

DATA MINING

SISTEM INFORMASI KELAS A

Dosen: Dr. Dian Eka Ratnawati, S.Si., M.Kom.

Project Data Mining



DISUSUN OLEH:

Nandana Rifqi Irfansyah

PROGRAM STUDI SISTEM INFORMASI
DEPARTEMEN SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BRAWIJAYA
MALANG
2023

DAFTAR ISI

Project Data Mining	1
DAFTAR ISI	2
BAB I	3
A. Latar Belakang	3
B. Tujuan Penelitian	3
BAB II	4
A. Proses Klasterisasi Data Market Campaign dengan Algoritma K-Means di Weka	4
1.1. Langkah 1	4
1.2. Langkah 2	5
1.3. Langkah 3	6
1.4. Langkah 4	7
1.5. Langkah 5	8
1.6. Langkah 6	9
1.7. Langkah 7	10
1.8. Langkah 8	11
1.9. Langkah 9	12
1.10. Langkah 10	14
B. Analisis Clustering dengan Algoritma K-Means	17
BAB III	19
Kesimpulan	19

BAB I

PENDAHULUAN

A. Latar Belakang

K-means merupakan salah satu algoritma yang bersifat unsupervised learning. K-Means memiliki fungsi untuk mengelompokkan data ke dalam data cluster. Algoritma ini dapat menerima data tanpa ada label kategori. K-Means Clustering Algoritma juga merupakan metode non-hierarchy. Metode Clustering Algoritma adalah mengelompokkan beberapa data ke dalam kelompok yang menjelaskan data dalam satu kelompok memiliki karakteristik yang sama dan memiliki karakteristik yang berbeda dengan data yang ada di kelompok lain. Cluster Sampling adalah teknik pengambilan sampel di mana unit-unit populasi dipilih secara acak dari kelompok yang sudah ada yang disebut “cluster”. K-Means Clustering adalah suatu metode penganalisaan data atau metode Data Mining yang melakukan proses pemodelan unsupervised learning dan menggunakan metode yang mengelompokkan data berbagai partisi. K Means Clustering memiliki objective yaitu meminimalisasi object function yang telah diatur pada proses klasterisasi. Dengan cara minimalisasi variasi antar 1 cluster dengan maksimalisasi variasi dengan data di cluster lainnya.

B. Tujuan Penelitian

1. Salah satu tujuan utama adalah untuk mempelajari secara mendalam tentang algoritma k-Means, termasuk prinsip kerjanya, langkah-langkah yang terlibat, dan metrik yang digunakan dalam perhitungan jarak.
2. Tujuan penelitian bisa mencakup penggunaan algoritma k-Means dalam analisis data nyata. Pada tugas ini berfokus pada perhitungan k-Means dengan menggunakan sebuah perangkat lunak bernama “Weka”.

BAB II

Isi

A. Proses Klasterisasi Data Market Campaign dengan Algoritma K-Means di Weka

1.1. Langkah 1

1.1.1. Screenshot

Marketing Campaign

Data Card Code (61) Discussion (6)

251 New Notebook

Detail Compact Column 10 of 29 columns

About this file

A response model can provide a significant boost to the efficiency of a marketing campaign by increasing responses or reducing expenses. The objective is to predict who will respond to an offer for a product or service.

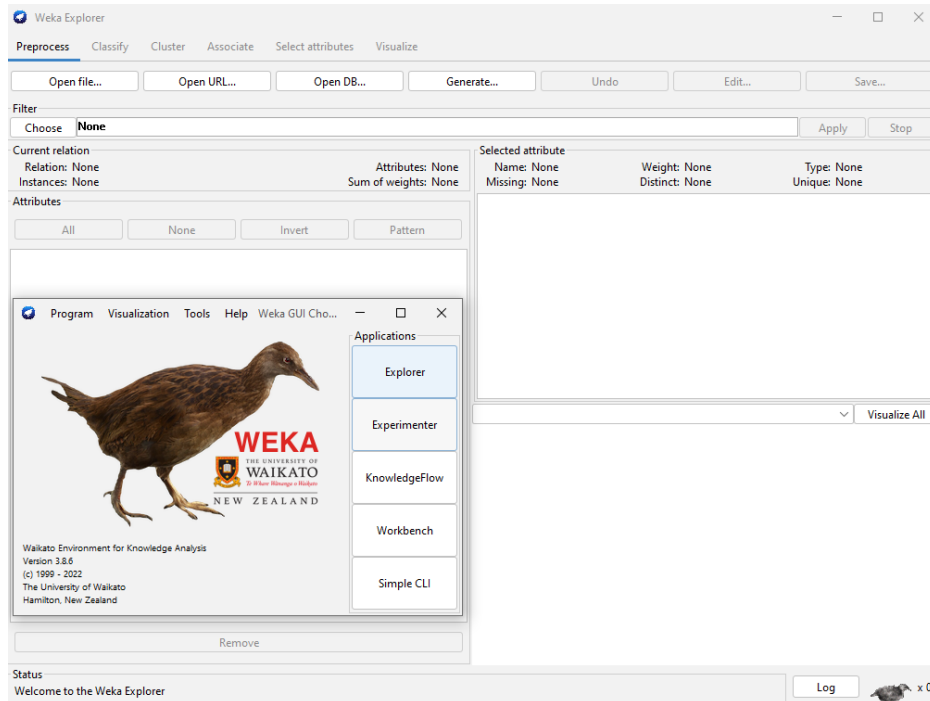
# Kidhome	# Teenhome	Dt_Customer	# Recency	# MntWines
Number of small children in customer's household	Number of teenagers in customer's household	Date of customer's enrolment with the company	Number of days since the last purchase	Amount spent on wine products in the last 2 years
2240 total values	2240 total values	2240 total values	2240 total values	2240 total values
0	0	2012-09-04	58	635
1	1	2014-03-08	38	11
0	0	2013-08-21	26	426
1	0	2014-02-10	26	11
1	0	2014-01-19	94	173
0	1	2013-09-09	16	520
0	1	2012-11-13	34	235
1	0	2013-05-08	32	76
1	0	2013-06-06	19	14
1	1	2014-03-13	68	28
1	0	2013-11-15	11	5

