PREDIKSI *CLUSTERING, CALCULATION* DAN *CLASSIFICATION FRUIT AND VEGETABLE CONSUMPTION*

Adriyendi

Program Studi Manajemen Informatika IAIN Batusangkar
Jl. Sudirman No. 137 Kuburajo Lima Kaum Batusankgar Indonesia
Email:suratkudisini@gmail.com

ABSTRACT

Prediction model using combination of K-Means Clustering, Excel Function, and Naïve Bayes Classifier. Process is dataset, clustering, calculation, classification and prediction. Dataset source on BPS 2013 about consumption of fruit and vegetable. Clustering using K-Means Clustering. Clustering by output Cluster 1, Cluster 2, and Cluster 3. Calculation using Excel. Calculation by output Priority Yes and Priority No. Classification using Naïve Bayes Classifier. Classification by output Class Good and Class Bad. All data processing for clustering, calculation, and classification using Excel. Experimental results on BPS 2013 Dataset show percentage of fruit consumption 42,42% Class Good (class above average) and percentage fruit consumption of fruit consumption 57,58% Class Bad (class below average). Percentage of vegetable consumption 45,45% Clas Good (class above average) and percentage of vegetable consumption 54,55% Class Bad (class below average). Clustering, calculation and classification can be combined becamed prediction model.

Key words: clustering, calculation, classification, fruit and vegetable consumption

PENDAHULUAN

Sumber gizi dan konsumsi bahan pangan yang baik salah satunya adalah buah dan sayur (S.M. Perdana et. al, 2013). Konsumsi buah dan sayur menjadi kebutuhan utama saat ini. Konsumsi buah dan sayur memberikan kontribusi penting bagi tubuh manusia (A.A. Candra et. al, 2013). Buah sebagai sumber zat gizi dan vitamin, sayur sebagai sumber zat gizi dan komponen bioaktif yang merupakan elemen nutrisi penting bagi kesehatan manusia (M. Fasitasari et. al, 2013). Kekurangan konsumsi buah dan sayur sebagai salah satu faktor yang dapat meningkatkan kondisi tubuh tidak sehat.

Peningkatan konsumsi buah dan sayur telah direkomendasikan (X. Wang et. al, 2014) sebagai komponen kunci kesehatan dengan meng-konsumsi buah dan sayur yang beragam, sehat, bergizi, seimbang dan aman. Buah dan sayur pada konsumsi harian, dapat meningkatkan kesehatan individual dan untuk mendukung kesehatan masyarakat (O. Stackelberg et. al, 2013). Konsumsi buah dan sayur yang dikembangkan lebih luas menjadi sinyal awal yang menandakan kesehatan tubuh secara umum berjalan dengan baik (T.S. Conner et. al, 2014). Untuk itu, perlu ditentukan klaster konsumsi buah dan sayuran agar dapat diklasifikasi untuk diprediksi konsumsi buah dan sayur. Hasil prediksi yang akurat sangat penting dan berguna untuk membuat kebijakan nasional.

Prediksi data diklasifikasikan berdasarkan perilaku atau nilai yang diperkirakan pada masa yang akan datang. Dalam proses prediksi, satu-satunya cara untuk memeriksa ketepatan hasil adalah dengan menunggu dan memperhatikan. Dalam melakukan prediksi, dengan menggunakan data sampel, dimana nilai dari variabel yang akan diprediksikan sudah diketahui. Hal ini sama dengan data historis untuk data sampel tersebut. Data historis ini bisa digunakan untuk membuat sebuah model yang berguna untuk menjelaskan perilaku yang sedang

diamati. Apabila model ini diaplikasikan pada data masukan, akan menghasilkan prediksi di masa yang akan datang.

Clustering adalah proses untuk melakukan segmentasi atas sebuah populasi yang heterogen menjadi beberapa sub kelompok atau cluster yang homogen. *Clustering* berbeda dengan proses klasifikasi, karena tidak bergantung pada kelas-kelas yang sudah ditetapkan sebelumnya (S. Shukla et. al, 2014), maupun sampel data. Data akan dikelompokkan berdasarkan kemiripan karakteristik. Analisis *clustering* biasanya berdasarkan klasifikasi, hirarki, populasi, dan sejenisnya (G. Liu et. al, 2014). Salah satu hal yang sangat penting dalam *clustering* adalah penggunaan ukuran kemiripan (similarity). Jika datanya numerik, fungsi kemiripan (similarity function) berdasarkan jarak. Jarak yang sering digunakan: *Euclidean Metric* (Euclidean distance), *Minkowsky Metric*, *Manhattan Metric*, dan sejenisnya. *Clustering* yang baik bergantung pada pengukuran kesamaan. Kesamaan di dalam kelas (intraclass similarity) yang tinggi dan kesamaan antar kelas (interclass similarity) yang rendah. Kelas-kelas yang belum ditentukan sebagai metode *unsupervised learning*. Metode ini tidak melibatkan tahap pembelajaran, melainkan bergantung pada penggunaan algoritma untuk mendeteksi pola-pola, seperti asosiasi, *sequences*, yang ada pada data masukan, berdasarkan kriteria penting yang telah ditentukan. Pendekatan ini mengarah kepada pembuatan aturan-aturan yang menggambarkan asosiasi, klaster, dan segmen yang telah ditemukan. Banyak algoritma yang telah diusulkan untuk *data clustering* (Y. Kumar et. al, 2014), satu di antaranya menggunakan *K-Means Algorithm*, dengan kelebihan, sederhana, efisien dan *ease convergence*.

Calculation menggunakan Microsoft Excel 2010 (Excel). Excel adalah *scientific learning tool*yang sangat potensial. Kemampuan optimal Exceldalam menampilkan kalkulasi numerik, mengelola data, dan menarik dalam penyajian data (K.K. Manjusha et. al, 2014). Kemampuan Excel sebagai alat untuk mengolah data diterapkan hampir pada semua algoritma. Klasifikasi melibatkan proses pemeriksaan karakteristik suatu obyek dan kemudian memasukkannya ke dalam salah satu kelas yang sudah didefinisikan sebelumnya. *Classification* atau klasifikasi melibatkan pendefinisian kelas-kelas dan sampel data yang berisi contoh obyek yang sudah diklasifikasi sebelumnya. Tujuannya untuk membuat sebuah model yang dapat diaplikasikan pada data yang belum terklasifikasi.

