Association Pattern Analysis of Global Company Market Capitalization Using the FP-Growth Algorithm with Load Balancing Constraint

Stenly Ibrahim Adam[1], **Stenly Richard Pungus**[2], **Wilsen Grivin Mokodaser***[3]
Fakultas Ilmu Komputer, Universitas Klabat, Airmadidi, Sulawesi Utara
E-mail : stenly.pungus@unklab.ac.id[1], stenly.adam@unklab.ac.id[2], wilsenm@unklab.ac.id[3]*
**Corresponding author*
Received 9 October 2025; Revised 20 October 2025; Accepted 7 November 2025

Abstract - This research focuses on analyzing the global company market capitalization dataset using the FP-Growth algorithm combined with a load-balancing constraint approach. The main objective is to identify association patterns among different market capitalization categories Small, Medium, Large, Mega, and Ultra to understand their distribution and interrelationships. The study begins with data preprocessing, cleaning, and categorization of companies based on their market values. The FP-Growth algorithm is applied with a minimum support threshold of 0.02, and a load balancing constraint is introduced by filtering rules with support ≥ 0.05 and lift > 1, ensuring balanced and significant association patterns. The analysis results show that the most dominant categories are Medium and Small, representing the majority of companies worldwide, while Large, Mega, and Ultra categories are relatively rare. The strongest rule indicates that countries with "Large" companies are very likely to also have "Small" and "Medium" companies. Evaluation metrics show an average lift of 1.171 and an average confidence of 1.000, confirming strong and reliable associations. Overall, this study provides insights into global market capitalization patterns and demonstrates the effectiveness of FP-Growth with constraints in revealing meaningful, balanced relationships within large-scale business data.

Keywords – FP-Growth, Load Balancing Constraint, Market Capitalization, Association.

1. INTRODUCTION

In the modern global economy, market capitalization serves as a crucial indicator for assessing the size, financial strength, and stability of companies across countries and sectors. The analysis of market capitalization data can provide valuable insights into global economic patterns, investment opportunities, and country-level economic performance.[1] However, given the large volume and diversity of such data, identifying meaningful patterns or associations manually becomes increasingly complex and time-consuming.[2]

Market capitalization, as a measure of a company's total market value, reflects not only the confidence of investors but also the overall health of national and global economies.[3] The aggregation of market capitalization data across multiple countries enables comparative economic analysis and helps identify regions with dominant or emerging economic power.[4] Understanding these global market capitalization trends is essential for policymakers, financial analysts, and investors, as it supports decision-making in areas such as portfolio diversification, investment risk assessment, and economic policy development.[5] Therefore, exploring patterns in global market capitalization data can offer a more comprehensive understanding of international financial dynamics.

Data mining techniques, particularly association rule mining, have emerged as powerful tools for uncovering hidden relationships among variables within large datasets.[6] The FP-Growth (Frequent Pattern Growth) algorithm is widely recognized for its efficiency in extracting frequent itemsets[7] without the need for candidate generation, as required by Apriori. This makes

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i4.14885

FP-Growth suitable for analyzing high-dimensional financial datasets such as global company market capitalization records.[8][9] Its ability to handle large volumes of data efficiently makes it an effective method for identifying underlying associations between market categories across countries.

Despite its advantages, one of the challenges in association rule mining lies in maintaining balanced rule generation across different data segments.[10] Some categories, such as companies with large or ultra-large market capitalizations, tend to dominate the pattern discovery process, resulting in skewed outcomes that overlook smaller but significant relationships.[11] To address this issue, a load-balancing constraint can be applied to ensure a fair distribution of rules across different market capitalization levels or geographical regions.[12] This constraint helps prevent data bias and ensures that all company categories contribute meaningfully to the resulting association patterns[13].

