

Using Big Data Analytics for Decision Making: Analyzing Customer Behavior using Association Rule Mining in a Gold, Silver, and Precious Metal Trading Company in Indonesia

Wecka Imam Yudhistyra^{a,1,*}, Evri Marta Risal^{b,2}, I-soon Raungratanaamporn^{a,3}, Vatanavongs Ratanavaraha^{a,4}

^aTransportation Engineering, Suranaree University of Technology, Thailand

^bProject and Enterprise Management School, University College London, United Kingdom

¹ D6040154@g.sut.ac.th; ² evri.risal.18@ucl.ac.uk; ³ isoonr@g.sut.ac.th; ⁴ vatanavongs@g.sut.ac.th

* corresponding author

ARTICLE INFO

Article history

Received May 10, 2020

Revised June 7, 2020

Accepted June 10, 2020

Keywords

big data analytics

customers

knowledge discovery

data science

decision making

ABSTRACT

Indonesia is facing many challenges in the fourth industrial revolution (4IR) era. One of them is related to big data technologies and implementation that can be seen clearly from Indonesia Industry Readiness Index (INI) 4.0. Therefore, focusing on implementing big data analytics in a gold, silver, and precious metal trading company is the objective of this manuscript to support daily business operations. To be more specific, the aim is to discover meaningful patterns and ensure high quality of knowledge discovery from the big data available in a company in Indonesia. It is needed to support the Making Indonesia 4.0 as a roadmap to implement industrial digitalization in Indonesia. The methodology used for the big data implementation in this manuscript is the combination of the CRISP-DM framework and key steps for customer analytics. The result of this research is a list of recommendations that facilitate strategic planning based on evidence of measurable big data analytics to achieve the business goals of a company.

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



1. Introduction

The fourth industrial revolution (*forth* 4IR) is a high-tech strategy deriving from technological advancement and disruptive development in the industrial sector worldwide that fundamentally alter the industrial environment work and relate to one another [1, 2]. With technological innovation, it has the potential to raise global income levels and improve the quality of the industrial sector around the world [3, 4]. Some consider it as the integration of emerging technologies such as the Internet of Things (IoT), Big Data, and Cloud technologies [5]. On the other hand, there are scientific arguments that the fourth industrial revolution is not only regarding technology integration but also concerning the whole concepts of 'how to acquire, share, organize data and resource that make and increase the product or service delivery faster, cheaper, more efficiency, more productivity, more effective, and more sustainable' [6-8]. Since technologies integration are considered the core of the 4IR, there some technologies frequently discussed in the literature: IoT, Big Data Analytics (BDA), Cloud, 3D printing, robotic, Artificial Intelligence (AI), Machine Learning (ML), and 5G technology [6, 9].

In response to this, President Joko Widodo has declared 'Making Indonesia 4.0' as a road map to implement industrial digitalization that can positively support industry performances in Indonesia [10]. It is followed by Indonesia Industry Readiness Index (INDI) 4.0 by the Ministry of Industry to support the government measuring progress on 4IR in Indonesia [11]. As the interest in the fourth industrial revolution is growing rapidly, this manuscript does not limit the focus on 4IR itself but intends to implement BDA as interconnection in supporting 4IR [12]. It is essential since the current research paradigm is at the fourth level, where big data is analyzed using various statistical methods for real-world needs in real-time settings [13]. However, the scientific manuscripts published in the journal related to BDA implementation are still deficient [14, 15], where many manuscripts explain that BDA can improve business strategy [16] and organizational performance [17].

Moreover, there is a big gap between big data available now and the capability to analyze it while there will be a big demand for data analytics skills at this time [18, 19]. Thus, this study aims to discover meaningful patterns and ensure high quality of knowledge discovery from the big data available in a company in Indonesia. It can provide new insight into what a company should offer as better incentives to its customers and what strategies a company should develop to allocate the resources appropriately. As a result, the main contributions of this manuscript are mentioned as follows:

- a) it presents analytics with a large amount of data to help a company strengthen its position to stay ahead of the competition;
- b) it minimizes any investment risk on researching and testing the market, product, or ideas;
- c) it facilitates strategic planning based on evidence of BDA to achieve the business goals of a company; and
- d) it helps the company in spotting emerging trends on customer needs and demands.

In addition, another contribution of this manuscript is that it helps a company to improve their performances, which are assessed by INDI 4.0 in Indonesia and increase the average score of 2.14 (scale 4) where the lowest score comes from metal companies in Indonesia [20].

2. Literature Reviews

2.1. Related works

There are few manuscripts published articles in journals related to BDA implementation. For instance, Batarseh and Latif in 2016 and Alani and team in 2016 assessed the quality of service and healthcare organization using BDA in healthcare industries in the United States and Iraq [21]; Moyne and Iskandar in 2017 implemented BDA in a manufacturing company [22]; [23]; Naimur Rahman and his team in 2016 used BDA to predict total electricity forecast in the United States [24]; Honarvar and Sami in 2019 tried to find a suitable solution for urban development by using the opportunities of big data and present data related to urban computing with the aim of assessing the knowledge that can be obtained through integration of multiple independent data sources in Smart Cities [25]. Meanwhile, the position of this manuscript that makes it different from previous related works is that rarely were studies found in the literature covering BDA in a gold, silver, and precious metal industry. This manuscript also explores the customer behaviors and habits based on a large amount of data analysis in conducting the transaction in a developing country like Indonesia.