Classification merupakan suatu kegiatan dalam *distillation of knowledge* dengan pendekatan *learning supervised*. Pendekatan ini melibatkan fase pembelajaran, yang terjadi ketika data-data historis yang karakteristiknya dipetakan ke hasil keluaran, diproses melalui algoritma. Proses tersebut akan melatih algoritma untuk mengenali variabel kunci dan nilai-nilai yang akan dijadikan sebagai dasar pembuatan prediksi. Berdasarkan kelompok pendekatan numerik, satu pendekatan probabilistik adalah *Naïve Bayes Classifier* (Bustami, 2014). Klasifikasi menggunakan *Naïve Bayes Classifier* merupakan klasifikasi dengan metode probabilistik dan statistik, menghitung peluang untuk suatu hipotesis, menghitung peluang suatu kelas dari masing-masing kelompok atribut yang ada, dan menentukan kelas mana yang paling optimal, dikenal dengan *Teorema Bayes*. Teorema tersebut dikombinasikan dengan *Naïve* dimana diasumsikan kondisi antar atribut saling bebas (V. Shukla et.al., 2014). Klasifikasi *Naïve Bayes Classifier* diasumsikan bahwa *ada atau tidak* ciri tertentu dari sebuah kelas, tidak ada hubungannya dengan ciri dari kelas lainnya. *Naïve Bayes Classifier* memiliki keunggulan, sederhana, cepat dan akurasi tinggi. Output *Naïve Bayes Classifier* dalam klasifikasi relatif sangat baik.

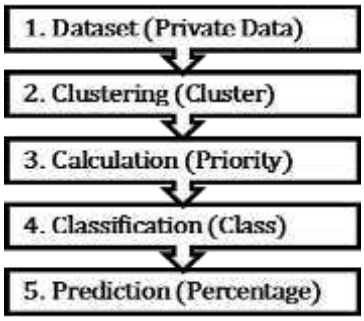
Penelitian Terkait

Pendekatan yang diajukan dalam paper (V. Shukla et. al., 2012), mengusulkan penelitian dengan kombinasi tiga metode pendekatan (K-Means Clustering, Apriori dan Decision Tree), untuk mereduksi biaya, memperbaiki waktu dan efisiensi eksekusi dalam *Intrusion Detection System* (IDS). Hasilnya menunjukkan bahwa IDS yang diusulkan lebih baik dalam akurasi dan efisiensi. Dalam paper (W. Yassin et.al., 2013), mengusulkan *integrated machine learning algorithm* berdasarkan *K-Means Clustering* dan *Naïve Bayes Classifier* untuk melakukan

anomaly-based detection dalam memperbaiki tanda bahaya palsu pada tingkat akurasi dan deteksi maksimal. Hasilnya menunjukkan implementasi *K-Means Clustering* dan *Naïve Bayes Classifier* akurat, meningkat secara signifikan dalam tingkat deteksi berkelanjutan, dan mengurangi tanda bahaya palsu. Proposal paper (M. Banerjee et. al, 2013), melakukan studi tentang *K-Means Clustering* melalui *Naïve Bayes Classification* untuk *anomaly based network intrusion detection*. Hasilnya pada KDD'99 Dataset menampilkan pendekatan baru dalam *detecting network intrusion*. Metode yang diusulkan lebih baik dalam *detection rate* saat diterapkan pada KDD'99 Dataset dibandingkan pendekatan *Naïve Bayes*. Berdasarkan analisis dalam paper (Y. Emami et. Al, 2014), mengusulkan metode gabungan pada *intrusion detection system.* Prinsip utama dari pendekatan ini adalah memberikan bobot *K-Means Clustering* dan *Naïve Bayes Classification*. Algoritma C5.0 digunakan untuk mengatur atribut, atribut diberi bobot, digunakan *K-Means Clustering,* karena itu *accuracy of clustering* meningkat. Membuat model klasifikasi tipe kelompok dan jenis serangan dan membantu administrator dalam melakukan identifikasi jenis serangan lebih awal serta menolak lebih cepat terhadap efek serangan. Paper (N.O.F. Elssied et. al, 2014), mengusulkan *hybrid scheme,* pada klasifikasi surat elektronik, berdasarkan pada *Naïve Bayes* dan *K-Means Clustering,* untuk meningkatkan akurasi dan mereduksi *misclassification rate of spam detection*. Hasil eksperimen dari skema yang diusulkan mampu memisahkan spam pada dataset, *Naïve Bayes* (KNavie) dengan output yang signifikan dalam *spam detection methods*. Eksperimen dalam paper (N. Sharma et. al, 2013), dengan kombinasi *clustering and classification* menghasilkan akurasi optimal saat *dataset is containing missing values*.

METODE PENELITIAN

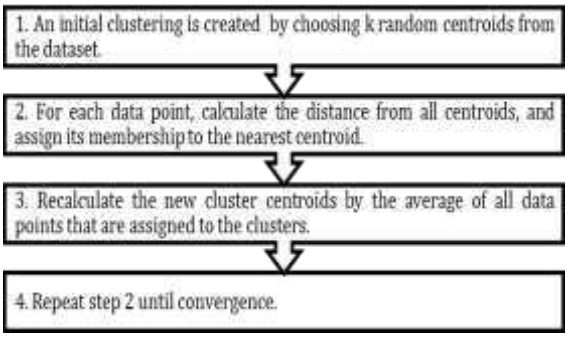
Metode penelitian dalam bentuk kerangka kerja (framework) ditampilkan pada Gambar 1.



Gambar 1. Framework

Step 1: Dataset bersumber pada data privat.

Step 2: Clustering dengan *K-Means Clustering* ditampilkan pada Gambar 5.2.



Gambar 2. Algoritma K-Means Clustering

Pada Gambar 2, tahap 1, inisialisasi k pusat klaster (centroid) secara acak. Pilih jumlah klaster k yang diinginkan. Nilai k random diambil dari dataset (data privat). Tahap 2, tempatkan setiap data atau obyek ke cluster terdekat. Kedekatan dua obyek ditentukan berdasar jarak. Jarak yang dipakai pada algoritma K-Means adalah *Euclidean distance* (d). Tahap 3, Hitung kembali pusat klaster dengan keanggotaan klaster yang sekarang. Pusat klaster adalah rata-rata (mean) dari semua data atau obyek dalam klaster tertentu. Tahap 4, hitung kembali (recalculate) sampai data stabil dan *convergence*. *Clustering* menggunakan *Euclidean distance* pada persamaan 1.

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \qquad (1)$$

d = jarak, x = atribut, x_1, x_2, ... x_n, y = atribut, y_1, y_2, ... y_n, dan n = jumlah atribut.

Step 3: Pengolahan data menggunakan Microsoft Excel 2010 (Excel) dengan formula ditampilkan pada Tabel 1 dan Tabel 2.