1.1.2. Penjelasan

Pada langkah pertama dilakukan pembukaan file yang ingin dilakukan praktik K-means pada microsoft excel. Pada langkah ini kami mengambil sebuah dataset dari kaggle yang bernama “marketing_campaign.csv”

1.2. Langkah 2

1.2.1. Screenshot

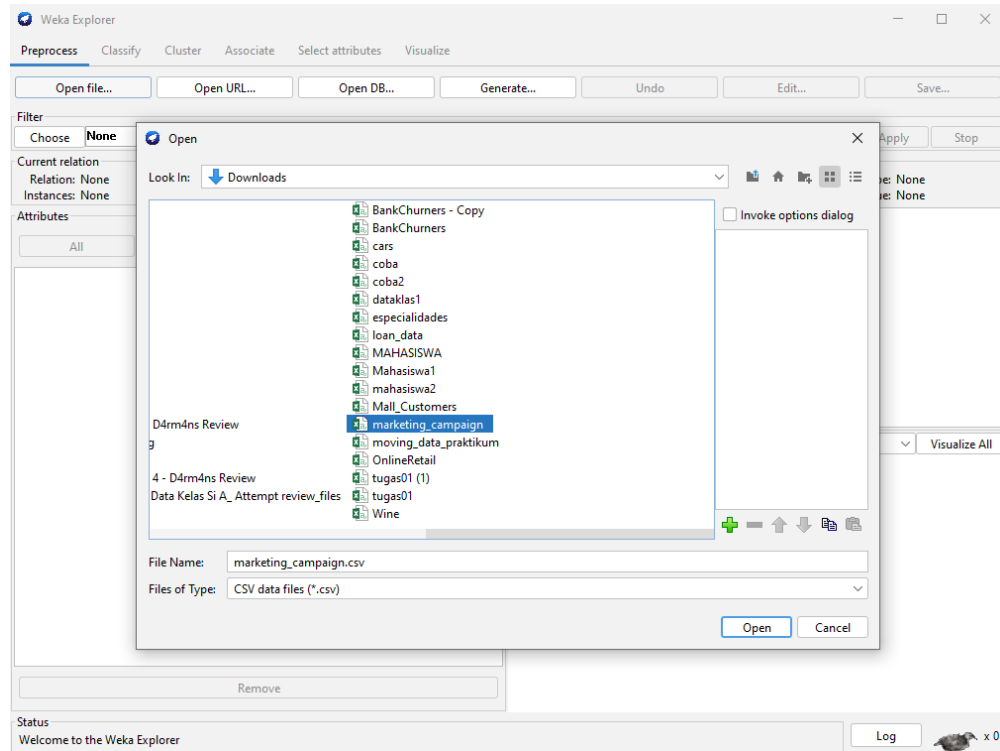


1.2.2. Penjelasan

Pada langkah ini dilakukan pembukaan aplikasi weka pada komputer

1.3. Langkah 3

1.3.1. Screenshot

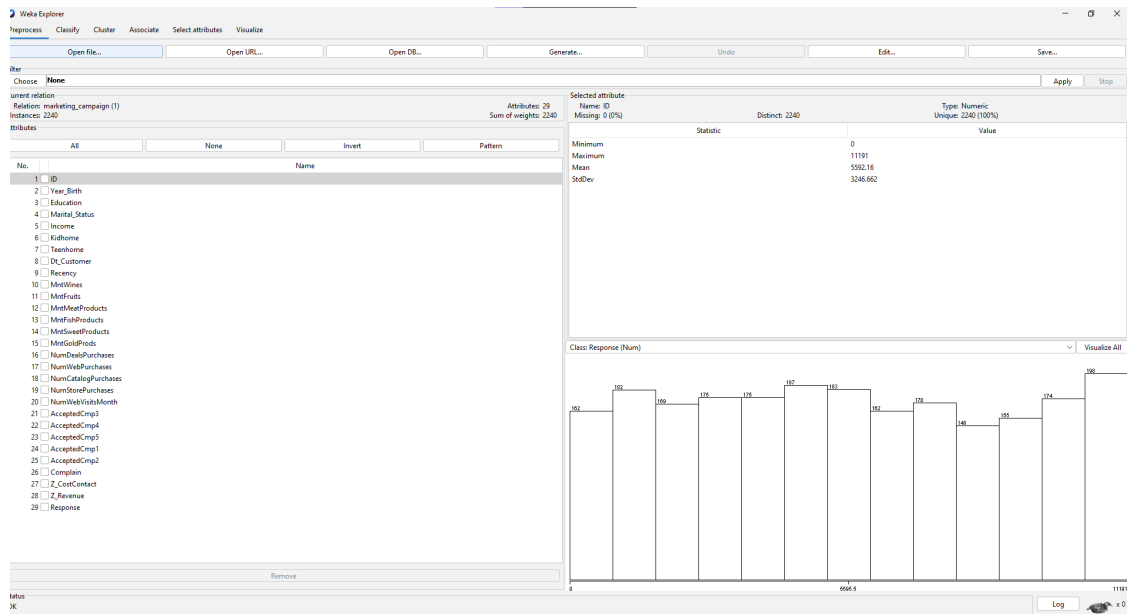


1.3.2. Penjelasan

Pada langkah selanjutnya dilakukan pembukaan file “marketing_campaign.csv” pada weka.