In order to support this research, a review was carried out related to related research that had been carried out previously. Previous research has been conducted on the layout of goods in a store. This store receives around 450 transactions every month, and from April 2017 to August 2019, this store received 14,512 transactions with 74 items. This store continues to arrange the layout of goods based on groups and types of goods. This has an impact on service and item search, because customers will spend more time searching for items. The FP-Growth algorithm is used in this final project. From 6,366 processed transaction data, it was found that Men's T-shirts and Men's Shorts items were often purchased together, with a confidence value of 15% and support of 2%. Based on these findings, this researcher suggests that Men's T-shirts and Men's Shorts items be placed close together in the layout of the Doa Bunda Store[14][15]. Other research involved collecting transaction data from Toko Lia, using the FP-Growth algorithm, and analyzing the results to identify relevant transaction patterns. The results showed that the use of the FP-Growth algorithm efficiently optimized transaction data, allowing for the identification of transaction patterns that are crucial for the success of Toko Lia's marketing and inventory management strategies[16]. In other research, there were 34 association rules with lift values above 1, and 8 association rules were generated from the 5% minimum support test; 3 of these had confidence levels above 50%. The higher the minimum support and confidence values, the fewer combinations of association rules were generated [17]. Other research study used 150 transaction data and 17 products. The results showed that both algorithms can be used to determine association rules to determine product relationships at the Emyra Bedding Store. The association rules generated by the Apriori algorithm are also generated by the FP-Growth algorithm. The Apriori algorithm produces 2 association rules and the FP-Growth algorithm produces 10 association rules with a minimum support of 0.05 and a minimum confidence of 0.7. The resulting relationship patterns can help the Emyra Bedding Store monitor the stock of items frequently purchased by customers to prevent supply shortages [18]. Other research used a secondary data collection method on cash transaction data from the meatball and otak-otak sales factory at PT Kusno Baso Ciledug from May 11, 2021, to May 11, 2023, consisting of 600 transaction data points and 21 attributes that were cleaned without missing values. The results of this study provide an association pattern that explains which products are most frequently purchased at the same time, which can be used as a reference in sales strategies with a cross-selling system. The best experimental research results using the FP-Growth algorithm were obtained with a minimum Support parameter setting of 20%, a minimum Confidence of 70%, and a lift ratio value > 1, which resulted in 9 recommended association rules with 11 product items[19]. Further research using FP-Growth generated a data tree structure, better known as a frequent pattern tree. This system achieved optimal results in analyzing warehouse distribution patterns using the FP-Growth algorithm, which is more efficient and effective in making decisions with minimal risk[20]. In data mining analysis, there is the FP-Growth algorithm, which is a data mining technique for finding association rules from lots of data. Further research obtained this, known as the Association Rule, where the rule will be determined from the minimum support and

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i4.14885

confidence results. The rules that are formed produce 10 rules out of 51 existing data; these rules help determine the export sales strategy at the XYZ store[21].

Based on the discussion of the background and research that has been carried out previously, in this study, by integrating the FP-Growth algorithm with load balancing constraints, this research aims to identify balanced and representative association patterns from global market capitalization data. The results are expected to enhance the understanding of inter-country market capitalization trends, reveal potential correlations among company size categories, and contribute to more equitable financial data analysis. Furthermore, this approach provides a foundation for applying constraint-based data mining techniques in broader economic and financial analytics, offering valuable insights for global economic strategy and investment planning..

2. RESEARCH METHOD

The method used in this research will go through several stages, which can be seen in the figure below:

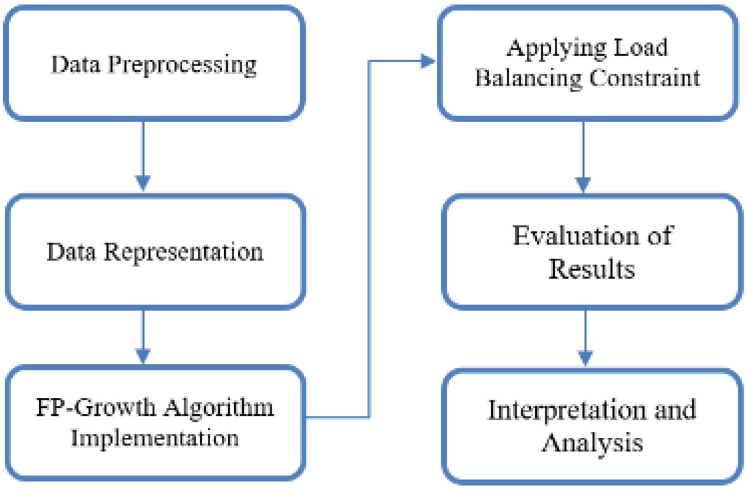


Figure 1. Research Steps

2.1. Data Preprocessing

In this stage, the raw dataset containing global company market capitalization information is cleaned and transformed into a structured format suitable for analysis.[22] The preprocessing steps include handling missing values, removing duplicates, standardizing numerical and categorical data, and converting market capitalization units into consistent values.[23] This ensures that the dataset is accurate, consistent, and ready for subsequent analytical processes.