Tabel 1. Excel Function 1

IF Function	AVERAGE Function	COUNTIF Function
Formula: IF (logical_test, [value_if_true], [value_if_false])	Formula: AVERAGE (number1, [number2], ...)	Formula: COUNTIF (range, criteria)

Tabel 2. Excel Function 2

MIN Function	SUM Function
Formula: MIN(number1, [number2], ...)	Formula: SUM(number1,[number2],...])

Step 4: *Classification* menggunakan *Naïve Bayes Classifier* berdasarkan *Teorema Bayes* dengan: Probabilitas (B terhadap A) sama dengan Probabilitas (A dan B) dibanding Probabilitas (A) berdasarkan persamaan 2.

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (2)$$

Keterangan:
X :Data dengan kelas yang belum diketahui
H : Hipotesis data X merupakan suatu kelas spesifik
P(H|X) : Probabilitas hipotesis H berdasar kondisi X (posteriori probability)

P(H) : Probabilitas hipotesisH (prior probability)
P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H
P(X) : Probabilitas X

Proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, *Teorema Bayes* disesuaikan menjadi persamaan 3.

$$P(C|F1 \dots Fn) = \frac{P(C)P(F1 \dots Fn|C)}{P(F1 \dots Fn)}$$

(3)

Keterangan:
C : Representasi kelas
F1 : Representasi karakteristik petunjuk dalam melakukan klasifikasi
P : Peluang

Peluang masuknya sampel karakteristik tertentu dalam kelas C (Posterior) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, disebut Prior), dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas C (disebut Likelihood), dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global (disebut Evidence). Karena itu, rumus sebelumnya ditulis menjadi persamaan 4.

$$Posterior = \frac{Likelihood * Prior}{Evidence}$$

(4)

Nilai *Evidence* selalu tetap untuk setiap kelas pada satu sampel. Nilai dari *Posterior* akan dibandingkan dengan nilai *Posterior* kelas lainnya, untuk menentukan pada kelas apa suatu sampel akan diklasifikasikan.

Step 5: Prediksi menggunakan kombinasi *Clustering*, *Calculation*, dan *Classification*. Semua proses *clustering*, *calculation* dan *classification* dilakukan dengan Excel.

HASIL DAN PEMBAHASAN

Eksperimen menggunakan dataset (Badan Penelitian dan Pengembangan Kementerian Kesehatan Republik Indonesia, Riset Kesehatan Dasar 2013, hal. 208-236) yang ditampilkan pada Tabel 3.

Table 3. Fruit And Vegetable Consumption

No	Province	FC	VC
1	Aceh	0,5	1,0
2	Sumatera Utara	0,5	1,3
3	Sumatera Barat	0,4	0,7
4	Riau	0,5	1,0
5	Jambi	0,5	1,0
6	Sumatera Selatan	0,4	1,0
7	Bengkulu	0,4	1,2
8	Lampung	0,5	1,6
9	Bangka Belitung	0,5	0,9

10	Kepulauan Riau	0,5	1,0
11	DKI Jakarta	0,7	1,0
12	Jawa Barat	0,5	0,9
13	Jawa Tengah	0,5	1,5
14	DI Yogyakarta	0,7	1,8
15	Jawa Timur	0,5	1,4
16	Banten	0,5	1,0
17	Bali	0,5	1,4
18	Nusa Tenggara Barat	0,4	1,4
19	Nusa Tenggara Timur	0,3	1,7
20	Kalimantan Barat	0,4	1,3
21	Kalimantan Tengah	0,4	1,2
22	Kalimantan Selatan	0,4	0,9
23	Kalimantan Timur	0,5	1,4
24	Sulawesi Utara	0,5	1,2
25	Sulawesi Tengah	0,4	1,2
26	Sulawesi Selatan	0,4	1,1
27	Sulawesi Tenggara	0,5	1,2
28	Gorontalo	0,5	1,0
29	Sulawesi Barat	0,3	1,1
30	Maluku	0,5	1,3
31	Maluku Utara	0,4	1,1
32	Papua Barat	0,4	1,5
33	Papua	0,4	1,5

1.1 *Step 1 (Dataset)*

Pada Tabel 3, rata-rata konsumsi buah (Fruit Consumption = FC) dan sayur (Vegetable Consumption = VC), jumlah porsi per hari dalam seminggu, penduduk umur ≥ 10 tahun ke atas, pada provinsi di Indonesia tahun 2013 adalah konsumsi buah atau sayur setiap hari tanpa memperhitungkan jumlah porsi. Selanjutnya proses *clustering*.

Tabel 4. Centroid 1

k	x	y
Data-3 as Centroid Cluster 1	0,4	0,7
Data-25 as Centroid Cluster 2	0,4	1,2
Data-33 as Centroid Cluster 3	0,4	1,5

1.2 *Step 2 (Clustering)*

Pada Tabel 4, pusat klaster (Centroid) ditentukan nilai k secara random dengan k =3, ada 3 klaster yang dibentuk. Salah satunya dengan nilai atribut x yang sama yaitu data ke-3 (Prov. Sumatera Barat, buah (x) = 0,4 dan sayur (y) = 0,7), data ke-25 (Prov. Sulawesi Tengah, buah (x) = 0,4 dan sayur (y) = 1,2), dan data ke-33 (Prov. Papua, buah (x) = 0,4 dan sayur (y) = 1,5). Selanjutnya menghitung jarak klaster paling dekat.

Tabel 5. Iteration 1

No	C1	C2	C3	d
1	0,3162	0,2236	0,5099	0,2236
2	0,6083	0,1414	0,2236	0,1414
3	0,0000	0,5000	0,8000	0,0000
4	0,3162	0,2236	0,5099	0,2236
5	0,3162	0,2236	0,5099	0,2236
6	0,3000	0,2000	0,5000	0,2000
7	0,5000	0,0000	0,3000	0,0000
8	0,9055	0,4123	0,1414	0,1414
9	0,2236	0,3162	0,6083	0,2236
10	0,3162	0,2236	0,5099	0,2236
11	0,4243	0,3606	0,5831	0,3606
12	0,2236	0,3162	0,6083	0,2236
13	0,8062	0,3162	0,1000	0,1000
14	1,1402	0,6708	0,4243	0,4243
15	0,7071	0,2236	0,1414	0,1414
16	0,3162	0,2236	0,5099	0,2236
17	0,7071	0,2236	0,1414	0,1414
18	0,7000	0,2000	0,1000	0,1000
19	1,0050	0,5099	0,2236	0,2236
20	0,6000	0,1000	0,2000	0,1000
21	0,5000	0,0000	0,3000	0,0000
22	0,2000	0,3000	0,6000	0,2000
23	0,7071	0,2236	0,1414	0,1414
24	0,5099	0,1000	0,3162	0,1000
25	0,5000	0,0000	0,3000	0,0000
26	0,4000	0,1000	0,4000	0,1000
27	0,5099	0,1000	0,3162	0,1000
28	0,3162	0,2236	0,5099	0,2236
29	0,4123	0,1414	0,4123	0,1414
30	0,6083	0,1414	0,2236	0,1414
31	0,4000	0,1000	0,4000	0,1000
32	0,8000	0,3000	0,0000	0,0000
33	0,8000	0,3000	0,0000	0,0000

