Matrik: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer
Vol. 21, No. 2, Maret 2022, pp. 463∼476
ISSN: 2476-9843, accredited by Kemenristekdikti, Decree No: 200/M/KPT/2020
DOI: 10.30812/matrik.v21i2.1432

Analyzing the use of Social Media by Fashion Designers with K-Means and C45

Abulwafa Muhammad , Sarjon Defit
Universitas Putra Indonesia YPTK Padang, Indonesia

Article Info

Article history:

Received September 03, 2021
Revised February 03, 2022
Accepted February 14, 2022

Keywords:

C45
Classification
Data mining
K-means
Social media

ABSTRACT

Social media is one part of digital marketing that is used for the development of marketing business products known as social-marketing. The use of social media as social marketing is still managed conventionally and has not implemented business social media. This study was conducted to analyze the clusters and classifications of the use of social media by fashion designers in West Sumatra in marketing their products. This analysis uses the k-Means algorithm and c45 uses the Rapidminer application for the fashion designer industry in West Sumatra. Data is collected from Instagram and Facebook of fashion designers. The data analyzed by K-Means resulted in 3 clusters of social media use, namely 3 less active clusters, 12 active clusters and 1 very active, then classification using the C45 method resulted in a decision tree that described the most and the least in using social media. This study resulted in grouping and classifying variables from whether or not the use of social media in social marketing for the fashion designer industry players in West Sumatra was good or not. The results of this study can be used as a reference for developing integrated marketing for West Sumatra fashion designers.

Corresponding Author:

Abulwafa Muhammad,
Department of Informatics Engineering,
Universitas Putra Indonesia YPTK Padang,
Email: abulwafa@upiyptk.ac.id

1. INTRODUCTION

Along with the development of Information Technology in the activities of today's society, social media applications are applications that are used by almost all levels of society with various uses [1, 2], integrated with other supporting applications and can be used to optimize services to consumers [3, 4]. Not only workers, business owners, students, students and even housewives are not left behind using social media for various purposes [5, 6]. For some community groups, Social Media is used as a means to conduct various kinds of business starting with promoting and selling their products and services [7]. Likewise, for the fashion creative economy players in West Sumatra, they use social media as a place to introduce products through posts, then hope that the products or services posted are liked by consumers (like), and increase followers on social media [8, 9] and get other positive praise, and what is no less important is the increasing awareness of the products / services being sold so that there are more opportunities to increase sales transactions [10].

Social media-based marketing is currently very developed compared to other types of digital marketing [11, 12]. Social media has also developed from only used for personal, now it can be used for business services [13]. With social media services for business, it is open opportunities for business people for further optimize the function of social media in marketing products and services as a form of business process digitalization process making it easier to analyze consumer behavior [12]. Currently, not all business people are able to optimize the use of social media. For this reason, it is necessary to analyze the current level of social media, so that it can be seen the level of active use of social media, and related parties can choose what things can be done for optimal results [14, 15].

The grouping of superior classes at MAN-1 Mataram can be grouped using the K-Means algorithm, equipped with a web application in the form of a web cloud or web that can facilitate the administration of school students in class grouping [16]. The k-means method can also be used to form clusters of fake news spread on social media so that it will be easier to detect [16, 17]. K-means can also be combined with other methods or algorithms to produce better decision support [18]. Based on the test results in the grouping of social assistance recipients, from 260 data there are 196 data thatinc luding cluster 1 with social assistance recipient status on target and 61 data including cluster 2 with social assistance status not on target [19]. The system can automatically detect suspected lesions by combining the k-means and ELM algorithms starting with features such as area, circularity, pixel intense correlation, eccentricity and entropy of the extracted intensity distribution, The extracted features are passed on to hybrid k-means and ELM, then ends with the classification with SVM can know whether the type of cancer is normal, benign or malignant in a shorter period of time and higher accuracy [18].

One way to combine the k-means method with other methods is to use the C4.5 method. The C4.5 method has a better accuracy rate than other methods [20]. The C4.5 method can also be used to classify insurance customers by comparing them with the Naive Bayes algorithm [21]. The C4.5 method successfully diagnoses the COVID-19 surveillance category and is modeled into a decision tree with PDP, ODP, and OTG classifications, then the testing process on a confusion matrix with 3 (three) classes produce an accuracy rate of 92.86% which is classified as superior or very good [22]. Many researchers compare one method with another such as the c4.5 with the Naive Bayes Classifier, the comparison can find out which method is more accurate [23, 24]. Telecommunication customer data as many as 54.979 records were processed using Rapidmanir software to produce costumers who are still active and have stopped using several variables [25, 26]. In other studies, it is known that the C4.5 Algorithm can be used to classify prospective students by making decision trees based on existing data. From the explanation above, it means that k-means and c4.5 if used simultaneously will complement each other so as to produce a classification with a better level of accuracy [29- 30].

From the problems and descriptions above, this study aims to analyze data on the use of social media Instagram and Facebook by fashion designers in West Sumatra. This study will use the K-Means algorithm to classify the types of social media usage, and use the C4.5 method for classification of factors that play a role in the success of fashion designer's social media marketing.

The structure of this paper is organized as follows: the Sect.1 is an introduction that discusses the background and reasons for conducting the research, Sect.2. Explains the methodology of data collection, analysis and the steps involved in producing research findings, Sect.3. Analysis and results discusses data processing with algorithms k means and c45 as well as using the rapidminer application for the implementation of the analysis, Sect.4. Discussing the conclusions of the research and explaining the advantages and disadvantages of the research as well as opportunities for further research. The last part before the reference there is a acknowledgements.

