



Employee Education and Training Recommendations using the Apriori Algorithm

Arief Wibowo¹, Vasthu Imaniar Ivanoti^{2*}, Megananda Hervita Permata Sari³

^{1,2,3}Master of Computer Science, Faculty of Information Technology, Universitas Budi Luhur, Jakarta, Indonesia

¹arief.wibowo@budiluhur.ac.id, ²2111602294@student.budiluhur.ac.id, ³2111602153@student.budiluhur.ac.id

Abstract

The Ministry of Finance (MoF) aims to enhance employee performance through suitable education and training opportunities. Based on the data of education and training implementation in 2022 at the Central ICT Department in MoF, only 27.35% of employees participated in education and training according to proposed needs for both positions and individuals. This is partly due to mandatory training that must be attended by some or all employees, urgent needs in the current year, or the substitute participants that is not from the same team or function. To address this issue, the association method of data mining techniques can be utilized to analyze historical data of employees. The study used the apriori algorithm to analyze historical data of employee positions, organizations, and education and training from 2011 to 2021. This research involved comparing various minimum support values, assuming that employees attended at least 2, 3, and 4 training courses, to calculate the corresponding minimum support values. The evaluation results of the model show that the best rules are generated with a minimum support value of 0.013 and a minimum confidence value of 0.6, which is a total of 10 rules. One of the training recommendations is that if an employee has taken the Enterprise Service Bus (ESB)-API Management training, they will take the ESB API Integration Platform training. Furthermore, it can be used by the Human Resource Unit to provide education and training aligned with organizational needs and improve employee competency in line with their duties and functions, leading to better overall organizational performance.

Keywords: training; competency; data mining; association rule; apriori algorithm

1. Introduction

In order to improve the performance of employees at the Ministry of Finance (MoF), each employee needs to be given the opportunity to improve their competency through education and training (training) that is appropriate to their duties and functions. MoF, through the Financial Education and Training Agency (BPPK) [1], provides various types of diklat, both soft skills and hard skills, that can be attended by all MoF employees. Available training include Distance Learning conducted through video conference since the pandemic, E-learning and microlearning for self-paced learning through video tutorials and quizzes, workshops and seminars via video conference, and technical training with classical methods, namely face-to-face in the classroom. In addition to diklat provided by BPPK, MoF also provides a budget for improving employee competency by attending training activities organized by educational institutions outside of MoF.

The mechanism for employees to attend training can be through proposals from the Human Resource (HR) unit

carried out every year or by attending available e-learning independently. In terms of training proposals, employees can propose the desired diklat to the HR unit at the end of the year for the following year's learning needs plan, and every month the HR unit will offer available training to employees.

Employee-initiated training proposals are done by selecting training from the list provided by BPPK in the learning module on the e-performance application, where currently there are 1444 available training from various categories. Next, employees need to read the Program Reference Framework (KAP) documents for each training to find out the objectives, targets, expected competency standards, and participant requirements. As for training needs that are not available in BPPK, employees actively seek information on training organized by educational institutions outside of MoF according to their duties and functions. However, training proposal submissions to the HR unit are made in less than a week, making it impossible for employees to fully understand the entire KAP document or training syllabus. On the other hand, every month the HR unit

confirms the participation of training participants and informs them of the available training offers for that month.

Based on the monthly monitoring of employee training data conducted by the HR unit at the MoF's Central ICT Department in 2022, only 27.35% of employees attended training that was proposed in accordance with their position and individual needs. Meanwhile, the rest is divided into 44.66% of employees attending mandatory training, 25.57% of employees attending training according to their duties and functions but not included in the initial proposal, and 2.42% of employees attending training that is not in accordance with the proposal or their duties and functions. It can be concluded that employees who attended training that did not match the proposed needs were either required to attend mandatory training, had sudden training needs in the current year, or were replacements for other training participants who were not from the same team or function.

To overcome the inefficiency in the training proposal selection process, an analysis is conducted by utilizing data mining techniques on employee training data, resulting in a training recommendation model. The intended training recommendation model can be used by the HR unit to provide training according to the needs of employees. It is expected that this model will enhance the competency of employees based on their duties and responsibilities.

The literature review covers various research studies on data mining using recommendation techniques. One study [2] utilized the apriori algorithm and association rules to generate sales strategies based on sales transaction data. The strongest association rules were determined using lift ratio. [3] conducted research on a book recommendation system for librarians using an association approach with an apriori algorithm. Other study [4] on a book recommendation system also adopts an association approach based on book borrowing data with an enhanced apriori algorithm to create a personalized book recommendation system. The system's performance using the apriori algorithm was compared to that using hybrid cooperative and K-means algorithms, and results showed that CPU usage, memory, and response time were minimal with 50 clients running the system simultaneously. A course recommendation system [5] was also developed using the apriori algorithm with minimum support and confidence thresholds of 7 and 80%, resulting in a high accuracy rate for study plan recommendations. A modified apriori algorithm was also proposed in another study [6] by combining combination reduction techniques and iteration limitation, where the iteration limit is taken from the maximum set size of the most frequently occurring transactions because the probability of getting the best association rules is

higher, to improve time efficiency. Other studies utilize the FP-Growth algorithm in the data mining association method. Several studies utilize this algorithm for various purposes, such as examining violence patterns with a minimum support of 50% and a minimum confidence of 60% to identify frequent acts of violence. This enables DP3AKB to implement suitable preventive or intervention measures [7]. Additionally, the algorithm is used to aid students in selecting elective courses [8] and optimize product placement on shelves based on customer search frequency levels [9]. While [10] compared three methods for determining self-development training, namely the C4.5 algorithm, the combination of PCA and C4.5, and the combination of C4.5, discretization, and PCA. The test results show that the combination of PCA, discretization, and C4.5 gives better performance than the other two methods with an average accuracy rate of 86.6%. Finally, another study [11] compared the apriori algorithm and FP-Growth in analyzing sales transaction data and found that while the FP-Growth algorithm generated association rules faster, the apriori algorithm was superior in terms of itemset variations.

After reviewing the literature, it is evident that association techniques are used in several fields, ranging from recommending sales strategies [2], [6], [9], [11], book recommendations [3], [4], study plan recommendations [5], [8], and even violence prevention recommendations [7]. However, research on training recommendations using data mining is very limited [10]. Therefore, the researcher opted to use association data mining techniques with the Apriori method to design training recommendations. The aim was to determine the correlation between training, using historical employee data for the period 2011 to 2021, which will ultimately produce a training recommendation model that can be utilized by the personnel unit in training planning.