Formula Excel:

d_{11} = (((D11-F4)^2) + ((E11 G4)^2))^0,5

= (((0,5-04)^2) + ((1,0-0,7)^2))^0,5

= 0,3162

d_{12} = (((D11-F5)^2) + ((E11-G5)^2))^0,5

= (((0,5-0,4)^2) + ((1,0-1,2)^2))^0,5

= 0,2236

d_{13} = (((D11-F6)^2) + ((E11-G6)^2))^0,5

= (((0,5-0,4)^2) + ((1,0-1,5)^2))^0,5

= 0,5099

Pada Tabel 5, pada iterasi 1, Cluster 1 (C1), Cluster 2 (C2), Cluster 3 (C3) dan *distance* (*d*), Iterasi 1, jarak paling dekat (minimum) dengan nilai (d_{11} = 0,3162), (d_{12} = 0,2236) dan (d_{13} = 0,5099). Pilih jarak minimum (distance) yaitu 0,2236. Hal yang sama dilakukan pada Excel dengan formula = MIN (N10 : P10) = 0,2236. *Centroid* (C) ditandai *Italic Font*. *Distance* (d) ditandai *Italic Font*. Selanjutnya pengelompokkan pusat klaster.

Tabel 6. Centroid Cluster 1

No	C1	C2	C3	No	C1	C2	C3	No
1		*		12	*			23
2		*		13			*	24
3	*			14			*	25
4		*		15			*	26
5		*		16		*		27
6		*		17			*	28
7		*		18			*	29
8			*	19			*	30
9	*			20		*		31
10		*		21		*		32
11		*		22	*			33

Pada Tabel 5.6, jarak paling dekat ditandai dengan *Asterisk Symbol*. Iterasi 1 menghasilkan 3 Cluster (C1, C2, C3). Cluster 1 (3, 9, 12, 22). Cluster 2 (1, 2, 4, 5, 6, 7, 10, 11, 16, 20, 21, 24, 25, 26, 27, 28, 29, 30, 31). Cluster 3 (8, 13, 14, 15, 17, 18, 19, 23, 32, 33). Selanjutnya pengelompokkan data tiap klaster.

Tabel 7 (a). Cluster Group

No	FC	VC	C1	C2	C3
1	0,5	1,0		*	
2	0,5	1,3		*	
3	0,4	0,7	*		
4	0,5	1,0		*	
5	0,5	1,0		*	
6	0,4	1,0		*	
7	0,4	1,2		*	
8	0,5	1,6			*
9	0,5	0,9	*		
10	0,5	1,0		*	
11	0,7	1,0		*	
12	0,5	0,9	*		
13	0,5	1,5			*
14	0,7	1,8			*
15	0,5	1,4			*
16	0,5	1,0		*	
17	0,5	1,4			*
18	0,4	1,4			*
19	0,3	1,7			*
20	0,4	1,3		*	
21	0,4	1,2		*	
22	0,4	0,9	*		
23	0,5	1,4			*
24	0,5	1,2		*	
25	0,4	1,2		*	

Tabel 7 (b). Cluster Group

No	FC	VC	C1	C2	C3
26	0,4	1,1		*	
27	0,5	1,2		*	
28	0,5	1,0		*	
29	0,3	1,1		*	
30	0,5	1,3		*	
31	0,4	1,1		*	
32	0,4	1,5			*
33	0,4	1,5			*
Total Data / Count Data in Cluster 1					0,4500
Total Data / Count Data in Cluster 1					0,8500
Total Data / Count Data in Cluster 2					0,4632
Total Data / Count Data in Cluster 2					1,1158
Total Data / Count Data in Cluster 3					0,4700
Total Data / Count Data in Cluster 3					1,5200

Pada Tabel 7, dilakukan penjumlahan data Fruit Consumption (FC) pada Cluster 1, dibagi banyak data FC tiap Cluster 1. Excel dengan Formula = SUM (AG12 + AG18 + AG21 + AG31) / 4 = 0,4500. Penjumlahan data Vegetable Consumption (VC) pada Cluster 1, dibagi banyak data VC tiap Cluster 1. Excel dengan Formula = SUM (AH12 + AH18 + AH21 + AH31) / 4 = 0,8500. Hal yang sama dilakukan pada Cluster 2 dan Cluster 3. Selanjutnya menentukan pusat klaster baru.

Tabel 8. Centroid 2

k	x	y
New Centroid Cluster 1	0,4500	0,8500
New Centroid Cluster 2	0,4632	1,1158
New Centroid Cluster 3	0,4700	1,5200

Pada Tabel 5.8, ditentukan nilai k = 3 berdasarkan Cluster Group yang baru. Nilai atribut x (x_1 = 0,4500, x_2 = 0,4632, x_3 = 0,4700) dan y (y_1 = 0,8500, y_2 = 1,1158, y_3 = 1,5200). Centroid 2 dijadikan pusat klaster pada proses

iterasi 2, untuk menghasilkan pusat klaster baru. Selanjutnya melakukan iterasi 2.

Tabel 9 (a). Iteration 2

No	C1	C2	C3	d
1	0,1581	0,1215	0,5209	0,1215
2	0,4528	0,1878	0,2220	0,1878
3	0,1581	0,4206	0,8230	0,1581
4	0,1581	0,1215	0,5209	0,1215
5	0,1581	0,1215	0,5209	0,1215
6	0,1581	0,1319	0,5247	0,1319
7	0,3536	0,1053	0,3276	0,1053
8	0,7517	0,4856	0,0854	0,0854
9	0,0707	0,2189	0,6207	0,0707
10	0,1581	0,1215	0,5209	0,1215
11	0,2915	0,2636	0,5686	0,2636
12	0,0707	0,2189	0,6207	0,0707