1.4. Langkah 4

1.4.1. Screenshot

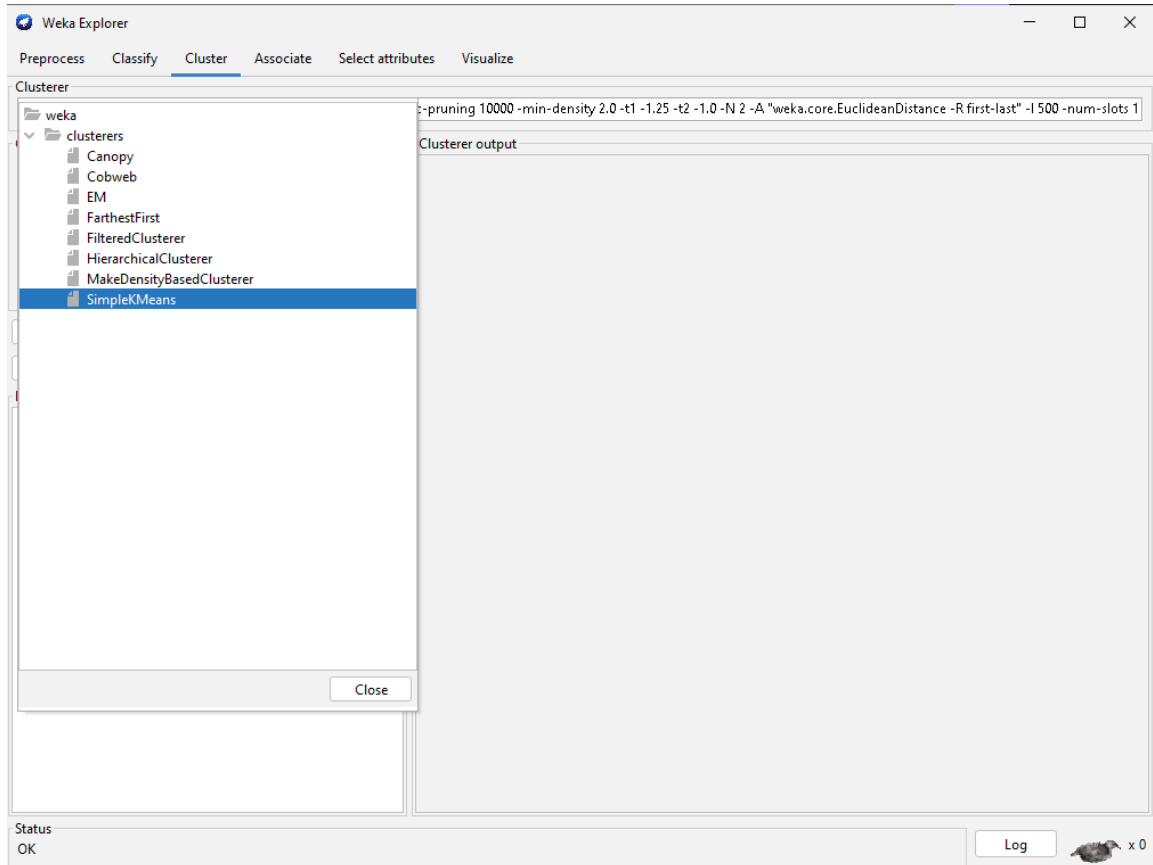


1.4.2. Penjelasan

Pada langkah ini dapat dilihat bahwa file “marketing_campaign.csv” berhasil dibuka pada weka.

