

BAB II
KAJIAN TEORI

2.1 Penelitian Terdahulu

Penelitian terdahulu adalah sebuah kumpulan penelitian yang akan dijadikan bahan referensi dan pertimbangan yang diharapkan bisa membantu dalam penyelesaian pembuatan sistem yang baru. Adapun penelitian sebelumnya tercantum dalam tabel seperti di bawah ini:

Tabel 2. 1 Penelitian Terdahulu

No	Judul	Penulis (Tahun)	Metode	Hasil penelitian
1.	Penerapan <i>Data Mining</i> Pemilihan Siswa Kelas Unggulan dengan <i>Metode K- Means</i> <i>Clustering</i> di SMP N 02 Tasikmadu [4]	Teguh wibowo (2018)	<i>K-Means</i>	<i>K-Mmeans</i> <i>Clustering</i> memudahkan dalam penyeleksian dan pengelompokan siswa kelas unggulan sesuai karakteristiknya dengan hasil cluster ke-1 didominasi kelas 7A dan 7D, Cluster ke-2 didominasi kelas 7A, 7B dan 7F, Cluster ke-3 didominasi kelas 7C dan 7D.

2.	Penerapan Metode K-Means Clustering dalam menentukan predikat kelulusan mahasiswa untuk [5] menganalisa kualitas lulusan	Venny Novita Sari, Yupianti, Dewi Maharani (2018)	K-Means	K-Means Clustering adalah metode penerapan yang baik dalam menganalisa mengelompokkan kualitas kelulusan mahasiswa yang berpotensi pada program studi yang di minati dengan bantuan Tools Rapid Miner.
3.	Aplikasi pembentukan kelas dengan K-Means Clustering sebagai alat bantu pemilihan siswa kelas unggulan di man 3 kediri [6]	Amri Muhammad (2016)	K-Means	Metode K-Means dapat digunakan untuk pengelompokan dengan hasil yang cepat dan akurat sesuai kebutuhan sekolah.
4.	Penentuan Tingkat Minat Belanja	Anita, M.Kom, Yulya	Clustering K-Means	Metode Clustering K-Means digunakan untuk membantu

Online Melalui Media Sosial Menggunakan Metode	Muharmi, M.Kom, Sri Tria Siska, M.Kom (2018)		tingkat minat belanja online berdasarkan tingkat kesamaannya dengan bantuan <i>Tools Rapid Miner</i> .
<i>Clustering K-Means</i> [7]			
5. Sistem pendukung keputusan menggunakan Metode	Restu Akbar Nugraha, Defri Kurniawan (2017)	Metode pengembangan sistem <i>waterfall</i>	Suatu sistem pengambil keputusan untuk menilai kinerja guru dengan menggunakan kriteria dalam pengumpulan data untuk menghasilkan keputusan cepat dan efektif.
<i>Algoritma K-Means</i> untuk penilaian kinerja guru (PKG) Di smkn 1 demak [8]			

Berdasarkan penelitian diatas, maka dapat ditarik kesimpulan penelitian ini dengan penelitian terdahulu dalam pengerjaan pengolahan data adanya perbedaan. Sistem pengolahan data dalam penelitian terdahulu masih manual sedangkan sistem dalam penelitian ini mengefisienkan proses penyeleksian dan pengelompokkan siswa kelas.

2.2 Data Mining

Data Mining adalah pengambilan keputusan suatu informasi dengan pola yang terdapat pada sekumpulan data yang ada di database dengan jumlahnya cukup besar. Terdapat juga di suatu jurnal ilmiah, KDD adalah nama lain dari data mining. *Data Mining* terdapat bidang ilmu, yaitu *Statistik, Database System, Data Warehousing, Information Retieval, dan Machine Learning*. [9]

Selain pengertian di atas terdapat macam-macam pengertian dari *Data Mining*. Macam-macam pengertian tersebut diantaranya adalah sebagai berikut : [10]

1. *Data Mining* adalah suatu analisa otomatis dengan data besar yang bertujuan menemukan pola-pola baru dan bersifat penting.
2. *Knowledge Discovery In Databases* (KDD) adalah suatu informasi yang belum diketahui sebelumnya yang mengandung manfaat. Pemrosesan ini memiliki metode teknis yang tidak sama, seperti halnya *Data Summarization*, *Clusterinn* dan *Learning Classification Rules*

KDD atau *Data Mining* adalah hasil analisis beberapa kumpulan data untuk menemukan keterkaitan yang tidak di duga dengan meringkas data menggunakan cara yang berbeda dari sebelumnya, sehingga menghasilkan data yang mudah di pahami dan berguna. “*Data Mining* merupakan bidang dari beberapa keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar” perkembangan yang pesat dari bidang *Data mining* disebabkan karena beberapa factor pendorong yaitu: [11]

1. Banyaknya data yang diolah

Peroses data mining merupakan informasi yang melandasi perintah tertentu sehingga informasi itu menjadi lebih signifikan, serta semakin besar data yang digunakan, maka hasil dari pemrosesanya pun semakin valid.

2. Sudah dilakukan proses *Data Warehousing*

Untuk mendapatkan hasil yang sempurna maka sumber data yang akan digunakan dalam proses data mining adalah data gabungan dari banyaknya unit, seperti data nilai mata pelajaran.

3. Kemampuan komputasi yang semakin terjangkau

Pada intinya proses data mining menjalankan banyak akses terhadap data yang sangat besar. Selain itu juga menjalankan proses komputasi yang memerlukan sumber daya yang sangat besar. Semakin berkembangnya teknologi yang semakin cepat menjadikan peroses data mining cocok untuk dilakukan secara komersial.