2. Research Methods

2.1. Research Flow

This research begins with problem identification, followed by a literature review to determine the approach chosen for analyzing data, in this case, the chosen approach is association. The next steps are in accordance with the stages in the CRISP-DM framework [12]–[14]. It begins with the Business Understanding stage, which involves conducting document studies, interviews, and observations to gain a comprehensive understanding of the business context. In the Data Understanding stage, data related to the problem is collected over a span of 10 years, including employee training data, employee organizational data, and employee position data. The next step is Data Preparation, where the attributes to be used in the analysis are determined, and the collected data is

transformed accordingly. Following that is the Modeling stage, where a model is created using the Apriori method, which helps in identifying patterns or associations within the data. Once the model is created, it moves to the Evaluation stage, where its effectiveness is assessed to ensure it meets the business requirements. Finally, in the Deployment stage, conclusions are drawn from the research results, and a research report is prepared to present the findings obtained from the analysis.

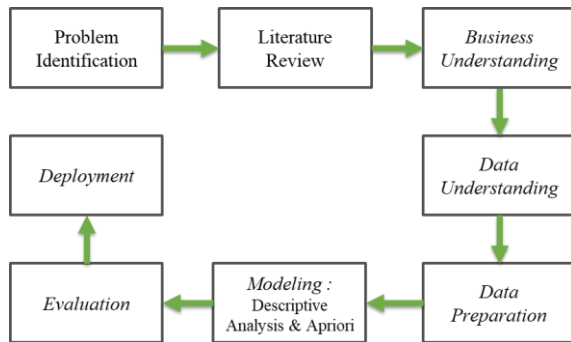


Figure 1. The Research Stages [14]

Figure 1 presents the flow of the research stages starting from the problem identification stage to the deployment stage which has been described previously.

2.2. Data Mining

Data mining is a process carried out to extract information and discover knowledge without explicit assumptions from large data sources, where the obtained information must have three characteristics, which are previously unknown, effective, and practical [15]. Data mining can be divided into two main categories, namely predictive and descriptive. Descriptive patterns describe the general nature of the data, for example, through association analysis and clustering analysis [16]. Meanwhile, predictive patterns summarize current data, for example, through classification and regression. As of now [17], the Cross Industry Standard Process for Data Mining (CRISP-DM) is the most commonly used framework for conducting data analysis, mining, and science work [18], [19]. CRISP-DM [17] consists of 6 phases which include (1) Business understanding - identifying business objectives and data mining goals, (2) Data understanding - collecting, exploring, and validating data, (3) Data preparation - cleaning, transforming, and integrating data, (4) Modeling - selecting modeling techniques, creating models, and evaluating models, (5) Evaluation - evaluating modeling results against business objectives, and (6) Deployment - planning for deployment, monitoring, and maintenance.

Some data mining techniques that are commonly used include [9] (1) Classification, which categorizes data into predefined categories, (2) Clustering, which groups data into several subsets or groups with high similarity

within one group and low similarity between groups, (3) Regression, which predicts the value of a certain continuous variable based on other variable values, assuming a linear or nonlinear dependency model, and (4) Association rules, which detect sets of attributes that frequently occur together and form a number of rules.

2.3. Association Rules

Association rule is a data mining technique used to analyze or detect associations between items in a set of item combinations. The goal is to find the relationship between items in itemsets [20]. Association rules are also known as market basket analysis. An illustration is the analysis of products purchased in a clothing store resulting in the likelihood of customers buying pants and shirts together. This can help the store owner to organize their inventory or offer promotion by providing special discounts for frequently purchased item combinations, which will increase sales [6]. The basic stages of association rules are divided into two [3], [21]. First is frequent pattern analysis. In this stage, the search for item combinations that meet the minimum support value criteria is performed, which is the value of the occurrence of product combinations in each transaction. The support value of an item is calculated by Formula 1:

$$\text{support}(A) = \frac{\sum \text{transaction with item } A}{\sum \text{transaction}} \quad (1)$$

While the formula for calculating the support value for two items is in Formula 2 and Formula 3:

$$\text{support } A, B = P(A \cap B) \quad (2)$$

$$\text{support } A, B = \frac{\sum \text{transaction with item } A \text{ and } B}{\sum \text{transaction}} \quad (3)$$

Second is association rule formation. After the frequent pattern is found, association rules that meet the minimum confidence value criteria are sought by calculating the confidence value of the "If A then B" rule using Formula 4.

$$\text{confidence} = P(B|A) = \frac{\sum \text{transaction with item } A \text{ and } B}{\sum \text{transaction with item } A} \quad (4)$$

2.4. The Apriori Method

The Apriori algorithm is a basic and popular algorithm in the application of association rules, introduced by Agrawal and Srikant in 1994. In the Apriori algorithm, each transaction is considered as an itemset where the algorithm will identify the items that are at least a minimum threshold subset as a new itemset. The approach used is "bottom-up" where the itemset is determined one by one, called candidate generation. This algorithm uses a breadth-first search and a hash tree structure to efficiently calculate the candidate itemsets. A group of candidates is tested against the data, which will be pruned if the candidate has infrequent subpatterns. This process is repeated until no more extensions are found. The Apriori algorithm is

considered a brute-force method because it considers every k-itemset as a frequent itemset candidate [22]. Although Apriori is a clear and simple algorithm, its main weakness is the time it takes to manage a large number of candidates, which is very long due to the repetitive process for each candidate. Therefore, when the available resources are limited, in this case, CPU and memory, the use of this algorithm is not efficient [23].

3. Results and Discussions

3.1. Business Understanding

The business understanding phase is a process to understand how the proposal process for training and its supporting applications work. The process includes document studies, such as the Training Needs Analysis request memorandum and the previous year's training program list, interviews with employees in the HR unit and person in charge (PIC) of the workgroup, as well as observations. The supporting applications used in the training proposal process include the Learning module in the E-Performance application and the Employee Profile module in the HRIS application.

3.2. Data Understanding

The data understanding phase is carried out to understand and identify the required data and collect it.

Based on the identification results, the data used are the historical training data, historical employee position data, and historical organizational data were obtained from the HRIS application.

Historical data on employee training for the period 2011 to 2021 with a total of 7.615 records as shown in Table 1, contains information on the name of the education and training that each employee has participated in along with the date of implementation and the name of the education and training organizer.

Table 2 shows the employee position's historical data for the period 2011 to 2021 with a total of 4.950 records, which includes details on employee ID, position ID, start date, end date, and decree number.