2.2. Frequent patterns, association, and correlation

One famous method to analyze a large amount of data often used by experts and researchers is the association rule mining method. It is a common technique that leads to the discovery of associations and correlations among items in a large database using a variety of algorithms to generate and test the possible rules [15, 26]. A typical example of this method is market basket analysis, an analysis of customers buying habits by finding an association between different items

that customers place in their 'shopping basket' [27]. Using this analysis, a company can understand the purchase behavior of its customers, which can help a company to create better decision making for supporting its business goals [28]. For instance, if customers tend to purchase a computer and a printer together, then having a sale on printers may boost the sale of printers as well as computers, and at the same time is shown in the following association rule:

$$\text{computer} \Rightarrow \text{printer} [\text{support} = 20\%, \text{confidence} = 60\%] \quad (1)$$

where a function defining *support* for itemset x can be defined as [29]:

$$\text{support}(x) = \frac{\text{count}(x)}{N} \quad (2)$$

where N is the number of purchased transactions in the dataset and $\text{count}(x)$ is the number of transactions containing itemset x , and the confidence is a measurement of its predictive accuracy where the *support* of itemset containing both x and y divided by the *support* of the itemset only x as followed

$$\text{confident}(x - y) = \frac{\text{support}(x, y)}{\text{support}(x)} \quad (3)$$

Rule *support* and *confidence* are two measures the usefulness and certainty of discovered rules. A *support* of 20% means that 20% of all the transactions under analysis show that computers and printers are purchased together. A *confidence* of 60% shows that 60% of the customers who purchased a computer also bought the printer. Rules are recognized interesting if they satisfy both a minimum *support* threshold and a minimum *confidence* threshold where users or experts can set it. Additional analysis can be performed to discover the interesting statistical correlation between associated items to get better knowledge. Furthermore, there are various algorithms to perform association rule mining and to compute large datasets such as the Apriori algorithm and the Continuous Association Rule Mining Algorithm (CARMA).

2.3. Apriori algorithm

Apriori is an algorithm proposed by Agrawal and Srikant in 1994 for frequent mining itemset for Boolean association rules [30]. Sometimes, the term association rules and Apriori are often used in reverse. The first step of this algorithm simply counts the item occurrences to determine the large 1-itemsets. A subsequent pass, say pass k , consists of two phases. First, the large itemsets L_{k-1} found in the $(k-1)$ -th pass are used to generate the candidate itemsets C_k , using the Apriori-gen function. The Apriori-gen function takes as argument L_{k-1} , the set of all large $(k-1)$ -itemsets. It returns a superset of the set of all large k -itemsets. Afterward, the database is scanned, and the *support* of candidates in C_k is counted. For fast counting, we need to efficiently determine the candidates in C_k that are contained in a given transaction t .

2.4. Continuous Association Rule Mining Algorithm (CARMA)

This algorithm needs, at most, two scans of the transaction sequence to produce all large itemset where during the first scan, this algorithm continuously constructs a list of itemsets (a lattice) of all potentially a large itemset. Then, during the second scan, this algorithm determines the precise *support* of each set in the lattice and continuously remove all small itemset [31].

The more detail explanations about every phase in CARMA algorithm are explained by Huang et. al. (2009) as follows [32]. Phase I objective in this algorithm is to produce all potentially large itemsets. Then it is followed by storing the picked itemset and their related value vectors which it is handled by lattice V . The value vector is a three-dimensional vector composed of *count* (v), *maxMissed* (v), and *firstTrans* (v) where *count* (v) is the number of occurrences of v since its insertion in the lattice; *maxMissed* (v) is the upper bound on the number of occurrence of v before v is inserted in the lattice; and *firstTrans* (v) is the index of the transaction at which v is inserted in

the lattice. Both $count(v)$ and $maxMissed(v)$ can define the interval of estimated $support$ of v . The lower bound is $minSupport_i(v) = count_i(v) / i$ and the upper bound is $maxSupport_i(v) = (maxMissed_i(v) + count_i(v)) / i, j$ is the current transaction number. A $support$ threshold can be specified by user. This threshold makes a sequence $\sigma = (\sigma_1, \sigma_2, \sigma_3, \dots)$. Based on the sequence σ , a new index $[\sigma]_i$ can be constructed, namely the ceiling of θ up to i . $[\sigma]$ is the last monotone decreasing sequence which is up to i pointwise greater of equal to σ and θ otherwise. Assume that $i-1$ transactions have been processed and let V_{i-1} be the $support$ lattice up to $i-1$. It is reading the i -th transaction t_i and update the V_{i-1} to V_i . σ_i is specified for the i -th transactions. For any itemset v , if it has already been in the lattice v , then it updates the $count(v)$, increasing it by one. When the set of v is not in V , it considers the following rules to judge whether or not to let it in. If v is 1-item set then it is directly inserted with $count(v) = 1, firstTrans(v) = i, maxMissed(v) = 0$. If v is k -item set ($k > 1$), two requirements should be satisfied. The first is that all subsets w of v have already been contained in V and are potentially large. Phase I also includes a prune step —only working every $1/\sigma_i$ or every 500 transactions (default) — whichever is the larger.

Phase II removes trivially small itemsets based on the last specified $support$ threshold. It rescans the process again and determines the accurate $support$ and continues to remove the trivially small dataset. If the user do not need precise $support$, Phase II can be omitted [31].

