based on their co-occurrence frequency within the dataset. Each node in the graph corresponds to a specific itemset, while the edges connecting the nodes illustrate the associations between two itemsets. The strength of these relationships is encoded through the thickness of the edges: thicker edges indicate higher frequencies of co-occurrence, signifying stronger associations.

This visual structure enables the identification of central itemsets that frequently appear across multiple associations, thereby highlighting their relative importance within the dataset. Peripheral nodes with fewer or thinner connections may represent less influential patterns, though they still contribute to the overall mapping of the association space. The layout of the network allows for intuitive clustering, where tightly connected subgroups of itemsets may suggest thematic groupings or shared characteristics. Such visual insights support not only interpretability but also communication of analytical results to stakeholders without technical backgrounds.

This graphical representation enables an intuitive understanding of how itemsets interact and cluster within the dataset. Additionally, the network graph includes nodes that represent single-item itemsets, such as *('EDUCATION_SMA')*, to ensure a comprehensive visualization. Even in cases where these single-item itemsets lack connections to other nodes, their inclusion is vital for capturing the entirety of the dataset's structure. By visualizing the dataset in this manner, Fig. 4 provides valuable insights into the dynamics and patterns underlying the relationships among various employee attributes. This approach emphasizes the effectiveness of network graphs in visualizing complex relational data, making abstract associations easier to interpret and communicate [25].

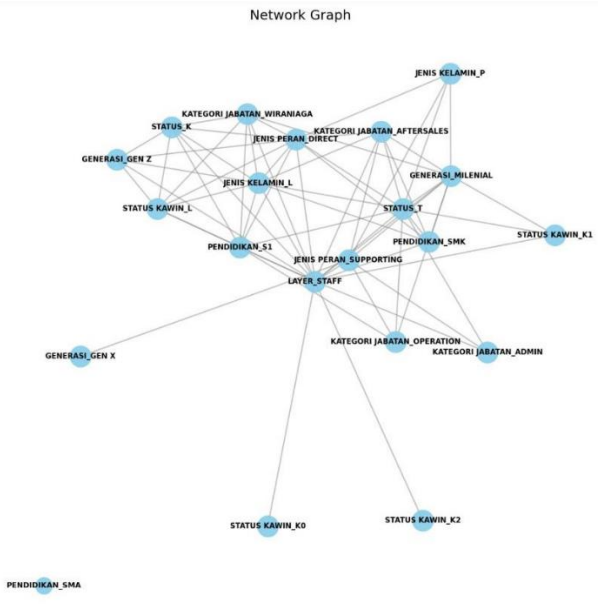

Figure 4. *Network Graph*

The heatmap matrix depicted in **Figure 5** serves as a comprehensive visualization tool for examining the relationships and associations between various categorical variables within the dataset. Using a gradient color scale, this visualization effectively highlights the strength of associations, where darker shades represent stronger relationships, and lighter shades indicate weaker or negligible connections. This structured approach transforms complex data into an accessible and interpretable format, enabling users to uncover key insights that might otherwise remain hidden in raw numerical data.

Specifically, the heatmap reveals notable patterns among variables such as job categories, educational levels, marital status, gender, generational cohorts, and role classifications. For example, the matrix may indicate stronger associations between certain job categories and specific generational groups or highlight the correlation between educational levels and organizational roles. These patterns are crucial in understanding workforce characteristics and dynamics, providing a foundation for actionable insights.

In organizational contexts, these insights can inform a wide range of strategic decisions. For instance, understanding how educational attainment correlates with job roles can guide recruitment and talent development strategies, while identifying generational clusters may help shape policies for engagement and retention. By leveraging this heatmap as an analytical tool, organizations can better align their workforce strategies with broader goals, ensuring data-driven decision-making and more efficient resource allocation.

Furthermore, the matrix format simplifies the interpretation of complex associations by presenting them in a clear, color-coded layout, enhancing accessibility for both technical and non-technical stakeholders. The application of matrices in data visualization is well-established, combining structural clarity and interpretability through advanced methods [26], [27].

Figure 5. *Matrix Pole*

Results and Discussion

The employee data analyzed in this study consist of demographic, educational, and employment-related variables, as summarized in **Table 1**.