Upon examining the table, it becomes evident that the date format used in the start date and end date columns differ from one another. Consequently, this inconsistency must be resolved during the data preparation phase.

In Table 3, there are records of the employee reassignments within the company covering the period from 2011 to 2021 with a total of 2.701 records, comprised of details on Employee ID, Position ID, Position Name, Unit Name, and Echelon Name.

Table 1. Historical Data On Employee Training

No	EmployeeID	TrainingName	Training Date	Organizer
0	73477	National Procurement Expert Certification	2011-02-01	LKPP
1	79579	advanced risk management implementation	2011-02-01	itjen
2	80875	workshop advanced risk management implementation	2011-02-01	ITJEN
3	118511	Specialization in Financial Management of Central Government Units	2011-02-07	Public Finance Training Center, MoF
4	118511	INTELLIGENCE ANALYSIS	2011-02-07	Public Finance Training Center, MoF
...	...	...	...	...
7611	101828	Preparation of Technical Specifications and Self-Estimated Prices	2021-09-27	BPPK
7612	118318	Communications Skills	2021-09-28	BPPK
7613	118511	HCIA - ROUTING & SWITCHING	2021-12-08	Telkom Prima Cipta
7614	128342	Huawei Certified Ict Associate (HCIA) International Certification Training	2021-12-08	PT Telkom Prima Cipta Certifia

Table 2. Historical Employee Position Data

No.	EmployeeID	PositionID	StartDate	EndDate	DecreeNo
0	67536	461	2011-02-14 00:00:00.000	2016-03-29 00:00:00.000	KEP-25/SJ/UP.11/2011
1	88385	480	2011-02-14 00:00:00.000	2014-02-11 00:00:00.000	KEP-25/SJ/UP.11/2011
2	88520	470	2011-02-14 00:00:00.000	2016-03-29 00:00:00.000	KEP-25/SJ/UP.11/2011
3	97063	584	2011-02-14 00:00:00.000	2017-10-03 00:00:00.000	KEP-24/SJ/UP.11/2011
4	42689	469	2011-02-16 00:00:00.000	2013-01-15 00:00:00.000	83/KM.1/UP.11/2011
...	...	...	...	...	...
4928	115937	302108	30-03-22 0:00	30-06-22	KMK 321.1/KM.1/UP.11/2022
4930	122132	302108	30-03-22 0:00	30-06-22	KMK 321.1/KM.1/UP.11/2022
4931	115169	681042	30-12-21 0:00	30-06-22	KEP-157/SJ/2021
4947	117997	302108	30-12-21 0:00	30-06-22	1540/KM.1/UP.11/2021
4949	118006	302108	30-12-21 0:00	30-06-22	1540/KM.1/UP.11/2021

Table 3. Historical Employee Position Data

No.	Employee ID	Position ID	Position Name	Unit Name	Echelon1	Echelon2	Echelon3	Echelon4
0	130187	0	Staff	ICT Management Section	Secretariat General	Central ICT Department	Medan ICT and State Property Management Office	ICT Management Section
1	128342	404719	ICT Administrator Level III	ICT Management Section	Secretariat General	Central ICT Department	Medan ICT and State Property Management Office	ICT Management Section
2	128342	673362	ICT Administrator Level IV)	ICT Management Section	Secretariat General	Central ICT Department	Medan ICT and State Property Management Office	ICT Management Section
3	127402	404719	ICT Administrator Level III	ICT Management Section	Secretariat General	Central ICT Department	Medan ICT and State Property Management Office	ICT Management Section
...	...	...	...	...	...	...	...	...
2697	88933	0	Head of IS Implementation Sub Division	IS Implementation Sub Division	Secretariat General	Central ICT Department	IT Service and Support Division	IS Implementation Sub Division
2698	88521	0	Head of IT Development Sub Division	Infrastructure Development Sub Division	Secretariat General	Central ICT Department	IT Service and Support Division	IT Development Sub Division
2699	99166	0	Staff	IT Infrastructure Development Sub Division	Secretariat General	Central ICT Department	IT Service and Support Division	IT Infrastructure Development Sub Division
2700	88521	0	Head of IT Infrastructure Development Sub Division	IT Infrastructure Development Sub Division	Secretariat General	Central ICT Department	IT Service and Support Division	IT Infrastructure Development Sub Division

3.3. Data preparation

The next step is data preparation, also known as data preprocessing, which aims to remove or improve certain characteristics so that the data is ready for use in the analysis of training recommendations. Data preprocessing is performed using the Python programming language by first importing the data from Ms. Excel format to Jupyter Notebook [24].

The following are the data preprocessing steps that were applied which include (1) cleaning training data by changing the date data type, standardizing training names, removing mandatory trainings, and removing e-learning trainings; (2) cleaning job data by changing the date data type and filling in null values in the completed job date column; (3) applying stopword

removal by filtering out common words and selecting important words by removing SK values of "PRIN" and "PPK"; (4) cleaning organization data by standardizing the naming of level 3 and level 4 positions and assigning role labels; (5) Integrating the three tables and removing duplicate data to obtain the role and job position data of employees during the training period as shown in Table 4; (6) creating a basket that contains that contains information about employees, roles, and training names; and (7) converting the data format in the basket to a binary table of 312 transactions as shown in Table 5, where the columns are transaction items containing true or false values. The transactions refers to one employee in one subrole, and the item column is the name of the training.

Table 4. Mapping Of Employee Position Role At The Time Of The Training

No	Employee ID	TrainingName	Training Date	Position ID	StartDate Decree	EndDate Decree	Position Name	...	Role	SubRole
0	104471	Cisco BGP Advanced	2011-04-27 00.00.00	56033	2011-02-18 00.00.00	2017-11-10 00.00.00	CPNS (Civil Servant Candidate)	...	IFPD	OP DC
1	88385	Project Management	2011-06-06 00.00.00	480	2011-02-14 00.00.00	2014-02-11 00.00.00	Head of Subsection	...	IFPD	SD

