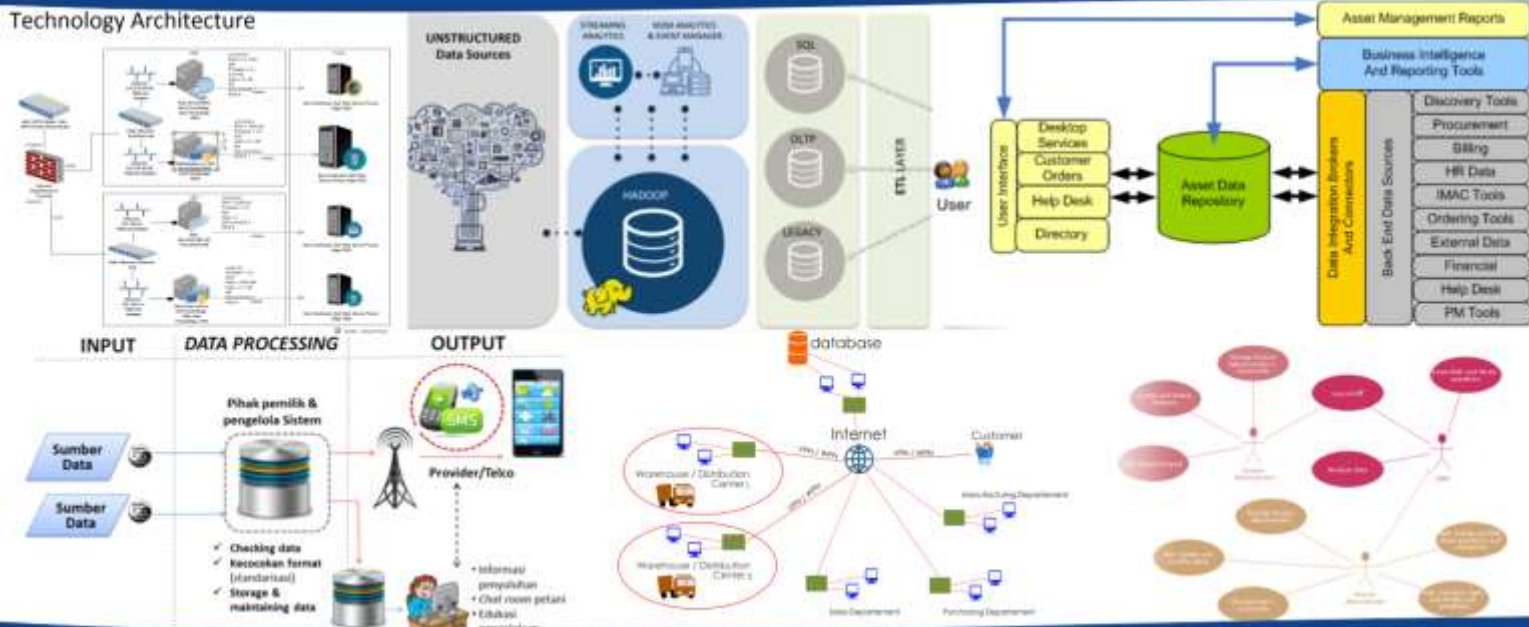




# MISI

## JURNAL MANAJEMEN INFORMATIKA & SISTEM INFORMASI

### Technology Architecture



Diterbitkan Oleh LPPM STMIK Lombok  
Jln. Basuki Rahmat No.105 Praya, Lombok Tengah - NTB  
Telp dan Fax (0370) 654310 - e-journal.stmiklombok.ac.id/jsi  
email. lppm@stmiklombok.ac.id



**DEWAN REDAKSI**

**JURNAL MISI ( JURNAL MANAJEMEN INFORMATIKA DAN SISTEM INFORMASI)**

**Jurnal Manager**

**Wire Bagye, S.Kom.,M.Kom** (STMIK Lombok, SINTA ID : 5992010)

**Reviewer :**

**Resad Setyadi, S.T., S.Si., MMSI., Ph.D** (cand)- Institut Teknologi Telkom Purwokerto  
SCOPUS ID 57204172534, SINTA ID : 6113570