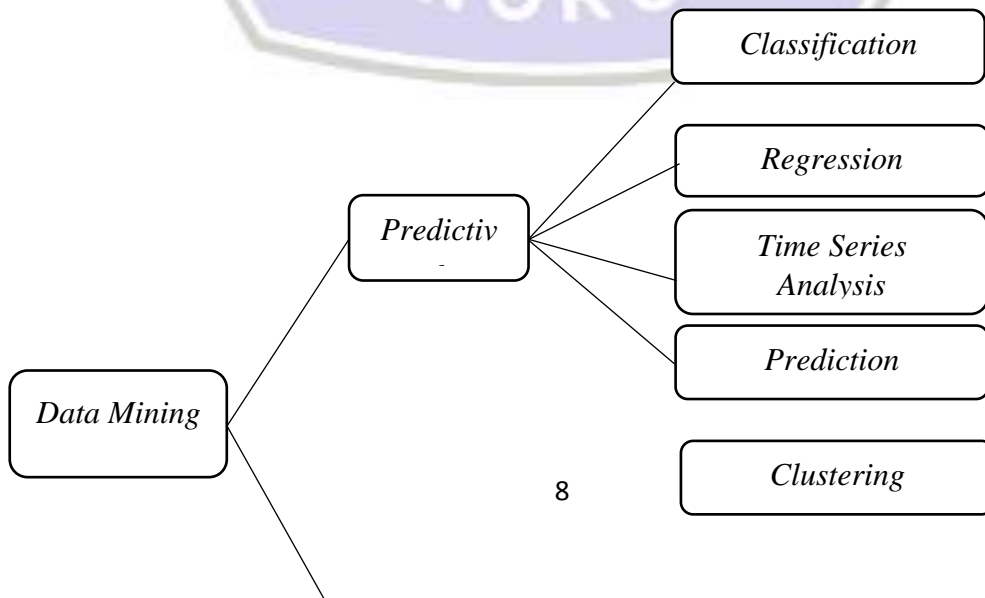
4. Persaingan sekolah yang semakin ketat

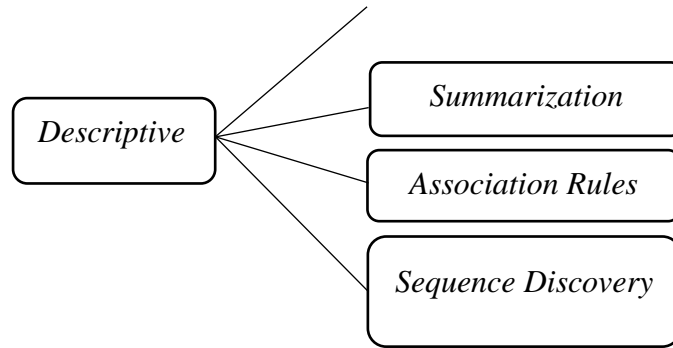
Munculnya sekolah baru dengan unggulan yang bagus semakin mendorong untuk selalu berinovasi agar mampu meningkatkan daya saing.

Berdasarkan definisi-definisi yang telah disebutkan di atas terdapat, hal penting yang berkaitan dengan data mining yaitu [12] :

1. Mengolah suatu informasi dari suatu data, yang sebelumnya tidak diketahui.
2. Memproses suatu data yang besar dengan otomatis untuk menemukan data yang riil, data baru dan data yang berguna.
3. Mengekstrak suatu kumpulan data yang sangat besar.
4. Eksplorasi & analisis, secara otomatis dari kumpulan data yang besar diharapkan mampu memberikan pola data yang penting.

Ahmet menyatakan ada dua kategori dalam data mining yaitu prediksi dan deskripsi. Teknik prediksi yaitu penggunaan data untuk menyimpulkan sesuatu kejadian masa depan. Sedangkan untuk teknik deskripsi yaitu menemukan pola pada data yang menyediakan sebagian informasi tentang hubungan interval yang tersembunyi. [13] :





Gambar 2. 1 Teknik Gambar *Mining* [13]

Menurut Kumar dan Saurafah dalam Komarudin[13], menyatakan terdapat teknik yang di dapat digunakan dalam data mining meliputi :

1. Klasifikasi

Klasifikasi adalah penerapan pada data mining yang paling umum. Keputusan pohon (*Decision Tree*) atau neural network berbasis algoritma klasifikasi merupakan pendekatan yang sering digunakan .

2. Prediksi

Data mining independent variabel merupakan atribut yang sudah dikenali dan respon variabel apa yang kita inginkan untuk diprediksi.

3. Neural network

Jaringan saraf merupakan serangkaian unit untuk menghubungkan input dengan output untuk menghasilkan penghubung yang berbobot. Selama fase learning, jaringan saraf memberikan bobot untuk memprediksi kelas dari setiap data yang di masukan.

4. Nearest Neighbor Method

Teknik K-Nearest Neighbor merupakan teknik klasifikasi sekumpulan data dari kompok data berdasarkan sebuah kombinasi suatu kelas k data yang sama dalam sebuah kumpulan data historis (sama halnya dengan $k > 1$).

5. Clustering

Clustering adalah pengelompokan suatu kelas yang memiliki kemiripan data.