2. RESEARCH METHOD

The first step of this research is to identify the conditions of using social media in the West Sumatran fashion industry. After being identified, the next step is to collect data on the use of social media by fashion designers in West Sumatra. The data that has been collected will be analyzed first to find out what types of digital marketing are most widely used. The researcher chose the 2 types of social media that were most widely used and at the same time determined the variables for the use of social media. After the data variable is determined, it is carried out to collect data or complete data based on the variable. After completing the data, data analysis was carried out using k-means clustering to find out which fashion designers were very active, active and less active. Then the data is set back to be classified using the C4.5 algorithm to classify any variables that play a role in digital marketing. Data set for processing with C4.5 are also analyzed or tested using rapidminer. So that after the analysis process is carried out, the research results are obtained in the form of clusters and classifications.

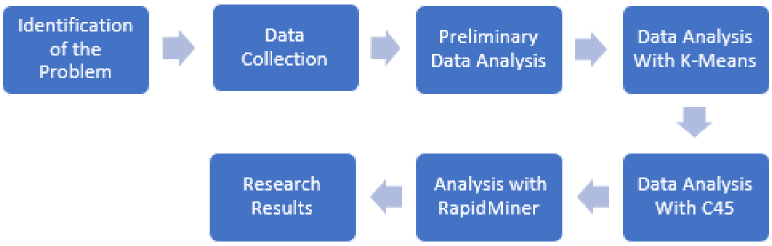

Figure 1. Research Illustration

2.1. Identification of the Problem

The initial stage of this research is to analyze the fashion creative industry in West Sumatra which is sourced from a report book issued by the West Sumatra Province Tourism and Creative Economy Office in 2019 and combined with data obtained from survey results in the field. From these sources, it is known the type and location of the fashion designer industry in West Sumatra, the use of digital marketing data for fashion designers in West Sumatra. Then the researchers looked for the types of digital marketing used by fashion designer industry players through the google search engine, from this activity it was known what types of digital marketing they had and the types of digital marketing that were most widely used by fashion designers. In Table 1 it can be seen the type of digital marketing used, based on this data the type of digital marketing used in almost all fashion designers is social media, namely Facebook and Instagram, so in this study data from social media Facebook and Instagram fashion designers were used.

Table 1. Data on the Use of Digital Marketing Fashion Designers in West Sumatra

No.	List Of Creative Industri of Fashion Designer Sumatera Barat	List Of Digital Marketing					
		Website	Sosial Media	email	Online Advertising	Search engine	Video Marketing
1	Syakira	https://syakira-fashion.business.site/	ig :syakira.fashionscrarf, facebook/mine.syakira.1, WA. 081266506130	yes	yes	yes	
2	Yadir Fashion Designer		ig :yadirsyah.zunur				
3	Ria Miranda Padang	riamiranda.com	ig :riamiranda_pdg				
4	Batik Tanah Liek Hj. Wirda Hanim	https://batiktanahliek.co.id/					
5	Novia Hertini Fashion Designer		https://www.facebook.com/public/Novia-Hertini, https://www.picburn.com/noviahertinifashion/, https://picpanzee.com/noviahertinifashion				
6	De Irma Fashion Designenr	https://de-irma.business.site/	deirmadek				
7	Rumah Tenun Nelvi Pandai Sikek Silungkang	https://tokonelvipandaisikek.business.site/	https://id-id.facebook.com/nelvirumahtenun				
8	Heni Adli Minangkabau Craft Village	https://www.henniadli.com/	https://www.facebook.com/pages/Henni-Adli-Minang-Kabau-Craft-Village/464563290247816, https://www.instagram.com/explore/locations/714788616/minangkabau-craft-village-henni-adli				

No.	List Of Creative Industri of Fashion Designer Sumatera Barat	List Of Digital Marketing					
		Website	Sosial Media	email	Online Advertising	Search engine	Video Marketing
9	Doni Rahman Fashion Desigener	https://doni-rahman-fashion-desainer.business.site/	https://www.instagram.com/doni.rahman0706/, https://ms-my.facebook.com/watch/Doni-rahman-SPd-167338353927112/				https://www.youtube.com/watch?v=9fuiXxShujU
10	Indra Collection	https://website–10548695072608184114-store.business.site/					
11	Sakura Ranti Manda Art Gallery						
12	House of Katy						
13	Galery of Riko Keket	https://galleryofrikokeket2.business.site/	https://www.instagram.com/rikokeket/, https://hi-in.facebook.com/pages/Galeri-Riko-Keket/340197779686922?rf=364735383921732				
14	Shio Collection and Zirc Management						
15	Rumah Jahit Neri		https://www.instagram.com/neriliawati/?hl=en				
16	Designer Yori		https://www.instagram.com/mryogiepratama/?hl=en				

Source : processed from various sources

2.2. Data Collection

The data sources in this study are social media data, namely Facebook and Instagram belonging to fashion designers in West Sumatra. The amount of data used is 16 records with variables (posts, likes, followers, comments, types of social media). Researchers visited Instagram and Facebook fashion designers to see and count the number of posts, likes, followers, comments and the number of social media used, and can be seen in Table 2.