No	Employee ID	TrainingName	Training Date	Position ID	StartDate Decree	EndDate Decree	Position Name	...	Role	SubRole
2	97060	Information Technology Project Management	2011-06-13 00.00.00	472	2007-07-18 00.00.00	2016-04-01 00.00.00	Head of Subsection	...	APL	APL DEV
3	88520	Information Technology Project Management	2011-06-20 00.00.00	470	2011-02-14 00.00.00	2016-03-29 00.00.00	Head of Subsection	...	APL	APL DEV
4	104471	Networking Advance	2011-06-30 00.00.00	56033	2011-02-18 00.00.00	2017-11-10 00.00.00	CPNS (Civil Servant Candidate)	...	IFPD	OP DC
...	...	...	...	...	...	...	...	...	...	...
3169	104471	Enterprise Service Bus: API Integration Platform	2021-04-26 00.00.00	59662 4	2020-07-21 00.00.00	2021-12-30 00.00.00	Head of Subsection	...	IFPD	ARCHITECTURE
3170	77750	Retirement Preparation	2021-09-27 00.00.00	16888	2016-08-04 00.00.00	2022-01-03 00.00.00	Staff	...	ADM	GENERAL
3171	113813	Government Procurement of Goods/Services	2021-09-27 00.00.00	40475 3	2021-02-23 00.00.00	2022-01-01 00.00.00	State Property Staff Lv.I	...	KPTIK	BMN MANAGEMENT
3178	104782	HCIA - ROUTING & SWITCHING	2021-12-08 00.00.00	40472 4	2019-02-02 00.00.00	2022-01-01 00.00.00	IT Operational Staff Lv.I	...	KPTIK	KPTIK MANAGEMENT
3179	118511	HCIA - ROUTING & SWITCHING	2021-12-08 00.00.00	40482 8	2020-09-14 00.00.00	2022-01-03 00.00.00	IT Operational Staff Lv.I	...	IFPD	KPTIK MANAGEMENT
3180	135810	HCIA - ROUTING & SWITCHING	2021-12-08 00.00.00	60446 2	2020-07-22 00.00.00	2022-01-03 00.00.00	Staff	...	NET	OP NET

The results of the fifth step in the data preparation stage can be seen in Table 4 where the table provides information on mapping the roles of employee positions when the training is carried out with a total of 3.181 records. For example, an employee with ID 104471

attended the Cisco BGP Advanced and Networking Advance training when he was in the DCOPS role and attended the Enterprise Service Bus: API Integration Platform training when he was in the ARCHITECTURE role.

Table 5. Binary Table From Preprocessing Data

Training Name	40-hour TOEFL Preparation Course	Government Accounting	Activity Management	Computer Network Administration	...	Zheng Hes Art of Collaboration	Implementing Cisco IP Routing	Implementing Cisco IP Switched Networks	Procurement Negotiation & Contracting Strategy
EmployeeID-SubRole									
101056-KPTIK	false	false	false	false	...	false	false	false	false
101816-APL	false	false	false	false	...	false	false	false	false
101816-JARKOMDAT	false	false	false	false	...	false	false	false	false
102522-KPTIK	false	false	false	true	...	false	false	false	false
104803-BD	false	false	false	false	...	false	false	false	false
104925-PKLTIK	false	false	false	true	...	false	false	false	false
...	...	...	...	...	...	...	...	...	...
104944-IFPD	false	false	false	false	...	false	false	false	false
114043-TU	true	false	false	false	...	true	false	false	false
140517-KI	false	false	false	false	...	false	false	false	false
75956-APL	false	false	false	false	...	false	true	true	false
76640-TU	false	false	false	false	...	false	false	false	true
88385-IFPD	false	false	true	false	...	false	false	false	false

Training Name	40-hour TOEFL Preparation Course	Government Accounting	Activity Management	Computer Network Administration	...	Zheng Hes Art of Collaboration	Implementing Cisco IP Routing	Implementing Cisco IP Switched Networks	Procurement Negotiation & Contracting Strategy
EmployeeID-SubRole									
97042-KPTIK	false	true	false	false	...	false	false	false	false

The binary table resulting from data preparation stage contains 312 records as seen in Table 5. The information presented in Table 5 illustrates how the employee ID and SubRole correspond to various training programs. The information is displayed in binary table format, where a "true" value indicates that the employee participated in the training, and a "false" value indicates that they did not attend. As an illustration, an employee with the ID number 75956 who belongs to the APL subrole attended both the Implementing Cisco IP Routing and Implementing Cisco IP Switched Networks training sessions.

3.4. Descriptive Analysis

After going through the data preparation stage, descriptive analysis was conducted by displaying visualization of training uptake trends per year, the most frequently taken trainings, as well as the number of employees with excess or insufficient amount of training.

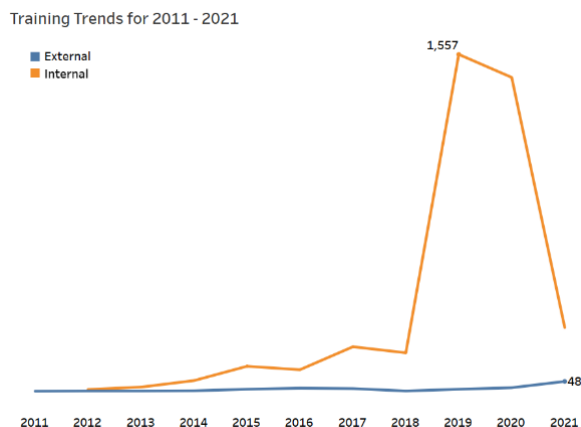


Figure 2. Training Uptake Trends Per Year

Figure 2 shows the trend in the number of trainings taken per year during the period from 2011 to 2021, where the highest number of internal trainings was taken in 2019 with a total of 1.557 trainings and the highest number of external trainings was taken in 2015 with a total of 48 trainings.

Next, an analysis was conducted on the 5 most frequently internal trainings taken by employees during the same period. Based on Figure 3, it can be seen that 4 of the 5 internal training that are most often attended by employees are in the form of E-learning.

To determine the adequacy of trainings attended by employees, both internal and external, classification is

carried out based on the number of trainings attended, which is divided into 3 categories which include (a) insufficient, if the number of internal/external trainings attended by employees is less than 2 trainings per year; (b) adequate, if the number of internal/external trainings attended by employees is 2 trainings per year; and (c) excess, if the number of internal/external trainings attended by employees is more than 2 trainings per year.

