



**TUGAS BESAR DATA MINING
CLUSTERING NUTRITION FACTS FOR FOOD OF STARBUCKS**

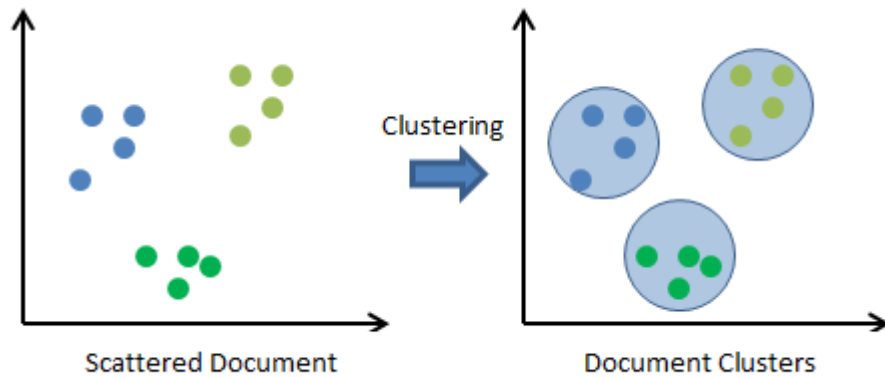
Disusun Oleh:

Rachmat Fauzan	3311801036
Abduraffi Naufal	3311801039
Andre Tamini Putra	3311801042
Rezki Janturi Pratama	3311801055

**POLITEKNIK NEGERI BATAM
BATAM
2019**

Pengertian Clustering

Apa itu Clustering ? Clustering merupakan proses partisi satu set objek data ke dalam himpunan bagian yang disebut dengan cluster. Objek yang di dalam cluster memiliki kemiripan karakteristik antar satu sama lainnya dan berbeda dengan cluster yang lain. Partisi tidak dilakukan secara manual melainkan dengan suatu algoritma clustering. Oleh karena itu, clustering sangat berguna dan bisa menemukan group atau kelompok yang tidak dikenal dalam data.