**Yesaya Tommy Paulus, S.Kom., MT., Ph.D.** - STMIK Dipanegara Makassar  
SCOPUS ID 57202829909, SINTA ID : 6002004

**Lalu Mutawalli, S.Kom., M.I.Kom., M.Kom** - STMIK Lombok  
SCOPUS ID : 57205057118, SINTA ID : 6659709

**Saruni Dwiasnati, ST., MM., M.Kom** - Universitas Mercu Buana  
SCOPUS ID : 57210968603, SINTA ID : 6150854

**Ida Bagus Ary Indra Iswara, S.Kom., M.Kom** - STMIK STIKOM Indonesia  
SCOPUS ID 57203711945, SINTA ID : 183498

**Erlin Windia Ambarsari** - Universitas Indraprasta PGRI  
SCOPUS ID : 56242503900, SINTA ID : 5998887

**Wafiah Murniati, ST., MT.** - STMIK Lombok  
SCOPUS ID : 56242503900, SINTA ID : 5998887

**Yuliadi, S.Kom., M.Kom** - Universitas Teknologi Sumbawa  
SINTA ID : 6730786

**Fachrudin Pakaja, S.Kom, M.T** - Universitas Gajayana  
SINTA ID : 6164357

**Ahmad Jufri, S.Kom., M.T** - Sekolah Tinggi Teknologi STIKMA Internasional  
SINTA ID : 172241

**Mohammad Taufan Asri Zaen, ST., MT** - STMIK Lombok  
SINTA ID : 5992087

**Hairul Fahmi, S.Kom., M.Kom - STMIK Lombok**  
SINTA ID : 5983160

**I Ketut Putu Suniantara, S.Si., M.Si** - ITB STIKOM Bali  
SINTA ID : 6086221

**Nawassyarif S. Kom., M.Pd.** - Universitas Teknologi Sumbawa  
SINTA ID : 6722660

**Muhamad Malik Mutoffar, ST., MM., CNSS** - Sekolah Tinggi Teknologi Bandung  
SINTA ID : 6013819

**Editor :**

**Saikin, Skom., M.Kom.** - STMIK Lombok

**Vrestanti Novalia Santosa, M.Pd.** - IKIP Budi Utomo Malang

**Desain Grafis & Web Maintenance**

**Jihadul Akbar, S.Kom** - STMIK Lombok

**Secretariat**

**Maulana Ashari, M.Kom** - STMIK Lombok

## DAFTAR ISI

- 1** ANALISIS CLUSTERING PROVINSI DI INDONESIA BERDASARKAN TINGKAT KEMISKINAN MENGGUNAKAN ALGORITMA K-MEANS **1 - 8**  
*Achmad Bahauddin<sup>1</sup>, Agustina Fatmawati<sup>2</sup>, Febrianti Permata Sari<sup>3</sup>*
- 2** PEMBOBOTAN MENGGUNAKAN *PAIRWISE COMPARISON* PADA *CASE BASED REASONING* REKOMENDASI HOTEL **9 - 18**  
*Kukuh Tri Nur Iman<sup>1</sup>, Setyawan Wibisono<sup>2</sup>*
- 3** IMPLEMENTASI METODE AHP PADA SISTEM PENDUKUNG KEPUTUSAN PENYELEKSIAN NASABAH PINJAMAN KREDIT **19 - 27**  
*Irfak Lahumu Darajat<sup>1</sup>, Wiwien Hadikurniawati<sup>2</sup>*
- 4** SELEKSI PENERIMAAN BEASISWA BIDIKMISI PADA STMIK INDONESIA PADANG MENGGUNAKAN METODE (AHP) **28 - 35**  
*Heru Saputra<sup>1</sup>, Efendi Mardiono<sup>2</sup>, Ilfa Stephane<sup>3</sup>, Ratih Purwasih<sup>4</sup>*
- 5** PENGELOMPOKAN JENIS RUMPUT LAUT MENGGUNAKAN FUZZY C-MEANS BERBASIS CITRA **36 - 44**  
*Franki Yusuf Bisilisin<sup>1</sup>, Remerta Noni Naatonis<sup>2</sup>*
- 6** SISTEM REKOMENDASI PRODUCT EMINA COSMETICS DENGAN MENGGUNAKAN METODE CONTENT - BASED FILTERING **45 - 54**  
*Fatoni Batari Agung Larasati<sup>1</sup>, Herny Februariyanti<sup>2</sup>*
- 7** SISTEM INFORMASI BOOKING (STUDI KASUS: REGGAENERASI INK STUDIO) **55 - 62**  
*Ni Wayan Yesi Mertha Sari<sup>1</sup>, Ni Luh Putu Ning Septyarini Putri Astawa<sup>2</sup>, I Nyoman Yudi Anggara Wijaya<sup>3</sup>*
- 8** PENERAPAN METODE SMART DALAM SISTEM PENDUKUNG KEPUTUSAN PEMBERIAN SANKSI PELANGGARAN TATA TERTIB SISWA (Studi Kasus: SMK Negeri 1 Pujut) **63 - 72**  
*Mohammad Taufan Asri Zaen<sup>1</sup>, Baiq Daniatan Janiah<sup>2</sup>, Sofiansyah Fadli<sup>3</sup>*
- 9** RANCANGAN SISTEM INFORMASI PERHITUNGAN PENYUSUTAN *FIXED ASSETS* MENGGUNAKAN *STRAIGHT LINE METHOD* PADA PT FIF GROUP PEMATANGSIANTAR **73 - 77**  
*Ayu Tiara Defi<sup>1</sup>, Dedi Suhendro<sup>2</sup>*
- 10** PERANCANGAN SIMPLE STATELESS AUTENTIKASI DAN OTORISASI LAYANAN REST-API BERBASIS PROTOKOL HTTP **78 - 87**  
*I Gusti Ngurah Ady Kusuma*