Table 2. The social media active level

NO.	Kab/Kota	List of Fashion Designer Creatif Industri Sumatera Barat	Variables				
			Posting	Like	Followers	Comment	Kind of social media
1	Padang	Syakira	163	562	253	22	2
2	Padang	Yadir Fashion Designer	75	675	2218	132	1
3	Padang	Ria Miranda Padang	2081	462	47000	100	2
4	Padang	Batik Tanah Liek Hj. Wirda Hanim	28	101	763	3	2
5	Padang	Novia Hertini Fashion Designer	472	987	1738	324	1
6	Padang	De Irma Fashion Designenr	245	897	3436	234	1
7	Padang	Rumah Tenun Nelvi Pandai Sikek Silungkang	9	102	1044	2	1
8	Padang	Heni Adli Minangkabau Craft Village	495	986	1222	223	1
9	Padang	Doni Rahman Fashion Desigener	1654	2200	20400	46	1
10	PasBar	Indra Collection	120	764	239	456	1
11	Solok Selatan	Sakura Ranti Manda Art Gallery	78	673	176	89	1
12	Solok	House of Katy	89	654	324	134	1
13	Solok	Galery of Riko Keket	4323	18765	22600	890	2
14	Padang Panjang	Shio Collection and Zirc Management	234	543	324	213	1
15	Sijunjung	Rumah Jahit Neri	250	346	872	156	1
16	Dharmasraya	Designer Yori	400	1324	13800	154	1

Source : processed from various sources

2.3. Preminary Data Analysis / Data Cleaning

This stage removes inconsistent and incomplete data and transforms the data into data that can be processed using the K-Means and C4.5 algorithms. The conversion variables used are shown in Table 3.

Table 3. Variables used

No	Variable	Convertion	Information
		0	Aktif
1	Posting	1	Sangat Aktif
		2	Kurang Aktif
		0	Aktif
2	Like	1	Sangat Aktif
		2	Kurang Aktif
		0	Aktif
3	Followers	1	Sangat Aktif
		2	Kurang Aktif
		0	Aktif
4	Comment	1	Sangat Aktif
		2	Kurang Aktif
		0	Aktif
5	Media Sosial	1	Sangat Aktif
		2	Kurang Aktif

Table 3 shows the variables that will be used to calculate the level of active use of social media Instagram and Facebook using the K-Means algorithm.

The basic methodology in analyzing data using the K-Means algorithm and clustering methods are as follows:

1. Determine the value of k as the number of clusters to be formed
2. Generate a random initial k centroid (cluster center point).
3. Calculate the distance of each data to each cetroid using the correlation formula between two objects, namely Euclidean Distance.
4. Group each data based on the shortest distance between the data and the centoid.
5. Determine the position of the new centroid (Ck) by calculating the average value of the data on the same centroid using the formula (1).

$$C_k = (\frac{1}{n_k}) \sum d_t \quad (1)$$

   Where n_k is the number of documents in cluster k and d, is the document in cluster k.
6. Return to step 3 if the position of the new centroid with the old centroid is not the same.
   (a) Determine the initial center of the cluster "Centroid", For the determination of the initial center is taken randomly from the value of:
      i. 1st cluster center: 9; 102; 1044; 2; 1
      ii. Center of the 2nd cluster: 495; 986; 1222; 223; 1
      iii. 3rd cluster center: 1654; 2200; 20400; 467; 1
   (b) Calculation of the cluster center distance Euclidian Distance is used to measure the distance between the data and the cluster center, then the distance matrix will be obtained as follows formula (2):

$$D_e = \sqrt{(M_{ix} - C_{ix})^2 + (M_{iy} - C_i y)^2} \quad (2)$$

   Note:
   D_e = Euclidian Distance
   $M_i x$ = Matrix to x
   $M_i y$ =Matrix to y
   $C_i x$ = Centroid to x
   $C_i y$ = Centorid to y

Number of clusters (k=3)
First iteration

1st cluster center: 9; 102; 1044; 2; 1
2nd cluster center: 495; 986; 1222; 223; 1
3rd cluster center: 1654; 2200; 20400; 467; 1

From the first iteration manual calculation, the classification of the cluster calculation results for each data object is shown in Table 4.

Table 4. Variables used

	Group	Distance C1	Distance C2	Distance C3
F1	0	938.11	1602.81	20368.40
F2	0	1324.48	722.62	18414.23
F3	1	46014.19	45293.18	26560.19
F4	0	281.54	1326.81	19825.73
F5	0	1258.23	0	18740.75
F6	0	2542.27	1617.82	17063.24
F7	0	0	1268.13	19438.78
F8	0	1057.93	526.29	19452.11
F9	2	19638.78	18640.76	0
F10	0	1132.23	1661.41	20374.36
F11	0	1044.88	1657.97	20242.74
F12	0	920.28	1514.29	20296.33
F13	1	28850.78	27.684.478	16933.28
F14	0	898.90	1505.16	20294.93
F15	0	413.42	1112.81	19666.31
F16	2	12821.26	12068.10	6775.81

Obtained cluster membership:
C1 Member = F1, F4, F7, F10, F11, F12, F14, F15
C2 Member = F2, F5, F6, F8
C3 Member = F3, F9, F13, F16
Number of clusters (k=3)
So that the centroid value for the 2nd iteration is obtained
1st cluster center : 121,375; 468,125; 499,375; 134,375; 1.25
2nd cluster center : 321.75; 886.25; 2153.5; 228.5; 1
3rd cluster center : 2114.5; 5687.75; 25950; 297.5; 1.5

From the 2nd iteration manual calculation, the classification of the cluster calculation results for each data object is shown in Table 5.