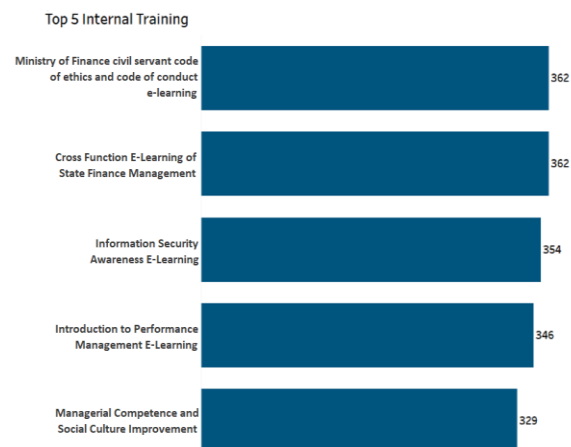


Figure 3. Top 5 Internal Trainings

Based on Figure 4, it can be seen that in the period from 2011-2021, the adequacy of employees in attending training with the 'Insufficient' status was 69.07%. Meanwhile, employees with 'Adequate' status were 12.36%, and 'Excess' status were 18.57%.

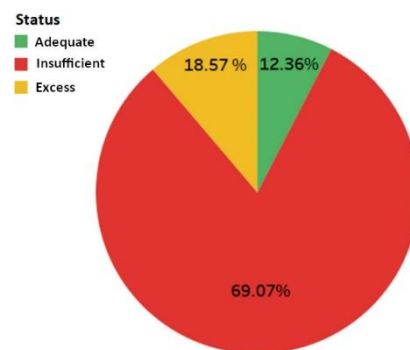


Figure 4. Status Of Training Uptake By Employees

3.5. Modeling with Apriori Algorithm

When creating a model using the Apriori algorithm, it is important to establish the minimum support and confidence values beforehand, as they will have an impact on the number of rules generated. The greater the support and confidence values used, the fewer the

number of rules obtained, and vice versa [5], [25]. This research involved comparing various minimum support values, assuming that employees attended at least 2, 3, and 4 training courses. The following is the calculation of each minimum support value: (1) employees attended at least 2 training course = $\frac{2}{312} = 0,006$; (2) employees attended at least 3 training course = $\frac{3}{312} = 0,009$; and (3) employees attended at least 4 training course = $\frac{4}{312} = 0,013$.

Then, the itemset results for each minimum support value were used to calculate the average confidence value.

Table 6 shows the calculation using a minimum support value of 0.006, resulting in 252 rows of itemsets.

Association rules were then created on the prepared baskets, resulting in a table with 436 rules as shown in Table 7. Based on Table 7, one of the recommendation is that if an employee has taken the Working Load Analysis and 40-hour TOEFL Preparation Course training, they will take the Job Analysis training with a probability of 100%.

Table 6. Calculation With Minimum Support Of 0.006

No	support	itemsets
0	0.006410256	(40-hour TOEFL Preparation Course)
1	0.060897436	(Computer Network Administration)

No	support	itemsets
2	0.006410256	(Expert in Government Procurement of Goods/Services)
3	0.073717949	(Acceleration and Implementation of Internal Compliance Unit)
4	0.016025641	(Workload Analysis)
...	...	...
247	0.009615385	(Prediction test of TOEIC', 'Acceleration and Implementation of Internal Compliance Unit', 'IT Service Management Foundation', 'Visionary Leadership Training Program)
248	0.006410256	(Government Goods/Services Procurement', 'Acceleration and Implementation of Internal Compliance Unit', 'Fiscal Policy and Latest Economic Developments', 'Implementation of Standard Budgeting Cost Structure for Work Units within the Ministry of Finance)
249	0.006410256	(Risk Management', 'Acceleration and Implementation of Internal Compliance Unit', 'Implementation of Standard Budgeting Cost Structures for Work Units within the Ministry of Finance', 'Government Procurement of Goods/Services)
250	0.006410256	(Prediction test of TOEIC', 'Information Technology Project Management', 'IT Service Management Foundation', 'Visionary Leadership Training Program)
251	0.006410256	(Prediction test of TOEIC', 'Filling Competency Gaps for Echelon IV Officials', 'IT Service Management Foundation', 'Examination of Employee Discipline Violations)

Table 7. Result Of Apriori Algorithm With Minimum Support 0,006

No	antecedents	consequents	antecedent support	consequents support	support	confidence	lift	leverage	conviction
2	(40-hour TOEFL Preparation Course)	(Job Analysis)	0.00641	0.00641	0.00641	1	156	0.006369	inf
3	(Job Analysis)	(40-hour TOEFL Preparation Course)	0.00641	0.00641	0.00641	1	156	0.006369	inf
188	(Seminar on Tax Incentives for Taxpayers Affected by the Covid-19 Pandemic)	(Digital Seminar on Excise Facilitation for Industry in the context of Economic Recovery Affected by COVID-19)	0.00641	0.00641	0.00641	1	156	0.006369	inf
189	(Digital Seminar on Excise Facilitation for Industry in the context of Economic Recovery Affected by COVID-19)	(Seminar on Tax Incentives for Taxpayers Affected by the Covid-19 Pandemic)	0.00641	0.00641	0.00641	1	156	0.006369	inf

No	antecedents	consequents	antecedent support	consequents support	support	confidence	lift	leverage	conviction
190	(Working Load Analysis', '40-hour TOEFL Preparation Course)	(Job Analysis)	0.00641	0.00641	0.00641	1	156	0.006369	inf
...	...	...	...	...	...	...	...	...	...
261	(Procurement of Goods/ Information Technology Services)	(Prediction test of TOEIC', 'Acceleration and Implementati on of Internal Compliance Unit)	0.102564	0.022436	0.00641	0.0625	2.785714	0.004109	1.042735
365	(Procurement of Goods/ Information Technology Services)	(Prediction test of TOEIC', 'Acceleration and Implementati on of Internal Compliance Unit', 'IT Service Management Foundation)	0.102564	0.022436	0.00641	0.0625	2.785714	0.004109	1.042735
122	(Procurement of Goods/ Information Technology Services)	(Implementati on of the Standard Budgeting Cost Structure for Work Units within the Ministry of Finance)	0.102564	0.028846	0.00641	0.0625	2.166667	0.003452	1.035897
159	(Procurement of Goods/ Information Technology Services)	(Government Procurement of Goods/ Services)	0.102564	0.057692	0.00641	0.0625	1.083333	0.000493	1.005128
15	(Procurement of Goods/ Information Technology Services)	(Computer Network Administra- tion)	0.102564	0.060897	0.00641	0.0625	1.026316	0.000164	1.001709

Calculation using a minimum support value of 0.009 resulted in 118 rows of itemsets, as shown in Table 8.

