Similarity of Competitive Merchant Behavior Using the Jaccard Coefficient

Marischa Elveny*

1,2,3,Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Padang Bulan 20155 USU, Medan, Indonesia

Article Info	ABSTRACT
Article history:	Advanced installment innovation has brought a business pattern that makes it simpler to see income conditions or what is known as a productive contrast. Different new ways are utilized to make progress, one of which is with an electronic-based business, yet with a particularly number of varieties, vulnerability in business is likewise progressively hard to foresee. Particularly foreseeing in the following not many years what exercises will regularly happen. Forecast itself is an interaction of efficiently assessing something that is well on the way to occur later on dependent on at various times data. To stay aware of the advancement of the organization, it is important to enhance the measurements for the business. Similarity involves the process of characterizing each object or describing in detail the features of the object. A way to measure similarity by characterizing each object so as to produce a similarity in behavior. used the jaccard method in finding similarities.
Keywords:	
business intelligence, prediction, similarity, jaccard Coefficient	

Corresponding Author:

Marischa Elveny,
Faculty of Computer Science and Information Technology,
Universitas Sumatera Utara, Medan, Indonesia
Email: marischaelveny@usu.ac.id

1. INTRODUCTION

Business measurements permit the economy to become quicker, and better. Since in business the information measurements acquired should be substantial for all individuals from the association, workers for hire and providersm (El Sawy, 2013). Also, the dispersed business measurements reflect piece of the pie and markets that change quickly so they should adjust to changes (J. Paulo Davim, 2020). Thusly, a system is expected to expand piece of the pie, deals, development and item improvement (ST Asah, 2014).

Enhancement can be characterized as a cycle used to accomplish most extreme outcomes and is deciphered as a type of expanding something that as of now exists, or planning and making something new or better (Melo, 2019). Expectation is an interaction of assessing methodically and primarily what may occur later on dependent on data/information in the at various times, so blunders that happen in the present can be determined so they don't happen later on . Forecast shows what will occur in a specific circumstance which turns into a contribution for the arranging and dynamic interaction (Syah, 2020).

In comparability procedures, extricating data includes the estimation of the distance between the items that can be utilized to follow the inception of the data, or it very well may be utilized to foresee social conduct (Bardicchia, 2020). Similaritas are searching for similitudes between the information source (Zhang, 2015). The comparability is utilized to search for linkages or linkages between client conduct on shipper upper hand. The similitude of the two items is estimated dependent on the idea of closeness, which is utilized to offer restricted benefit (Elveny, 2020). There are numerous similitudes estimation, yet it includes an arrangement of segments of the article requires an exceptional methodology (Dijkman, 2020). This

examination means to uncover something in like manner, where each item is gathered by distance closeness, similitude and distinction (Nasution, 2020).

2. RESEARCH METHODE

In the exploration approach, the means that will be taken are gathering information from versatile wallets dependent on dealers and clients enlisted in serious installment channels, this is done to see client conduct while executing at a shipper (R. Syah, 2020). The information is gathered dependent on exchanges that have been produced using recently enlisted clients. After that the information is ordered and gathered dependent on comparable dealers, shippers of various kinds and traders with a similar area. After that it is planned and afterward we search for anomalies that happen between exchanges, which vendor gets the greatest outcome utilizing powerful relapse.
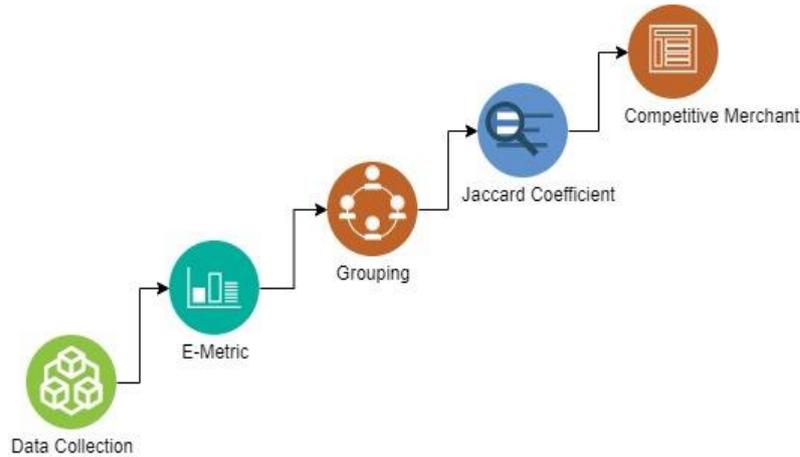