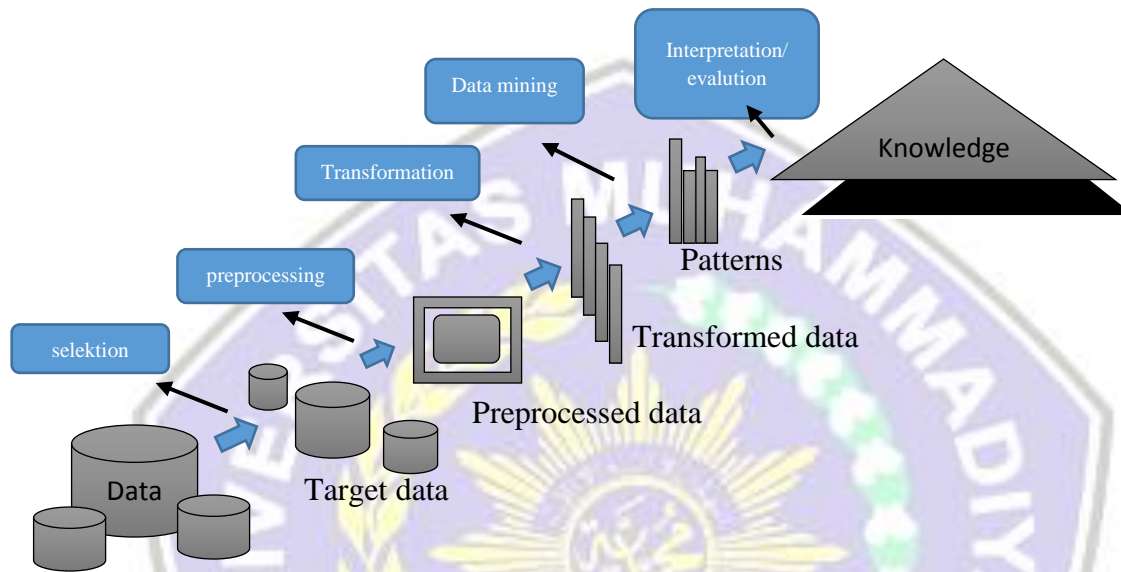
6. Association rule

Digunakan untuk mengetahui kelakuan dari kejadian-kejadian khusus atau proses yang dimana link asosiasi muncul pada setiap kejadian.

7. Decision trees

Decision trees merupakan struktur tree-shaped yang mewakili setiap keputusan

Metode analisis data yang di gunakan yaitu clustering dengan menggunakan penerapan tahapan *Knowledge Discovery In Databases* (KDD) yang terdiri dari beberapa tahapan diantaranya [14]:



Gambar 2. 2 Tahapan Data Mining

1. Pembersihan data

Dipergunakan untuk penyeleksian data yang tidak pasti. Selain itu data warehouse juga memungkinkan tipe analisa seperti OLAP.

2. Preprocessing,

Proses data yang tidak berguna di buang dan untuk menjamin format tetap konsisten.

3. Transformasi data

Proses pemetaan data yang kompleks

4. Aplikasi Teknik Data Mining

Proses pembentukan pola dari data yang ada atau proses utama KDD

5. Evaluasi pola yang ditemukan

Proses teknik data mining berupa pola-pola yang unik maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai.

6. Knowledge

Tujuan utama proses KDD adalah untuk memperoleh pengetahuan atau informasi yang bermanfaat dan mudah di mengerti.

2.3 Clustering

Clustering termasuk kelompok homogen, yaitu setiap bagian mempunyai kesamaan satu sama lain. kualifikasi dari kluster harus memiliki data yang homogen dan tidak sama dengan data yang lainnya, sedangkan data yang berbeda akan berkumpul dengan kluster yang berbeda pula. Adapun contoh Aplikasi Clustering antara lain: *Medicine Sciences, Bioinformatics, Social Sciences, Engineering, Data dan Web Mining*. [15]

2.4 Algoritma K-Means

Pengelompokan data adalah teknik eksplorasi suatu kumpulan data yang memiliki objek karakteristik yang sama dimana di kelompokkan menjadi satu untuk mengklasifikasi pengolahan data lebih lanjut. [16]

Algoritma k-means adalah metode cluster analysis non hirarki yang yang mengelompokkan suatu data berdasarkan karakteristiknya kedalam satu atau lebih cluster, jadi data yang memiliki kesamaan dikelompokkan dalam suatu cluster yang sama dan sebaliknya data yang memiliki kesamaan yang tidak sama dikelompokkan kedalam cluster yang tidak sama pula.

Dalam *Algoritma K-Means* terdapat beberapa pendekatan dalam mengembangkan metode clustering, dua pendekatan utama yaitu *Algoritma Parsial* dan *Algoritma Hierarki*. *Algoritma Parsial* adalah algoritma yang Membuat beberapa partisi untuk di kelompokkan menjadi beberapa objek berdasarkan kriteria tertentu sedangkan algoritma hierarki adalah algoritma yang Membuat suatu kelompok objek berdasarkan kesamaanya. Misalnya = tua-muda, merokok-tidak merokok. [17]

Berikut adalah rumus penghitungan jarak dalam *Algoritma K-Means* :

$$D(i,j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \dots\dots(1)$$

dimana:

$D(i,j)$ = Jarak data ke i ke pusat cluster j

X_{ki} = Data ke i pada atribut data ke k

X_{kj} = Titik pusat ke j pada atribut ke k

Untuk langkah awal dilakukannya tahapan klasterisasi data yaitu dengan penerapan *Algoritma K-Means* lalu memilih suatu titik centroid (c_j) yang umumnya pemilihan pusat centroid di pilih dengan cara acak. Jumlah centroid (c_j) yang di pilih sesuai jumlah cluster yang telah di pilih sejak awal. Kalau sudah terbentuk (k) centroid lalu penghitungan setiap jarak data (x_i) dengan centroid (j) sampai (k), ditulis dengan notasi ($d(x_i, c_j)$). Dalam tolak ukuran suatu jarak yang di terapkan ialah kecocokan dari instance suatu data, salah satunya yaitu jarak Euclid. Rumus penjumlahan jarak Euclid

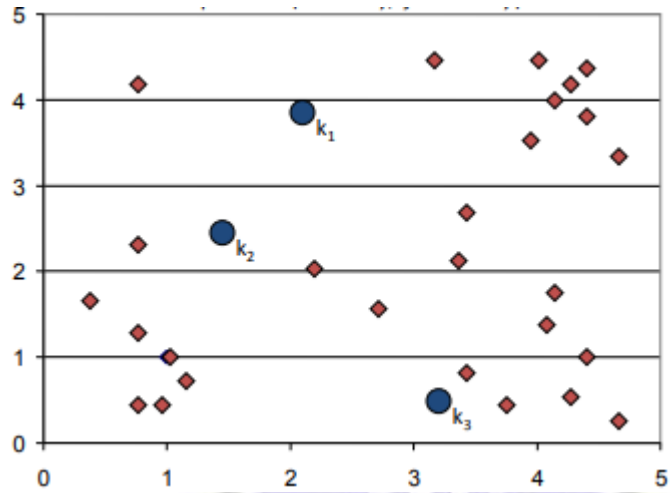
$$d(X_i, C_j) = \sqrt{\sum_{i=1}^N (X_i - C_j)^2} \dots\dots\dots(2)$$

$d(X_i, C_j)$ semakin kecil, kecocokan data akan semakin dekat. Persyaratan dalam menerapkan jarak Euclid yaitu Kalau suatu data set tidak cocok maka menerapkan suatu konsep jarak Mahalanobis. selanjutnya jarak dicari yang paling dekat untuk berkelompok dengan centroid terdekat. Langkah selanjutnya yaitu memilih ulang centroid dengan cara penghitungan rata-rata seluruh jarak pada data centroid. Penghitungan akan berulang ke proses awal dan penghitungan terus diulang hingga mendapatkan centroid tetap maksudnya pusat centroid tidak mengalami perubahan. Atau penghitungan terpaksa berhenti jika penjumlahan maksimal ditentukan.[18]

Contoh Penerapan Algoritma K-Means[19].

Step 1 dan 2

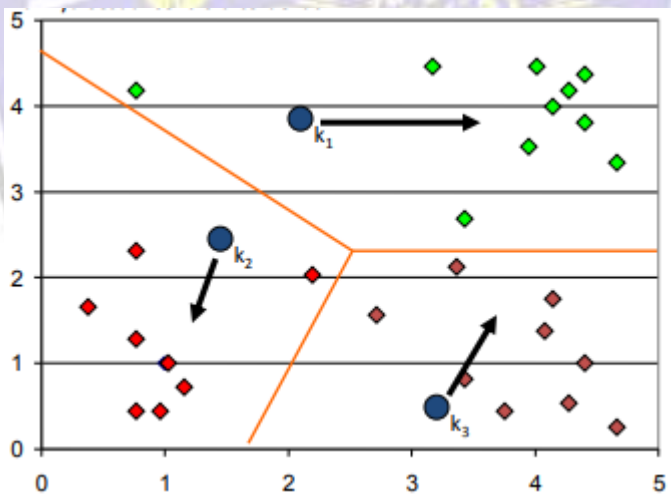
Menentukan nilai K dan Pemilihan pemilihan centroid pada tiap klaster



Gambar 2. 3 Penentuan Nilai K

Step 3

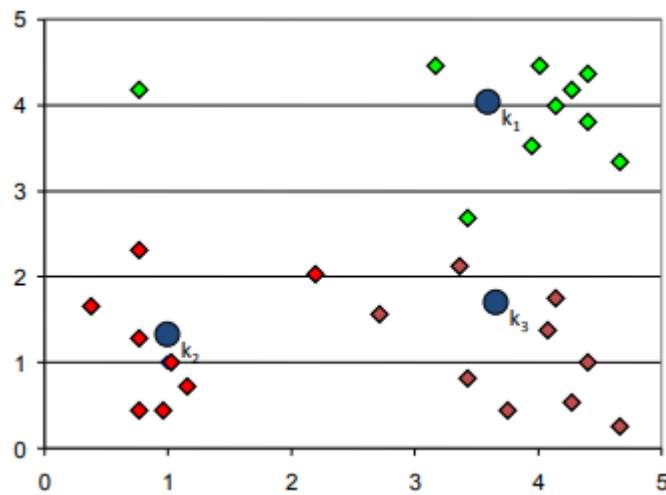
Mengitung jarak suatu objek centroid pada setiap kluster dengan menggunakan kelompok claster jarak minimum pada suatu objek.