1.5. Langkah 5

1.5.1. Screenshot

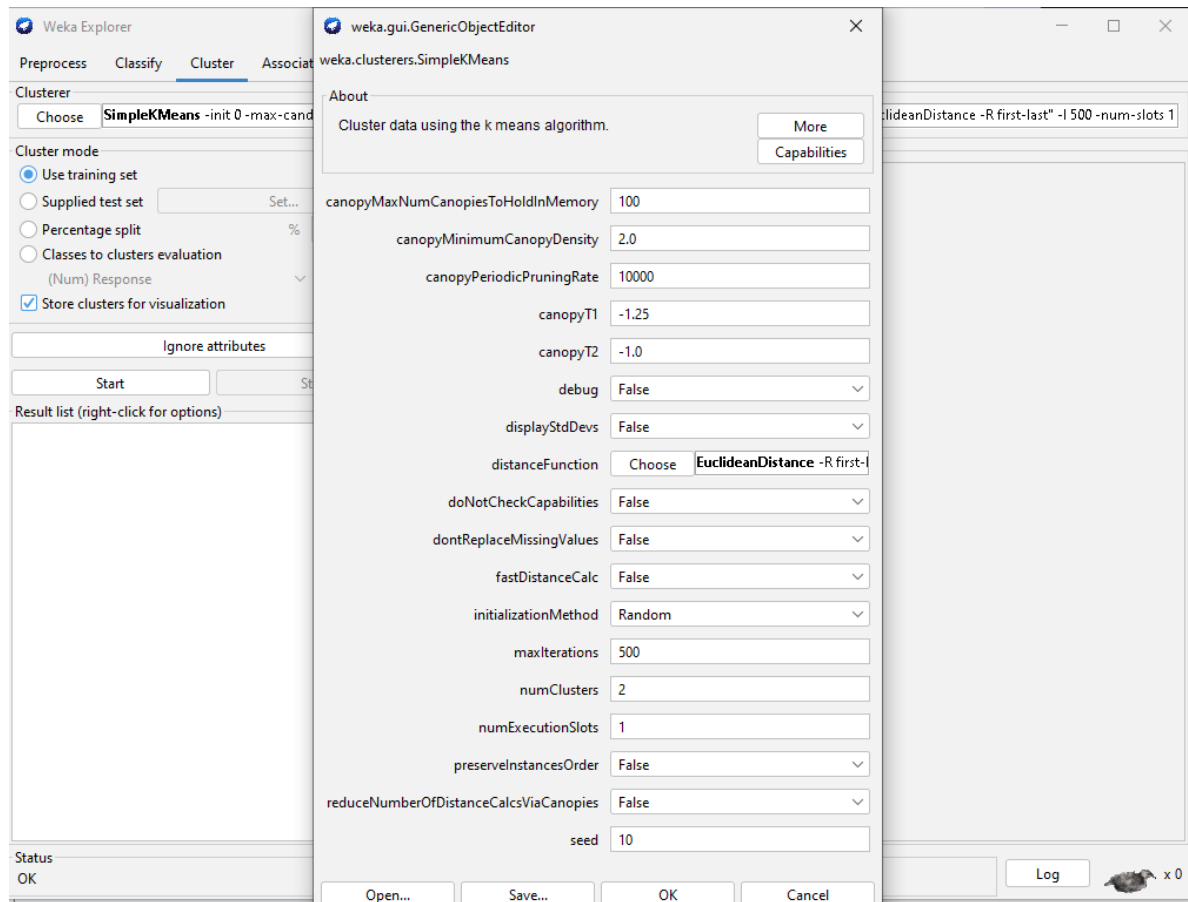


1.5.2. Penjelasan

Pada langkah selanjutnya akan dilakukan perhitungan kMeans. Pada weka untuk melakukan perhitungan kMeans dapat menuju ke bar “Cluster” kemudian pilih folder “clusterers” dan pilih file “simpleKmeans”.

1.6. Langkah 6

1.6.1. Screenshot



1.6.2. Penjelasan

Pada langkah ini dilakukan konfigurasi simpleKmeans, dimana pada konfigurasi ini kami menggunakan pengukuran dengan euclidean kemudian untuk iterasi sebanyak 500 kali dan membagi menjadi 2 cluster.

1.7. Langkah 7

1.7.1. Screenshot

Final cluster centroids:

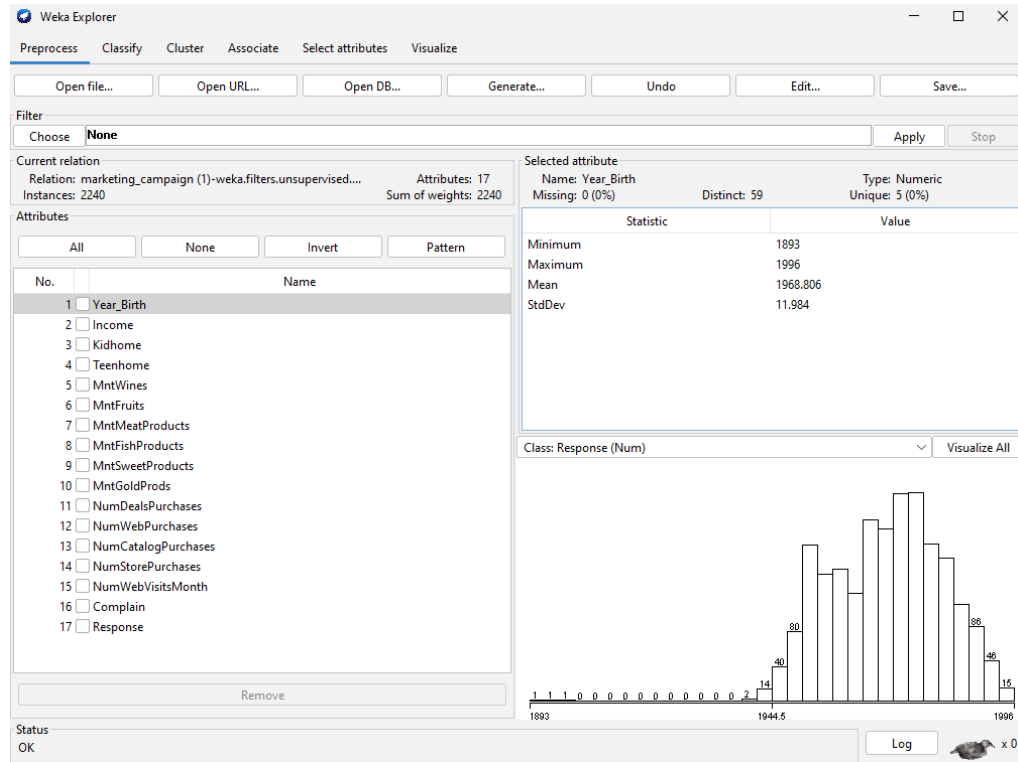
Attribute	Cluster#		
	Full Data (2240.0)	0 (911.0)	1 (1329.0)
ID	5592.1598	5535.3557	5631.0978
Year_Birth	1968.8058	1966.8782	1970.1272
Education	Graduation	Graduation	Graduation
Marital_Status	Married	Married	Married
Income	52247.2514	70339.5799	39845.3618
Kidhome	0.4442	0.0593	0.7081
Teenhome	0.5062	0.4369	0.5538
Dt_Customer	2012-08-31	2012-08-31	2014-03-23
Recency	49.1094	49.697	48.7065
MntWines	303.9357	607.0779	96.1384
MntFruits	26.3022	54.5357	6.9488
MntMeatProducts	166.95	345.2887	44.7028
MntFishProducts	37.5254	77.8068	9.9135
MntSweetProducts	27.0629	55.4533	7.602
MntGoldProds	44.0219	73.8211	23.5952
NumDealsPurchases	2.325	2.0209	2.5335
NumWebPurchases	4.0848	5.6806	2.991
NumCatalogPurchases	2.6621	5.0944	0.9947
NumStorePurchases	5.7902	8.6707	3.8157
NumWebVisitsMonth	5.3165	3.8617	6.3138
AcceptedCmp3	0.0728	0.0779	0.0692
AcceptedCmp4	0.0746	0.1625	0.0143
AcceptedCmp5	0.0728	0.1789	0
AcceptedCmp1	0.0643	0.1537	0.003
AcceptedCmp2	0.0134	0.0307	0.0015
Complain	0.0094	0.0077	0.0105
Z_CostContact	3	3	3
Z_Revenue	11	11	11