Table 5. Calculation Results of 2 Clusters of Each Data Object

	Group	Distance C1	Distance C2	Distance C3
F1	0	279.61	1945.45	26377.23
F2	0	1741.65	344.87	14341.77
F3	1	46641.91	44683.18	22689.86
F4	0	478.85	1659.24	24885.05
F5	0	1300.82	464.17	25718.72
F6	0	2872.02	1274.85	24093.94
F7	0	668.83	1422.45	26613.08
F8	0	958.41	954.72	26223.16
F9	2	21034.74	18443.08	6475.88
F10	0	518.70	1932.37	25254.56
F11	0	377.92	2018.67	25337.01
F12	0	267.59	1851.20	25194.64
F13	1	29207.54	27461.91	14691.81
F14	0	234.00	1853.55	25205.02
F15	0	414.25	1374.44	26708.68
F16	2	13431.06	11655.22	14024.00

Obtained cluster membership:
C1 Members = F1, F4, F7, F10, F11, F12, F14, F15
C2 Members = F2, F5, F6, F8, F16
C3 Members = F3, F9, F13
Number of clusters (k=3)
Centroid value for 3rd iteration
1st cluster center: 121,375; 468,125; 499,375; 134,375; 1.25
2nd cluster center: 337.4; 973.8; 4482.8; 213.4; 1
3rd cluster center: 2686; 7142.33; 30000; 345.33; 1.67

From the 3rd iteration manual calculation, the classification of the cluster calculation results for each data object is shown in Table 6.

Table 6. Calculation Results of 3 Clusters of Each Data Object

	Group	Distance C1	Distance C2	Distance C3
M1	0	279.61	4357.67	31572.12
M2	0	1631.65	2320.88	26644.87
M3	1	45541.91	42456.16	17277.11
M4	0	489.85	3849.1	31192.12
M5	0	1500.82	2650.35	29109.15
M6	0	2872.02	1153.87	27497.46
M7	0	688.83	3669.02	29821.57
M8	0	978.41	3364.64	29410.83
M9	2	21034.74	17019.43	11850.86
M10	0	518.70	4361.45	31544.99
M11	0	397.92	4426.86	32629.89
M12	0	267.59	4279.22	31488.56
M13	1	29107.54	26711.91	13786.06
M14	0	245	4282.33	30499.93
M15	0	423.25	3766.46	30110
M16	2	13431.06	9424.17	17465.35

Obtained cluster membership:
Members C1 = F1, F2, F4, F5, F7, F8, F10, F11, F12, F14, F15
Member C2 = F6, F16
Member C3 = F3, F9, F13
Number of clusters (k=3)
Centroid value for 4th iteration
1st cluster center: 183; 581.18; 833.90; 159.45; 1.18
2nd cluster center: 322.5; 1110.5; 8618; 194; 1
3rd cluster center: 2686; 7142.33; 30000; 345.33; 1.67
From the 4th iteration manual calculation, the classification of the cluster calculation results for each data object is shown in Table 7.

Table 7. Calculation Results of 4 Clusters of Each Data Object

	Grup	Distance C1	Distance C2	Distance C3
F1	0	587.59	8486.24	30572.12
F2	0	1491.73	64 19.87	28644.87
F3	1	45205.28	38427.84	18.277.118
F4	0	543.01	7927.37	30192.12
F5	0	1035.30	6883.95	29009.15
F6	0	2622.97	5187.12	27397.46
F7	0	563.42	7649.68	29921.57
F8	0	654.88	7399.11	29510.83
F9	2	19788.30	11907.86	10850.86
F10	0	693.27	8392.69	30544.99
F11	0	675.21	8457.51	30629.89
F12	0	525.20	8310.05	30488.56
F13	1	28572.06	22883.75	13886.06
F14	0	516.65	8313.88	30499.93
F15	0	247.51	7784.06	30010
F16	2	12989.16	5187.12	17365.35

Obtained cluster membership:
Anggota C1 = F1, F2, F4, F5, F6, F7, F8, F10, F11, F12, F14, F15
Anggota C2 = F16
Anggota C3 = F3, F9, F13
Number of clusters (k=3)
Centroid value for 5th iteration
1st cluster center : 188.67; 607.5; 1050.75; 165.67; 1,167
2nd cluster center : 400; 1324; 13800; 154; 1
3rd cluster center : 2686; 7142.33; 30000; 345.33; 1.67

From the 5th iteration manual calculation, the classification of the cluster calculation results for each data object is shown in Table 8.

Table 8. Calculation Results of 5 Clusters of Each Data Object

	Group	Distance C1	Distance C2	Distance C3
F1	0	666.72	13471.12	30472.12
F2	0	2157.29	11304.74	28744.87
F3	1	46861.31	18377.11	33353.74
F4	0	1262.79	13200.39	30292.12
F5	0	1612.10	12168.11	29109.15
F6	0	3302.17	10474.25	27497.46
F7	0	1424.71	12921.26	29821.57
F8	0	1087.15	12483.08	29510.83
F9	2	20293.99	11850.86	6.765.806
F10	0	729.47	16578.80	30644.99
F11	0	656.42	13543.49	30729.89
F12	0	684.81	13396.24	30588.56
F13	1	28839.01	13786.06	19838.90
F14	0	683.56	13499.76	30599.93
F15	0	1071.14	12865.80	30110.00
F16	2	13639.50	0	17465.35

Obtained cluster membership:
Anggota C1 = F1, F2, F4, F5, F6, F7, F8, F10, F11, F12, F14, F15
Anggota C2 = F9, F16
Anggota C3 = F3, F13
Number of clusters (k=3)
Centroid value for the 6th iteration

1st cluster center : 188.67; 607.5; 1050.75; 165.67; 1,167
2nd cluster center : 1027; 1762; 17100; 100; 1
3rd cluster center : 3202; 963.12; 34800; 495.2; 2

From the 6th iteration manual calculation, the classification of the cluster calculation results for each data object is shown in Table 9.