Gambar 2. 4 Penghitungan Jarak Suatu Kluster

Step 4

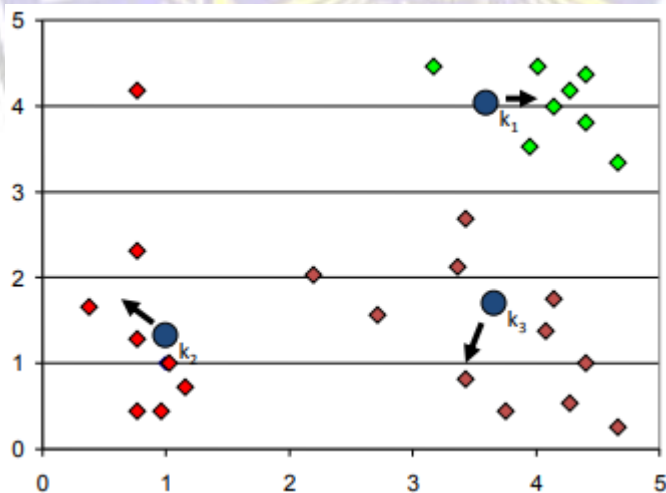
Setelah kluster dan kelompok sudah terbentuk, kemudian menghitung mean pada setiap kluster dan menjadikan centroid yang baru



Gambar 2. 5 Penghitungan Mean Untuk Centroid Baru

Step 5

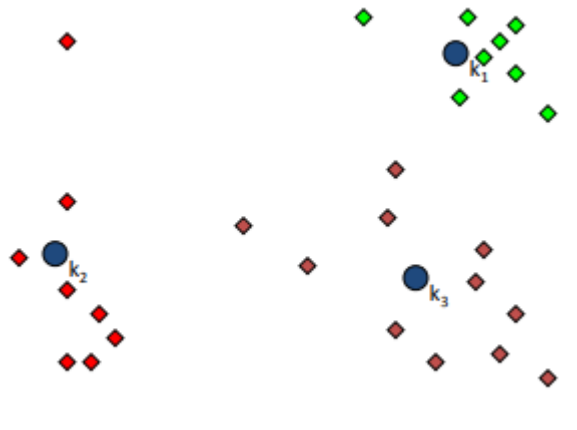
centroid baru dengan centroid yang lama tidak sama,
perlu di perbarui kelompok pada objeknya



Gambar 2. 6 Penghitungan jarak suatu Kluster

Finish

Ulangi step 3-5 hingga tak ada perubahan pada centroid dan objek tidak berpindah kelas lagi, sebagaimana pada Gambar 2.4 sampai Gambar 2.6.



Gambar 2. 7 Finish Algoritma K-Means

Dalam algoritma k-means juga terdapat kelebihan dan kekurangan. Adapun kelebihanya yaitu :

1. Kelebihan
 - a. Selalu dapat melakukan clustering
 - b. Menggunakan operasi matematika sederhana
 - c. Beban komputasi relatif lebih ringan sehingga clustering bisa dilakukan dengan cepat walaupun relatif tergantung pada banyak jumlah data dan jumlah kluster yg ingin dicapai.[19]
2. Kekurangan
 - a. Ketika jumlah data tidak banyak, pengelompokan awal akan menentukan cluster secara signifikan.
 - b. Jumlah gugus, K, harus ditentukan sebelumnya.
 - c. Kita tidak pernah tahu cluster nyata, menggunakan data yang sama, jika itu dimasukkan dengan cara yang berbeda dapat menghasilkan cluster berbeda jika jumlah data sedikit.

- d. Kita tidak pernah tahu atribut mana yang berkontribusi lebih besar pada proses pengelompokan karena kita berasumsi bahwa masing-masing atribut memiliki bobot yang sama. [20]

2.5. Microsoft Excel 2016

Microsoft Corporation memproduksi dan menjual sebuah *Software Microsoft Office* yang di dalamnya terdapat *Software* berupa *Microsoft Excel*, *Microsoft Word*, *Microsoft Outlook*, *Microsoft PowerPoint*, dan lainnya. Dalam dunia kerja *Microsoft Excel* menjadi sebuah *Software* yang terkenal sebagai multifungsi dalam mengolah data berupa angka, di dalam *Microsoft Excel* juga terdapat kombinasi cell untuk memudahkan user dalam mengolah data. selain itu *Microsoft Excel* mendapat pengakuan dari berbagai profesi akan kemudahan dan kecepatan dalam menangani sebuah masalah yang di hadapi user dalam mengolah data berupa angka entah itu pekerjaan yang ringan maupun pekerjaan berat sekalipun untuk mendapatkan hasil yang maksimal.[21]

2.6. Pemograman R dan Rstudio

Pemograman R merupakan GNU komputasi statistik dan grafik yang dikembangkan oleh *Bell Laboratories* oleh *John Chamber* dan teman-temannya yang sebelumnya di buat pada tahun 1992 di Universitas Auckland, New Zealand oleh Ross Ihaka dan Robert Gentleman. Dalam pemograman R terdapat berbagai tools statistik linier, non linier, clustering, klasifikasi, uji statistik klasik dan Tool teknis grafis yang menampilkan data secara visual berbentuk grafik. Pemograman R juga merupakan program yang open source dalam mengolah data Statistika berbentuk grafik yang sangat berguna untuk penelitian maupun industri. [22]

Rstudio adalah GUI untuk R yang memiliki kelebihan yaitu dapat digunakan dan dijalankan pada browser, sehingga memudahkan user untuk tidak menginstall program R, kecuali paket (package) pemrograman sesuai kebutuhan user.[23] Pada gambar 2.8 sampai gambar 2.15 memuat contoh penerapan *K-Means* menggunakan Rstudio. Terdapat data sebuah hotel bintang yang memiliki 3 variabel yaitu: [24].