2.2. Data Representation

After cleaning, the dataset is represented in a transactional format that can be processed by the FP-Growth algorithm. Companies are categorized based on market capitalization levels such as Small, Medium, Large, Mega, and Ultra. These categories are then grouped by country to form transactions, enabling the analysis of relationships between different market capitalization classes within and across nations.

2.3. FP-Growth Algorithm Implementation

In this phase, the FP-Growth algorithm is applied to discover frequent itemsets within the dataset.[24] The algorithm constructs an FP-tree structure that efficiently stores item relationships and identifies patterns of co-occurrence among market capitalization categories.[25] This step aims to uncover the most significant combinations of company capitalization classes that

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i4.14885

frequently appear together in different countries. application of algorithm calculations using the following formula Support Count to measure how often an item or itemset appears:

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

Minimum support threshold an itemset is considered frequent if:

$$Support(X) \geq MinSupport$$

Frequent itemset growth rule FP-Growth generates combinations from the FP-Tree. For any item A:

$$FrequentItemsets = \sum(ConditionalPatternBase(A)) \Rightarrow ConditionalFPTree(A)$$

Meaning that extract conditional pattern base of item A and construct conditional fp-tree and generate new frequent patterns by combining $A$with frequent prefix paths

2.4. Applying Load Balancing Constraint

To address potential bias toward dominant categories, a load-balancing constraint is introduced during pattern mining.[26] This constraint ensures that the discovered association rules are distributed more evenly across different capitalization levels and countries. By doing so, the analysis becomes more representative, highlighting not only large-cap companies but also the important patterns among small and medium ones. The goal is to ensure tasks are evenly distributed across available resources (servers, processors, or nodes).

$$\left| L_i - \bar{L} \right| \leq \epsilon \quad \forall i$$

Where: L_i= Load on resource i, $\bar{L}$= Average load across all resources ϵ= Allowed imbalance threshold (tolerance). And average load:

$$\bar{L} = \frac{\sum_{i=1}^{n} L_i}{n}$$

Meaning Each machine's load must not deviate too much from the system's average load.

2.5. Evaluation of Results

Once the frequent itemsets and association rules are generated, they are evaluated using standard metrics such as support, confidence, and lift.[27] These metrics measure how strong and relevant each discovered rule is in representing relationships among market capitalization categories. The evaluation helps identify the most valuable and reliable association rules that reflect meaningful patterns in the dataset.[28]

2.6. Interpretation and Analysis

In the final stage, the discovered association rules and their evaluation results are interpreted to draw meaningful insights about global market capitalization patterns. Visualizations and narrative analyses are used to explain relationships between countries and company size categories. This interpretation helps reveal global economic tendencies and provides a foundation for future research or decision-making in international finance and investment analysis.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i4.14885

3. RESULTS AND DISCUSSION

3.1. Data Preprocessing

Dataset berhasil dimuat! Jumlah baris: 7988

	Rank	Company Names	Company Code	Marketcap	Stock Price	Origin Flag	Country
0	1.0	Microsoft	MSFT	$3.033 T	$407.21	us	USA
1	2.0	Apple	AAPL	$2.951 T	$190.92	us	USA
2	3.0	Saudi Aramco	2222.SR	$2.026 T	$8.34	sa	S. Arabia
3	4.0	Alphabet (Google)	GOOG	$1.909 T	$153.46	us	USA
4	5.0	Amazon	AMZN	$1.653 T	$160.05	us	USA

Figure 1. Dataset stucture

In the Data Preprocessing stage, one of the essential steps is the cleaning and conversion of the “Marketcap” column. This process ensures that all market capitalization values are represented in a consistent numerical format suitable for quantitative analysis. The function convert_marketcap(value) was developed to handle various market capitalization formats, such as values expressed in trillions (T), billions (B), millions (M), or thousands (K). It systematically removes currency symbols ($), commas, and extra spaces, then converts each value into a float by multiplying the base number by its corresponding scale factor for example, “1.2T” becomes 1.2×10^{12}, and “500M” becomes 5.0×10^{8}. Any values that cannot be converted or do not match the expected pattern are replaced with NaN to mark them as missing data.