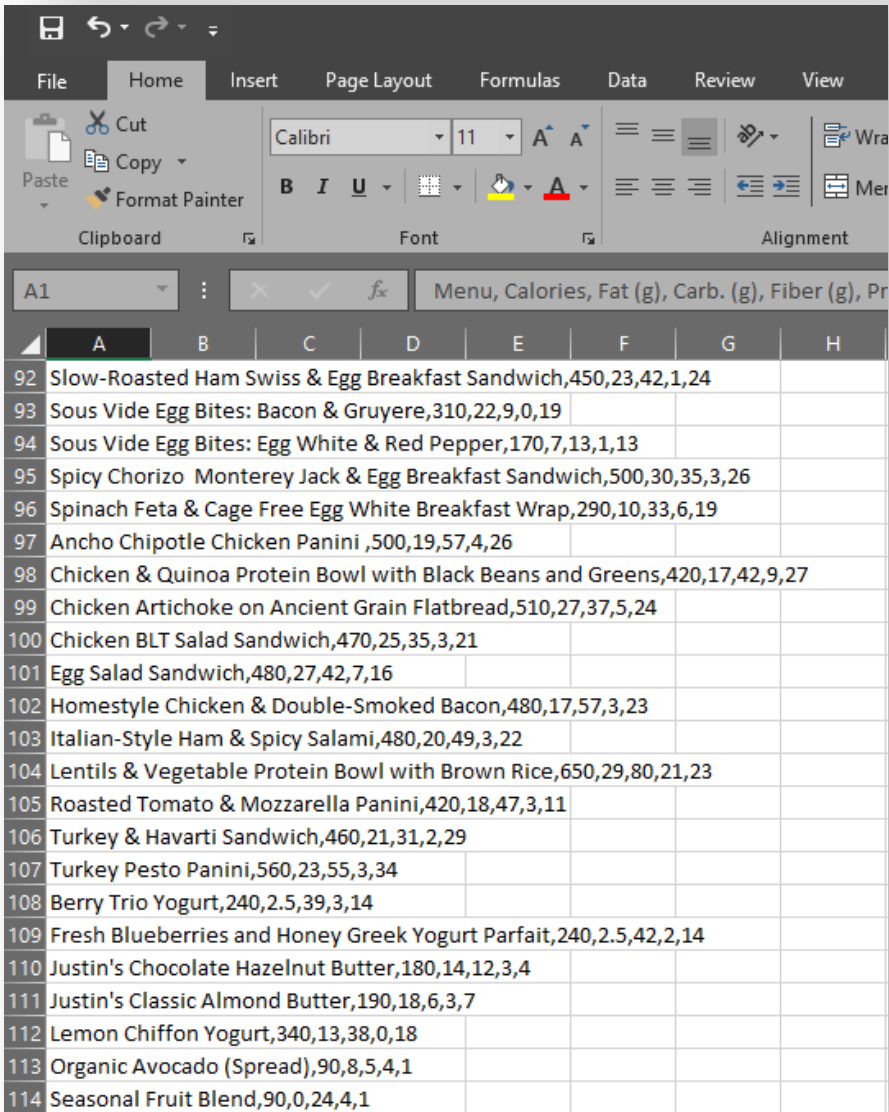


Gambar.1 Pengertian Clustering

Dataset

Dataset Awal

Jumlah Data = 114 data



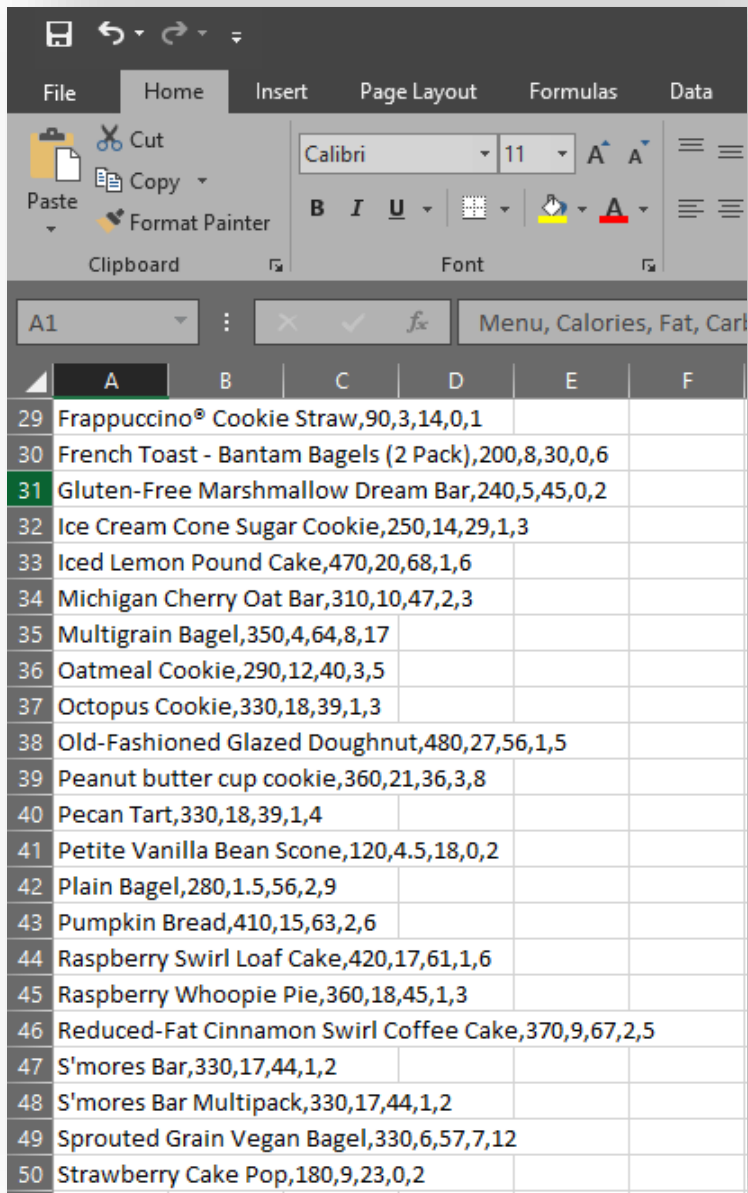
The screenshot shows the Microsoft Excel interface with the 'Home' tab selected. The ribbon includes 'Clipboard', 'Font', and 'Alignment' groups. The active cell is A1, and the formula bar shows the text 'Menu, Calories, Fat (g), Carb. (g), Fiber (g), Pr'. The data is organized in a table with columns A through H. The first column (A) contains item names, and the subsequent columns (B-H) contain numerical values representing nutritional information.