Table 9. Calculation Results of 6 Clusters of Each Data Object

	Group	Distance C1	Distance C2	Distance C3
F1	0	822.24	16911.93	35845.23
F2	0	1175.14	14952.00	33932.17
F3	1	45888.49	15297.24	29946.81
F4	0	625.66	16451.76	35486.92
F5	0	839.68	15393.27	34278.29
F6	0	2414.39	13714.32	32587.75
F7	0	561.77	16173.95	35318.96
F8	0	518.63	15906.32	34375.23
F9	2	19470.29	16276.29	3377.91
F10	0	878.84	16918.59	35818.88
F11	0	887.40	16985.53	35888.15
F12	0	735.63	16838.73	35768.76
F13	1	28490.12	15297.14	18289.01
F14	0	732.56	16849.29	35673.68
F15	0	322.88	16308.26	35296.23
F16	2	12771.13	22752.66	3387.91

C1 Member = F1, F2, F4, F5, F6, F7, F8, F10, F11, F12, F14, F15

C2 Member = F9, F16

C3 Member = F3, F13

Because the cluster members generated from the 6th iteration are the same as the cluster members generated in the 5th iteration calculation, the calculation process is stopped. In this study, data processing using the C4.5 algorithm was not done manually, but directly processed using rapidminer software.

3. RESULT AND ANALYSIS

Testing data on the use of Facebook and Instagram social media by Fashion Designer as many as 16 records using the Rapid-Miner tool.

3.1. Analysis with the K-Means Algoritm

From this data on the use of Facebook and Instagram social media by Fashion Designer, a dataset table is created which will be processed using the K-Means algorithm. Researchers used Rapid Maner software for data testing, can be seen in Figure 2. The test results are, to determine the use of social media, a data set is formed, a data set is made based on the number of posts, likes, followers, social media, and comments. After processing the data using rapid maner, 3 clusters are generated, where the results are not much different from manual calculations. Cluster 0 represents active social media users, then cluster 1 represents very active social media users, and then cluster 2 represents less active users, can be see in Table 3 and the graph of k-means cluster formed by rapidminer can be seen in Figure 3.

Data on the use of social media for fashion designers in West Sumatra is processed using the Rapidminer application. The data to be processed is prepared using Microsoft Excel. Data processing using Rapidminer can be seen in Table 6.

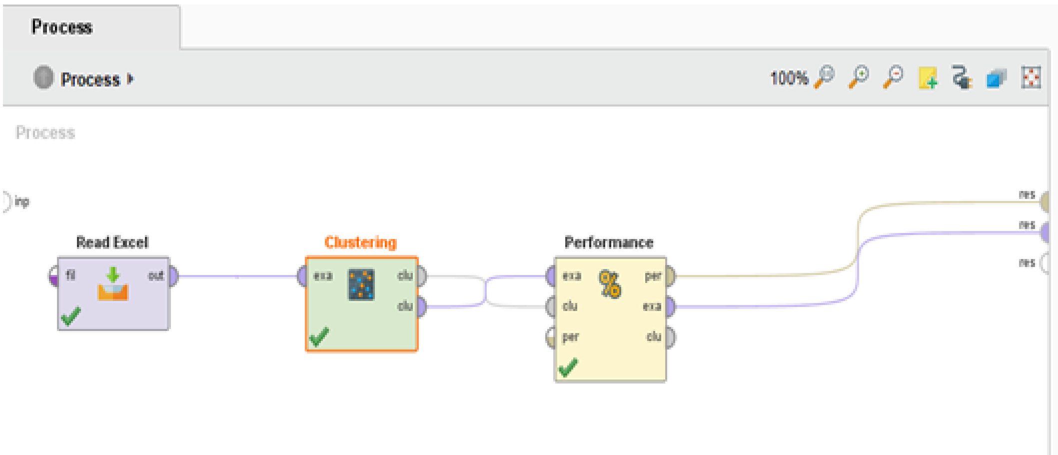

Figure 2. The Process of Calculating K-Means with Rapidminer

Data processing of fashion designer social media users in West Sumatra is processed using the rapidminer application, resulting in clusters of social media usage. There are 3 clusters formed, namely cluster-0, cluser-1 and cluster-2. The results of fashion designer data processing based on clusters can be seen in Table 1. The three clusters formed have similarities with the results of manual counting, where cluster 0 represents active cluster, cluster 1 represents very active, cluster 2 represents less active.

Table 10. Results of the Rapidminer Cluster

City	List of Names	Posting	Like	Followers	Comment	Kind of Media	Cluster
Padang	Syakira	163.0	562.0	253.0	22.0	2.0	cluster_0
Padang	Yadir Fashion Designer	75.0	675.0	2218.0	132.0	1.0	cluster_0
Padang	Ria Miranda Padang	2081.0	462.0	47000.0	100.0	2.0	cluster_1
Padang	Batik Tanah Hj. Wirda Hanim	28.0	101.0	763.0	3.0	2.0	cluster_0
Padang	Novia Hertini Fashion Designer	472.0	987.0	1738.0	324.0	1.0	cluster_0
Padang	De Irma Fashion Designer	245.0	897.0	3436.0	234.0	1.0	cluster_0
Padang	Rumah Tenun Nelvi Silungkang Pandai Sikek	9.0	102.0	1044.0	2.0	1.0	cluster_0
Padang	Yonk Ricardo	495.0	986.0	1222.0	223.0	1.0	cluster_0
Padang	Doni Rahman	1654.0	2200.0	20400.0	46.0	1.0	cluster_2
Pasbar	Indra Collection	120.0	764.0	239.0	456.0	1.0	cluster_0
Solsel	Sakura Ranti Manda Art Gallery	78.0	673.0	176.0	89.0	1.0	cluster_0
Solok	House Of Katy	89.0	654.0	324.0	134.0	1.0	cluster_0
Solok	Galery Of Riko Keket	4323.0	18765.0	22600.0	890.0	2.0	cluster_2
Padang Panjang	Zhio Collection And Zircmanagement	234.0	543.0	324.0	213.0	1.0	cluster_0
Sijunjung	Rumah Jahit Neri	250.0	346.0	872.0	156.0	1.0	cluster_0
Dharmasraya	Desainer Yori	400.0	1324.0	13800.0	154.0	1.0	cluster_2

Clusters that are formed in the data processing process using the Rapidminer application are also displayed in graphical form. The graph depicting the cluster can be seen in Figure 3. In the graph, you can see which fashion designers are included in cluster-0, to cluster-1, and to cluster-2.