Table 1. Dataset of Kalla Toyota's Employee

No	Variable	Value	Type
1	*'EDUCATION'* (EDUCATION)	SD, SMP, SMA, SMK, MA, BACHELOR DEGREE, S2, S3	Categorical
2	*'JENIS KELAMIN'* (GENDER)	P and L	Categorical
3	*'GENERATION'* (GENERATION)	BABY BOOMER, GEN X, MILLENNIAL, GEN Z	Categorical
4	*'ROLE TYPE'* (ROLE TYPE)	DIRECT and SUPPORTING	Categorical
…	…	…	…

No	Variable	Value	Type
79	*'STATUS KAWIN'* (MARITAL STATUS)	L, K2, K1, K0, K3, K4, C2, K5, C1, K6 , C3	Categorical
80	*'USIA'* (AGE)	19 – 62	Numerical
81	*'STATUS'* (EMPLOYMENT TYPE)	MG, K, T	Categorical
82	*'GOLONGAN DARAH'* (BLOOD TYPE)	A, B, B+, AB, O	Categorical

These variables represent the core attributes used to examine structural patterns within the workforce and to support subsequent association analyses. Statistical associations among categorical variables were evaluated using the Chi-Square test, Kendall's Tau, and Cramér's V. The Chi-Square results Table 2.

Table 2. Chi-Square Test Result

Column	Chi-Square Statistic	Degrees of Freedom	P-value
STATUS and *LAYER*	122.888	24	2.96×10^{-15}
EDUCATION and *LAYER*	834.104	132	2.22×10^{-102}
LAYER and *GENERATION*	651.131	36	6.247×10^{-114}
…	…	…	…
Column	**Chi-Square Statistic**	**Degrees of Freedom**	**P-value**
EDUCATION and *STATUS*	124.762	22	2.36×10^{-16}
EDUCATION and *GENERATION*	1084.08	33	1.86×10^{-206}
GENERATION and *STATUS*	157.37	6	2.13×10^{-31}

Show that several variable pairs are significantly associated, indicating non-random relationships among employee characteristics. Kendall's Tau analysis in Table 3.

Table 3. Kendall's Tau Test Result

Column	Kendall's Tau (t)	P-value	Pair Type
STATUS and *LAYER*	0.280	4.613×10^{-21}	Concordant
EDUCATION and *LAYER*	0.250	2.574×10^{-18}	Concordant
LAYER and *GENERATION*	0.158	8.862×10^{-8}	Concordant
…	…	…	…
EDUCATION and *STATUS*	-0.209	5.355×10^{-13}	Discordant
EDUCATION and *GENERATION*	-0.040	1.674×10^{-1}	Discordant
GENERATION and *STATUS*	0.36	4.0615×10^{-33}	Concordant

Further reveals both positive and negative monotonic relationships, reflecting alignment and divergence across selected attributes. The magnitude of these relationships, measured using Cramér's V in Table 4.

Table 4. Cramér's V Result

Column	Cramér's V	Association
STATUS and *LAYER*	0.251	Weak
EDUCATION and *LAYER*	0.278	Weak
LAYER and *GENERATION*	0.471	Moderate
…	…	…
EDUCATION and *STATUS*	0.253	Weak
EDUCATION and *GENERATION*	0.608	Strong
GENERATION and *STATUS*	0.283	Weak

Suggests that while most associations are weak to moderate, certain variable pairs demonstrate stronger dependencies. Frequent itemsets were identified using Apriori, FP-Growth, and ECLAT algorithms with a minimum support threshold of 0.1. The resulting itemsets from each algorithm are presented in Table 5, Table 6 and Table 7.