1.7.2. Penjelasan

Pada langkah ini dapat dilihat hasil dari konfigurasi pada langkah sebelumnya, bahwa masih terdapat beberapa atribut yang dianggap kurang berpengaruh yang artinya hasil dari centroidnya mirip sehingga tidak membentuk karakteristik saat diklusterisasi, dan juga ada atribut yang tidak termasuk target yang ingin kami analisis karena hasil clusterisasinya terhadap analisis terlalu kecil. Sehingga beberapa atribut tersebut kami lakukan eliminasi .

1.8. Langkah 8

1.8.1. Screenshot



1.8.2. Penjelasan

Pada langkah ini dilakukan sudah dilakukan penghapusan beberapa atribut yang kurang kontribusi, dimana pada langkah sebelumnya terdapat 29 atribut menjadi 17 atribut.

1.9. Langkah 9

1.9.1. Screenshot

2 Cluster :

Clusterer output			
Initial starting points (random):			
Cluster 0: 1961,45938,0,0,273,11,178,62,83,29,2,8,4,6,6,0,0			
Cluster 1: 1974,44989,0,1,98,0,106,49,10,106,5,5,1,5,6,0,0			
Missing values globally replaced with mean/mode			
Final cluster centroids:			
Attribute	Cluster#		
	Full Data	0	1
	(1792.0)	(795.0)	(997.0)
Year_Birth	1970	1967	1972
Income	52463	70356	38200
Kidhome	0	0	1
Teenhome	0	0	1
MntWines	182.5	524	31
MntFruits	8	35	3
MntMeatProducts	69	271	18
MntFishProducts	12	59	4
MntSweetProducts	9	38	3
MntGoldProds	25	49	12
NumDealsPurchases	2	1	2
NumWebPurchases	4	5	2
NumCatalogPurchases	2	5	1
NumStorePurchases	5	8	3
NumWebVisitsMonth	6	3	7
Complain	0	0	0
Response	0	0	0
Time taken to build model (percentage split) : 0.01 seconds			
Clustered Instances			
0	176 (39%)		
1	272 (61%)		

3 Cluster :

Clusterer output				
Cluster 0: 1961,45938,0,0,273,11,178,62,83,29,2,8,4,6,6,0,0				
Cluster 1: 1974,44989,0,1,98,0,106,49,10,106,5,5,1,5,6,0,0				
Cluster 2: 1982,70038,0,0,587,54,348,71,54,130,1,4,4,8,2,0,0				
Missing values globally replaced with mean/mode				
Final cluster centroids:				
Attribute	Cluster#			
	Full Data (1792.0)	0 (566.0)	1 (713.0)	2 (513.0)
Year_Birth	1970	1964	1974	1969
Income	52463	55733	33462	75278
Kidhome	0	0	1	0
Teenhome	0	1	0	0
MntWines	182.5	291	18	571
MntFruits	8	8	3	51
MntMeatProducts	69	80	15	405
MntFishProducts	12	11	4	84
MntSweetProducts	9	9	3	51
MntGoldProds	25	30	10	54
NumDealsPurchases	2	3	2	1
NumWebPurchases	4	5	2	5
NumCatalogPurchases	2	2	0	5
NumStorePurchases	5	6	3	9
NumWebVisitsMonth	6	6	7	3
Complain	0	0	0	0
Response	0	0	0	0
Time taken to build model (percentage split) : 0.03 seconds				
Clustered Instances				
0	151 (34%)			
1	192 (43%)			
2	105 (23%)			

5 Cluster :

Final cluster centroids:						
Attribute	Cluster#					
	Full Data (1792.0)	0 (131.0)	1 (530.0)	2 (282.0)	3 (168.0)	4 (681.0)
Year_Birth	1970	1963	1964	1971	1969	1975
Income	52463	61825	55760	75243.5	77089.5	33249
Kidhome	0	0	0	0	0	1
Teenhome	0	0	1	0	0	0
MntWines	182.5	180	296.5	715.5	509.5	16
MntFruits	8	26	7	37.5	103	2
MntMeatProducts	69	134	79.5	414	431.5	14
MntFishProducts	12	38	11	63	158	3
MntSweetProducts	9	28	8	44	92	2
MntGoldProds	25	41	31	43.5	90	9
NumDealsPurchases	2	1	3	1	1	2
NumWebPurchases	4	4	5	5	4	2
NumCatalogPurchases	2	3	2	5	6	0
NumStorePurchases	5	6	6	10	7	3
NumWebVisitsMonth	6	4	6	3	2	7
Complain	0	0	0	0	0	0
Response	0	0	0	0	0	0
Time taken to build model (percentage split) : 0.02 seconds						
Clustered Instances						
0	32 (7%)					
1	137 (31%)					
2	44 (10%)					
3	49 (11%)					
4	186 (42%)					