Figure 1 Research Methodology

a. Data Collection

This investigation begins by social event data and information about the amount of customers and the amount of sellers taking all things together zones in North Sumatra (Syah, 2020). The data is discovered using an adaptable wallet that has been enrolled at various vendors all through North Sumatra.
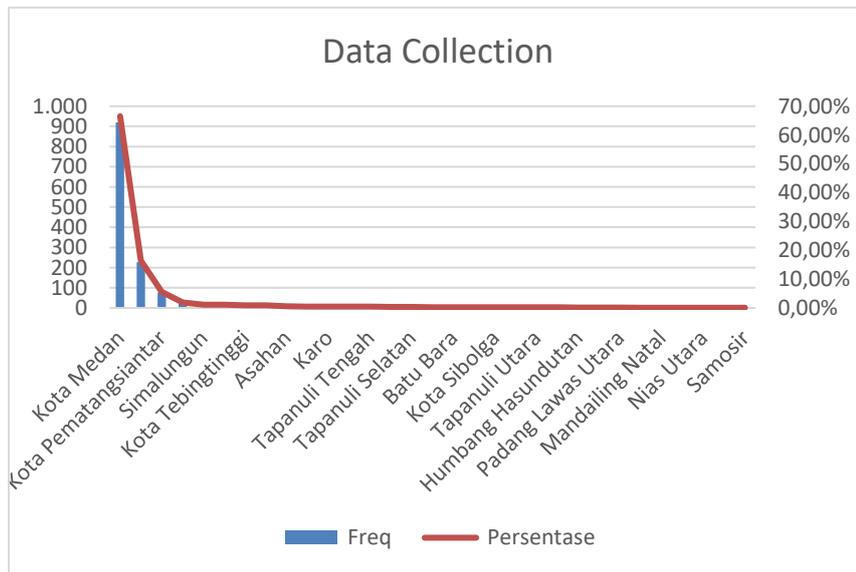

Figure 2 Data Collection

With an assortment of vendors in using computerized business improvement electronically, it can make it simpler to see income conditions or what is known as a beneficial contrast. With an enormous number and different varieties, vulnerability in business is likewise harder to anticipate. At this stage the trader variations are gathered and ordered so later there will be anomalies or contrasts that can be utilized as a bit of leeway at the vendor.

b.   E-Metric Data

Electronic-Metrics or purported e-measurements are electronic-based client conduct (e-client conduct). The mechanical period has carried enormous changes to society, this can be seen in introducing everything electronically which gives comfort about how a client carries on electronically and gives data that we can apply (M Elveny, 2020).

Table 1 E-Metric Data

No Transaction	No Reference	Customer's name	Transaction	Merchant Name
J190100028785	900400148349	"Darryl Felix Picauly	Rp 15.000	SOSIS GORENG ALKUDUS NL
J190100063799	901100217341	"Darryl Felix Picauly	Rp 30.000	MILALA BENGKEL
J190100063816	901100218155	"Darryl Felix Picauly	Rp 15.000	TOKO DEDI LK
J190100063824	901100218518	"Darryl Felix Picauly	Rp 10.000	WARUNG TIARA LK
J190100080855	901500038426	"Darryl Felix Picauly	Rp 15.000	TOKO BAYU LK
J190100081017	901500045168	"Darryl Felix Picauly	Rp 10.000	TOKO DEDI LK
J190100092427	901600115773	"Darryl Felix Picauly	Rp 10.000	TOKO DEDI LK
J190100092448	901600116257	"Darryl Felix Picauly	Rp 15.000	SATE BANG JON LK
J190300191290	921900369521	" Darryl Felix Picauly	Rp 15.000	SATE BANG JON LK
J190100040006	900700103494	"Putri Riennika Aprillia	Rp 10.000	TOKO DEDI LK
J190100056472	901000160879	"Putri Riennika Aprillia	Rp 15.000	TOKO DEDI LK
J190100072976	901400108316	"Sumiati	Rp 10.000	MILALA BENGKEL
J190100072976	901400108316	"Sumiati	Rp 15.000	BURGER BU NENG LK
J190100045712	900800159641	"Hendra Sinaga	Rp 15.000	IR.ONE S
J190100056460	901000160649	"Eli Nur Cahaya Purba	Rp 10.000	SOSIS GORENG ALKUDUS NL
J190200056756	910600375534	"Armansyah Saragih	Rp 10.000	MARIANA BR SINAGA
J190200088182	911100030125	"Armansyah Saragih	Rp 15.000	TOKO DEDI LK
J190200103267	911200106098	"Armansyah Saragih	Rp 10.000	WARUNG TIARA LK
J190200110868	911200407167	"Armansyah Saragih	Rp 15.000	TOKO BAYU LK
J190200062211	910700209730	"Alham Kinanda	Rp 18..500	WARUNG TIARA LK
J190200073364	910800213607	"Alham Kinanda	Rp 10.000	WARUNG JUSS PAK YADI LK
J190200077048	910800324374	"Setiawan	Rp 10.000	SOSIS GORENG ALKUDUS NL
J190200115407	911300143342	"Setiawan	Rp 45.000	MILALA BENGKEL
J190200091150	911100120416	"Suzana	Rp 30.000	MILALA BENGKEL
J190200116210	911300181560	"Suzana	Rp 10.000	TOKO DEDI LK
J190200097370	911100309671	"Elfira Suci Dayanti	Rp 15.000	TOKO DEDI LK
J190200102079	911200063726	"Elfira Suci Dayanti	Rp 10.000	WARUNG TIARA LK
J190200059781	910700100207	"Elfira Suci Dayanti	Rp 60.000	MILALA BENGKEL
J190100111769	901800169735	"Chairatun Aulia	Rp 15.000	BURGER BU NENG LK