Table 5. Frequent Itemsets for Apriori

No	Support	Itemsets
1	0.608	*'LAYER_STAFF', 'GENDER_MALE'*
2	0.494	*'GENERATION_MILLENNIAL', 'LAYER_STAFF'*
3	0.483	*'STATUS_PERMANENT', 'GENDER_MALE'*
4	0.348	*'EDUCATION_BACHELOR DEGREE', 'GENERATION_MILLENNIAL'*
..	…	...
372	0.233	*'GENERATION_MILLENNIAL', 'ROLE TYPE_SUPPORTING', 'GENDER_MALE'*
373	0.118	*'EDUCATION_BACHELOR DEGREE', 'GENERATION_GEN Z', 'LAYER_STAFF'*
374	0.117	*'ROLE TYPE_DIRECT', 'GENDER_FEMALE'*
375	0.101	*'STATUS_CONTRACT','GENERATION_GEN Z', 'LAYER_STAFF'*

Table 6. Frequent Itemsets for FP-Growth

No	Support	Itemsets
1	0.494	*'GENERATION_MILLENNIAL', 'LAYER_STAFF'*
2	0.388	*'LAYER STAFF', 'ROLE TYPE DIRECT', 'JENIS KELAMIN L'*
3	0.348	*'EDUCATION BACHELOR DEGREE', 'GENERATION MILLENNIAL'*
4	0.316	*'GENERATION_MILLENNIAL', 'ROLE TYPE_SUPPORTING'*
..	…	...
372	0.309	*'STATUS KAWIN L', 'LAYER STAFF'*
No	**Support**	**Itemsets**
373	0.255	*'STATUS_T', 'ROLE TYPE_SUPPORTING', 'EDUCATION_BACHELOR DEGREE'*
374	0.244	*'EDUCATION BACHELOR DEGREE', 'GENERATION MILLENNIAL', 'JENIS KELAMIN L'*
375	0.207	*'STATUS KAWIN L', 'GENERATION GEN Z'*

Table 7. Frequent Itemsets for ECLAT

No	Support	Itemsets
1	0.478	*'JENIS KELAMIN_L', 'GENERATION_MILLENNIAL'*
2	0.400	*'GENERATION MILLENNIAL', 'LAYER STAFF'*
3	0.380	*'LAYER STAFF', 'EDUCATION BACHELOR DEGREE'*
4	0.357	*'JENIS KELAMIN_L', 'EDUCATION_BACHELOR DEGREE'*
..	…	...
352	0.312	*'JENIS KELAMIN L', 'GENERATION MILLENNIAL', 'LAYER STAFF'*
353	0.244	*'JENIS KELAMIN_L', 'GENERATION_MILLENNIAL', 'EDUCATION_BACHELOR DEGREE'*
354	0.242	*'JENIS KELAMIN L', 'LAYER_STAFF', 'KATEGORI JABATAN AFTERSALES'*
355	0.207	*'STATUS KAWIN L', 'GENERATION GEN Z'*

Across the three methods, recurring patterns emerge involving generation, education level, organizational layer, job category, and employment status, indicating consistent structural tendencies within the dataset. To ensure robustness, the frequent itemsets generated by the three algorithms were intersected, retaining only patterns identified consistently across all methods. The resulting set of stable associations is shown in **Table 8**

Table 8. The Intersection from All Algorithms

No	Itemsets	Association
1	*'LAYER STAFF', 'KATEGORI JABATAN ADMIN'*	Individuals working in the staff layer tend to work in the administrative job category.
2	*'STATUS T', 'EDUCATION BACHELOR DEGREE'*	Individuals with 'T' (permanent) status are more likely to have a bachelor's degree background.

No	Itemsets	Association
3	*'GENERATION MILLENNIAL', 'LAYER STAFF', 'KATEGORI JABATAN AFTERSALES'*	Millennial generation individuals working in staff positions tend to be in the aftersales job category.
4	*'STATUS KAWIN L', 'EDUCATION BACHELOR DEGREE'*	Individuals with a single status (L) and a bachelor's degree are frequently found in the dataset.
…	...	…
81	*'STATUS K', 'GENERATION GEN Z', 'LAYER STAFF'*	Generation Z with a temporary contract status is frequently found working at the staff level.
82	*'GENERATION_MILLENNIAL', 'KATEGORI JABATAN_WIRANIAGA'*	Millennials are frequently found in sales positions.
83	*'EDUCATION SMK', 'KATEGORI JABATAN AFTERSALES'*	Vocational school graduates are commonly found in the aftersales job category.
84	*'STATUS K', 'STATUS KAWIN L', 'ROLE TYPE_DIRECT'*	Employees with contract status (K), single (L), and holding direct roles are prominent in this pattern.