	A	B	C	D	E	F	G	H
92	Slow-Roasted Ham Swiss & Egg Breakfast Sandwich,	450	23	42	1	24		
93	Sous Vide Egg Bites: Bacon & Gruyere,	310	22	9	0	19		
94	Sous Vide Egg Bites: Egg White & Red Pepper,	170	7	13	1	13		
95	Spicy Chorizo Monterey Jack & Egg Breakfast Sandwich,	500	30	35	3	26		
96	Spinach Feta & Cage Free Egg White Breakfast Wrap,	290	10	33	6	19		
97	Ancho Chipotle Chicken Panini ,	500	19	57	4	26		
98	Chicken & Quinoa Protein Bowl with Black Beans and Greens,	420	17	42	9	27		
99	Chicken Artichoke on Ancient Grain Flatbread,	510	27	37	5	24		
100	Chicken BLT Salad Sandwich,	470	25	35	3	21		
101	Egg Salad Sandwich,	480	27	42	7	16		
102	Homestyle Chicken & Double-Smoked Bacon,	480	17	57	3	23		
103	Italian-Style Ham & Spicy Salami,	480	20	49	3	22		
104	Lentils & Vegetable Protein Bowl with Brown Rice,	650	29	80	21	23		
105	Roasted Tomato & Mozzarella Panini,	420	18	47	3	11		
106	Turkey & Havarti Sandwich,	460	21	31	2	29		
107	Turkey Pesto Panini,	560	23	55	3	34		
108	Berry Trio Yogurt,	240	2.5	39	3	14		
109	Fresh Blueberries and Honey Greek Yogurt Parfait,	240	2.5	42	2	14		
110	Justin's Chocolate Hazelnut Butter,	180	14	12	3	4		
111	Justin's Classic Almond Butter,	190	18	6	3	7		
112	Lemon Chiffon Yogurt,	340	13	38	0	18		
113	Organic Avocado (Spread),	90	8	5	4	1		
114	Seasonal Fruit Blend,	90	0	24	4	1		

Gambar.2 Dataset Original

Dataset yg Diambil

Jumlah Data = 50 Data



	A	B	C	D	E	F
29	Frappuccino®	Cookie Straw,	90,3,14,0,1			
30	French Toast - Bantam	Bagels (2 Pack),	200,8,30,0,6			
31	Gluten-Free Marshmallow	Dream Bar,	240,5,45,0,2			
32	Ice Cream Cone Sugar	Cookie,	250,14,29,1,3			
33	Iced Lemon Pound	Cake,	470,20,68,1,6			
34	Michigan Cherry Oat	Bar,	310,10,47,2,3			
35	Multigrain Bagel,	350,4,64,8,17				
36	Oatmeal Cookie,	290,12,40,3,5				
37	Octopus Cookie,	330,18,39,1,3				
38	Old-Fashioned Glazed	Doughnut,	480,27,56,1,5			
39	Peanut butter cup	cookie,	360,21,36,3,8			
40	Pecan Tart,	330,18,39,1,4				
41	Petite Vanilla Bean	Scone,	120,4.5,18,0,2			
42	Plain Bagel,	280,1.5,56,2,9				
43	Pumpkin Bread,	410,15,63,2,6				
44	Raspberry Swirl Loaf	Cake,	420,17,61,1,6			
45	Raspberry Whoopie	Pie,	360,18,45,1,3			
46	Reduced-Fat Cinnamon	Swirl Coffee Cake,	370,9,67,2,5			
47	S'mores Bar,	330,17,44,1,2				
48	S'mores Bar Multipack,	330,17,44,1,2				
49	Sprouted Grain Vegan	Bagel,	330,6,57,7,12			
50	Strawberry Cake	Pop,	180,9,23,0,2			