After the conversion, the function is applied to the entire column using the command df['Marketcap'] = df['Marketcap'].apply(convert_marketcap), and invalid entries are removed with df = df.dropna(subset=['Marketcap']) to maintain data consistency. This transformation not only standardizes the numerical representation of company market capitalization but also enhances the reliability of subsequent analyses, such as pattern discovery using the FP-Growth algorithm. By integrating this conversion process into the preprocessing pipeline, the dataset becomes clean, uniform, and analytically ready, reducing potential errors that could arise from inconsistent data formats in later stages of the study.

.

3.2. Data Representation analysis

In the Data Representation stage, the dataset is prepared in a format suitable for association pattern analysis by categorizing companies based on their market capitalization levels. This categorization allows the data to be structured in a way that reflects meaningful economic groupings, making it easier to identify relationships across company sizes and countries. In this step, a series of value ranges (bins) is defined to classify market capitalization into five categories: Small, Medium, Large, Mega, and Ultra. The code accomplishes this using pd.cut(), where the boundaries are set as [0, 1e9, 1e11, 5e11, 1e12, np.inf]. Each company’s market capitalization is then assigned a label corresponding to its range, stored in a new column named “Marketcap_Category.” This transformation converts raw numerical values into categorical data, which is essential for the FP-Growth algorithm that relies on item-based or categorical representations.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i4.14885

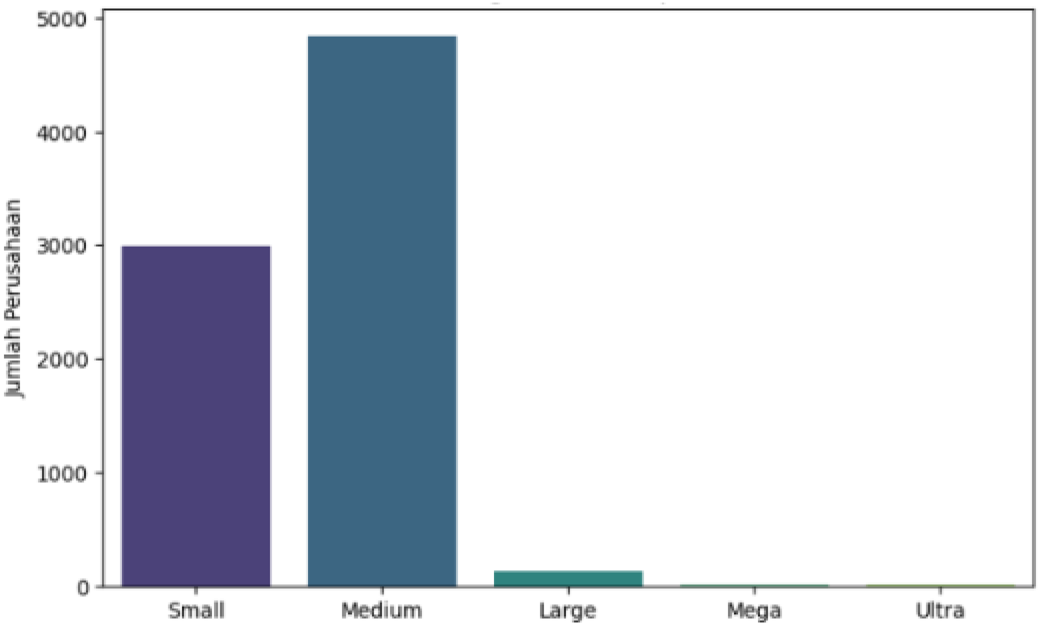


Figure 3. Company market capitalization category

After categorization, a visualization step is carried out to better understand the distribution of companies across these market capitalization categories. Using Seaborn's countplot(), the number of companies in each category is plotted as a bar chart, giving a clear overview of how many companies fall into the Small, Medium, Large, Mega, or Ultra groups. This visual representation helps identify the dominant company size categories within the dataset, highlights potential imbalances in data distribution, and provides insight into the global economic landscape before performing association rule mining in later stages.