The performance comparison among the three association mining algorithms is summarized in **Table 9**. The ECLAT algorithm demonstrated the fastest processing time (0.32 seconds) and the lowest memory consumption (0.02 MB), followed closely by Apriori (0.39 seconds and 0.04 MB). In contrast, FP-Growth exhibited a significantly longer runtime (2.87 seconds) and the highest memory usage (0.57 MB), although it remained efficient relative to the dataset size. These findings highlight ECLAT's superior efficiency for the analyzed employee dataset.

Table 9. Comparison of Runtime and Memory Usage for All Algorithms

Algortihm	Runtime (second)	Memory Usage (MB)
Apriori	0.39	0.04
FP-Growth	2.87	0.57
ECLAT	0.32	0.02

The findings of this study align with the objectives outlined in the *Introduction*, demonstrating that association analysis can effectively uncover meaningful patterns within employee data at Kalla Toyota. By leveraging advanced frequent itemsets mining algorithms—Apriori, FP-Growth, and ECLAT—frequent patterns and their intersections were successfully identified. These patterns provide actionable insights, such as the relationships between employee demographics and job roles, which can serve as a foundation for improving workforce strategies [28], as can be seen in **Figure 2**. These patterns provide actionable insights, such as the relationships between employee demographics and job roles, which can serve as a foundation for improving workforce strategies.

The comparison of frequent itemsets can be seen in **Figure 3**, revealed intersections patterns among the three algorithms, validating the robustness of the results. For example, a high correlation was found between specific generational groups and particular job roles, suggesting opportunities for targeted training and retention strategies. Furthermore, the use of visualization techniques, such as network graphs and matrix plots, made the identified patterns more accessible and interpretable, facilitating their practical application in decision-making processes.

From the data in **Tables 1**, **2**, and **3**, the use of statistical techniques like Cramér's V, Kendall's Tau, and Chi-Square to pre-filter relevant variables proved instrumental in enhancing the efficiency of the analysis. These methods ensured that only significant attributes were considered, reducing noise and improving the quality of the results.

Conclusion

The results of this study demonstrate that employing association analysis techniques such as Apriori, FP-Growth, and ECLAT effectively reveals meaningful patterns within Kalla Toyota's employee data. As anticipated in the "Introduction" section, these algorithms addressed the limitations of traditional analytical methods by uncovering complex relationships among employee attributes. The application of statistical filters, including Cramér's V, Kendall's Tau, and Chi-Square, ensured that only relevant variables were considered, thereby enhancing the reliability and interpretability of the findings.

The analysis yielded 84 frequent itemsets as the final result, derived by intersecting the outputs of all three algorithms. This approach ensured that only the most significant and consistent patterns were retained. The discovered patterns were visualized using network graphs and matrix plots, offering actionable insights to optimize human

resource management strategies. For example, relationships between job roles and generational demographics were identified, providing valuable guidance for designing targeted training and development programs.

Nevertheless, this study is subject to several limitations. The dataset employed, while providing valuable insights, is relatively limited in scope and may not comprehensively capture the heterogeneity of employee characteristics across all operational units of Kalla Toyota. Furthermore, inherent biases—such as generational overrepresentation or disproportionality in job roles—may have affected the prevalence and strength of the identified association patterns.

Future research may benefit from integrating these association rule models into real-time human resource analytics frameworks to support dynamic decision-making processes. Additionally, the application of more sophisticated machine learning techniques—such as clustering ensembles or deep learning architectures—could offer comparative perspectives and uncover deeper insights. Expanding the dataset to incorporate longitudinal records or cross-company samples would further enhance the generalizability and practical relevance of the findings.

Acknowledgement

We sincerely express our heartfelt gratitude to the Faculty of Computer Science at Universitas Muslim Indonesia. Their exceptional expertise, insightful guidance, and unwavering encouragement have been integral to the successful completion of this research. This work is a testament to their dedication to fostering academic excellence and innovation.

References

[1] R. Gulia. and V. Rastogi, "The Impact of Information Technology On Data-Driven Decision Making In Human Resource Management," *International Journal For Multidisciplinary Research*, vol. 6, no. 6, Nov. 2024, doi: 10.36948/ijfmr.2024.v06i06.30314.