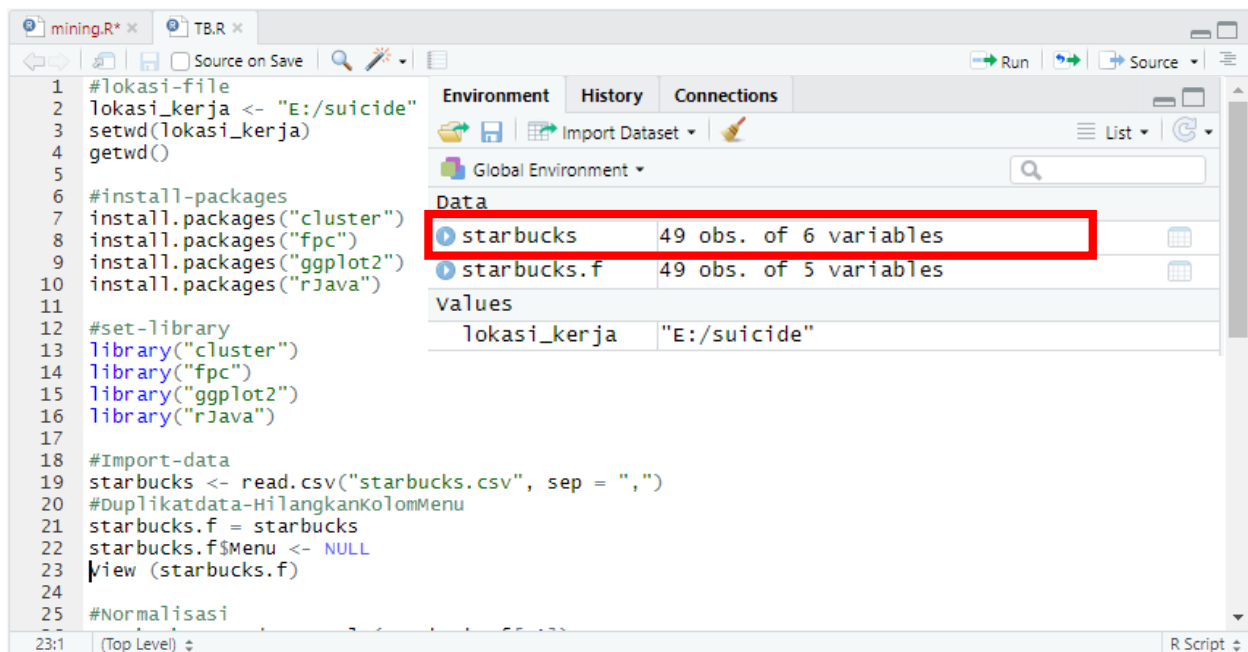
Gambar.3 Dataset Diambil

Proses Data Mining

Langkah Kerja :

1. Set lokasi_kerja
2. Instalasi paket yang diperlukan
3. Set-library
4. Import-data
5. Duplikat data untuk menghilangkan string di table Menu
6. Normalisasi data
7. Membuat cluster dengan metode K-means
8. Buat plot diagram
9. Buat point pada plot diagram agar mudah dibaca
10. Tampilkan tabel menu untuk melihat ada di cluster yg mana

Import DataSet



Gambar.4 Import Dataset

Normalisasi

```

25 #Normalisasi
26 starbucks.stand <- scale(starbucks.f[-1])
27
28 #view
29 view(starbucks.stand)
30 head(starbucks.stand)
31 starbucks.stand$Menu

```

Gambar.4 Normalisasi Data

View Data

	Menu	Calories	Fat	Carb	Fiber	Protein
1	Chonga Bagel	300	5.0	50	3	12
2	8-Grain Roll	380	6.0	70	7	10
3	Almond Croissant	410	22.0	45	3	10
4	Apple Fritter	460	23.0	56	2	7
5	Banana Nut Bread	420	22.0	52	2	6
6	Blueberry Muffin with Yogurt and Honey	380	16.0	53	1	6
7	Blueberry Scone	420	17.0	61	2	5
8	Butter Croissant	240	12.0	28	1	5
9	Butterfly Cookie	350	22.0	38	0	2
10	Cheese Danish	320	16.0	36	1	8
11	Chewy Chocolate Cookie	170	5.0	30	2	2
12	Chocolate Chip Cookie	310	15.0	42	2	4
13	Chocolate Chunk Muffin	440	21.0	60	2	7