Provinsi	Akomodasi	Kamar	Tempat Tidur
ACEH	21	1603	2654
BALI	551	52927	78801
BANTEN	113	10607	15507
BENGKULU	9	538	845
DI YOGYAKARTA	143	14328	23477
DKI JAKARTA	326	46899	60849
GORONTALO	8	535	804
JAMBI	37	2461	3780
JAWA BARAT	463	43034	62725
JAWA TENGAH	291	23516	33530
JAWA TIMUR	231	26585	38540
KALIMANTAN BARAT	41	4269	6204
KALIMANTAN SELATAN	52	4075	6488
KALIMANTAN TENGAH	13	968	1496
KALIMANTAN TIMUR	63	6438	9293
KALIMANTAN UTARA	1	81	113
BANGKA BELITUNG	46	2927	4426
KEP. RIAU	132	13756	19425
LAMPUNG	19	1997	2477
MALUKU	22	1193	1865
MALUKU UTARA	2	116	116
NUSA TENGGARA BARAT	83	5678	8328
NUSA TENGGARA TIMUR	24	1830	2541
PAPUA	32	2187	3156
PAPUA BARAT	16	980	1348
RIAU	87	6988	9802
SULAWESI BARAT	5	164	260
SULAWESI SELATAN	139	10418	15757
SULAWESI TENGAH	16	1492	2361
SULAWESI TENGGARA	19	1439	2122
SULAWESI UTARA	35	3072	3484
SUMATERA BARAT	74	4493	7717
SUMATERA SELATAN	76	6161	8811
SUMATERA UTARA	124	10296	15509

Gambar 2. 8 Data Penerapan K-Means Menggunakan Rstudio

Langkah pertama install package ggplot2 dan factoextra, selanjutnya aktifkan package menggunakan perintah library.

```
install.packages("ggplot2")
install.packages("factoextra")
library(factoextra)
library(ggplot2)
```

Selanjutnya import data ke RStudio menggunakan syntax berikut

```
data=read.csv(file.choose(), header=TRUE)
data
```

Kemudian akan muncul output seperti ini:

```
> data
```

	Provinsi	Akomodasi	Kamar	Tempat.Tidur
1	Aceh	20	1503	2579
2	Sumatera Utara	116	10732	16418
3	Sumatera Barat	61	3653	6015
4	Riau	58	5206	7832
5	Jambi	31	2187	2885
6	Sumatera Selatan	64	4781	6800
7	Bengkulu	9	523	786
8	Lampung	18	1954	2847
9	kepulauan Bangka Belitung	42	2528	3653
10	kepulauan Riau	88	11393	15836
11	DKI Jakarta	232	39806	55293
12	Jawa Barat	314	29053	46013
13	Jawa Tengah	217	17246	26878
14	D.I Yogyakarta	94	9256	14432
15	Jawa Timur	158	17014	26393
16	Banten	58	5939	8844
17	Bali	317	33599	49549
18	Nusa Tenggara Barat	75	4525	6721
19	Nusa Tenggara Timur	23	1646	2652
20	Kalimantan Barat	36	3821	5264
21	Kalimantan Tengah	12	1051	1548
22	Kalimantan Selatan	49	3584	5254
23	Kalimantan Timur	51	5468	8048
24	Kalimantan Utara	10	461	710
25	Sulawesi Utara	35	2576	3752
26	sulawesi Tengah	8	766	1182
27	Sulawesi Selatan	93	6475	9420
28	sulawesi Tenggara	18	1295	1959
29	Gorontalo	8	471	750
30	sulawesi Barat	7	374	587
31	Maluku	22	1110	1566
32	Maluku Utara	3	313	492
33	Papua Barat	15	1014	1477
34	Papua	25	1684	2524

Gambar 2. 9 Tampilan Data di Rstudio

Langkah pembuatan cluster menggunakan data numerik dengan syntax berikut.

```
data.numerik<-data[2:4]
```

```
data.numerik
```

```

> data.numerik
  Akomodasi Kamar Tempat.Tidur
1         20  1503         2579
2        116 10732        16418
3         61  3653         6015
4         58  5206         7832
5         31  2187         2885
6         64  4781         6800
7          9   523          786
8         18  1954         2847
9         42  2528         3653
10        88 11393        15836
11       232 39806        55293
12       314 29053        46013
13       217 17246        26878
14        94  9256        14432
15       158 17014        26393
16        58  5939         8844
17       317 33599        49549
18        75  4525         6721
19        23  1646         2652
20        36  3821         5264
21        12  1051         1548
22        49  3584         5254
23        51  5468         8048
24        10   461          710
25        35  2576         3752
26         8   766         1182
27        93  6475         9420
28        18  1295         1959
29         8   471          750
30         7   374          587
31        22  1110         1566
32         3   313          492
33        15  1014         1477
34        25  1684         2524