Table 8. Calculation With Minimum Support Of 0.009

No	support	itemsets
0	0.060897	('Computer Network Administration')
1	0.073718	('Acceleration and Implementation of Internal Compliance Unit')
2	0.016026	('Workload Analysis')
3	0.009615	('Basic Level Company Financial Statement Analysis')
4	0.012821	('Basic Level Financial Statement Analysis')
...	...	...
113	0.009615	('Fulfillment of Competency Gap for Echelon IV Officials', 'IT Service Management Foundation', 'Prediction test of TOEIC')

No	support	itemsets
114	0.009615	('Examination of Employee Discipline Violations', 'IT Service Management Foundation', 'Prediction test of TOEIC')
115	0.009615	('Procurement of Goods/ Information Technology Services', 'IT Service Management Foundation', 'Prediction test of TOEIC')
116	0.016026	('Visionary Leadership Training Program', 'IT Service Management Foundation', 'Prediction test of TOEIC')
117	0.009615	('Acceleration and Implementation of Internal Compliance Unit', 'Visionary Leadership Training Program', 'IT Service Management Foundation', 'Prediction test of TOEIC')

Table 9. Result Of Apriori Algorithm With Minimum Support 0,009

No	antecedents	consequents	antecedent support	consequents support	support	confidence	lift	leverage	conviction
34	('IT Service Management Foundation')	('Prediction test of TOEIC')	0.038461538	0.038461538	0.038461538	1	26	0.036982249	inf
35	('Prediction test of TOEIC')	('IT Service Management Foundation')	0.038461538	0.038461538	0.038461538	1	26	0.036982249	inf
60	('Acceleration and Implementation of Internal Compliance Unit', 'IT Service Management Foundation')	('Prediction test of TOEIC')	0.022435897	0.038461538	0.022435897	1	26	0.021572978	inf
61	('Acceleration and Implementation of Internal Compliance Unit', 'Prediction test of TOEIC')	('IT Service Management Foundation')	0.022435897	0.038461538	0.022435897	1	26	0.021572978	inf
67	('Acceleration and Implementation of Internal Compliance Unit', 'Visionary Leadership Training Program')	('IT Service Management Foundation')	0.009615385	0.038461538	0.009615385	1	26	0.009245562	inf
...	...	...	...	...	...	...	...	...	...
50	('Procurement of Information Technology Goods/Services')	('Prediction test of TOEIC')	0.102564103	0.038461538	0.009615385	0.09375	2.4375	0.005670611	1.061007958
99	('Procurement of Information Technology Goods/Services')	('IT Service Management Foundation', 'Prediction test of TOEIC')	0.102564103	0.038461538	0.009615385	0.09375	2.4375	0.005670611	1.061007958
18	('Procurement of Information Technology Goods/Services')	('Communication Skills')	0.102564103	0.041666667	0.009615385	0.09375	2.25	0.00534188	1.057471264
52	('Procurement of Information Technology Goods/Services')	('Information Security Management System')	0.102564103	0.044871795	0.009615385	0.09375	2.089285714	0.005013149	1.053934571

Association rules were then created on the prepared baskets, resulting in a table with 122 rules as shown in Table 9. Based on the information provided in the Table 9, it can be inferred that employees who participate in the Procurement of Information Technology Goods/Services training are also likely to participate in the Communication Skills training. Then, the calculation was performed using a minimum support value of 0.013, resulting in 54 rows of itemsets as shown in Table 10.

Table 10. Calculation With Minimum Support Of 0.013

No	support	itemsets
0	0.060897	('Computer Network Administration')
1	0.073718	('Acceleration and Implementation of Internal Compliance Unit')
2	0.016026	('Workload Analysis')
3	0.016026	('Architecture Core')
4	0.016026	('Basic Level ICT Audit')
...	...	...
49	0.016026	('IT Service Management Foundation', 'Visionary Leadership Training Program')

No	support	itemsets	No	support	itemsets
50	0.016026	('Technical Reporting of Annual Tax Returns for State Civil Apparatus', 'Human Resource Planning')	53	0.016026	Management Foundation', 'Prediction test of TOEIC')
51	0.016026	('Visionary Leadership Training Program', 'Prediction test of TOEIC')			('Visionary Leadership Training Program', 'IT Service Management Foundation', 'Prediction test of TOEIC')
52	0.022436	('Acceleration and Implementation of Internal Compliance Unit', 'IT Service			

Table 11. Result Of Apriori Algorithm With Minimum Support 0,013

No	antecedents	consequents	antecedent support	consequents support	support	confidence	lift	leverage	conviction
6	('IT Service Management Foundation')	('Prediction test of TOEIC')	0.038462	0.038462	0.038462	1	26	0.036982	inf
7	('Prediction test of TOEIC')	('IT Service Management Foundation')	0.038462	0.038462	0.038462	1	26	0.036982	inf
14	('Acceleration and Implementation of Internal Compliance Unit', 'IT Service Management Foundation')	('Prediction test of TOEIC')	0.022436	0.038462	0.022436	1	26	0.021573	inf
15	('Acceleration and Implementation of Internal Compliance Unit', 'Prediction test of TOEIC')	('IT Service Management Foundation')	0.022436	0.038462	0.022436	1	26	0.021573	inf
20	('IT Service Management Foundation', 'Visionary Leadership Training Program')	('Prediction test of TOEIC')	0.016026	0.038462	0.016026	1	26	0.015409	inf
22	('IT Service Management Foundation', 'Prediction test of TOEIC')	('Visionary Leadership Training Program')	0.038462	0.022436	0.016026	0.416667	18.57143	0.015163	1.675824
0	('Acceleration and Implementation of Internal Compliance Unit')	('IT Service Management Foundation')	0.073718	0.038462	0.022436	0.304348	7.913043	0.019601	1.382212
2	('Acceleration and Implementation of Internal Compliance Unit')	('Prediction test of TOEIC')	0.073718	0.038462	0.022436	0.304348	7.913043	0.019601	1.382212
17	('Acceleration and Implementation of Internal Compliance Unit')	('IT Service Management Foundation', 'Prediction test of TOEIC')	0.073718	0.038462	0.022436	0.304348	7.913043	0.019601	1.382212

Table 11 contains a total of 36 association rules that were derived from the basket in Table 10. The information presented in Table 11 suggests that if an employee participates in the IT Service Management Foundation training, then they are also likely to enroll in the Prediction test of TOEIC training.

Table 12 provides a concise summary of the outcomes that were achieved as a result of implementing the Apriori algorithm model.