Gambar.5 Data Original

	Calories	Fat	Carb	Fiber	Protein
1	300	5.0	50	3	12
2	380	6.0	70	7	10
3	410	22.0	45	3	10
4	460	23.0	56	2	7
5	420	22.0	52	2	6
6	380	16.0	53	1	6
7	420	17.0	61	2	5
8	240	12.0	28	1	5
9	350	22.0	38	0	2
10	320	16.0	36	1	8
11	170	5.0	30	2	2
12	310	15.0	42	2	4
13	440	21.0	60	2	7
14	330	18.0	38	1	6

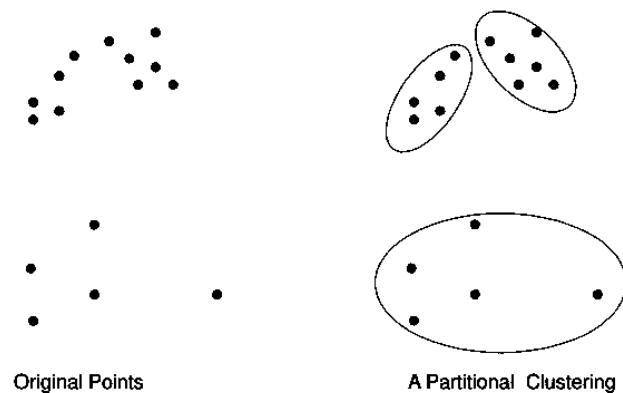
Gambar.6 Hilangkan Kolom Menu

	Fat	Carb	Fiber	Protein
1	-1.27517728	0.220960122	0.66918862	1.94453833
2	-1.14223327	1.636260248	3.01134880	1.32177769
3	0.98487097	-0.132864910	0.66918862	1.32177769
4	1.11781498	0.645550160	0.08364858	0.38763673
5	0.98487097	0.362490134	0.08364858	0.07625641
6	0.18720688	0.433255141	-0.50189147	0.07625641
7	0.32015089	0.999375191	0.08364858	-0.23512392
8	-0.34456918	-1.335870017	-0.50189147	-0.23512392
9	0.98487097	-0.628219954	-1.08743151	-1.16926488
10	0.18720688	-0.769749967	-0.50189147	0.69901705
11	-1.27517728	-1.194340004	0.08364858	-1.16926488
12	0.05426286	-0.345159929	0.08364858	-0.54650424
13	0.85192695	0.928610185	0.08364858	0.38763673

Gambar.7 Normalisasi Data

Algoritma K-Means

Apa itu K-Means ? suatu metode penganalisaan data atau metode Data Mining yang melakukan proses pemodelan tanpa supervisi (unsupervised) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi.



Gambar.8 Partisi Data Cluster

Membuat Cluster

```
33 #Membuat-Cluster-dengan-K-Means
34 set.seed (8953)
35 hasil <- kmeans(starbucks.stand, 3)
36 hasil
37 |
38 #View-Satu2
39 hasil$cluster
40 #Untuk Melihat Jumlah anggota di tiap cluster
41 hasil$size
42 hasil$centers
```

Gambar.9 Code Membuat Cluster

Membuat Plot

```
47 #plot1
48 plot(starbucks.stand, col = hasil$cluster)
49 #points
50 points(hasil$centers, col = 2:6, pch = 8, cex = 2)
51 #plot2
52 clusplot(starbucks.stand, hasil$cluster, color = T, shade = T, table = 2, lines = 0)
```