The bar chart illustrates the distribution of companies based on their market capitalization categories: Small, Medium, Large, Mega, and Ultra. The data reveals a strong imbalance, with the Medium category dominating at around 5,000 companies, followed by the Small category with roughly 3,000 companies. In contrast, the Large, Mega, and Ultra categories contain significantly fewer companies, almost negligible in comparison. This indicates that the global corporate landscape is heavily concentrated in small and medium-sized enterprises, suggesting that these two segments form the backbone of most national economies, while only a few firms reach very large capitalization levels.

3.3. FP-Growth Algorithm Implementation and Applying Load Balancing Constraint

Table 1. Transaction data sample ready for FP-Growth

Marketcap Category	**Small**	**Medium**	**Large**	**Mega**	**Ultra**
Argentina	1	1	0	0	0
Australia	1	1	1	0	0
Austria	1	1	0	0	0
Bahamas	1	1	0	0	0
Belgium	1	1	1	0	0

The transaction table prepared for FP-Growth represents how different countries are associated with various market capitalization categories of companies. Each row corresponds to a country, while the columns (Small, Medium, Large, Mega, Ultra) indicate whether that country has companies in each respective category, where "1" means presence and "0" means absence.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i4.14885

For example, Argentina has companies in the Small and Medium categories but none in the Large, Mega, or Ultra categories. Similarly, Australia has a broader distribution, with companies present in Small, Medium, and Large categories, while Austria and the Bahamas show similar patterns to Argentina. This table structure allows the FP-Growth algorithm to identify frequent co-occurrences or associations between countries and company size categories.

This process represents the core implementation phase of the FP-Growth algorithm, where frequent itemsets and association rules are derived from the transactional dataset. First, the fpgrowth() function is applied to the prepared basket data to identify combinations of items (in this case, country–market capitalization categories) that frequently occur together, using a minimum support threshold of 0.02. This means that only item combinations appearing in at least 2% of all transactions are considered significant. The resulting output, stored in frequent_itemsets, lists these common patterns along with their corresponding support values. Next, the association_rules() function generates association rules from these frequent itemsets, evaluating their strength based on the confidence metric with a minimum threshold of 0.4. Confidence measures how likely an item (or group of items) appears in a transaction given the presence of another item. The printed output of frequent_itemsets.head() displays the top detected patterns, serving as the foundation for further analysis and evaluation of relationships among company market capitalization categories across different countries.

Table 2. Frequent Itemsets after FP-Growth Implementation

Support	**Item sets**
0.945205	(Medium)
0.821918	(Small)
0.246575	(Large)
0.027397	(Ultra)
0.027397	(Mega)

The results of the FP-Growth algorithm implementation show the frequent itemsets combinations of market capitalization categories that appear most often across different countries in the dataset. Each itemset is accompanied by its support value, which represents the proportion of transactions (countries) containing that particular category. From the output, the Medium market capitalization category has the highest support value of 0.945, meaning it appears in approximately 94.5% of all countries, followed by the Small category with 82.2%, and the Large category with 24.7%. In contrast, the Mega and Ultra categories have much lower support values of 2.7%, indicating they are relatively rare and concentrated in only a few countries. These results suggest that most countries have companies categorized as Small and Medium marketcap, while very few have Mega or Ultra companies. This insight highlights the overall global distribution pattern of company sizes, serving as a foundation for generating association rules in the next stage of the analysis.