Tabel 9 (b). Iteration 2

No	C1	C2	C3	d
13	0,6519	0,3860	0,0361	0,0361
14	0,9823	0,7240	0,3624	0,3624
15	0,5523	0,2866	0,1237	0,1237
16	0,1581	0,1215	0,5209	0,1215
17	0,5523	0,2866	0,1237	0,1237
18	0,5523	0,2911	0,1389	0,1389
19	0,8631	0,6066	0,2476	0,2476
20	0,4528	0,1947	0,2309	0,1947
21	0,3536	0,1053	0,3276	0,1053
22	0,0707	0,2249	0,6239	0,0707
23	0,5523	0,2866	0,1237	0,1237
24	0,3536	0,0919	0,3214	0,0919
25	0,3536	0,1053	0,3276	0,1053
26	0,2550	0,0651	0,4258	0,0651
27	0,3536	0,0919	0,3214	0,0919
28	0,1581	0,1215	0,5209	0,1215
29	0,2915	0,1640	0,4531	0,1640
30	0,4528	0,1878	0,2220	0,1878
31	0,2550	0,0651	0,4258	0,0651
32	0,6519	0,3894	0,0728	0,0728
33	0,6519	0,3894	0,0728	0,0728

Tabel 10. Comparation of Iteration 1and Iteration 2

Iteration 1				Iteration 2			
No	C1	C2	C3	No	C1	C2	C3
1	0,3162	0,2236	0,5099	1	0,1581	0,1215	0,5209
2	0,6083	0,1414	0,2236	2	0,4528	0,1878	0,2220
3	0,0000	0,5000	0,8000	3	0,1581	0,4206	0,8230
4	0,3162	0,2236	0,5099	4	0,1581	0,1215	0,5209
5	0,3162	0,2236	0,5099	5	0,1581	0,1215	0,5209
6	0,3000	0,2000	0,5000	6	0,1581	0,1319	0,5247
7	0,5000	0,0000	0,3000	7	0,3536	0,1053	0,3276
8	0,9055	0,4123	0,1414	8	0,7517	0,4856	0,0854
9	0,2236	0,3162	0,6083	9	0,0707	0,2189	0,6207
10	0,3162	0,2236	0,5099	10	0,1581	0,1215	0,5209
11	0,4243	0,3606	0,5831	11	0,2915	0,2636	0,5686
12	0,2236	0,3162	0,6083	12	0,2236	0,3162	0,6083
13	0,8062	0,3162	0,1000	13	0,8062	0,3162	0,1000
14	1,1402	0,6708	0,4243	14	1,1402	0,6708	0,4243
15	0,7071	0,2236	0,1414	15	0,7071	0,2236	0,1414
16	0,3162	0,2236	0,5099	16	0,3162	0,2236	0,5099
17	0,7071	0,2236	0,1414	17	0,7071	0,2236	0,1414
18	0,7000	0,2000	0,1000	18	0,7000	0,2000	0,1000
19	1,0050	0,5099	0,2236	19	1,0050	0,5099	0,2236
20	0,6000	0,1000	0,2000	20	0,6000	0,1000	0,2000
21	0,5000	0,0000	0,3000	21	0,5000	0,0000	0,3000
22	0,2000	0,3000	0,6000	22	0,2000	0,3000	0,6000
23	0,7071	0,2236	0,1414	23	0,5523	0,2866	0,1237
24	0,5099	0,1000	0,3162	24	0,3536	0,0919	0,3214
25	0,5000	0,0000	0,3000	25	0,3536	0,1053	0,3276
26	0,4000	0,1000	0,4000	26	0,2550	0,0651	0,4258
27	0,5099	0,1000	0,3162	27	0,3536	0,0919	0,3214
28	0,3162	0,2236	0,5099	28	0,1581	0,1215	0,5209
29	0,4123	0,1414	0,4123	29	0,2915	0,1640	0,4531
30	0,6083	0,1414	0,2236	30	0,4528	0,1878	0,2220
31	0,4000	0,1000	0,4000	31	0,2550	0,0651	0,4258
32	0,8000	0,3000	0,0000	32	0,6519	0,3894	0,0728
33	0,8000	0,3000	0,0000	33	0,6519	0,3894	0,0728

Pada Tabel 9, pada iterasi 2 data, menghasilkan pusat klaster yang sama dengan iterasi 1. Iterasi 2 dilakukan untuk perbandingan setiap iterasi dengan pegelompokkan kembali (re-grouping) untuk menentukan pusat klaster yang konsisten dan data konvergen. Data konvergen artinya pemusatan data pada satu titik (jarak terdekat). *Distance* (d) ditandai dengan *Italic Font*. Pusat klaster (C) ditandai dengan *Italic Font*. Selanjutnya melakukan perbandingan iterasi 1 dan iterasi 2 berdasarkan proses *Clustering*.

Pada Tabel 10, pusat klaster baru pada iterasi 2 sama dengan pusat klaster sebelumnya (iterasi 1). Pusat klaster (C) pada klaster 1 (C1), klaster 2 (C2), klaster 3 (C3) ditandai dengan *Italic Font*. Hal ini menunjukkan data *convergence*, dengan *distance* yang memusat pada titik yang sama, maka proses *clustering* selesai. Selanjutnya melakukan proses *priority calculation*.

Tabel 11. Priority

No	FC	VC	FP	VP
1	0,5	1,0	No	Yes
2	0,5	1,3	No	No
3	0,4	0,7	Yes	Yes
4	0,5	1,0	No	Yes
5	0,5	1,0	No	Yes
6	0,4	1,0	Yes	Yes
7	0,4	1,2	Yes	No
8	0,5	1,6	No	No
9	0,5	0,9	No	Yes
10	0,5	1,0	No	Yes
11	0,7	1,0	No	Yes
12	0,5	0,9	No	Yes
13	0,5	1,5	No	No
14	0,7	1,8	No	No

15	0,5	1,4	No	No
16	0,5	1,0	No	Yes
17	0,5	1,4	No	No
18	0,4	1,4	Yes	No
19	0,3	1,7	Yes	No
20	0,4	1,3	Yes	No
21	0,4	1,2	Yes	No
22	0,4	0,9	Yes	Yes
23	0,5	1,4	No	No
24	0,5	1,2	No	No
25	0,4	1,2	Yes	No
26	0,4	1,1	Yes	Yes
27	0,5	1,2	No	No
28	0,5	1,0	No	Yes
29	0,3	1,1	Yes	Yes
30	0,5	1,3	No	No
31	0,4	1,1	Yes	Yes
32	0,4	1,5	Yes	No
33	0,4	1,5	Yes	No
ANFC				0,5
ANVC				1,2

1.3 *Step 3 (Calculation)*

Pada Tabel 11, Average Fruit Consumption National (AFCN), formula Excel=AVERAGE(E7:E40)=0,5. Average Vegetable Consumption National (AVCN), formula Excel=AVERAGE(F7:F40)=1,2. Bila Fruit Consumption (FC) di bawah AFCN, maka Fruit Priority (FP) Yes. Bila FP di atas AFCN, maka FP No, formula Excel=IF(E7<0,5;"Yes";"No"). Hal yang sama dilakukan terhadap Vegetable Consumption (VP), bila VC di bawah AVCN, maka VP Yes. Bila VP di atas AVCN, maka VP No, formula Excel=IF(F7<1,2;"Yes";"No"). FP ditandai dengan *Red Font*. VP ditandai dengan *Blue Font*. Selanjutnya proses *Calculation 1*.