```

Gambar 2. 10 Output Data Numerik

Selanjutnya data distandartkan. Dengan tujuannya menyamakan satuan data. Data terapkan memiliki satuan yang berbeda sehingga perlu standarisasi agar satuan data menjadi sama menggunakan syntax berikut.

```

data.stds<-scale(data1)
data.stds

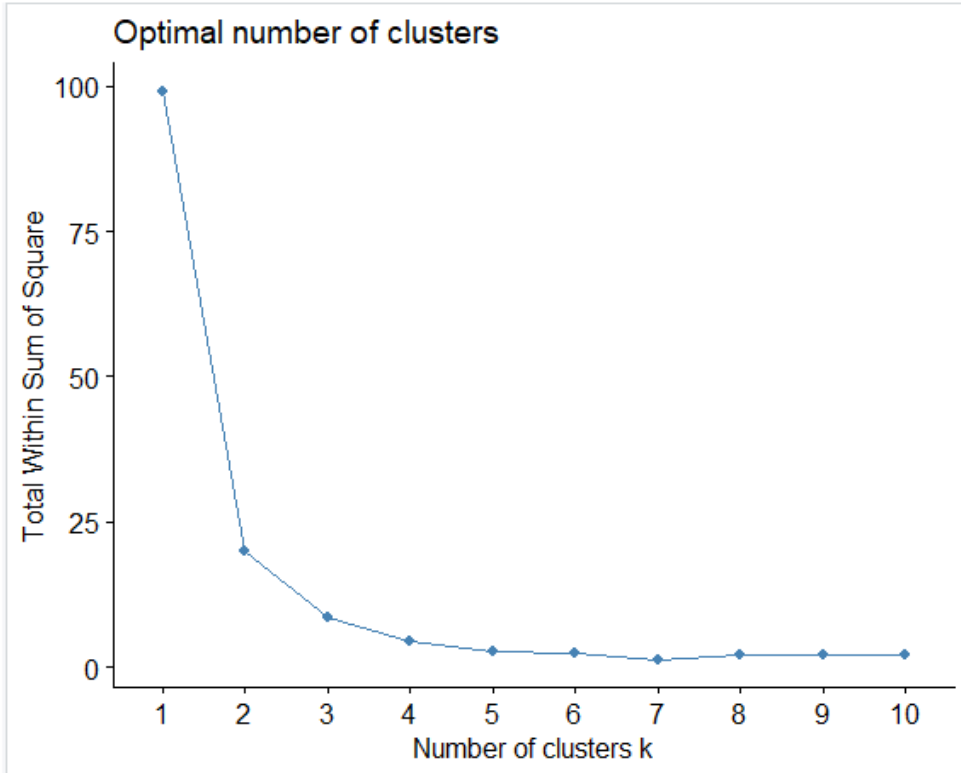
```

Langkah berikutnya, menentukan nilai K menggunakan dua pilihan metode yaitu *Metode Elbow* dan *Metode Silhouette*. dengan menggunakan library *factoextra* di RStudio

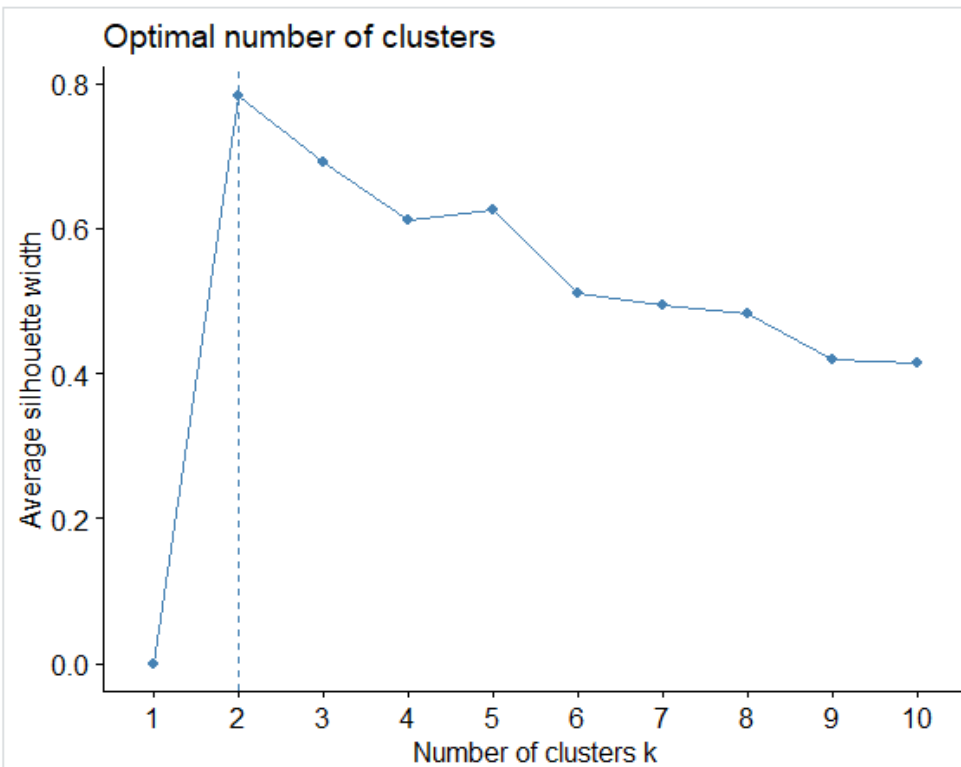
```

fviz_nbclust(data.stds, kmeans, method = "wss")
fviz_nbclust(data.stds, kmeans, method = "silhouette")

```

Gambar 2. 11 Grafik Cluster K Metode Elbow



Gambar 2. 12 Grafik Cluster K Metode Silhouette

Metode Elbow memiliki nilai K yang membentuk siku pada K=2. Kemudian, *Metode Silhouette* memperoleh K=2. Jadi nilai K yang optimal dari dua metode tersebut adalah 2. Jumlah K=2 terbilang terlalu dalam melakukan *Clustering*, dengan itu dibuat 3 *Clustering* (K=3). Dengan *Metode K-Means* dengan nilai K = 3, maka gunakan syntax berikut:

```
Clustering=kmeans (data.stds,centers=3,nstart=25)
Clustering
```