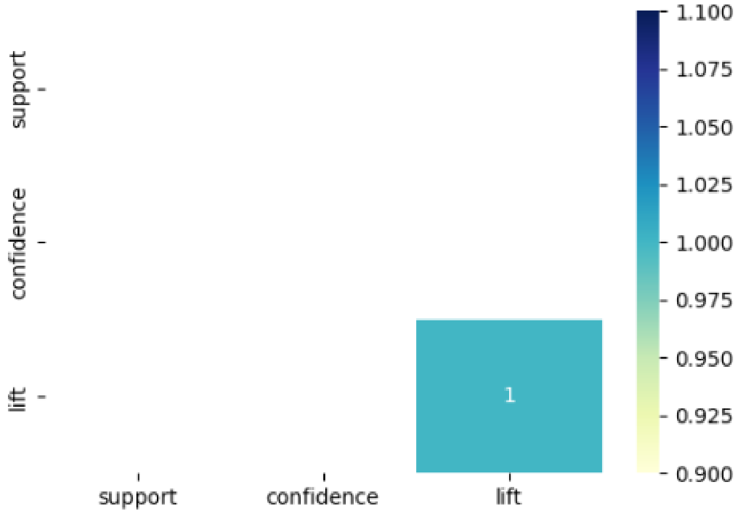
3.4. Evaluation of Results

This process represents the evaluation stage of the FP-Growth results, where the generated frequent itemsets are further analyzed to extract meaningful association rules and assess their strength. Using the association_rules() function, the algorithm calculates key evaluation metrics support, confidence, and lift to quantify the relationships between itemsets. The confidence threshold is set to 0.6, meaning that only rules with at least 60% reliability (where the consequent appears in at least 60% of cases when the antecedent is present) are retained.

After generating the initial rules, a constraint-based filtering step (referred to as load balancing) is applied to ensure that only strong and well-balanced associations are considered. Specifically, rules must have a support value of at least 0.05 (appearing in at least 5% of all transactions) and a lift greater than 1, indicating a positive correlation between the antecedent and consequent. Rules that meet these criteria are considered both statistically significant and

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i4.14885

practically relevant. The printed output displays the top association rules, including their antecedents, consequents, and evaluation metrics. This step allows researchers to interpret which market capitalization categories frequently co-occur across countries and to identify strong inter-category relationships supported by the data.




```
Jumlah Aturan Asosiasi Ditemukan: 5
Rata-rata Lift: 1.171
```

Figure 4. Evaluation matric corelation

The visualization illustrates the correlation between the main evaluation metrics support, confidence, and lift derived from the association rules generated by the FP-Growth algorithm. In this heatmap, the color intensity represents the strength of the correlation, while the numerical values indicate the degree of relationship between metrics. The result shows that only lift has a measurable correlation value (1.0) with itself, while the correlations between support, confidence, and lift are effectively zero or negligible. This indicates that in this dataset, the variations in support and confidence do not strongly influence the lift value, meaning each metric provides distinct information about the quality of association rules.

The summary below the heatmap reports that a total of 5 association rules were identified after applying filtering constraints, with an average lift value of 1.171. Since a lift value greater than 1 suggests a positive relationship between antecedents and consequents, these results imply that the discovered rules represent meaningful and non-random associations within the dataset. Overall, this evaluation confirms that while the relationships between metrics are weakly correlated, the rules produced are statistically valid and reflect consistent inter-category patterns.

3.5. Interpretation and Analysis

The interpretation and analysis of the FP-Growth results provide meaningful insights into the relationships between company market capitalization categories across different countries. From the analysis, a total of 5 association rules were discovered, with an average lift of 1.171 and an exceptionally high average confidence of 1.000, indicating that the identified rules are highly reliable and strongly supported by the data.

The strongest rule reveals that if a country has companies categorized as Large, it is very likely (100% confidence) to also have companies in the Small and Medium categories, with a support of 0.247 and a lift value of 1.304. This means that such a relationship occurs more frequently than expected by random chance, emphasizing a consistent co-occurrence between these company size levels. In other words, countries hosting large corporations tend to also have a well-balanced presence of smaller and medium-sized enterprises, suggesting a diversified economic structure.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i4.14885

The distribution of market capitalization categories further reinforces this interpretation the Medium (4,837) and Small (2,986) categories dominate globally, while Large (132), Mega (6), and Ultra (7) companies are significantly less common. This indicates that while a few countries possess very large corporations, the majority of markets are driven by medium and small-sized companies, forming the backbone of the global corporate landscape. Overall, these findings highlight a hierarchical yet interconnected structure in company size distribution across nations.