3. Case Study

The combination of key steps for customer analysis [33] and the CRISP-DM framework [34, 35] are fused to ensure the high quality of knowledge discovery in this case study. This case study is initially divided into three parts: pre-processing steps, processing-steps, and post-processing steps. Then these three parts is combined with the CRISP-DM framework. First parts, the pre-processing steps are business understanding, data understanding, and data preparation. In the second part, the processing steps are the modeling and evaluation; and the post-processing parts are the final results, such as strategic planning and policies developed in an effective way. The comprehensive framework for this case study is shown in Figure 1 as follows.

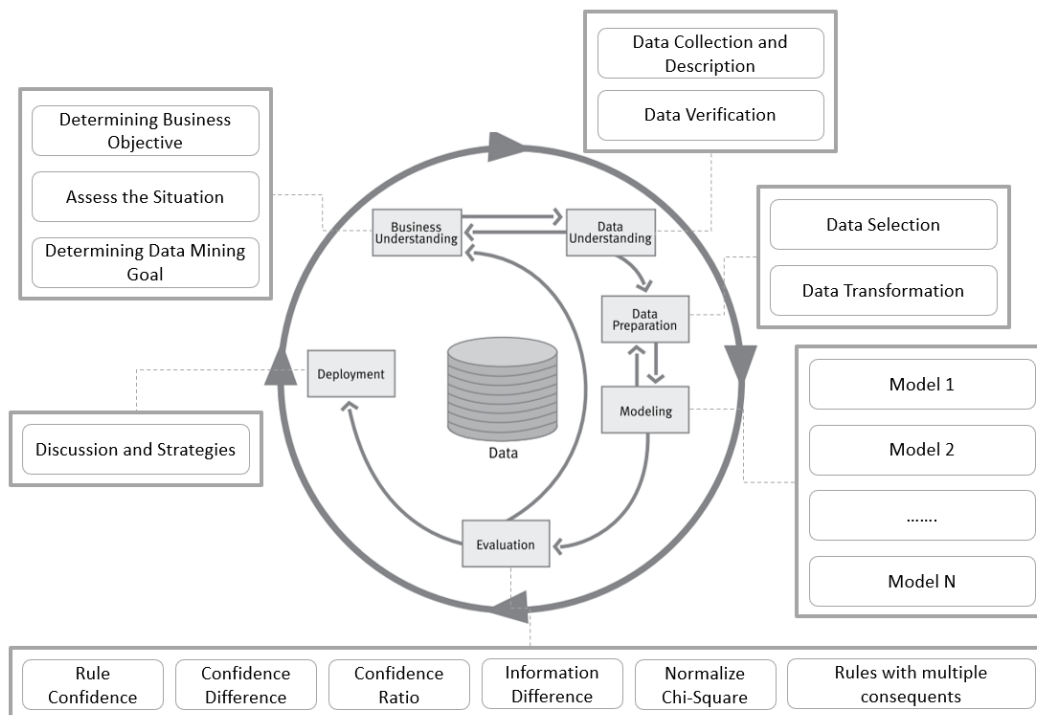


Fig 1. Combination of CRISP-DM and keys steps for developing company strategic planning

3.1. Business understanding

The important factor in this step is the current business situation in the gold, silver, and precious metal trading company, which intends to use BDA to gain knowledge about the behaviors of its customers in conducting a transaction. It is used to develop the company's shelf management and campaign strategies in offering the company's products. To be more detailed, the intention of using BDA in this case study covers a) finding frequent itemset patterns on the company's historical transaction dataset using b) recognizing peculiarities on customers' transactional data; c) identifying useful and actionable pattern to be aligned with business strategies.

Detail fact-findings found to support BDA implementation, such as the resource availability (software, hardware) or legal issues in which the data is allowed to be utilized, had been completed before beginning the project plan. A list of possible project plans to be performed during the rest of the project is listed as follows.

- a) A collection of the dataset is collected from various resources available at the company.
- b) Association rule mining techniques are utilized to discover meaningful relationships or patterns from the dataset.
- c) Data mining software applications such as R, Modeler, Tableau, or RapidMiner are considered to be used for analyzing the dataset and calculating the accuracy of predictive models.
- d) The data mining goal criteria are categorized as successful outcomes when the accuracy of the models' percentage is above 50%.
- e) The repetition of the modeling and evaluation process is adjusted based on rule confidence, confidence difference, confidence ratio, information difference, normalize chi-square, and rules with multiple consequents.

3.2. Data understanding and preparation

3.2.1. Data collection

The datasets were provided by the company, which was exported from multiple resources of company database covering July 2010 to October 2019. It contains 3,986,872 observations from 248,856 customers. There are seven variables on the dataset, and the description of datasets is provided by the company shown in Table 1 as follows.

Table 1. Initial data

Attribute Name	Description
CUSTOMERS_ID	The ID of the customers
TRANS_DATE	The date of the transaction
TRANS_ID	The invoice number of the order transaction
CUSTOMERS_ADDRESS	The address of the customers
DESCRIPTION	Description of the product
QUANTITY_TOTAL	Quantity total of the product
PRICE_TOTAL	Price total of the product

3.2.2. Data verification

Exploratory data analysis was performed to verify the quality of the data. It covered exploring the structure of the data, measuring central tendency, removing the null of the tuple in the datasets (cleaning), sorting, hashing, exploring the relationship, grouping, aggregation, merging, visualizing the relationship, examining relationship, checking the collinearity, outlier's detection, etc. Afterward, there is a list of some unsatisfied results found in the data quality verification steps as follows.