3.   RESULT AND ANALYSIS

The idea of comparability prompts a type of balanced like a line or a triangle, specifically:
$$I^2 = m^2 + n^2 \qquad (1)$$

Whenever set $I = \{m_i \mid I = 1, ..., n\}$ addresses the article with which ai is a part of I, so every segment has a relationship with another segment. Assume $|I| = n$, at that point we get $(n - 1)$ connections or there is a bunch of connections $I = (m_j \mid n = 1, ..., m)$ for $m_i \in I$, $n = 1, ..., n$. Any relationship can be acknowledged in paired $\{0,1\}$ to be authorized from two parts in I. Accordingly, every segment has at any rate one connection with another (Nasution, 2017).

a.   Jaccard Coefficient

Estimating the Jaccard similitude coefficient between two informational collections is the consequence of division between the quantity of highlights that are normal to all separated by the quantity of properties as demonstrated beneath (Cheng, 2019). There are equations vectors $m, n \in E$ and $m \cap n \in E$ where Jaccard coefficient J_c as a measurement of $x (m, n)$ then

$$j_c = m \cap n / m \cup n \qquad (2)$$

Where: $m \cup n = m + n - m \cap n$

$$J(X,Y) = \frac{\sum_{i}^{p} = 1 x_i\, y_i}{\sum_{j}^{p} = 1 x_i^2 + \sum_{j}^{p} = 1 y_i^2 - \sum_{i}^{p} = 1 x_i y_i} \qquad (3)$$

b.   Performance Evaluation

We assessed the similitude execution of search words by utilizing the exactness, review, and F-measure (Niwattanaku, 2017). Illustration of accuracy test results Jaccard's similarity coefficient with transaction similarity based on merchant can be seen in table 2 and transaction similarity based on customer can be seen in table 3.

Table 2 Jaccard Similarity Coefficient with Transaction Merchant

Merchant Name	Difference in value	Value Comparison
Sosis Goreng Alkudus; Warung Tiara	17.1	35.2
Sate bng jon; Burger bu neng	22.5	27.2
Milala bengkel; Toko Dedi	12.4	19.5
Toko Bayu; ir one	5.5	7.2

Table 3 Jaccard Similarity Coefficient with Transaction Customer's

Customer's Name	Difference in value	Value Comparison
Darryl Felix Picauly; Sumiati	12.3	21.1
Putri Riennika Aprillia; Hendra Sinaga	22.8	34.1
Eli Nur Cahaya Purba; Armansyah Saragih	11.8	21.3
Alham Kinanda; Setiawan	6.7	15.6
Suzana; Elfira Suci Dayanti	25.7	15.6
Chairatun Aulia; Darryl Felix Picauly	34.9	26.4
Sumiati; Putri Riennika Aprillia	22.9	31.4
Eli Nur Cahaya Purba; Darryl Felix Picauly	28.2	32.2
Putri Riennika Aprillia; Armansyah Saragih	7.5	11.8
Sumiati; Eli Nur Cahaya Purba	11.3	23.6

c.   Competitive Merchant

The process is carried out in three ways, first looking for completeness that compares the labels and attributes attached to process elements (Syah, 2020). Second, look for structural groups that compare label elements and topology of the process model. Third, look for behaviors that compare label elements and relationships in the model process.
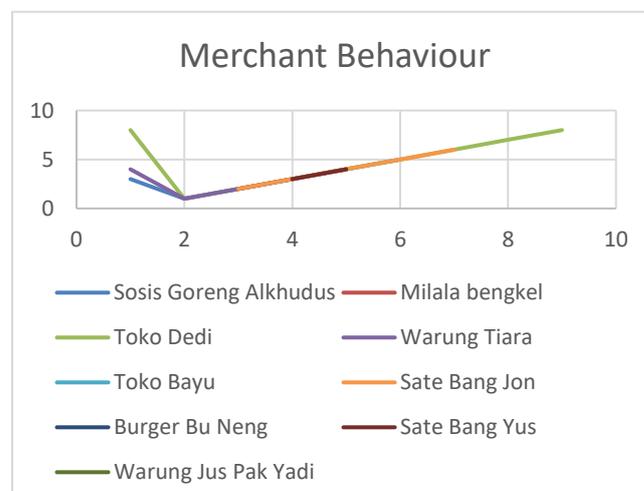