4. CONCLUSION

The overall process of applying the FP-Growth algorithm to the Companies Market Capitalization Dataset successfully revealed meaningful patterns and associations between different categories of company sizes across countries. Through systematic data preprocessing and transformation, the "Marketcap" column was standardized into a clean numerical format, ensuring analytical consistency. The subsequent categorization process translated these values into five meaningful economic groups—Small, Medium, Large, Mega, and Ultra—providing a structured representation for association rule mining.

During the FP-Growth Algorithm Implementation, frequent itemsets were extracted, revealing that Medium and Small market capitalization categories dominate globally, while Mega and Ultra companies are rare. The application of a Load Balancing Constraint (support ≥ 0.05 and lift > 1) ensured that only statistically strong and balanced rules were retained. The evaluation metrics showed that 5 significant association rules were found, with an average lift of 1.171 and average confidence of 1.000, confirming the reliability and positive correlation among the identified patterns.

From the interpretation and analysis, the strongest rule indicated that countries with Large companies almost always also have Small and Medium companies, reflecting a diversified and stable economic structure. This suggests that the presence of large corporations tends to coexist with smaller enterprises, implying that economic ecosystems are supported by multiple company sizes. Overall, the integration of FP-Growth with load balancing constraints provided not only statistically sound but also economically interpretable insights, emphasizing the hierarchical yet interconnected structure of global company distributions.

.

REFERENCES

[1] W. Chen and L. Wang, "The Global Economic Trends and Their Impact on National Economies," *Int. J. Accounting, Financ. Econ. Stud.*, vol. 2, no. 1, 2024.

[2] O. A. Bello, A. Folorunso, O. E. Ejiofor, F. Z. Budale, K. Adebayo, and O. A. Babatunde, "Machine learning approaches for enhancing fraud prevention in financial transactions," *Int. J. Manag. Technol.*, vol. 10, no. 1, pp. 85–108, 2023.

[3] G. Zhou, L. Liu, and S. Luo, "Sustainable development, ESG performance and company market value: Mediating effect of financial performance," *Bus. Strateg. Environ.*, vol. 31, no. 7, pp. 3371–3387, 2022.

[4] K. Hopewell, "Global Power Shifts and the Cotton Subsidy Problem: How Emerging Powers Became the New Kings of Cotton Subsidies," *Glob. Stud. Q.*, vol. 4, no. 2, p. ksae012, 2024.

[5] L. P. L. Cavaliere *et al.*, "The impact of portfolio diversification on risk management practices," *Nveo-Natural Volatiles Essent. Oils Journal| NVEO*, pp. 8447–8469, 2021.

[6] M. Soleimani, F. Campean, and D. Neagu, "Integration of Hidden Markov Modelling and Bayesian Network for fault detection and prediction of complex engineered systems," *Reliab. Eng. Syst. Saf.*, vol. 215, p. 107808, 2021.

[7] W. Thurachon and W. Kreesuradej, "Incremental association rule mining with a fast incremental updating frequent pattern growth algorithm," *IEEE access*, vol. 9, pp. 55726–55741, 2021.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i4.14885