- a) *Incorrect data*. Some of the data contain common errors, such as error product names and biased customer data.
- b) *Exogenous force*. Some of the data contain a contradiction with the facts and circumstances such as production date or year for some products.
- c) *Duplication*. It is a common mistake made by people when inputting the data, such as similar customer data with different ID recorded in the systems.
- d) *Inconsistent naming standard*. When it is seemingly simple, this kind of disorganization often wreaks havoc on data analysis.

All of these mistakes can lead to measurement errors that cannot be fitted with existing theory. They will inaccurately skew the dataset, resulting in corrupt predictions and poor decision-making because of the unrepresentative dataset. There is some solution to minimize these mistakes such as automate data entry to reduce human error, use consistent and organized naming standard, schedule training events for end-users for the systems to make they understand the value of data, and consider moving to the cloud systems which the systems can see time-stamped logs that users have made throughout the day.

3.2.3. Data transformation

The datasets were reduced to be suitable with the variables used for the algorithms based on practical knowledge. The datasets are transformed into 80 variables shown in Table 2 as follows. The product variables are transformed to become tabular data that is suitable for building association rules mining models.

Table 2. Data transformation

Attribute name	Description
TRANS_ID	The ID of the transaction
PRODUCT_N (1-79)	The tabular data of products (1-79) that have items represented by separate flags, where each flag field represents the presence or absence of a specific item

3.3. Modeling, evaluation, and deployment

After the data preparation completed, initial data analysis for finding associations and checking interrelation among the products in every transaction, which can be seen in Figure 2 as follows. Figure 2 denotes the number of times that the two items occurred in the same transaction. For instance, the number of times that 'product 8' is purchased with 'product 11' is 1,007 times. This observation will provide an initial overview of customer behavior in their shopping cart items.

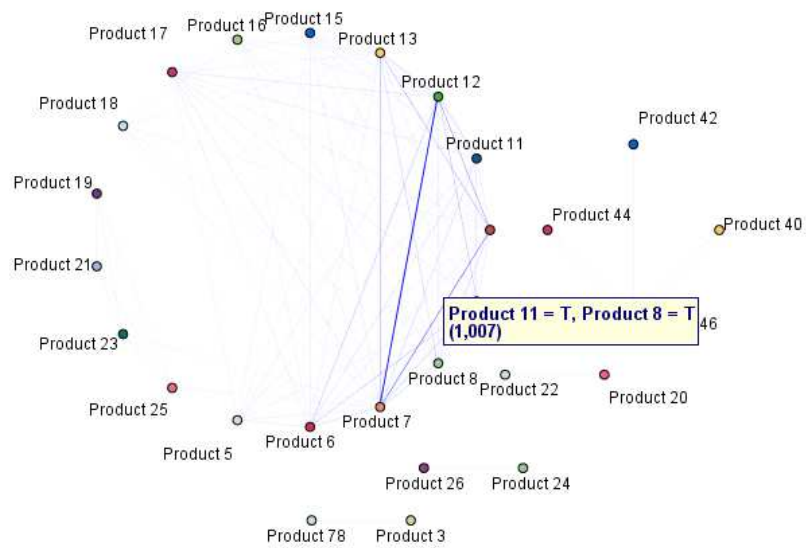


Fig 2. A web graph shows the number of times that the two items occurred in the same transaction

To be more specific, the relationship of each transaction is divided into three parts (i.e., strong link, medium link, and weak link) of item frequency in the data, as shown in Figure 3 (the thicker the line, the stronger the relationships).

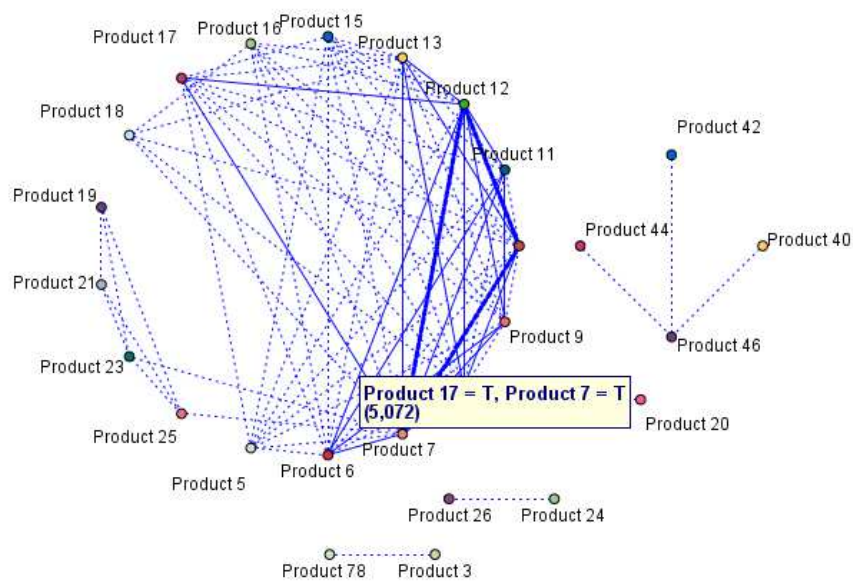


Fig 3. Dividing the item frequency relationships

Information from Figure 3 shown that some products have a strong relationship based on the company transaction history if the items occurred in the same transaction more than 15,000 times. Some products that have strong relationships are ‘product 7’ and ‘product 12’ with 38,123 times; ‘product 10’ and ‘product 7’ with 23,367 times; ‘product 10’ and ‘product 12’ with 15,530 times; and so forth. The products have a weak relationship if the items occurred in the same transaction less than 5,000 times. Some products that have weak relationships are ‘product 23’ and ‘product 7’ with 291 times; ‘product 18’ and ‘product 8’ with 532 times; ‘product 21’ and ‘product 25’ with 296 times; and so forth. It is also sufficient that there are no customers purchased ‘product 19’ and ‘product 18’ in the same transaction so that the company should not promote these products at the same time for its customers.