Figure 3. Graph of K-Means Cluster Formed by Rapidminer

3.2. Analysis with the C45 Algoritm

After the data on social media users, fashion designers are processed using mens papers which classify social media users into 3 clusters. The next stage is data processing using C45. The results of C45 processing can be seen in Table 7.

Table 11. Data Set for Rapid Maner Process

City /Regency	Fashion Creative Industry Of west Sumatera	Posting	Like	Followers	Comment	Kind of Social Media
Padang	Syakira	Many	Many	Many	Few	Many
Padang	Yadir Fashion Designer	Few	Many	So many	Many	Few
Padang	Ria Miranda Padang	So Many	Many	So Many	Many	Many
Padang	Batik Tanah Hj. Wirda Hanim	Few	Many	Many	Few	Many
Padang	Novia Hertini Fashion Designer	Many	Many	So Many	Many	Few
Padang	De Irma Fashion Designer	Many	Many	So Many	Many	Few
Padang	Rumah Tenun Nelvi Silungkang Pandai Sikek	Few	Many	So Many	Few	Few
Padang	Yonk Ricardo	Many	Many	So Many	Many	Few
Padang	Donny Rahman Fahshion Designer	So Many	So many	So Many	Few	Few
Pasbar	Indra Collection	Many	Many	Many	Many	Few
Solsel	Sakura Ranti Manda Art Gallery	Few	Many	Many	Few	Few
Solok	House Of Katy	Few	Many	Many	Many	Few
Solok	Galery Of Riko Keket	So many	So Many	So many	Many	Many
Padang Panjang	Zhio Collection And Zircmanagement	Many	Many	Many	Many	Few
Sijunjung	Rumah Jahit Neri	Many	Many	Many	Many	Few
Dharmasraya	Desainer Yori	Many	Many	So Many	Many	Few

Table 11 is a data set that will be processed using rapidminer software which shows a list of the creative industry of fashion designers in West Sumatra with the variable number of posts, likes, followers, comments and the number of social media used. These variables have been cleaned with 3 classifications, namely many, very many and few.

A Decision Tree That Formed

From the data set process using rapidminer software, a decision tree format is generated as follows, where if a fashion designer social media user has many followers and has many posts, it can be seen that the owner of the fashion designer is syakira. Then for many followers and a few posts, namely Batik Tanah Hj. Wirda Hanim, as can be seen in the section below

.Followers = Many

| Posting = Many : Syakira

Syakira=1, Yadir Fashion Designer=0, Ria Miranda Padang=0, Batik Tanah Hj. Wirda Hanim=0, Novia Hertini Fashion Designer=0, De Irma Fashion Designer=0, Rumah Tenun Nelvi Silungkang Pandai Sikek=0, Yonk Ricardo=0, Donny Rahman Fahshion Designer=0, Indra Collection=1, Sakura Ranti Manda Art Gallery=0, House Of Katy=0, Galery Of Riko Keket=0, Zhio Collection And

Zircmanagement=1, Rumah Jahit Neri=1, Desainer Yori=0
| Posting = Few : Batik Tanah Hj. Wirda Hanim
Syakira=0, Yadir Fashion Designer=0, Ria Miranda Padang=0, Batik Tanah Hj. Wirda Hanim=1, Novia Hertini Fashion Designer=0, De Irma Fashion Designer=0, Rumah Tenun Nelvi Silungkang Pandai Sikek=0, Yonk Ricardo=0, Donny Rahman Fahshion Designer=0, Indra Collection=0, Sakura Ranti Manda Art Gallery=1, House Of Katy=1, Galery Of Riko Keket=0, Zhio Collection And Zircmanagement=0, Rumah Jahit Neri=0, Desainer Yori=0

The social media user of the West Sumatran fashion designer industry who has a lot of followers is the Gallery of Riko Keket as seen in the following section
Followers = So Many: Galery Of Riko Keket Syakira=0, Yadir Fashion Designer=0, Ria Miranda Padang=0, Batik Tanah Hj. Wirda Hanim=0, Novia Hertini Fashion Designer=0, De Irma Fashion Designer=0, Rumah Tenun Nelvi Silungkang Pandai Sikek=0, Yonk Ricardo=0, Donny Rahman Fahshion Designer=0, Indra Collection=0, Sakura Ranti Manda Art Gallery=0, House Of Katy=0, Galery Of Riko Keket=1, Zhio Collection And Zircmanagement=0, Rumah Jahit Neri=0, Desainer Yori=1