Gambar.10 Code Membuat Plot

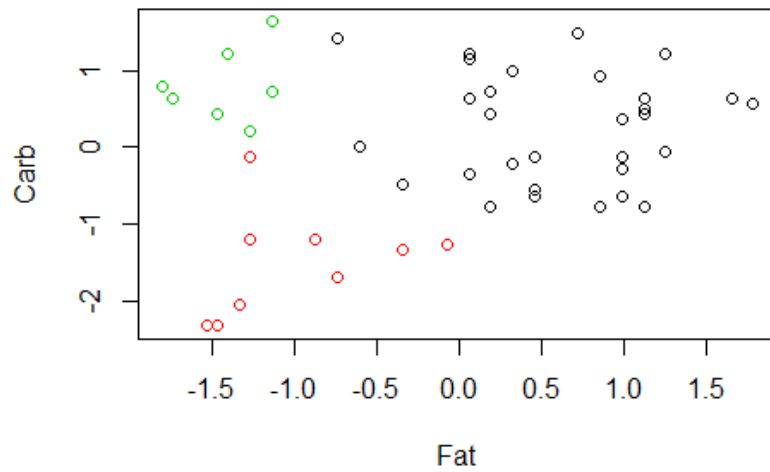
Melihat Tabel Clustering

```
44 #Tampilkan-Menu
45 table(starbucks$Menu, hasil$cluster)
46
```