A predictive model was extracted using an Apriori algorithm to get deeper knowledge and to discover more association rules in the data as a market basket analysis. In this case, if the company wants to send a promotional campaign through multiple channels for existing customers, actionable strategies can be developed better after knowing the recommendations of a strong association in its sales products. Using the Apriori algorithm, there are 21 rules found from the analysis result, as shown in Table 3 as follows.

Table 3. Rules developed by Apriori algorithm

Consequent	Antecedent	Support (%)	Confidence (%)	Lift
Product 13	Product 8	1.3	63.026	5.547
	Product 10			
Product 13	Product 8	1.354	58.763	5.171
	Product 7			
Product 7	Product 8	1.3	63.375	1.905
	Product 10			
Product 10	Product 8	1.354	60.838	3.58
	Product 7			
Product 12	Product 8	1.354	52.002	2.152
	Product 7			
Product 9	Product 11	1.259	57.367	6.674
	Product 6			
Product 12	Product 11	1.453	50.418	2.087
	Product 9			
Product 12	Product 11	1.259	51.275	2.122
	Product 6			
Product 12	Product 11	1.526	56.452	2.337
	Product 7			
Product 12	Product 9	1.7	59.902	2.48
	Product 7			
Product 12	Product 6	1.606	61.653	2.552
	Product 7			
Product 10	Product 13	1.5	61.593	3.624
	Product 12			
Product 7	Product 13	2.461	50.737	1.525
	Product 10			
Product 10	Product 13	2.252	55.446	3.262
	Product 7			
Product 7	Product 13	1.5	69.608	2.092

	Product 12			
Product 7	Product 10	2.816	63.284	1.902
	Product 12			
	Product 13			
Product 8	Product 10	1.249	50.632	5.998
	Product 7			
	Product 13			
Product 8	Product 12	1.044	51.285	6.076
	Product 7			
	Product 9			
Product 6	Product 12	1.018	56.187	6.109
	Product 7			
	Product 13			
Product 12	Product 10	1.249	62.756	2.598
	Product 7			
	Product 13			
Product 10	Product 12	1.044	75.061	4.417
	Product 7			

When constructing the rules, the minimum antecedent *support* of 1% and the minimum rule of *confidence* of 50% were adjusted using this algorithm. Using three ways (*support* percentage, *confidence* percentage, and *lift* value) to measure the association, the first rule on the table above informs as, “if a customer buys ‘product 8’ and ‘product 10’ (with 1.3% frequency of transaction), they will also buy ‘product 13’ (with 63.026% *confidence*). The second rule informs that customers will buy ‘product 13’ if they previously buy ‘product 8’ and ‘product 7’. This rule covers 1.354 percent of the transactions and is correct in 58.763 percent of purchases involving ‘product 13’ (and so on until the last rule in Table 3 above). In addition, *lift* values in Table 3 above indicate the strength of a rule over the random occurrence of the antecedent item(s) and the *consequent* item(s). It basically mentions the ratio of the *confidence* and the *support* (the strength of any rule). A *lift* value higher than 1 measures that the *consequent* item(s) is likely to be bought if the *antecedent* item(s) is bought, while a value less than 1 means that *consequent* item(s) is unlikely to be bought if the antecedent item(s) is bought. A larger *lift* value is, therefore, a strong indicator that a rule is important and reflects a true connection between the items.

Furthermore, a CARMA algorithm was also performed to be compared with the Apriori results in order to improve the product recommendations. Another advantage of using CARMA over Apriori is that ‘if rules with many *consequents* are desired, which it can generate association rules with multiple *consequents*, and it also requires less space and time than using the Apriori algorithm. When constructing the rules, the minimum antecedent *support* of 1% and the minimum rule of *confidence* of 50% were also adjusted using this algorithm. The result, unfortunately, has no multiple *consequents* rules like it is expected before. There are five rules which are similar to the Apriori rules, as shown in Table 4.

Table 4. Rules developed by CARMA algorithm

Consequent	Antecedent	Support (%)	Confidence (%)	Lift
Product 7	Product 12	1.5	69.608	2.092
	Product 13			
Product 7	Product 10	2.816	63.284	1.902
	Product 12			
Product 12	Product 7	1.7	59.902	2.48
	Product 9			
Product 10	Product 13	2.252	55.446	3.262
	Product 7			
Product 7	Product 10	2.461	50.737	1.525
	Product 13			

After rerunning the models several times and comparing the results, unique findings are to identify the most useful rules quickly, so the attention is paid directly to the ones with the highest *support*, *confidence*, or *lift* values. The best five rules, according to the *lift* statistic, are denoted in Table 5 as follows.