## ANALISIS CLUSTERING PROVINSI DI INDONESIA BERDASARKAN TINGKAT KEMISKINAN MENGGUNAKAN ALGORITMA K-MEANS

Achmad Bahauddin<sup>1</sup>, Agustina Fatmawati<sup>2</sup>, Febrianti Permata Sari<sup>3</sup>

<sup>1,2,3</sup>Jurusan Teknik Industri, Fakultas Teknik, Universitas Sultan Ageng Tirtayasa  
Jln. Jend. Sudirman KM. 3, Cilegon, Banten 42435

<sup>1</sup>[baha@untirta.ac.id](mailto:baha@untirta.ac.id), <sup>2</sup>[agustinafatmawati04@gmail.com](mailto:agustinafatmawati04@gmail.com), <sup>3</sup>[febriantips50@gmail.com](mailto:febriantips50@gmail.com)

---

### Abstract

*We can find out the percentage of poverty in all provinces of Indonesia using data taken from BPS (the Central Bureau of Statistics). However, in the data presented at the BPS, the percentage of poverty categories presented is only by province, whereas in an effort to eradicate existing poverty, the government needs to know which areas have high, medium or low levels of poverty so that the government can establish priority scales to reduce poverty. Therefore, it is necessary for grouping provinces in Indonesia based on their poverty levels so that the government can provide policies and approaches that are appropriate and quick to overcome the poverty that occurs based on the poverty levels of each province. This research was used the clustering method with the K-Means algorithm for grouping provinces based on their poverty level with the help of Weka software. The results showed that there were three provincial clusters in Indonesia based on their poverty level, namely Cluster 0 (a province with low poverty level), Cluster 1 (a province with moderate poverty level), and Cluster 2 (a province with high poverty level). Provinces included in the category of provinces with high poverty rates were Maluku, West Papua and Papua.*

**Keywords :** *clustering, poverty level, K-Means*

### Abstrak

Berdasarkan sumber informasi dari Badan Pusat Statistik (BPS), kita dapat mengetahui persentase kemiskinan yang ada di seluruh provinsi Indonesia. Namun didalam data yang tersaji di BPS tersebut, kategori persentase kemiskinan yang disajikan hanya berdasarkan provinsi sedangkan dalam upaya untuk memberantas kemiskinan yang ada, pemerintah perlu mengetahui daerah manakah yang memiliki tingkat kemiskinan tinggi, sedang atau rendah sehingga pemerintah dapat menetapkan skala prioritas untuk menanggulangi kemiskinan tersebut. Oleh karena itu, perlu dilakukan pengelompokan provinsi yang ada di Indonesia berdasarkan tingkat kemiskinannya agar pemerintah dapat memberikan kebijakan atau pendekatan yang tepat dan cepat untuk mengatasi kemiskinan yang terjadi berdasarkan tingkat kemiskinan masing-masing provinsi. Pada penelitian ini digunakan metode clustering dengan algoritma K-Means untuk mengelompokan provinsi berdasarkan tingkat kemiskinannya dengan bantuan software Weka. Hasil penelitian menunjukkan terdapat 3 cluster provinsi di Indonesia berdasarkan tingkat kemiskinannya yaitu Cluster 0 (provinsi dengan tingkat kemiskinan rendah), Cluster 1 (provinsi tingkat kemiskinan sedang), dan Cluster 2 (provinsi dengan tingkat kemiskinan tinggi). Provinsi yang termasuk dalam kategori provinsi dengan tingkat kemiskinan tinggi yaitu Maluku, Papua Barat dan Papua.