Table 12. Summary Of Modeling Results

Minimum support	Average confidence value	The number of associated training	The number of rules produced
0.006	0.456367	252	436
0.009	0.444433	118	122
0.013	0.62265	54	26

3.7. Evaluation

The model evaluation and validation were performed by calculating the average confidence value for each minimum support value so that the minimum confidence value was determined from the highest average confidence value of 0.6, as seen in Table 12, to obtain higher accuracy.

To obtain the rules with the best accuracy, a recalculation was performed using a minimum

confidence value of 0.6. The number of rules obtain from recalculation results can be seen in Table 13 in the last column. The details of the rules generated with a minimum support of 0.013 and minimum confidence of 0.6 are displayed in Table 14.

Table 13. Summary Of Modeling Results

Min. Support	Average confidence value	The number of associated training	The number of rules produced	The number of rules produced with min. confidence 0,6
0.006	0.456367	252	436	145
0.009	0.444433	118	122	35
0.013	0.62265	54	26	10

One of the training recommendations based on Table 14 is that if an employee has taken the Enterprise Service Bus (ESB)-API Management training, they will take the Enterprise Service Bus API Integration Platform training.

Table 14. Apriori Algorithm Calculation Results With A Minimum Support Of 0.013 And A Minimum Confidence Of 0.6

No	antecedents	consequents	antecedent support	consequents support	support	confidence	lift	leverage	conviction
6	('IT Service Management Foundation')	('Prediction test of TOEIC')	0.038462	0.038462	0.038462	1	26	0.036982	inf
7	('Prediction test of TOEIC')	('IT Service Management Foundation')	0.038462	0.038462	0.038462	1	26	0.036982	inf
14	('Acceleration and Implementation of Internal Compliance Unit', 'IT Service Management Foundation')	('Prediction test of TOEIC')	0.022436	0.038462	0.022436	1	26	0.021573	inf
15	('Acceleration and Implementation of Internal Compliance Unit', 'Prediction test of TOEIC')	('IT Service Management Foundation')	0.022436	0.038462	0.022436	1	26	0.021573	inf
20	('IT Service Management Foundation', 'Visionary Leadership Training Program')	('Prediction test of TOEIC')	0.016026	0.038462	0.016026	1	26	0.015409	inf
21	('Prediction test of TOEIC', 'Visionary Leadership Training Program')	('IT Service Management Foundation')	0.016026	0.038462	0.016026	1	26	0.015409	inf
9	('Visionary Leadership Training Program')	('IT Service Management Foundation')	0.022436	0.038462	0.016026	0.714286	18.57143	0.015163	3.365385

No	antecedents	consequents	antecedent support	consequents support	support	confidence	lift	leverage	conviction
12	('Visionary Leadership Training Program')	('Prediction test of TOEIC')	0.022436	0.038462	0.016026	0.714286	18.57143	0.015163	3.365385
23	('Visionary Leadership Training Program')	('IT Service Management Foundation', 'Prediction test of TOEIC')	0.022436	0.038462	0.016026	0.714286	18.57143	0.015163	3.365385
4	('Enterprise Service Bus (ESB)-API Management')	('Enterprise Service Bus: API Integration Platform')	0.032051	0.038462	0.022436	0.7	18.2	0.021203	3.205128

3.8. Deployment

Based on the results of the descriptive analysis, there are 69.07% of employees with an insufficient status, where employees participate in fewer than 2 internal/external training sessions per year. Therefore, HR needs to further analyze the causes of this occurrence.

Furthermore, according to the above Apriori model evaluation results, the 10 rules with the best accuracy can be utilized by the HR unit as input in determining training according to organizational needs and can improve employee competency according to their duties and functions.

4. Conclusion

This research was conducted using three datasets covering the period from 2011 to 2021. The datasets include historical employee training data with a total of 7.615 records, historical employee organizational data with a total of 2.701 records, and historical employee position data with a total of 4.950 records. These three datasets were processed and integrated during the data preparation stage to create a single binary table dataset consisting of 312 records, which encompassed employee information, roles, and training names.

In the modeling phase, a descriptive analysis was conducted based on historical employee training data to identify trends in training uptake per year, the most frequently taken trainings, and the number of employees with excess or insufficient amount of training. The results of the descriptive analysis revealed that the highest number of training sessions attended by employees for internal training occurred in 2019, with a total of 1,557 sessions. As for external training, the highest number of sessions attended was in 2021, with a total of 48 sessions. Additionally, it was found that the most commonly attended internal training format was E-Learning. In terms of training adequacy, there are 69.07% of employees with an insufficient status, where employees participate in fewer than 2 internal/external training sessions per year. Therefore, HR needs to further analyze the causes of this occurrence.

Next, modeling was conducted using the Apriori algorithm with the prepared binary table dataset to identify correlations between the trainings attended by employees. This research utilized three minimum support values, assuming that employees attended at least 2, 3, and 4 trainings. As a result, the obtained minimum support values were 0.006, 0.009, and 0.0013, respectively. Meanwhile, the minimum confidence value to achieve the highest accuracy was determined based on the highest average confidence value for each minimum support value, which was set at 0.6. Based on the model evaluation results, the best rule was generated with a minimum support value of 0.013 and a minimum confidence value of 0.6, resulting in 10 rules with strong correlations between antecedents and consequents, providing a potential accuracy of more than 60%. Further research can be conducted by comparing the performance of the Apriori algorithm with other algorithms to obtain the algorithm with the best performance.

References

- [1] Kementerian Keuangan, *Regulation of the Minister of Finance Number 217/PMK.01/2018 concerning the Organization and Work Procedure of the Ministry of Finance*. Jakarta, 2018. [Daring]. Tersedia pada: <https://jdih.kemenkeu.go.id/fullText/2018/217~PMK.01~2018Per.pdf>
- [2] R. Takdirillah, "Application of Data Mining Using the Apriori Algorithm for Transaction Data to Support Sales Strategy Information," *Edumatic J. Pendidik. Inform.*, vol. 4, no. 1, hal. 37–46, 2020, doi: 10.29408/edumatic.v4i1.2081.
- [3] Saefudin dan D. Fernando, "Application of Book Recommendation Data Mining Using the Apriori Algorithm," *JSiI (Jurnal Sist. Informasi)*, vol. 7, no. 1, hal. 50–56, 2020, doi: 10.30656/jsii.v7i1.1899.
- [4] Y. Zhou, "Design and Implementation of Book Recommendation Management System Based on Improved Apriori Algorithm," *Intell. Inf. Manag.*, vol. 12, hal. 75–87, 2020, doi: 10.4236/iim.2020.123006.
- [5] A. Syahrul dan A. Solichin, "Recommendations for Selecting Courses in Completing Student Study Plans with the Application of the Apriori Algorithm," *J. Tek. Elektro, Teknol. Inf. dan Komput.*, vol. 6, no. 1, hal. 79–88, 2022, doi: 10.31961/eltikom.v6i1.522.
- [6] A. W. O. Gama dan N. M. Widnyani, "Simple Modification for an Apriori Algorithm With Combination Reduction and Iteration Limitation Technique," *Knowl. Eng. Data Sci.*, vol. 3, no. 2, hal. 89–98, 2020, doi: 10.17977/um018v3i22020p89-98.