1.9.2. Penjelasan

Pada langkah ini dilakukan run simple k-means yang dimana untuk konfigurasinya adalah default settings, namun menggunakan metode Manhattan. Akan tetapi klasterisasi yang terlihat dari 2 hingga 5 Cluster, nampak bahwa hasil akurasi kurang akurat karena angka yang dihasilkan kebanyakan serupa (0, 0, 1, 0) di sebagian besar atribut. Menurut kami hal ini terjadi dikarenakan tidak ada pembobotan pada tiap dimensi dan Metode Manhattan menganggap perubahan dalam setiap dimensi sebagai langkah yang sama. Sehingga hasil menjadi sulit diinterpretasikan dan berdampak pada penentuan karakteristik Pelanggan yang kurang maksimal. Oleh karena itu kami menganggap Manhattan adalah metode yang kurang tepat untuk dataset dan tujuan target analisis kami, sehingga pada langkah selanjutnya kami menggunakan metode Euclidean pada Klasterisasi K-Means

1.10. Langkah 10

1.10.1. Screenshot

2 Cluster :

Final cluster centroids:			
Attribute	Full Data (1792.0)	Cluster#	
		0 (948.0)	1 (844.0)
Year_Birth	1968.8694	1972.8829	1964.3614
Income	52734.671	52973.5862	52466.316
Kidhome	0.4381	0.4515	0.423
Teenhome	0.4989	0.0137	1.0438
MntWines	310.8064	322.7046	297.4419
MntFruits	26.1987	34.3724	17.0178
MntMeatProducts	171.769	238.2321	97.1161
MntFishProducts	38.1016	51.4852	23.0687
MntSweetProducts	27.5179	35.3502	18.7204
MntGoldProds	44.0491	45.808	42.0735
NumDealsPurchases	2.3125	1.5812	3.1339
NumWebPurchases	4.101	3.7658	4.4775
NumCatalogPurchases	2.7204	3.1403	2.2488
NumStorePurchases	5.8359	5.7827	5.8957
NumWebVisitsMonth	5.2801	4.9177	5.6872
Complain	0.0089	0.0084	0.0095
Response	0.149	0.2057	0.0853

Time taken to build model (percentage split) : 0.01 seconds

Clustered Instances

0	226 (50%)
1	222 (50%)

3 Cluster :

Final cluster centroids:				
Attribute	Cluster#			
	Full Data	0	1	2
	(1792.0)	(522.0)	(696.0)	(574.0)
=====				
Year_Birth	1968.8694	1976.2548	1964.102	1967.9338
Income	52734.671	33347.8647	50273.342	73349.6498
Kidhome	0.4381	0.8448	0.4511	0.0523
Teenhome	0.4989	0.0402	1.0546	0.2422
MntWines	310.8064	46.6149	254.1078	619.8136
MntFruits	26.1987	7.6973	9.6825	63.0505
MntMeatProducts	171.769	43.1916	70.4928	411.5
MntFishProducts	38.1016	11.2414	13.6221	92.2108
MntSweetProducts	27.5179	9.2222	11.0072	64.176
MntGoldProds	44.0491	21.2241	35.8247	74.7787
NumDealsPurchases	2.3125	1.9866	3.1566	1.5854
NumWebPurchases	4.101	2.5287	4.1264	5.5
NumCatalogPurchases	2.7204	0.7375	1.819	5.6167
NumStorePurchases	5.8359	3.3391	5.3506	8.6951
NumWebVisitsMonth	5.2801	6.7337	5.7989	3.3293
Complain	0.0089	0.0096	0.0072	0.0105
Response	0.149	0.1015	0.0776	0.2787
=====				
Time taken to build model (percentage split) : 0 seconds				
Clustered Instances				
0	138 (31%)			
1	177 (40%)			
2	133 (30%)			

5 Cluster :

Final cluster centroids:						
Attribute	Cluster#					
	Full Data (1792.0)	0 (139.0)	1 (445.0)	2 (267.0)	3 (352.0)	4 (589.0)
=====						
Year_Birth	1968.8694	1968.9281	1963.4045	1969.7191	1968.2528	1972.9677
Income	52734.671	39704.3694	56250.9087	60333.5868	73377.3447	37371.9452
Kidhome	0.4381	0.0072	0.0854	0.3521	0.0795	1.0594
Teenhome	0.4989	0.0072	1.0652	0.3146	0.2244	0.4346
MntWines	310.8064	117.5036	382.6517	493.9625	589.8778	52.3379
MntFruits	26.1987	13.4532	15.7775	38.985	69.946	5.1392
MntMeatProducts	171.769	118.705	107.4719	299.8764	412.7216	30.798
MntFishProducts	38.1016	23.9209	21.4742	53.3333	104.392	7.489
MntSweetProducts	27.5179	19.2662	18.582	38.9064	70.875	5.1426
MntGoldProds	44.0491	35.1439	49.8742	62.0412	74.7699	15.2343
NumDealsPurchases	2.3125	1.4388	2.9573	2.4232	1.5028	2.4652
NumWebPurchases	4.101	3.0576	5.2225	5.0037	5.3153	2.365
NumCatalogPurchases	2.7204	1.9712	2.6517	4.2622	5.5	0.5891
NumStorePurchases	5.8359	3.9281	6.8315	6.0562	9.2727	3.3803
NumWebVisitsMonth	5.2801	5.4029	5.2831	5.2697	3.0824	6.5671
Complain	0.0089	0.0144	0.0045	0.0075	0.0114	0.0102
Response	0.149	0	0	1	0	0
=====						
Time taken to build model (percentage split) : 0.01 seconds						
Clustered Instances						
0	36 (8%)					
1	106 (24%)					
2	67 (15%)					
3	79 (18%)					
4	160 (36%)					