Table 5. The best five rules according to the *lift* statistic

Consequent	Antecedent	Support (%)	Confidence (%)	Lift
Product 9	Product 11	1.259	57.367	6.674
	Product 6			
Product 6	Product 9	1.018	56.187	6.109
	Product 12			
Product 8	Product 7	1.044	51.285	6.076
	Product 13			
Product 8	Product 12	1.249	50.632	5.998
	Product 7			
Product 13	Product 13	1.3	63.026	5.547
	Product 10			

These rules appear to be more sophisticated than those in Table 3. The first rule, with a *lift* of about 6.674, implies that customers who buy 'product 11' and 'product 6' are six times more likely to buy 'product 9' than the typical customers. Rule two and rule three are also similar to rule one, which is over six times more likely to be found in the dataset.

Although the *confidence* and the *lift* could give high values, some measurements were conducted to evaluate and compare the associations that provided clear and useful insight. A common approach is to take the association rules and divide them into three categories, as follows [29]:

- a) The aim of using association rule mining is to figure out valuable patterns that provide a meaningful insight from datasets that seem obvious once discover.
- b) Sometimes, the rules are trivial, not worth to be considered as valuable patterns. Although the rule is clear, sometimes it is not appropriate.
- c) When the connection between the items is ambiguous, the rule is called inexplicable. It means that figuring out how to use the insight from the patter is absurd.

In this case, suppose that giving the preceding rule, the marketing team has enthusiasm about the possibility of creating an advertisement to promote ‘*product 8*’ and ‘*product 10*’, which are now in season (i.e., Eid Mubarak season for Muslim people or Christmas for Christian people). Before finalizing the campaign, however, the rules reveal that ‘*product 8*’ and ‘*product 10*’ are often purchased with other items, as shown in Table 6 as follows.

Table 6. An actionable rule

Consequent	Antecedent	Support (%)	Confidence (%)	Lift
Product 13	Product 8	1.3	63.026	5.547
	Product 10			
Product 7	Product 8	1.3	63.375	1.905
	Product 10			

There are two rules involving ‘*product 8*’ and ‘*product 10*’. A rule with *lift* of 5.547 seems to be interesting enough to be called an actionable rule than another one. Meanwhile, the combination of ‘*product 13*’ and ‘*product 12*’ are also purchased frequently with ‘*product 10*’ with *lift* of 3.624, but the ‘*product 7*’ with *lift* of 2.092 (lower than the *lift* of ‘*product 10*’) shown in Table 7. It is a hidden gem because it has a stronger interrelationship than the ‘*product 10*’ with ‘*product 13*’ and ‘*product 12*’ on the web graph.

Table 7. Rules with lower *lift* score become a hidden gem after evaluation steps

Consequent	Antecedent	Support (%)	Confidence (%)	Lift
Product 10	Product 13	1.5	61.593	3.624
	Product 12			
Product 7	Product 13	1.5	69.608	2.092
	Product 12			

Meanwhile, the processing time (elapsed time for the model build) for both algorithms with the same setting have variation when executing the algorithms. In this case, the Apriori algorithm has better performance when it is compared with the CARMA algorithm shown in Figure 4 as follows.

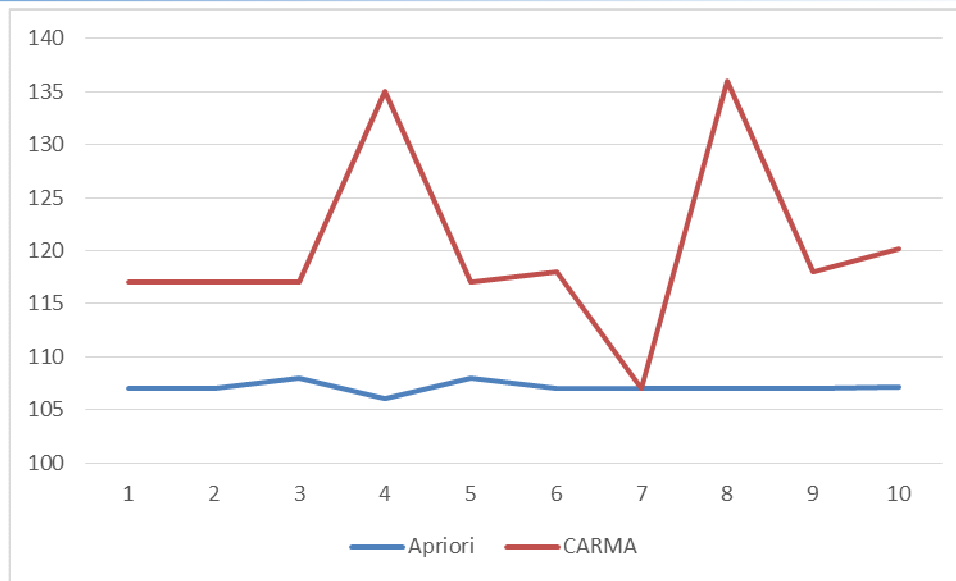


Fig 4. Performance comparison between the Apriori algorithm and the CARMA algorithm

Information from Figure 4 above shows that the range of elapsed time for the model build using the Apriori algorithm is between 106 seconds and 108 seconds (107 seconds for average time). It is faster than the CARMA algorithm, which has ranged between 107 and 136 seconds (120 seconds for average time). Furthermore, the identification of products in this case study depends on the datasets used, and the result of the number of product recommendation has no difference using both Apriori and CARMA algorithm. Although several researchers mention that the CARMA algorithm performs better than the Apriori algorithm [31, 32], the results in this case study show the opposite. The results using the CARMA algorithm deliver that it is slower, and it is not satisfying to satisfy business needs when it is compared to the Apriori algorithm. After finishing the general procedure to create the relevant model, some general strategies that can be proposed listed as followed.