Tabel 12 (a). Calculation 1

No	FP	VP	C	No	FP	VP	C
1	No	Yes	C2	12	No	Yes	C1
2	No	No	C2	13	No	No	C3
3	Yes	Yes	C1	14	No	No	C3
4	No	Yes	C2	15	No	No	C3
5	No	Yes	C2	16	No	Yes	C2
6	Yes	Yes	C2	17	No	No	C3
7	Yes	No	C2	18	Yes	No	C3
8	No	No	C3	19	Yes	No	C3
9	No	Yes	C1	20	Yes	No	C2
10	No	Yes	C2	21	Yes	No	C2
11	No	Yes	C2	22	Yes	Yes	C1

Tabel 12 (b). Calculation 1

No	FP	VP	C
23	No	No	C3
24	No	No	C2
25	Yes	No	C2
26	Yes	Yes	C2
27	No	No	C2
28	No	Yes	C2
29	Yes	Yes	C2
30	No	No	C2
31	Yes	Yes	C2
32	Yes	No	C3
33	Yes	No	C3

Pada Tabel 12, kelompokkan klaster (C) berdasarkan Fruit Priority (FP) dan Vegetable Priority (VP), poin data-1 ada pada Cluster 2 (C2), poin data-2 ada pada Cluster 2 (C2), poin data-3 ada pada Cluster 1 (C1), poin data-4 ada pada Cluster 2 (C2), poin data-5 ada pada Cluster 2 (C2), dan seterusnya, sampai pada poin data-33 ada pada Cluster 3 (C3). Poin data tersebut (C) ditandai dengan *Italic Font.* Selanjutnya *Calculation 2*.

Tabel 13 (a). Calculation 2

No	C	FP	VP	FCC	VCC
3	C1	Yes	Yes	Good	Good
9	C1	No	Yes	Bad	Good
12	C1	No	Yes	Bad	Good
22	C1	Yes	Yes	Good	Good
1	C2	No	Yes	Bad	Good
2	C2	No	No	Bad	Bad
4	C2	No	Yes	Bad	Good
5	C2	No	Yes	Bad	Good
6	C2	Yes	Yes	Good	Good
7	C2	Yes	No	Good	Bad
10	C2	No	Yes	Bad	Good
11	C2	No	Yes	Bad	Good
16	C2	No	Yes	Bad	Good
20	C2	Yes	No	Good	Bad
21	C2	Yes	No	Good	Bad
24	C2	No	No	Bad	Bad

25	C2	Yes	No	Good	Bad
26	C2	Yes	Yes	Good	Good
27	C2	No	No	Bad	Bad
28	C2	No	Yes	Bad	Good
29	C2	Yes	Yes	Good	Good
30	C2	No	No	Bad	Bad
31	C2	Yes	Yes	Good	Good
8	C3	No	No	Bad	Bad
13	C3	No	No	Bad	Bad
14	C3	No	No	Bad	Bad
15	C3	No	No	Bad	Bad
17	C3	No	No	Bad	Bad
18	C3	Yes	No	Good	Bad
19	C3	Yes	No	Good	Bad
23	C3	No	No	Bad	Bad
32	C3	Yes	No	Good	Bad
33	C3	Yes	No	Good	Bad

Pada Tabel 13, class disusun berdasarkan urutan mulai Cluster 1 (C1), Cluster 2 (C2) dan Cluster 3 (C3). Perbandingan tiap anggota Cluster (C) pada Fruit Priority (FP). Bila FP Yes, maka Fruit Consumption Class (FCC) Good. Bila FP No, maka FCC Bad. Excel dengan formula = IF (T7 = "Yes" ; "Good" ; "Bad"). Hal yang sama juga dilakukan pada Vegetable Priority (VP). Bila VP Yes, maka Vegetable Consumption Class (VCC) Good. Bila VP No, maka VCC Bad. Excel dengan formula = IF (U7 = "Yes" ; "Good" ; "Bad"). FCC dan VCC ditandai dengan *Italic Font*. Selanjutnya *Calculation 3*.

Tabel 14 (a). Calculation 3

No	Cluster	Class	No	Cluster	Class
3	Cluster 1	Good	19	Cluster 3	Good
22	Cluster 1	Good	32	Cluster 3	Good
6	Cluster 2	Good	33	Cluster 3	Good
7	Cluster 2	Good	9	Cluster 1	Bad
20	Cluster 2	Good	12	Cluster 1	Bad
21	Cluster 2	Good	1	Cluster 2	Bad
25	Cluster 2	Good	2	Cluster 2	Bad
26	Cluster 2	Good	4	Cluster 2	Bad
29	Cluster 2	Good	5	Cluster 2	Bad
31	Cluster 2	Good	10	Cluster 2	Bad
18	Cluster 3	Good	11	Cluster 2	Bad

Tabel 14 (b). Calculation 3

No	Cluster	Class
16	Cluster 2	Bad
24	Cluster 2	Bad
27	Cluster 2	Bad
28	Cluster 2	Bad
30	Cluster 2	Bad
8	Cluster 2	Bad
13	Cluster 3	Bad
14	Cluster 3	Bad
15	Cluster 3	Bad
17	Cluster 3	Bad
23	Cluster 3	Bad

Pada Tabel 14, Fruit Class dikelompokkan dan diurutkan mulai dari Class Good sampai Class Bad. Data ke-n (mulai dari poin data-33, data-22, data-6, sampai poin data-23). Poin data diurutkan berdasarkan Class Good (sebanyak 14 poin data) dan Class Bad (sebanyak 19 poin data). Poin data ke-n (No) ditandai dengan *Italic Font*. Urutan Class (Good, Bad) ditandai dengan *Italic Font*. Langkah selanjutnya *Calculation 4*.