Gambar.11 Code Menampilkan Tabel Cluster

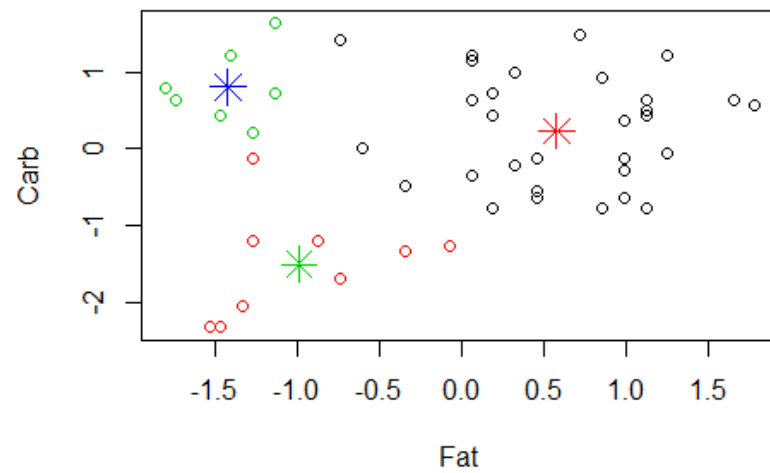
Hasil Data Mining

Plot Original



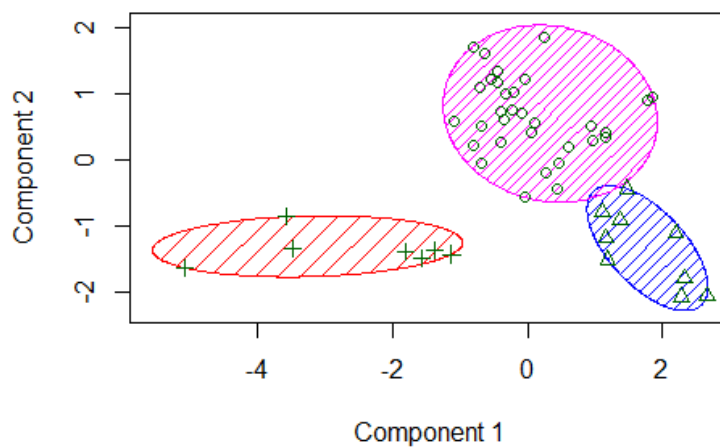
Gambar.11 Plot Ori

Plot Point



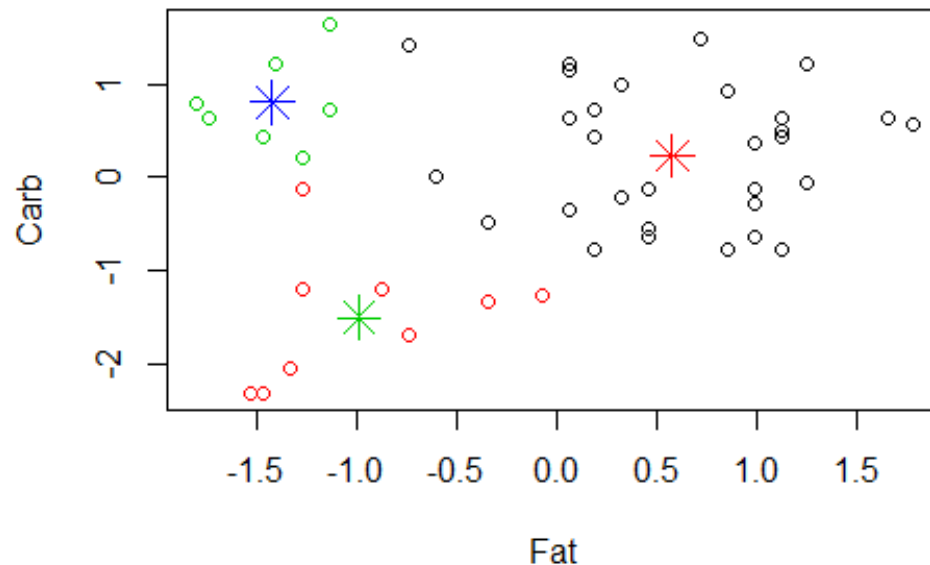
Gambar.12 Plot Point Original

Plot Partisi



Gambar.13 Plot Cluster

Kesimpulan



Gambar.12 Plot Point Original

Terdapat 3 Cluster pada nutrition fact menu of starbuck, yaitu :

Cluster 1 = 33 Lingkaran (Hitam)

Cluster 2 = 9 Lingkaran (Merah)

Cluster 3= 7 Lingkaran (Hijau)

Tabel Clustering

	1	2	3
8-Grain Roll	0	0	1
Almond Croissant	1	0	0
Apple Fritter	1	0	0
Banana Nut Bread	1	0	0
Blueberry Muffin with Yogurt and Honey	1	0	0
Blueberry Scone	1	0	0
Butter Croissant	0	1	0
Butterfly Cookie	1	0	0
Cheese Danish	1	0	0
Chewy Chocolate Cookie	0	1	0
Chocolate Chip Cookie	1	0	0
Chocolate Chunk Muffin	1	0	0
Chocolate Croissant	1	0	0
Chocolate Hazelnut Croissant	1	0	0
Chocolate Marble Loaf Cake	1	0	0
Chonga Bagel	0	0	1
Cinnamon Morning Bun	1	0	0
Cinnamon Raisin Bagel	0	0	1
Classic Coffee Cake	1	0	0
Cookie Butter Bar	1	0	0
Cranberry Orange Scone	1	0	0
Devil's Food Doughnut	1	0	0
Double Chocolate Brownie	1	0	0
Double Chocolate Chunk Brownie	1	0	0
Double Chocolate Loaf Cake	1	0	0
Everybody's Favorite - Bantam Bagel (2 Pack)	0	1	0
Everything Bagel with Cheese	0	0	1
Frappuccino® Cookie Straw	0	1	0
French Toast - Bantam Bagels (2 Pack)	0	1	0
Gluten-Free Marshmallow Dream Bar	0	1	0
Ice Cream Cone Sugar Cookie	0	1	0
Iced Lemon Pound Cake	1	0	0
Michigan Cherry oat Bar	1	0	0
Multigrain Bagel	0	0	1
Oatmeal Cookie	1	0	0
Octopus Cookie	1	0	0

Dengan Penjabaran Seperti Berikut :

Cluster 1 = Nutrisi Kurang Baik

Cluster 2 = Nutrisi Cukup Baik

Cluster 3 = Nutrisi Baik

Catatan :

Kalori yang tinggi tidak baik bagi kesehatan

Gambar.14 Tabel Menu Clustering

Referensi

Youtube : <https://www.youtube.com/watch?v=hVexp9fOXw&t=428s>
<https://www.youtube.com/watch?v=OuOMSsbkm00>
<https://www.youtube.com/watch?v=oFBYHlzUcLU>

Link Data : <https://www.kaggle.com/starbucks/starbucks-menu/data#> =

Link Github

Rachmat : https://github.com/rachmatfauzan/3311801036_datamining_polibatam

Abduraffi : https://github.com/raffinaufal/3311801039_datamining_polibatam

Andre : https://github.com/badbons/3311801042_datamining_polibatam

Rezki : https://github.com/arxrezky/3311801055_datamining_polibatam