The social media user of the West Sumatran fashion designer industry who has very many followers with many posts is Novia Hertini Fashion Designer, very many followers and very many posts is Ria Miranda Padang, then very many followers with few posts is Yadir Fashion Designer, this is described in the following section.
Followers = So Many
| Posting = Many: Novia Hertini Fashion Designer Syakira=0, Yadir Fashion Designer=0, Ria Miranda Padang=0, Batik Tanah Hj. Wirda Hanim=0, Novia Hertini Fashion Designer=1, De Irma Fashion Designer=1, Rumah Tenun Nelvi Silungkang Pandai Sikek=0, Yonk Ricardo=1, Donny Rahman Fahshion Designer=0, Indra Collection=0, Sakura Ranti Manda Art Gallery=0, House Of Katy=0, Galery Of Riko Keket=0, Zhio Collection And Zircmanagement=0, Rumah Jahit Neri=0, Desainer Yori=0
| Posting = So Many: Ria Miranda Padang Syakira=0, Yadir Fashion Designer=0, Ria Miranda Padang=1, Batik Tanah Hj. Wirda Hanim=0, Novia Hertini Fashion Designer=0, De Irma Fashion Designer=0, Rumah Tenun Nelvi Silungkang Pandai Sikek=0, Yonk Ricardo=0, Donny Rahman Fahshion Designer=1, Indra Collection=0, Sakura Ranti Manda Art Gallery=0, House Of Katy=0, Galery Of Riko Keket=0, Zhio Collection And Zircmanagement=0, Rumah Jahit Neri=0, Desainer Yori=0
| Posting = Few : Yadir Fashion Designer Syakira=0, Yadir Fashion Designer=1, Ria Miranda Padang=0, Batik Tanah Hj. Wirda Hanim=0, Novia Hertini Fashion Designer=0, De Irma Fashion Designer=0, Rumah Tenun Nelvi Silungkang Pandai Sikek=1, Yonk Ricardo=0, Donny Rahman Fahshion Designer=0, Indra Collection=0, Sakura Ranti Manda Art Gallery=0, House Of Katy=0, Galery Of Riko Keket=0, Zhio Collection And Zircmanagement=0, Rumah Jahit Neri=0, Desainer Yori=0

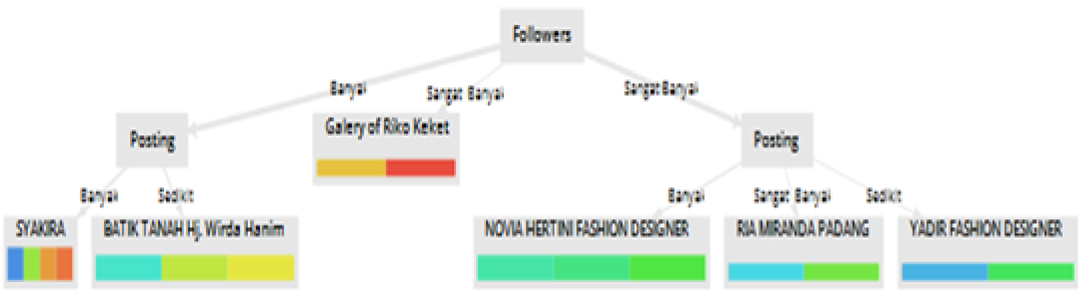

Figure 4. Decision Tree Formed by Rapidminer

4. CONCLUSION

Based on research conducted on grouping with the K-Means algorithm, it produces 3 clusters, namely Cluster 0 represents active social media users, then cluster 1 represents very active social media users, and then cluster 2 represents less active users. The dataset of the West Sumatran fashion designer industry has also been processed using rapidminer software with the C4.5 method to produce a decision tree with many, very many and few classifications. Post and followers variables are 2 variables that play a very important role in the use of social media for digital marketing. It can be seen that the social media user with a lot of followers with a lot of posts is Syakira and the few posts are batik Tanah Liek Hj Wirda Hanim. For social media users, the fashion designer with a lot of followers is Riko Keket's gallery. Then social media users with lots of followers and lots of posts are novia hertini fashion designer, the most posts are ria miranda padang, the fewest posts are yadir fashion designer.

The use of the K-Means method to determine the active, very active and less active user groups and the C4.5 method succeeded in showing the classification of social media users into three, namely many, very many and few. However, it is also necessary to try to use other methods to process data sets from the West Sumatra fashion designer industry. The data is collected from fashion designer industry players in West Sumatra Province, so the amount of data is still very limited. The hope is for further research to use larger amounts of data from other provinces in Indonesia or abroad.

ACKNOWLEDGEMENTS

Thank you to Universitas Putra Indonesia YPTK Padang for providing the funding for publication, fully supported by the Padang Computer College Foundation.

REFERENCES

[1] M. Yadav and Z. Rahman, "The Influence of Social Media Marketing Activities on Customer Loyalty: A Study of E-Commerce Industry," *Benchmarking*, vol. 25, no. 9, pp. 3882–3905, 2018.

[2] M. Hartemo, "Email Marketing in The Era of The Empowered Consumer," *Journal of Research in Interactive Marketing*, vol. 10, no. 3, pp. 212–230, 2016.

[3] K. J. Trainor, J. Andzulis, A. Rapp, and R. Agnihotri, "Social Media Technology Usage and Customer Relationship Performance: A Capabilities-Based Examination of Social CRM," *Journal of Business Research*, vol. 67, no. 6, pp. 1201–1208, 2014.

[4] T. H. Song, S. Y. Kim, and J. Y. Kim, "The Dynamic Effect of Customer Equity Across Firm Growth: The Case of Small and Medium-Sized Online Retailers," *Journal of Business Research*, vol. 69, no. 9, pp. 3755–3764, 2016.

[5] G. Piccoli, T. W. Lui, and B. Grün, "The Impact of IT-Enabled Customer Service Systems on Service Personalization, Customer Service Perceptions, and Hotel Performance," *Tourism Management*, vol. 59, pp. 349–362, 2017.

[6] K. C. Chang, "How Travel Agency Reputation Creates Recommendation Behavior," *Industrial Management and Data Systems*, vol. 115, no. 2, pp. 332–352, 2015.