- a) *Create a new retail store layout to increase sales.* Start with attractive window display and design effective shelf display or place the product optimally based on the best model which has been chosen. For example, 'product 9' should be put near 'product 11' and 'product 9'. Additional store layout ideas such as have a good lighting, use colour to convince customers (red colour indicates urgency and promotes sales), offer welcoming tones, and speed up the checkout process can be considered as business strategies to be implemented.
- b) *Use product categorization.* Offering a promotional campaign related product in customers basket based on the best model when a customer conducts the transaction with the company;
- c) *Response campaign documentation.* Documenting the promotional campaign whether a customer accepts or reject the promotional campaign for further analysis.
- d) *Information technology and system upgrade.* Know every detail about meaningful orientation to apply these techniques into real-time information systems with integration of logistic and distribution systems, retail systems, customer service systems, transportation, and warehouses systems in the company. In the COVID 19 pandemic era where some countries implement lockdown policy, conventional or traditional business models without sophisticated technology integration will lead companies into bankruptcy.

These strategies will also change the relationship between customers and a company in marketing their products, which can make promotional activities measurable than previously. The important things to be noted is to document the all implementation of business strategies as a whole and hold a sustainable action review with all stakeholders and go through lesson learned.

4. Conclusion

Knowledge is power. Using association rules which start with an unsupervised learning process, a company will have the capability of extracting knowledge from a large database without any prior knowledge of what patterns to seek. After conducting this case study, here is a list of meaningful conclusions from these models and recommendations for companies:

- a) This manuscript presents a novel implementation of big data analytics to change the way a company developing a new strategy for its customers, especially in a gold, silver, and precious metal trading company in a developing country like Indonesia.
- b) The challenge in implementing big data analytics is that the infrastructure and technology to analyze it. The more data collected by a company, the higher computational technologies and infrastructure are required.
- c) With the Apriori algorithm and CARMA algorithm for modeling in this manuscript, a company can explain in detail how the association of each product can be determined. In case of each segmentation using both procedures, the company can promote the possible products to its customers and deliver attractive offers or advanced service as promotional campaign strategies.

In addition, this manuscript is not a theoretical paper but a technical paper which can be useful guidance for managers, professional, or lecturer in implementing BDA. Further research should focus on observing the development of this big data analytics implementation and develop sustainability and real-time model for developing a new model. Once again, if companies really want to improve their business and performance, they should consider implementing big data analytics and embrace its challenges. In this 4IR era, it not only supports decision-makers to make the most valuable decision based on measurable data, but it also can achieve business goals easier and more effectively.

References

- [1] K. Schwab, "The Global Competitiveness Report," Geneva, 2016. [Online]. Available: http://www3.weforum.org/docs/GCR2016-2017/05FullReport/TheGlobalCompetitivenessReport2016-2017_FINAL.pdf
- [2] P. Dallasega, E. Rauch, and C. Linder, "Industry 4.0 as an enabler of proximity for construction supply chains: A systematic literature review," *Computers in Industry*, vol. 99, pp. 205-225, 2018, doi: 10.1016/j.compind.2018.03.039.
- [3] K. Schwab. "The Fourth Industrial Revolution: what it means and how to respond." <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/> (accessed).
- [4] K. Schwab, *The Fourth Industrial Revolution*. World Economic Forum, 2016, pp. 175-175.
- [5] A. C. Pereira and F. Romero, "A review of the meanings and the implications of the Industry 4.0 concept," *Procedia Manufacturing*, vol. 13, pp. 1206-1214, 2017, doi: 10.1016/j.promfg.2017.09.032.
- [6] M. Piccarozzi, B. Aquilani, and C. Gatti, "Industry 4.0 in Management Studies: A Systematic Literature Review," *Sustainability*, vol. 10, no. 10, 2018, doi: 10.3390/su10103821.
- [7] L. Koh, G. Orzes, and F. Jia, "The fourth industrial revolution (Industry 4.0): technologies disruption on operations and supply chain management," *International Journal of Operations & Production Management*, vol. 39, no. 6/7/8, pp. 817-828, 2019, doi: 10.1108/IJOPM-08-2019-788.
- [8] J. Benitez, J. Llorens, and J. Braojos, "How information technology influences opportunity exploration and exploitation firm's capabilities," *Information & Management*, vol. 55, no. 4, pp. 508-523, 2018, doi: 10.1016/j.im.2018.03.001.
- [9] S. S. Kamble, A. Gunasekaran, and S. A. Gawankar, "Sustainable Industry 4.0 framework: A systematic literature review identifying the current trends and future perspectives," *Process Safety and Environmental Protection*, vol. 117, pp. 408-425, 2018, doi: 10.1016/j.psep.2018.05.009.
- [10] P. Deputi Bidang Protokol-Pers-dan Media Sekretariat. "Presiden Jokowi Bahas Implementasi Peta Jalan Industri 4.0." <http://ksp.go.id/presiden-jokowi-bahas-implementasi-peta-jalan-industri-4-0/index.html> (accessed).