Tabel 15 (a). Calculation 4

No	Cluster	Class	No	Cluster	Class
3	Cluster 1	Good	26	Cluster 2	Good
9	Cluster 1	Good	28	Cluster 2	Good
12	Cluster 1	Good	29	Cluster 2	Good
22	Cluster 1	Good	31	Cluster 2	Good
1	Cluster 2	Good	2	Cluster 2	Bad
4	Cluster 2	Good	7	Cluster 2	Bad
5	Cluster 2	Good	20	Cluster 2	Bad
6	Cluster 2	Good	21	Cluster 2	Bad
10	Cluster 2	Good	24	Cluster 2	Bad
11	Cluster 2	Good	25	Cluster 2	Bad
16	Cluster 2	Good	27	Cluster 2	Bad

Tabel 15 (b). Calculation 4

No	Cluster	Class
30	Cluster 2	Bad
8	Cluster 3	Bad

13	Cluster 3	Bad
14	Cluster 3	Bad
15	Cluster 3	Bad
17	Cluster 3	Bad
18	Cluster 3	Bad
19	Cluster 3	Bad
23	Cluster 3	Bad
32	Cluster 3	Bad
33	Cluster 3	Bad

Pada Tabel 15, Vegetable Class dikelompokkan dan diurutkan dari Class Good sampai Class Bad. Caranya sama dengan Fruit Class. Data ke-n (No) ditandai dengan *Italic Font*. Urutan Class (Good, Bad) ditandai dengan *Italic Font*. Selanjutnya *Classification 1*.

Tabel 16. Classification 1

P(V=↓| ...	...| C=Good)	...| C=Bad)	%
Cluster 1	14,29%	10,53%	-
Cluster 2	57,14%	63,16%	-
Cluster 3	28,57%	26,32%	-
%	100%	100%	-
P(Good/Bad)	42,42%	57,58%	100%

1.4 *Step 4 (Classification)*

Pada Tabel 16, melakukan klasifikasi untuk Fruit Consumption dengan menghitung Probability Class Good tiap anggota Cluster 1 terhadap Probabilitas Cluster 1 dari semua Cluster, lalu dikalikan 100%. Probability Class Bad tiap anggota Cluster 1 terhadap Probability Cluster 1 dari semua Cluster, lalu dikalikan 100%. Hal yang sama juga dilakukan pada Cluster 1 dan Cluster 3. Klasifikasi pada Cluster 1 untuk Class Good dengan Excel formula = COUNTIF (Z7 : Z21 ; "Cluster 1") / 14 * 100% = 14,29%. Klasifikasi pada Cluster 2 untuk Class Good dengan Excel formula = COUNTIF (Z7 : Z21 ; "Cluster 2") / 14 * 100% = 57,14%. Klasifikasi pada Cluster 3 untuk Class Good dengan Excel formula = COUNTIF (Z7 : Z21 ; "Cluster 3") / 14 * 100% = 28,57%. Klasifikasi pada Cluster 1 untuk Class Bad dengan Excel formula = COUNTIF (Z22 : Z40 ; "Cluster 1") / 19 * 100% = 10,53%. Klasifikasi pada Cluster 2 untuk Class Bad dengan Excel formula = COUNTIF (Z22 : Z40 ; "Cluster 2 ") / 19 * 100% = 63,16%. Klasifikasi pada Cluster 3 untuk Class Bad dengan Excel formula = COUNTIF (Z22 : Z40 ; "Cluster 3") / 19 * 100% = 26,32%. Lalu, hitung Probability Class Good Total untuk Fruit Consumption terhadap semua Cluster. Excel dengan formula = COUNTIF (AA7 : AA40 ; "Good") / 33 * 100% = 42,42%. Lakukan hal yang sama terhadap Probability Class Bad Total Fruit Consumption terhadap semua Cluster. Excel dengan formula = COUNTIF (AA7 : AA40 ; "Bad") / 33 * 100% = 57,58%. Langkah selanjutnya *Classification 2*.

Tabel 17. Classification 2

P(V=↓| ...	...| C=Good)	...| C=Bad)	%
Cluster 1	26,67%	0,00%	-
Cluster 2	73,33%	44,44%	-
Cluster 3	0,00%	55,56%	-
%	100%	100%	-
P(Good/Bad)	45,45%	54,55%	100%

Pada Tabel 17, melakukan klasifikasi untuk Vegetable Consumption dengan menghitung Probability Class Good tiap anggota Cluster 1 terhadap Probability Cluster 1 dari semua Cluster, lalu dikalikan 100%. Probability Class Bad tiap anggota Cluster 1 terhadap Probability Cluster 1 dari semua Cluster, lalu dikalikan 100%. Hal yang sama juga dilakukan pada Cluster 1 dan Cluster 3. Klasifikasi pada Cluster 1 untuk Class Good dengan Excel formula = COUNTIF (AD7 : AD21 ; "Cluster 1") / 15 * 100% = 26,67%. Klasifikasi pada Cluster 2 untuk Class Good dengan Excel formula = COUNTIF(AD7 : AD21 ; "Cluster 2") / 15 * 100% = 73,33%. Klasifikasi pada Cluster 3 untuk Class Good dengan Excel formula = COUNTIF (AD7 : AD21 ; "Cluster 3") / 15 * 100% = 0,00%.

Tabel 18 (a). Prediction

No	Cluster 1	No	Cluster 2	No	Cluster 3
1	Sumatera	1	Aceh	1	Lampung

	Barat				
2	Bangka Belitung	2	Sumatera Utara	2	Jawa Tengah
3	Jawa Barat	3	Riau	3	DI Yogyakarta
4	Kalimantan Selatan	4	Jambi	4	Jawa Timur

Tabel 18 (b). Prediction

No	Cluster 1	No	Cluster 2	No	Cluster 3
		5	Sumatera Selatan	5	Bali
		6	Bengkulu	6	Nusa Tenggara Barat
		7	Kepulauan Riau	7	Nusa Tenggara Timur
		8	DKI Jakarta	8	Kalimantan Timur
		9	Banten	9	Papua Barat
		10	Kalimantan Barat	10	Papua
		11	Kalimantan Tengah		
		12	Sulawesi Utara		
		13	Sulawesi Tengah		
		14	Sulawesi Selatan		
		15	Sulawesi Tenggara		
		16	Gorontalo		
		17	Sulawesi Barat		
		18	Maluku		
		19	Maluku Utara		
FCCG = 14,29%		FCCG = 57,14%		FCCG = 28,57%	
FCCB = 10,53%		FCCB = 63,16%		FCCB = 26,32%	
TFCCG = 42,42%					
TFCCB = 57,58%					
VCCG = 26,67%		VCCG = 73,33%		VCCG = 0,00%	
VCCB = 0,00%		VCCB = 44,44%		VCCB = 55,56%	
TVCCG = 45,45%					
TVCCB = 54,55%					

Klasifikasi pada Cluster 1 untuk Class Bad dengan Excel formula = COUNTIF (AD22 : AD39 ; "Cluster 1") / 18 * 100% = 0,00%. Klasifikasi pada Cluster 2 untuk Class Bad dengan Excel formula =COUNTIF (AD22 : AD39 ; "Cluster 2") / 18 * 100% = 44,44%. Klasifikasi pada Cluster 3 untuk Class Bad dengan Excel formula = COUNTIF (AD22 : AD39 ; "Cluster 3") / 18 * 100% = 55,56%. Lalu, hitung Probability Class Good Total untuk Vegetable Consumption terhadap semua Cluster. Excel dengan formula = COUNTIF (AE7 : AE39 ; "Good") / 33 * 100% = 45,45%. Lakukan hal yang sama pada Probability Class Bad Total untuk Vegetable Consumption terhadap semua Cluster. Excel dengan formula = COUNTIF (AE7 : AE39 ; "Bad") / 33 * 100% = 54,55%. Langkah selanjutnya melakukan *Prediction*.