Figure 3 Competitive Merchant

4. CONCLUSIOON

This study looks for the similarity of behavior between merchants and customers. The results showed that the test method with Jaccard coefficients can work well in measuring similarity behavior by comparing each transaction made by the customer to the merchant. However, this method is not capable detect large amounts of data. In conclusion, Jaccard's similarity coefficient is quite suitable used in similarity measurements. But not in large amounts of data. Therefore, other algorithms are needed in order to get maximum results.

REFERENCES

[1] El Sawy, OA, & Pereira, F. (2013). Business modelling in the dynamic digital space: An ecosystem approach (Vol. 1, Springer Briefs in digital spaces). Heidelberg: Springer-Verlag

[2] Business Intelligence and Analytics in Small and Medium Enterprises. Manufacturing Design and Technology Series Series Editor: J. Paulo Davim, University of Aveiro, Portugal, (2020)

[3] ST Asah, AD Guerry, DJ Blahna, JJ Lawler, Perception, acquisition and use of ecosystem services: Human behaviour, and ecosystem management and policy implications. Ecosystem Services. 10, 180–186 (2014).

[4] Melo, PN, & Machado, C. (Eds.). (2019). Business Intelligence and Analytics in Small and Medium Enterprises. CRC Press

[5] Syah, R., Elveny, M., & Nasution, M. K. (2020, October). Clustering Large DataSet'to Prediction Business Metrics. In *Proceedings of the Computational Methods in Systems and Software* (pp. 1117-1127). Springer, Cham.

[6] *Bardicchia, Marco (2020).* Digital CRM: Strategies and Emerging Trends: Building Customer Relationship in the Digital Era. p.12

[7] Zhang, X., Qin, Z., Liu, X., Hou, Q., Zhang, B., & Wu, J. (2015). Hadoop-Based Similarity Computation System for Composed Documents

[8] Elveny, M., Nasution, M. K., Zarlis, M., & Zamzami, E. M. (2020, November). Similarity Approach Based to Customer Behavior for Trade Business Metrics. In *Journal of Physics: Conference Series* (Vol. 1641, No. 1, p. 012053). IOP Publishing.

[9] Dijkman, Remco M., Marlon Dumas, Boudewijn F. van Dongen, Reina Uba and Jan Mendling. "Similarity of business process models: Metrics and evaluation." Inf. Syst. 36 (2011): 498-516

[10] Nasution, M. K., Sitompul, O. S., Elveny, M., Syah, R., & Rahmat, R. F. (2020, July). A similarity for new meanings. In *2020 International Conference on Data Science, Artificial Intelligence, and Business Analytics (DATABIA)* (pp. 37-44). IEEE.

[11] *R. Syah, MKM Nasution, E. B Nababan, S. Efendi. Knowledge Acceleration Estimator (KAE) Model to Customer Behavior Using Business Metrics. Journal of Theoretical and Applied Information Technology (JATIT 2020)*

[12] M Elveny., R Syah., MKM, Nasution. 2020. Taxonomy Business Model to Customer Behavior in E-Metric Ecosystem. Journal of Advanced Research in Dynamical and Control Systems. 12 (7), pp. 99-104. DOI: 10.5373 / JARDCS / V12I7 / 20201989

[13] Syah, R., Elveny, M., Nasution, M. K., & Weber, G. M. (2020). Enhanced Knowledge Acceleration Estimator Optimally with MARS to Business Metrics in Merchant Ecosystem. In 2020 4rd International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM) (pp. 1-6). IEEE.

[14] Nasution, M. K., Sitompul, O. S., Nasution, S., & Ambarita, H. (2017). New similarity. In *IOP Conference Series: Materials Science and Engineering* (Vol. 180, No. 1, p. 012297).

[15] Cheng, J., & Zhang, L. (2019, April). Jaccard coefficient-based bi-clustering and fusion recommender system for solving data sparsity. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 369-380). Springer, Cham.

[16] Niwattanakul, S., Singthongchai, J., Naenudorn, E., & Wanapu, S. (2013, March). Using of Jaccard coefficient for keywords similarity. In *Proceedings of the international multiconference of engineers and computer scientists* (Vol. 1, No. 6, pp. 380-384).

[17] Syah, R., Elveny, M., & Nasution, M. K. (2020, June). Performance Knowledge Acceleration Optimization with MARS to Customer Behavior in Merchant Ecosystem. In *2020 3rd International Conference on Mechanical, Electronics, Computer, and Industrial Technology (MECnIT)* (pp. 178-182). IEEE.