[2] Marwan Marwan and F. Alhadar, "Effects Of HR Management Practices On Employee Innovative Work Behavior With Two Mediation," *Jurnal Manajemen*, vol. 28, no. 2, pp. 247–271, Jun. 2024, doi: 10.24912/jm.v28i2.1796.

[3] L. N. Hayati, A. N. Handayani, W. S. G. Irianto, R. A. Asmara, D. Indra, and M. Fahmi, "Classifying BISINDO Alphabet using TensorFlow Object Detection API," *ILKOM Jurnal Ilmiah*, vol. 15, no. 2, pp. 358–364, Aug. 2023, doi: 10.33096/ilkom.v15i2.1692.358-364.

[4] H. Li and P. C. Y. Sheu, "A scalable association rule learning heuristic for large datasets," *J Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00473-3.

[5] M. Sornalakshmi *et al.*, "An efficient apriori algorithm for frequent pattern mining using mapreduce in healthcare data," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 1, pp. 390–403, Feb. 2021, doi: 10.11591/eei.v10i1.2096.

[6] W. A. Pertiwi and J. S. P. Tyoso, "Inventory-based Business Strategy using FP-Growth at PT. Pinus Merah Abadi, Kendal Regency," *Journal of Business Management and Economic Development*, vol. 2, no. 02, pp. 895–911, Apr. 2024, doi: 10.59653/jbmed.v2i02.793.

[7] V. Srinadh, "Evaluation of Apriori, FP growth and Eclat association rule mining algorithms," *Int J Health Sci (Qassim)*, pp. 7475–7485, Apr. 2022, doi: 10.53730/ijhs.v6ns2.6729.

[8] P. Ranganathan and S. Hunsberger, "Handling missing data in research," *Perspect Clin Res*, vol. 15, no. 2, pp. 99–101, 2024, doi: 10.4103/picr.picr_38_24.

[9] R. L. Sapra and S. Saluja, "Understanding statistical association and correlation," *Curr Med Res Pract*, vol. 11, no. 1, pp. 31–38, Jan. 2021, doi: 10.4103/cmrp.cmrp_62_20.

[10] C. Shen, S. Panda, and J. T. Vogelstein, "The Chi-Square Test of Distance Correlation," *Journal of Computational and Graphical Statistics*, vol. 31, no. 1, pp. 254–262, 2022, doi: 10.1080/10618600.2021.1938585.

[11] S. T. Nihan, "Karl Pearsons chi-square tests," *Educational Research and Reviews*, vol. 15, no. 9, pp. 575–580, Sep. 2020, doi: 10.5897/err2019.3817.

[12] I. Gijbels and M. Matterne, "Study of partial and average conditional Kendall's tau," *Dependence Modeling*, vol. 9, no. 1, pp. 82–120, Jan. 2021, doi: 10.1515/demo-2021-0104.

[13] E. Szegedi-Hallgató, K. Janacsek, and D. Nemeth, "Different levels of statistical learning - Hidden potentials of sequence learning tasks," *PLoS One*, vol. 14, no. 9, Sep. 2019, doi: 10.1371/journal.pone.0221966.

[14] D. Dwiputra, A. M. Widodo, H. Akbar, G. Firmansyah, "Evaluating the performance of association rules in Apriori and FP-Growth algorithms : Market basket analysis to discover rules of item combination ," *Journal of World Science*, vol. 2, no. 8, Aug. 2023, doi: 10.58344/jws.v2i8.403.

[15] X. Wan and X. Han, "Efficient Top-k Frequent Itemset Mining on Massive Data," *Data Sci Eng*, vol. 9, no. 2, pp. 177–203, Jun. 2024, doi: 10.1007/s41019-024-00241-2.

[16] M. H. Santoso, "Application of Association Rule Method Using Apriori Algorithm to Find Sales Patterns Case Study of Indomaret Tanjung Anom," *Brilliance: Research of Artificial Intelligence*, vol. 1, no. 2, pp. 54–66, Dec. 2021, doi: 10.47709/brilliance.v1i2.1228.