1.5 *Step 5 (Prediction)*

Pada Tabel 18, Cluster 1, Cluster 2 dan Cluster 3 menunjukkan klasifikasi Fruit Consumption berurutan dengan Class Good (14,29% di atas rata-rata, 57,14% di atas rata-rata, 28,57% di atas rata-rata) dan Class Bad (10,53% di bawah rata-rata, 63,16% di bawah rata-rata, 26,32% di bawah rata-rata).

Total Fruit Good artinya konsumsi buah di atas rata-rata sebesar 42,42%. Total Fruit Bad artinya konsumsi buah di bawah rata-rata sebesar 57,58%. Total Vegetable Good artinya konsumsi sayur di atas rata-rata sebesar 45,45%. Total Vegetable Bad artinya konsumsi sayur di bawah rata-rata sebesar 54,55%.

KESIMPULAN

Clustering menghasilkan output Cluster 1 dengan 4 *object*, Cluster 2 dengan 19 *object*, dan Cluster 3 dengan *10 object*. Calculation menghasilkan output *Priority Yes* dan *Priority No*. *Priority Yes* artinya konsumsi buah dan sayur dengan *object di atas rata-rata*. *Priority No* artinya konsumsi buah dan sayur dengan *object di bawah rata-rata.* Classification menghasilkan output *Class Good*dan *Class Bad*. *Class Good* artinya *kelas di atas rata-rata*. *Class Bad* artinya *kelas di bawah rata-rata*. Prediksi konsumsi buah dan sayur menggunakan kombinasi *K-Means Clustering, Excel Calculation,* dan *Naïve Bayes Classifier* dengan output konsumsi buah sebesar 42,42%

untuk kelas di atas rata-rata dan konsumsi sayur sebesar 57,58% untuk kelas di bawah rata-rata. Perbandingan antara *Class Good* dan *Class Bad* menjadi indikasi bahwa *konsumsi buah dan sayur "*mayoritas di bawah rata-rata". Hasil penelitian dijadikan sebagai sumber data bagi kebijakan pangan dan gizi nasional.

DAFTAR KEPUSTAKAAN

S. M. Perdana, Hardinsyah, dan E. Damayanthi, (2013). *Alternative of balanced diet index to assess nutritional quality of diet in indonesian adult females,* Journal of Nutrient & Food, vol. 9, no. 1, pp. 43-50.

A. A. Candra, B. Setiawan, dan M. R. M. Damanik, (2013). *The effect of snack feeding, nutrition education, and iron suplementation to nutritional status, nutrition knowledge, and anemia status in elementary school students,* Journal of Nutrient & Food, vol. 8, no. 2, pp. 103-108.

M. Fasitasari, (2013). *Nutrition therapy in elderly with chronic obstructive pulmonary diseas,* Sains Medika Journal, vol. 5, no. 1, pp. 50-61.

X. Wang, Y. Ouyang, J. Liu, M. Zhu, G. Zhao, W. Bao, dan F. B. Hu, (2014). *Fruit and vegetable consumption and mortality from all causes, cardiovascular disease, and cancer: systematic review and dose-response meta-analysis of prospective cohort studies,* Bio. Med. Journal, pp. 1-14.

O. Stackelberg, M. Björck, S. C. Larsson, N. Orsini, dan A. Wolk, (2013), *Fruit and vegetable consumption with risk of abdominal aortic aneurysm,* Circulation: Journal of the American Heart Association, vol. 128, pp. 795-802.

T. S. Conner, K. L. Brookie, A. C. Richardson, dan M. A. Polak, (2014). *On carrots and curiosity: eating fruit and vegetable is associated with greather flourishing in daily life,* British Journal of Health Psychology, pp. 1-31.

S. Shukla dan S. Naganna, (2014). *A review on k-means data clustering approach,* Int. Journal of Information & Computation Technology, vol. 4, no. 17, pp. 1847-1860.

G. Liu, S. Huang, C. Lu, dan Y. Du, (2014). *An imporved k-means algorithm based on association rules,* Int. Journal of Computer Theory and Engineering, vol. 6, no. 2, pp. 146-149.

Y. Kumar dan G. Sahoo, (2014). *A new initialization method to originate initial cluster centers for k means algortihm,* Int. Journal of Advanced Science and Technology, vol. 62, pp. 43-54.

K. K. Manjusha, K. Sankaranarayanan dan P. Seena, (2014). *Prediction of different dermatological conditions using naïve bayesian classification,* Int. Journal of Advanced Research in Computer Science and Software Engineering, vol. 4, no. 1, pp. 864-868,

Bustami, (2014). *Implementing naïve bayes algorithm for classification customer data insurance,* Informatic Journal, vol. 8, no. 1, pp. 1-15.

V. Shukla dan S. Vashishtha, (2014). *New hybrid intrusion detection system based on data mining technique to enhanced performance,* Int. Journal of Computer Science and Information Security, vol. 12, no. 6, pp. 14-19.

W. Yassin, N. I. Udzir, Z. Muda, dan M. N. Sulaiman**,** (2013). *Anomaly-based intrusion detection through k-means clustering and naïves bayes classification,* Proceedings of the 4th International Conference on Computing and Informatics, 28-30 August, Sarawak, Malaysia, pp. 298-303.

M. Banerjee dan R. Soni, (2013). *Design and implementation of network intrusion detection system by using k-means clustering and naïve bayes,* Int. Journal of Science, Engineering and Technology Research, vol. 2, no. 3, pp. 756-760.

Y. Emami, M. Ahmadzadeh, M. Salehi dan S. Homayoun, (2014). *Efficient intrusion detection using weighted k-means clustering and naïve bayes classification,* Journal of Emerging Trends in Computing and Information Sciences, vol. 5, no. 8 pp. 620-623.

N. O. F. Elssied dan O. Ibrahim, (2014). *K-means clustering scheme for enhanced*

spam detection, Research Journal of Applied Sciences, Engineering and Technology, vol.7, no.10, pp. 940-1952.

N. Sharma dan S. Niranjan, (2013). *Performance enhancement using combinatorial approach of classification and clustering in machine learning,* Int. Journal of Application or Innovation in Engineering & Management, vol. 2, no. 4 pp. 71-78.