- [11] R. Aisyah. "Indonesia rolls out index to assess progress on Industry 4.0." <https://www.thejakartapost.com/news/2019/01/18/indonesia-rolls-out-index-to-assess-progress-on-industry-4-0.html> (accessed).
- [12] M. A. Kamarul Bahrin, M. F. Othman, N. H. Nor Azli, and M. F. Talib, "Industry 4.0: A Review on Industrial Automation and Robotic," *Jurnal Teknologi*, vol. 78, no. 6-13, pp. 2180-3722, 2016, doi: 10.11113/jt.v78.9285.
- [13] R. Kitchin, "Big Data, new epistemologies and paradigm shifts," *Big Data & Society*, vol. 1, no. 1, 2014, doi: 10.1177/2053951714528481.
- [14] S. Fosso Wamba, S. Akter, A. Edwards, G. Chopin, and D. Gnanzou, "How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study," *International Journal of Production Economics*, vol. 165, pp. 234-246, 2015, doi: 10.1016/j.ijpe.2014.12.031.
- [15] W. I. Yudhistyra, E. M. Risal, I. s. Raungratanaamporn, and V. Ratanavaraha, "Exploring Big Data Research: A Review of Published Articles from 2010 to 2018 Related to Logistics and Supply Chains," *Operations and Supply Chain Management: An International Journal*, vol. 13, no. 2, pp. 134-149, 2020, doi: 10.31387/oscm0410258.
- [16] S. Akter, S. F. Wamba, A. Gunasekaran, R. Dubey, and S. J. Childe, "How to improve firm performance using big data analytics capability and business strategy alignment?," *International Journal of Production Economics*, vol. 182, pp. 113-131, 2016, doi: 10.1016/j.ijpe.2016.08.018.
- [17] A. Gunasekaran *et al.*, "Big data and predictive analytics for supply chain and organizational performance," *Journal of Business Research*, vol. 70, pp. 308-317, 2017, doi: 10.1016/j.jbusres.2016.08.004.
- [18] B. Marr. "The 6 Top Data Jobs In 2018." <https://www.forbes.com/sites/bernardmarr/2018/05/09/the-6-top-data-jobs-in-2018/#6f7cf63e430d> (accessed).
- [19] B. Marr. "How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read." <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#6145c9b560ba> (accessed).
- [20] R. I. Kementrian Perindustrian, "Videografis Indonesia Industry 4.0 Readiness Index 4.0 (INDI 4.0)," ed, 2020.
- [21] F. A. Batarseh and E. A. Latif, "Assessing the Quality of Service Using Big Data Analytics," *Big Data Research*, vol. 4, pp. 13-24, 2016, doi: 10.1016/j.bdr.2015.10.001.
- [22] A. A. Alani, F. D. Ahmed, M. A. Majid, and M. S. Ahmad, "Big Data Analytics for Healthcare Organizations a Case Study of the Iraqi Healthcare Sector," *Advanced Science Letters*, vol. 24, no. 10, pp. 7783-7789, 2018, doi: 10.1166/asl.2018.13017.
- [23] J. Moyné and J. Iskandar, "Big data analytics for smart manufacturing: Case studies in semiconductor manufacturing," *Processes*, vol. 5, no. 3, 2017, doi: 10.3390/pr5030039.
- [24] M. Naimur Rahman, A. Esmailpour, and J. Zhao, "Machine Learning with Big Data An Efficient Electricity Generation Forecasting System," *Big Data Research*, vol. 5, pp. 9-15, 2016, doi: 10.1016/j.bdr.2016.02.002.
- [25] A. R. Honarvar and A. Sami, "Towards Sustainable Smart City by Particulate Matter Prediction Using Urban Big Data, Excluding Expensive Air Pollution Infrastructures," *Big Data Research*, vol. 17, pp. 56-65, 2019, doi: 10.1016/j.bdr.2018.05.006.
- [26] J. Manyika *et al.*, "Big Data: The Next Frontier for Innovation, Competition, and Productivity," Seoul, 2011. [Online]. Available: www.mckinsey.com/mgi.
- [27] D. A. Valarmathi, "Market Basket Analysis for Mobile Showroom," *International Journal for Research in Applied Science and Engineering Technology*, vol. V, no. X, pp. 1279-1284, 2017, doi: 10.22214/ijras.2017.10185.
- [28] M. Kaur and S. Kang, "Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining," *Procedia Computer Science*, vol. 85, no. Cms, pp. 78-85, 2016, doi: 10.1016/j.procs.2016.05.180.
- [29] B. Lantz, *Machine Learning with R*. Packt Publishing, 2015, pp. 417-417.
- [30] R. Agrawal and R. Srikant, "A fast algorithm for mining association rules in image," Santiago, 1994/06// 1994: IBM Almaden Research Center.
- [31] C. Hidber, "Online association rule mining," New York, New York, USA, 1999 1999: ACM Press, pp. 145-156, doi: 10.1145/304182.304195. [Online]. Available: <http://portal.acm.org/citation.cfm?doid=304182.304195>
- [32] Y. Huang, X. Wang, and B.-C. Shia, "Efficiency and Consistency Study on Carma," presented at the 2009 Fifth International Joint Conference on INC, IMS and IDC, 2009.
- [33] W. I. Yudhistyra, I. s. Raungratanaamporn, and V. Ratanavaraha, "Big Data Analytics Techniques Exploration for Customers Analysis to Gain Competitive Advantage," Bangkok, 2019 2019, pp. 58-63.

-
- [34] R. Wirth, "CRISP-DM : Towards a Standard Process Model for Data Mining," *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, no. 24959, pp. 29-39, 2000.
- [35] G. Mariscal, Ó. Marbán, and C. Fernández, "A survey of data mining and knowledge discovery process models and methodologies," *The Knowledge Engineering Review*, vol. 25, no. 2, pp. 137-166, 2010, doi: 10.1017/S0269888910000032.