[7] E. W. Ngai, K. L. K. Moon, S. S. Lam, E. S. Chin, and S. S. Tao, "Social Media Models, Technologies, and Applications: An Academic Review and Case Study," *Industrial Management and Data Systems*, vol. 115, no. 5, pp. 769–802, 2015.

[8] H. H. Chang, K. H. Wong, and P. W. Fang, "The Effects of Customer Relationship Management Relational Information Processes on Customer-Based Performance," *Decision Support Systems*, vol. 66, pp. 146–159, 2014.

[9] T. A. Monteiro, A. C. Giuliani, N. K. Pizzinatto, and E. Larios-Gomez, "Managing The Digital Consumer: Insights From Brazil and Spain," *Journal of Manufacturing Technology Management*, vol. 30, no. 8, pp. 1196–1215, 2019.

[10] I.-L. Kuo, "The Effectiveness of Environmental Interpretation at Resource-Sensitive Tourism Destinations," *International Journal of Tourism Research*, vol. 4, no. 2, pp. 87–101, 2002.

[11] P. Foroudi, S. Gupta, A. Nazarian, and M. Duda, "Digital Technology and Marketing Management Capability: Achieving Growth in SMEs," *Qualitative Market Research*, vol. 20, no. 2, pp. 230–246, 2017.

[12] G. Alfian, M. F. Ijaz, M. Syafrudin, M. A. Syaekhoni, N. L. Fitriyani, and J. Rhee, "Customer Behavior Analysis Using Real-Time Data Processing: A Case Study of Digital Signage-Based Online Stores," *Asia Pacific Journal of Marketing and Logistics*, vol. 31, no. 1, pp. 265–290, 2019.

[13] L. Choi and S. Lotz, "Motivations Leading to Customer Citizenship Behavior in Services: Scale Development and Validation," *Journal of Consumer Marketing*, vol. 33, no. 7, pp. 539–551, 2016.

[14] C. Wirawan, E. Khudzaeva, T. H. Hasibuan, Karjono, and Y. H. K. Lubis, “Application of Data Mining to Prediction of Timeliness Graduation of Students (A Case Study),” *2019 7th International Conference on Cyber and IT Service Management, CITSM 2019*, pp. 18–21, 2019.

[15] J. Y. M. Kang and J. Kim, “Online customer Relationship Marketing Tactics Through Social Media and Perceived Customer Retention Orientation of The Green Retailer,” *Journal of Fashion Marketing and Management*, vol. 21, no. 3, pp. 298–316, 2017.

[16] K. M. Yazdi, A. M. Yazdi, S. Khodayi, J. Hou, W. Zhou, and S. Saedy, “Improving Fake News Detection Using K-Means and Support Vector Machine Approaches,” vol. 14, no. 2, pp. 38–42, 2020.

[17] J. Majewska and S. Truskolaski, “Cluster-Mapping Procedure for Tourism Regions Based on Geostatistics and Fuzzy Clustering: Example of Polish Districts,” *Current Issues in Tourism*, vol. 22, no. 19, pp. 2365–2385, 2019.

[18] G. Chandrarathne, K. Thanikasalam, and A. Pinidiyaarachchi, “A Comprehensive Study on Deep Image Classification with Small Datasets,” *Lecture Notes in Electrical Engineering*, vol. 619, no. 7, pp. 93–106, 2020.

[19] Y. Alkhalifi, A. Zumarniansyah, R. Ardianto, N. Hardi, and A. E. Augustia, “Comparison of Naive Bayes Algorithm and C.45 Algorithm in Classification of Poor Communities Receiving Non Cash Food Assistance in Wanasari Village Karawang Regency,” *Jurnal Techno Nusa Mandiri*, vol. 17, no. 1, pp. 37–42, 2020.

[20] M. Irfan, N. Lukman, A. A. Alfauzi, and J. Jumadi, “Comparison of Algorithm Support Vector Machine and C4.5 for Identification of Pests and Diseases in Chili Plants,” *Journal of Physics: Conference Series*, vol. 1402, no. 6, 2019.

[21] S. Dwiasnati and Y. Devianto, “Utilization of Prediction Data for Prospective Decision Customers Insurance Using The Classification Method of C.45 and Naive Bayes Algorithms,” *Journal of Physics: Conference Series*, vol. 1179, no. 1, 2019.

[22] W. Wiguna and D. Riana, “Diagnosis of Coronavirus Disease 2019 (Covid-19) Surveillance Using C4.5 Algorithm,” *Jurnal Pilar Nusa Mandiri*, vol. 16, no. 1, pp. 71–80, 2020.

[23] Y. Doan, F. Dalkiliç, D. Birant, R. A. Kut, and R. Yilmaz, “Novel Two-Dimensional Visualization Approaches for Multivariate Centroids of Clustering Algorithms,” *Scientific Programming*, vol. 2018, 2018.

[24] M. A. Abdillah and A. Setyanto, “Implementasi Decision Tree Algoritma C4 . 5 untuk Memprediksi Kesuksesan Pendidikan Karakter,” vol. XV, pp. 59–69, 2020.

[25] I. N. Budi, I. Ranggadara, I. Prihandi, N. R. Kurnianda, and Suhendra, “Prediction Using C4.5 Method and RFM Method for Selling Furniture,” *International Journal of Engineering and Advanced Technology*, vol. 9, no. 1, pp. 535–541, 2019.

[26] S. Datta, S. Bhattacharjee, and S. Das, “Clustering with Missing Features: A Penalized Dissimilarity Measure Based Approach,” *Machine Learning*, vol. 107, no. 12, pp. 1987–2025, 2018.