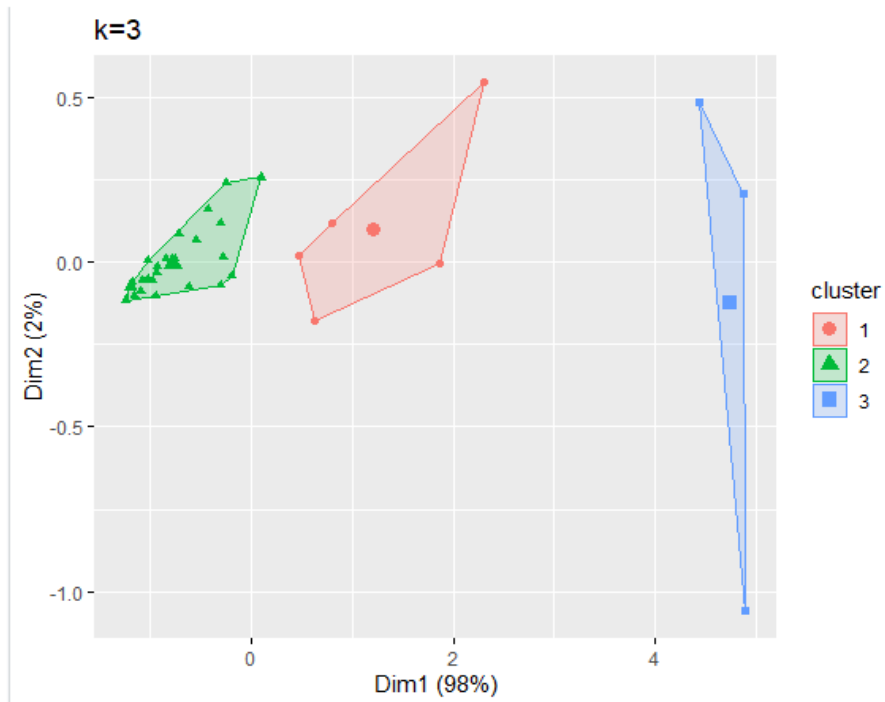
```
K-means clustering with 3 clusters of sizes 5, 26, 3
Cluster means:
  Akomodasi      Kamar Tempat.Tidur
1  0.7714867    0.6440601    0.6822221
2 -0.4489781   -0.4471624   -0.4535762
3  2.6053327    2.8019736    2.7939567
Clustering vector:
[1] 2 1 2 2 2 2 2 2 2 1 3 3 1 1 1 2 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
within cluster sum of squares by cluster:
[1] 2.964730 4.039241 1.494904
   (between_SS / total_SS =  91.4 %)
Available components:
[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
[6] "betweenss"    "size"         "iter"         "ifault"
```

output K-Means Clustering

Gambar 2. 13 Output K-Means

Dari output diperoleh anggota dengan cluster 1 terdiri 5 anggota, cluster 2 terdiri 26 anggota, dan cluster 3 terdiri 3 anggota. Setelah itu visualisasikan clustering dengan K = 3

```
fviz_cluster(clustering, geom = "point", data =  
data.stds)+ggtitle("k=3")
```



Gambar 2. 14 Visualisasi Cluster

Dari output gambar 2.14 cluster 1 merah, cluster 2 hijau, dan cluster 3 biru.

Berikut syntax untuk melihat data yang telah di cluster

```
final=data.frame(data,clustering$cluster)
```

```
View(final)
```

	Provinsi	Akomodasi	Kamar	Tempat.Tidur	clustering.cluster
1	Aceh	20	1503	2579	2
2	Sumatera Utara	116	10732	16418	1
3	Sumatera Barat	61	3653	6015	2
4	Riau	58	5206	7832	2
5	Jambi	31	2187	2885	2
6	Sumatera Selatan	64	4781	6800	2
7	Bengkulu	9	523	786	2
8	Lampung	18	1954	2847	2
9	Kepulauan Bangka Belitung	42	2528	3653	2
10	Kepulauan Riau	88	11393	15836	1
11	DKI Jakarta	232	39806	55293	3
12	Jawa Barat	314	29053	46013	3
13	Jawa Tengah	217	17246	26878	1
14	D.I Yogyakarta	94	9256	14432	1
15	Jawa Timur	158	17014	26393	1
16	Banten	58	5939	8844	2
17	Bali	317	33599	49549	3
18	Nusa Tenggara Barat	75	4525	6721	2
19	Nusa Tenggara Timur	23	1646	2652	2
20	Kalimantan Barat	36	3821	5264	2
21	Kalimantan Tengah	12	1051	1548	2
22	Kalimantan Selatan	49	3584	5254	2
23	Kalimantan Timur	51	5468	8048	2
24	Kalimantan Utara	10	461	710	2
25	Sulawesi Utara	35	2576	3752	2
26	Sulawesi Tengah	8	766	1182	2
27	Sulawesi Selatan	93	6475	9420	2
28	Sulawesi Tenggara	18	1295	1959	2
29	Gorontalo	8	471	750	2
30	Sulawesi Barat	7	374	587	2
31	Maluku	22	1110	1566	2
32	Maluku Utara	3	313	492	2
33	Papua Barat	15	1014	1477	2
34	Papua	25	1684	2524	2

Gambar 2. 15 Data yang telah ada Cluster

Berdasarkan output, anggota cluster 1 yaitu Provinsi Sumatera Utara, Jawa Tengah, DI Yogyakarta dan Jawa Timur, anggota cluster 2 dapat dilihat pada output gambar 2.15 dan 3 DKI Jakarta, Jawa Barat dan Bali.