8 Cluster :

Final cluster centroids:									
Attribute	Cluster#								
	Full Data (1792.0)	0 (138.0)	1 (126.0)	2 (141.0)	3 (327.0)	4 (314.0)	5 (410.0)	6 (256.0)	7 (80.0)
Year_Birth	1968.8694	1968.6884	1969.6746	1969.7589	1967.685	1978.449	1963.1634	1965.3672	1974.0375
Income	52734.671	40101.3092	41515.119	77150.0899	73857.3344	32968.8929	55844.4651	41829.0886	59367.8002
Kidhome	0.4381	0	0.7143	0.0284	0.0122	1.0287	0.0073	1.0859	1.0375
Teenhome	0.4989	0	0.5794	0.078	0.2324	0	1.0561	1.0391	0.4375
MntWines	310.8064	123.6014	199.119	757.4397	591.9052	30.449	382.5439	64.1953	395.3625
MntFruits	26.1987	13.6522	9.6508	65.1986	71.1346	4.9809	15.3585	2.7148	35.475
MntMeatProducts	171.769	120.6087	74.7143	501.0851	426.4924	24.1433	104.2171	23.1094	192.625
MntFishProducts	38.1016	24.4275	12.1667	90.1206	105.1315	7.5605	20.561	4.0938	55.4625
MntSweetProducts	27.5179	17.5435	10.8175	64.0071	72.107	5.1146	17.2415	3.1328	43.0875
MntGoldProds	44.0491	32.942	44.3095	77.8865	75.7339	13.5892	49.3341	12.7227	66.3625
NumDealsPurchases	2.3125	1.4565	3.6667	1.3121	1.4067	1.9013	2.6415	2.9414	5.0375
NumWebPurchases	4.101	2.8986	4.2381	5.6879	5.1957	2.1592	5.1195	2.2461	7.025
NumCatalogPurchases	2.7204	2.0145	2.0079	6.2766	5.6361	0.3949	2.6805	0.5898	3.025
NumStorePurchases	5.8359	4.0072	4.1825	7.7305	9.2446	3.1433	6.7878	3.3516	7.9625
NumWebVisitsMonth	5.2801	5.3551	7.246	3.5035	2.893	6.8121	5.2171	6.2891	6.025
Complain	0.0089	0.0145	0.0079	0.0071	0.0092	0.0096	0.0049	0.0117	0.0125
Response	0.149	0	1	1	0	0	0	0	0

Time taken to build model (percentage split) : 0.01 seconds

Clustered Instances

0	36 (8%)
1	31 (7%)
2	36 (8%)
3	76 (17%)
4	85 (19%)
5	89 (20%)
6	78 (17%)
7	17 (4%)

10 Cluster :

Final cluster centroids:											
Attribute	Cluster#										
	0 (1792.0)	1 (128.0)	2 (213.0)	3 (174.0)	4 (135.0)	5 (234.0)	6 (242.0)	7 (62.0)	8 (126.0)	9 (141.0)	10 (337.0)
Year_Birth	1968.8694	1969.0313	1964.3146	1967.546	1969.2593	1965.1538	1962.843	1967.9194	1969.6746	1969.7589	1978.6231
Income	52734.671	38653.4818	64139.2488	74112.0767	73803.8148	40801.426	50723.1612	60445.968	41515.119	77150.0899	33688.2849
Kidhome	0.4381	0	0	0.0977	0.0593	1.0855	0	1.0806	0.7143	0.0284	1.0237
Teenhome	0.4989	0	1.0423	0	0.2148	1.0342	1.0455	1.0323	0.5794	0.078	0
MntWines	310.8064	91.2109	610.8967	667.3333	489.2741	47.2265	215.8554	391.3065	199.119	757.4397	40.2493
MntFruits	26.1987	13.2813	34.2911	41.931	103.0074	2.0897	9.2066	21.7903	9.6508	65.1986	6.7211
MntMeatProducts	171.769	111.8594	192.338	434.4483	430.1778	18.6581	56.814	140.0323	74.7143	501.0851	35.5846
MntFishProducts	38.1016	23.5781	42.3192	67.5977	155.3852	3.2094	13.095	35.7742	12.1667	90.1206	9.2849
MntSweetProducts	27.5179	17.5156	38.4225	48.3448	95.3704	2.4701	9.8926	29.1613	10.8175	64.0071	7.2136
MntGoldProds	44.0491	33.1172	74.939	56.1667	95.2963	11.0812	32.6777	51.5806	44.3095	77.8865	17.3086
NumDealsPurchases	2.3125	1.2734	3.0141	1.1322	1.4222	2.6239	2.281	6.7097	3.6667	1.3121	2.1395
NumWebPurchases	4.101	2.8359	6.5822	4.9943	5.0593	1.9744	4.219	6.2419	4.2381	5.6879	2.451
NumCatalogPurchases	2.7204	1.8828	4.3286	5.5632	5.4963	0.4615	1.7066	2.4516	2.0079	6.2766	0.5668
NumStorePurchases	5.8359	3.9531	9.5869	9.6609	8.0963	3.1838	5.0165	7.2258	4.1825	7.7305	3.2997
NumWebVisitsMonth	5.2801	5.3359	4.9953	2.8218	2.8148	6.265	5.2934	6.2097	7.246	3.5035	6.8398
Complain	0.0089	0.0156	0.0141	0.0057	0	0.0128	0.0041	0.0161	0.0079	0.0071	0.0089
Response	0.149	0	0	0	0	0	0	0	1	1	0

Time taken to build model (percentage split) : 0.02 seconds

Clustered Instances

0	32 (7%)
1	51 (11%)
2	25 (6%)
3	41 (9%)
4	72 (16%)
5	55 (12%)
6	19 (4%)
7	31 (7%)
8	36 (8%)
9	86 (19%)

1.10.2. Penjelasan

Pada langkah ini dilakukan run k-means namun berbeda dengan langkah sebelumnya, pada langkah ini dilakukan dengan metode euclidean dimana menggunakan pembobotan untuk setiap dimensinya, sehingga hasil yang ditampilkan pada jumlah klusterisasi dari 2 hingga 10 kluster menjadi lebih akurat.