[8] M. A. Ben Ali and O. Hammami, "Association Rule Mining for Market Basket Analysis in Retail Data: Enhancing Automated Knowledge Discovery with Apriori and FP-Growth Algorithms," *J. Data Mining, Knowl. Discov. Decis. Support Syst.*, vol. 12, no. 3, pp. 1–18, 2022.
[9] W. Shi, X. Zhang, H. Chen, and X. Zhang, "High dimensional data differential privacy protection publishing method based on association analysis," *Electronics*, vol. 12, no. 13, p. 2779, 2023.
[10] Y. Yang, N. Tian, Y. Wang, and Z. Yuan, "A parallel FP-growth mining algorithm with load balancing constraints for traffic crash data," *Int. J. Comput. Commun. Control*, vol. 17, no. 4, 2022.
[11] A. Akter and R. Hasan, "Frequent pattern mining with improved Apriori and FP-growth algorithms for big data applications," *J. Data Min. Anal.*, vol. 1, no. 1, pp. 1–9, 2024.
[12] Z. Yao and Z. Ding, "Learning distributed and fair policies for network load balancing as markov potential game," *Adv. Neural Inf. Process. Syst.*, vol. 35, pp. 28815–28828, 2022.
[13] B. Van Giffen, D. Herhausen, and T. Fahse, "Overcoming the pitfalls and perils of algorithms: A classification of machine learning biases and mitigation methods," *J. Bus. Res.*, vol. 144, pp. 93–106, 2022.
[14] E. Munanda and S. Monalisa, "Penerapan Algoritma Fp-Growth Pada Data Transaksi Penjualan Untuk Penentuan Tataletak Barang," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 7, no. 2, pp. 173–184, 2021.
[15] A. Anggrawan, M. Mayadi, and C. Satria, "Menentukan Akurasi Tata Letak Barang dengan Menggunakan Algoritma Apriori dan Algoritma FP-Growth," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 1, pp. 125–138, 2021.
[16] A. H. Talia, N. Suarna, and D. Pratama, "Penerapan Algoritma FP-Growth Dalam Analisis Pola Transaksi Untuk Optimalisasi Pengelolaan Data Transaksi di Toko LIA," *J. Kecerdasan Buatan dan Teknol. Inf.*, vol. 3, no. 1, pp. 26–36, 2024.
[17] M. Y. Ardianto, S. Adinugroho, and I. Indriati, "Penentuan Tata Letak Produk menggunakan Algoritma FP-Growth pada Toko ATK," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 9, pp. 3826–3832, 2021.
[18] D. A. Istiqomah, Y. Astuti, and S. Nurjanah, "Implementasi algoritma FP-growth dan Apriori untuk persediaan produk," *J. Inform. Polinema*, vol. 8, no. 2, pp. 37–42, 2022.
[19] F. Prasetyo and H. Hasugian, "Analisis pola pembelian produk makanan menggunakan algoritma FP-Growth untuk strategi penjualan," *IDEALIS Indones. J. Inf. Syst.*, vol. 7, no. 1, pp. 11–20, 2024.
[20] B. Aktavera, S. Satrianansyah, E. Elmayati, and H. O. L. Wijaya, "IMPLEMENTASI DAN ANALISIS ASOSIASION RULE MENGGUNAKAN ALGORITMA APRIORI DAN ALGORITMA FP GROWTH," *J. Teknol. Inf. Mura*, vol. 16, no. 1, pp. 54–61, 2024.
[21] A. F. Boy, S. Yakub, I. Ishak, and Z. Azmi, "Implementasi Data Mining Pada Pengaturan Distribusi Barang Dengan Menggunakan Algoritma Fp-Growth," *J. Sci. Soc. Res.*, vol. 5, no. 2, pp. 431–435, 2022.
[22] D. Ather, N. Chaudhary, G. Singh, T. Beig, and R. Kler, "Enhancing used automobile valuations: A data-cleaning and linear regression approach for predicting prices in competitive market," in *2023 4th International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, 2023, pp. 1–5.
[23] L. O. Joel, W. Doorsamy, and B. S. Paul, "A review of missing data handling techniques for machine learning," *Int. J. Innov. Technol. Interdiscip. Sci.*, vol. 5, no. 3, pp. 971–1005, 2022.
[24] M. Shawkat, M. Badawi, S. El-ghamrawy, R. Arnous, and A. El-desoky, "An optimized FP-growth algorithm for discovery of association rules," *J. Supercomput.*, vol. 78, no. 4, pp. 5479–5506, 2022.
[25] M. Ghosh, A. Roy, P. Sil, and K. C. Mondal, "Frequent itemset mining using FP-tree: a CLA-based approach and its extended application in biodiversity data," *Innov. Syst. Softw.*

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i4.14885

Eng., vol. 19, no. 3, pp. 283–301, 2023.
[26] E. Gures, I. Shayea, M. Ergen, M. H. Azmi, and A. A. El-Saleh, "Machine learning-based load balancing algorithms in future heterogeneous networks: A survey," *IEEE Access*, vol. 10, pp. 37689–37717, 2022.
[27] T. M. Alam *et al.*, "A novel framework for prognostic factors identification of malignant mesothelioma through association rule mining," *Biomed. Signal Process. Control*, vol. 68, p. 102726, 2021.
[28] M. H. Santoso, "Application of association rule method using apriori algorithm to find sales patterns case study of indomaret tanjung anom," *Brill. Res. Artif. Intell.*, vol. 1, no. 2, pp. 54–66, 2021.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i4.14885