- [7] J. Zeniarja, D. L. A. Mangkawa, dan A. Salam, "Data Mining Applications for Violence Pattern Analysis with FP-Growth Algorithm," *J. Appl. Intell. Syst.*, vol. 6, no. 1, hal. 32–39, 2021, [Daring]. Tersedia pada: <http://publikasi.dinus.ac.id/index.php/jais/article/view/4444>
- [8] M. Tahir dan N. Sitompul, "The application of the FP-Growth algorithm in determining students' tendencies in selecting elective courses," *Netw. Eng. Res. Oper.*, vol. 6, no. 1, hal. 59, 2021, doi: 10.21107/nero.v6i1.216.
- [9] A. H. Nasyuha dkk., "Frequent pattern growth algorithm for maximizing display items," *TELKOMNIKA (Telecommunication, Comput. Electron. Control.)*, vol. 19, no. 2, hal. 390–396, 2021, doi: 10.12928/TELKOMNIKA.v19i2.16192.
- [10] H. Rahmawan, "Determining Self-Development Training Recommendations for Civil Servants Using the C4.5 Algorithm with Principal Component Analysis and Discretization," *J. Tekno Kompak*, vol. 14, no. 1, hal. 5, 2020, doi: 10.33365/jtk.v14i1.531.
- [11] S. Anas, N. Rumui, A. Roy, dan P. H. Saputro, "Comparison of Apriori Algorithm and FP-Growth in Managing Store Transaction Data," *Int. J. Comput. Inf. Syst.*, vol. 3, no. 4, hal. 158–162, 2022, [Daring]. Tersedia pada: <https://ijcis.net/index.php/ijcis/index>
- [12] A. Fachrurrozi, M. Junaedi, J. L. Putra, dan W. Gata, "Algorithm Implementation Of Interest Buy Apriori Data On Consumer Retail Sales In Industry," *JITE (Journal Informatics Telecommun. Eng.)*, vol. 4, no. 1, hal. 48–56, 2020, doi: 10.31289/jite.v4i1.3775.
- [13] Firmansyah dan A. Yulianto, "Market Basket Analysis for Books Sales Promotion using FP Growth Algorithm, Case Study: Gramedia Matraman Jakarta," *J. Informatics Telecommun. Eng.*, vol. 4, no. 2, hal. 383–392, 2021, doi: 10.31289/jite.v4i2.4539.
- [14] D. Kurniawan, M. S. Sipayung, R. Ismayanti, M. R. Ibrahim, dan S. A. Miranda, "Optimization of Stock Goods Inventory Fulfillment Strategies Using the FP-Growth Frequent Pattern Growth Algorithm," *Media Teknol. Inf. dan Komput.*, vol. 6, no. 2, hal. 104–114, 2022, doi: 10.47002/metik.v6i2.387.
- [15] J. Yang dkk., "Brief introduction of medical database and data mining technology in big data era," *J. Evid. Based. Med.*, vol. 13, hal. 57–69, 2020, doi: 10.1111/jebm.12373.
- [16] S. M. Ayyoubzadeh, S. M. Ayyoubzadeh, H. Zahedi, M. Ahmadi, dan S. R. Niakan Kalhori, "Predicting COVID-19 Incidence Through Analysis of Google Trends Data in Iran: Data Mining And Deep Learning Pilot Study," *JMIR Public Heal. Surveill.*, vol. 6, no. 2, hal. 1–6, 2020, doi: 10.2196/18828.
- [17] J. S. Saltz dan I. Krasteva, "Current approaches for executing big data science projects—a systematic literature review," *PeerJ Comput. Sci.*, 2022, doi: 10.7717/PEERJ-CS.862.
- [18] F. Martinez-Plumed dkk., "CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 8, hal. 3048–3061, 2019, doi: 10.1109/TKDE.2019.2962680.
- [19] J. S. Saltz dan N. Hotz, "Identifying the most Common Frameworks Data Science Teams Use to Structure and Coordinate their Projects," *Proc. - 2020 IEEE Int. Conf. Big Data (Big Data)*, hal. 2038–2042, 2020, doi: 10.1109/BigData50022.2020.9377813.
- [20] T. M. Afriyanti dan E. Retnoningsih, "Library Book Recommendation System using Frequent Pattern Growth Algorithm," *Techno.Com*, vol. 21, no. 2, hal. 292–310, 2022, doi: 10.33633/tc.v21i2.5789.
- [21] Sunarti, F. Handayanna, dan E. Irfani, "Analysis of Food Sales Patterns Using the Apriori Algorithm," *Techno.Com*, vol. 20, no. 4, hal. 478–488, 2021, doi: 10.33633/tc.v20i4.4715.
- [22] D. Wicaksono, M. I. Jambak, dan D. M. Saputra, "The Comparison of Apriori Algorithm with Preprocessing and FP-Growth Algorithm for Finding Frequent Data Pattern in Association Rule," *Sriwij. Int. Conf. Inf. Technol. Its Appl. (SICONIAN 2019)*, 2019, doi: 10.2991/aisr.k.200424.047.
- [23] M. Somalakshmi dkk., "An efficient apriori algorithm for frequent pattern mining using mapreduce in healthcare data," *Bull. Electr. Eng. Informatics*, vol. 10, no. 1, hal. 390–403, 2021, doi: 10.11591/eei.v10i1.2096.
- [24] I. A. Ashari, A. Wirasto, D. Nugroho Triwibowo, dan Purwono, "Implementation of Market Basket Analysis with Apriori Algorithm for Retail Business Income Analysis," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 3, hal. 701–709, 2022, doi: 10.30812/matrik.v21i3.1439.
- [25] A. Anggrawan, M. Mayadi, dan C. Satria, "Determination of Item Layout Accuracy using Apriori Algorithm and FP-Growth Algorithm," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 1, hal. 125–138, 2021, doi: 10.30812/matrik.v21i1.1260.