B. Analisis Clustering dengan Algoritma K-Means

Hasil Interpretasi Karakteristik dari 2 Cluster K-Means :

Pelanggan dalam **kluster 0** adalah kelompok yang berusia sekitar 55-an, memiliki pendapatan menengah, tidak memiliki tanggungan anak kecil atau remaja di rumah. Mereka cenderung menghabiskan lebih banyak untuk pembelian anggur dan produk daging. Respons terhadap tawaran dan promosi juga relatif rendah.

Sementara Pelanggan dalam **kluster 1** ini adalah kelompok yang sedikit lebih muda yaitu umur 50 awal, dengan pendapatan menengah, memiliki sedikit atau tidak ada anak kecil, dan tingkat remaja di rumah yang rendah. Mereka cenderung menghabiskan lebih banyak untuk pembelian anggur dan produk daging. Respons terhadap tawaran dan promosi cenderung sedikit lebih tinggi dibandingkan kluster lainnya, meskipun tidak terlalu tinggi.

Hasil Interpretasi Karakteristik dari 3 Cluster K-Means :

Dapat diinterpretasikan bahwa pelanggan dalam **kluster 0** cenderung memiliki profil yang lebih muda, pendapatan menengah rendah, dan respons terhadap promosi yang rendah. Pelanggan dalam **kluster 1** cenderung memiliki profil yang sedikit lebih tua, pendapatan menengah, dan respons terhadap promosi yang rendah. Sedangkan pelanggan dalam **kluster 2** cenderung memiliki profil yang lebih tua, pendapatan yang tinggi, dan respons terhadap promosi yang tinggi.

Hasil Interpretasi Karakteristik dari 5 Cluster K-Means :

Dapat diinterpretasikan bahwa kluster 0 terdiri dari pelanggan dengan karakteristik pendapatan menengah, sedangkan kluster 1, 2, dan 3 terdiri dari pelanggan dengan pendapatan menengah. Kluster 1 dan 3 memiliki respons yang lebih tinggi terhadap promosi daripada kluster lainnya. Kluster 2 memiliki jumlah anak kecil dan remaja di rumah yang tinggi. Kluster 4 terdiri dari pelanggan dengan pendapatan menengah rendah dan memiliki jumlah anak kecil yang tinggi.

Hasil Interpretasi Karakteristik dari 8 Cluster K-Means :

Dapat diinterpretasikan Berdasarkan hasil klasterisasi, terdapat delapan kluster pelanggan yang memiliki karakteristik yang berbeda. Kluster tersebut mencakup pelanggan dengan berbagai tingkat pendapatan, jumlah anak dan remaja di rumah, serta pola pembelian dan respons terhadap keluhan dan promosi. Misalnya, terdapat kluster pelanggan dengan pendapatan menengah yang cenderung menghabiskan jumlah yang sedang pada pembelian anggur, daging, dan produk manis, sementara kluster lainnya terdiri dari pelanggan dengan pendapatan tinggi yang memiliki kecenderungan pembelian yang tinggi dan respons terhadap promosi yang tinggi.

Hasil Interpretasi Karakteristik dari 10 Cluster K-Means :

Dapat diinterpretasikan bahwa terdapat pola umum dalam karakteristik pelanggan dalam 10 kluster yang terbentuk. Beberapa kluster cenderung memiliki pelanggan yang lebih tua dengan pendapatan menengah hingga tinggi, sementara respons terhadap promosi bervariasi antara rendah dan tinggi. Terdapat juga kluster dengan pelanggan yang lebih muda dan pendapatan menengah rendah yang memiliki respons terhadap promosi rendah.

BAB III

Penutup

Kesimpulan

Dapat disimpulkan bahwa pada K-means adalah algoritma clustering yang digunakan untuk mengelompokkan data menjadi beberapa kelompok yang serupa berdasarkan atribut yang dimiliki. Serta melakukan pengaturan konfigurasi (parameter yang digunakan) weka pada saat perhitungan K-means akan mempengaruhi hasil yang didapatkan pula.

Kemudian untuk hasil analisis dari seluruh Clustering yang dilakukan (2 - 10 Cluster) terhadap Dataset Marketing Campaign, maka jumlah kluster yang paling ideal untuk kasus ini adalah **3 Cluster**. Hal ini dikarenakan perusahaan bisa mendapat gambaran yang lebih sederhana dan ringkas tentang profil pelanggan. Ini memudahkan pemahaman dan interpretasi data bagi pengguna yang mungkin tidak memiliki pengetahuan yang mendalam tentang analisis kluster. Berikutnya 3 kluster mewakili pembagian yang cukup beragam dalam karakteristik pelanggan, seperti pendapatan, usia, dan respons terhadap promosi. Serta menggunakan jumlah kluster yang lebih sedikit membantu menghindari overfitting dan kompleksitas yang berlebihan pada model. Dengan demikian, hasil 3 kluster memberikan ringkasan yang cukup informatif dan cukup representatif tentang segmentasi pelanggan yang ada. Dimana hasil segmentasinya yaitu **kluster 0** cenderung memiliki profil yang lebih muda, pendapatan menengah rendah, dan respons terhadap promosi yang rendah. Pelanggan dalam **kluster 1** cenderung memiliki profil yang sedikit lebih tua, pendapatan menengah, dan respons terhadap promosi yang rendah. Sedangkan pelanggan dalam **kluster 2** cenderung memiliki profil yang lebih tua, pendapatan yang tinggi, dan respons terhadap promosi yang tinggi.