[17] N. L. Chusna and I. Permata Sari, "Application of Data Mining for Product Purchase Pattern Analysis with Frequent Pattern Growth (FP-Growth) Algorithm on Sales Transaction Data," *Technology, and Engineering (JSTE)*, vol. 1, no. 1, 2021, doi: 10.33603/jste.v1i1.6034.

[18] A. M. A. Al-Badani and R. A. H. Al-Dilami, "An improved efficient FP-growth algorithm using FP-TDA algorithm," *International Journal of Innovative Science and Research Technology*, vol. 10, no. 12, Dec. 2025, doi: 10.38124/ijisrt.25dec443.

[19] K. Gulzar, M. Ayoob Memon, S. M. Mohsin, S. Aslam, S. M. A. Akber, and M. A. Nadeem, "An Efficient Healthcare Data Mining Approach Using Apriori Algorithm: A Case Study of Eye Disorders in Young Adults," *Information (Switzerland)*, vol. 14, no. 4, Apr. 2023, doi: 10.3390/info14040203.

[20] R. Rachmania and R. Supriyanto, "Implementation of FP-Growth and Fuzzy C-Covering Algorithm based on FP-Tree for Analysis of Consumer Purchasing Behavior," 2020. doi: doi.10.5120-2020920171.

[21] I. Riadi, H. Herman, F. Fitriah, and S. Suprihatin, "Optimizing Inventory with Frequent Pattern Growth Algorithm for Small and Medium Enterprises," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 23, no. 1, pp. 169–182, Nov. 2023, doi: 10.30812/matrik.v23i1.3363.

[22] M. Man and M. A. Jalil, "Frequent itemset mining: Technique to improve ECLAT based algorithm," *International Journal of Electrical and Computer Engineering*, vol. 9, no. 6, pp. 5471–5478, 2019, doi: 10.11591/ijece.v9i6.pp5471-5478.

[23] B. Vo, T. Le, F. Coenen, and T. P. Hong, "Mining frequent itemsets using the N-list and subsume concepts," *International Journal of Machine Learning and Cybernetics*, vol. 7, no. 2, pp. 253–265, Apr. 2016, doi: 10.1007/s13042-014-0252-2.

[24] D. Zhu, Y. Wang, B. Wei, Z. Guo, and F. Wan, "Data Visualization Overview," in *2021 IEEE 3rd International Conference on Civil Aviation Safety and Information Technology (ICCASIT)*, IEEE, Oct. 2021, pp. 735–738. doi: 10.1109/ICCASIT53235.2021.9633610.

[25] T. Venturini, M. Jacomy, and P. Jensen, "What do we see when we look at networks: Visual network analysis, relational ambiguity, and force-directed layouts," *Big Data Soc*, vol. 8, no. 1, 2021, doi: 10.1177/20539517211018488.

[26] Z. Huang, R. Borgo, A. Kerren, and D. Archambault, "Matrix Snap&Go: Visualization of Paths on Matrices," *The Eurographics Association*, 2024, doi: 10.2312/evs.20241058.

[27] J. Graffelman and J. de Leeuw, "Improved Approximation and Visualization of the Correlation Matrix," *American Statistician*, vol. 77, no. 4, pp. 432–442, 2023, doi: 10.1080/00031305.2023.2186952.

[28] S. Sancoko, Z. S. Prayogi, B. Al Aufa, and R. Yuliawan, "Evaluation of Employee Acceptance of the IMS Application at PT Sarana Utama Adimandiri: TAM Approach," *ILKOM Jurnal Ilmiah*, vol. 14, no. 1, pp. 74–79, Apr. 2022, doi: 10.33096/ilkom.v14i1.1120.74-79.

[29] R. A. Putra, M. A. M. Putri, S. M. Sinaga, S.F. Octavia, and R.C. Rachman, "Implementation of association rules algorithm to identify popular topping combinations in orders," *Public Research Journal of Engineering, Data Technology and Computer Science*, vol. 1, no. 2, pp. 95–101, 2024, doi: 10.57512/predatecs.v1i2.863.

[30] C. Zhang and W. Han, "Ensembles of decision trees and gradientbased learning for employee turnover rate prediction," *PeerJ Comput Sci*, vol. 10, 2024, doi: 10.7717/PEERJ-CS.2387