**Kata kunci :** clustering, tingkat kemiskinan, K-Means

---

### 1. Pendahuluan

Kemiskinan adalah masalah yang banyak ditemui di negara berkembang. Indonesia sebagai salah satu negara berkembang juga mengalami permasalahan kemiskinan ini. Menurut [1]

seseorang dikatakan miskin ketika orang tersebut tidak mampu memenuhi kebutuhan minimal yang layak bagi kehidupannya. Kebutuhan dasar minimal yang dimaksud adalah yang berkaitan dengan kebutuhan pangan, sandang, perumahan dan kebutuhan sosial.

Kemiskinan merupakan permasalahan utama dalam pembangunan ekonomi di negara berkembang seperti Indonesia. Berdasarkan data dari BPS, angka kemiskinan di Indonesia pada September 2019 adalah sebesar 9,22 persen dan jumlah orang miskin di Indonesia mencapai 24,79 juta jiwa. Berdasarkan sumber informasi dari BPS, kita dapat mengetahui persentase kemiskinan yang ada di seluruh provinsi di Indonesia. Namun didalam data yang tersaji di BPS tersebut, kategori persentase kemiskinan yang disajikan hanya berdasarkan provinsi sedangkan dalam upaya untuk memberantas kemiskinan yang ada, pemerintah perlu mengetahui daerah manakah yang memiliki tingkat kemiskinan tinggi, sedang atau rendah sehingga pemerintah dapat menetapkan skala prioritas untuk menanggulangi kemiskinan tersebut. Oleh karena itu, penggolongan tingkat persen kemiskinan provinsi yang ada di Indonesia dapat dikelompokkan menggunakan metode yang terdapat didalam data mining yaitu metode *clustering*. Data mining merupakan suatu proses menggali data dari pengetahuan yang berbeda dan menghasilkan sebuah informasi-informasi penting yang dapat dipakai dalam meningkatkan keuntungan, memperkecil biaya pengeluaran, atau bahkan keduanya [2].

*Clustering* adalah metode yang membagi data ke dalam grup-grup yang mempunyai obyek dengan karakteristik yang sama. [3] menyatakan *clustering* adalah mengelompokkan item data ke dalam sejumlah kecil grup sedemikian rupa sehingga masing-masing grup mempunyai sesuatu persamaan yang esensial. [4] menjelaskan bahwa *clustering* adalah proses membagi data yang tidak berlabel menjadi kelompok-kelompok data yang memiliki kemiripan.

Penelitian ini bertujuan mengelompokkan provinsi manakah di Indonesia yang memiliki tingkat persentase kemiskinan yang tinggi, sedang, dan rendah. Dari hasil pengelompokkan tersebut maka akan diketahui provinsi manakah di Indonesia yang selayaknya mendapatkan prioritas dan perhatian khusus untuk mengatasi kemiskinan sehingga mengalami kemajuan dan perbaikan dari segi kehidupan dan juga ekonomi.

## 2. Tinjauan Pustaka

Penelitian sebelumnya yang melakukan analisis clustering menggunakan metode K-Means, diantaranya adalah penelitian [5]. Pada penelitian tersebut metode analisis clustering dengan algoritma K-Means digunakan untuk mengelompokkan seluruh desa di Kabupaten

Pangkep berdasarkan kriteria-kriteria kemiskinan kemudian tingkat keakuratan pengelompokkan tersebut diuji dengan menggunakan metode analisis diskriminan.

Penelitian lainnya adalah penelitian [1], dimana pada penelitian tersebut analisis clustering K-Means digunakan untuk mengelompokkan kabupaten dan kota di provinsi Jawa Barat berdasarkan informasi data kemiskinan.

Penelitian [6] menggunakan algoritma K-Means untuk mengelompokkan mahasiswa berdasarkan prestasi mereka di perkuliahan dengan menggunakan bantuan software WEKA.

Penelitian [7] menggunakan algoritma K-Means dengan bantuan Weka Interface untuk mengelompokkan mahasiswa berdasarkan nilai akademiknya. Diperoleh 4 kelompok mahasiswa berdasarkan nilai IPK dan atribut mata kuliah yang mereka ambil.

### 2.1. Analisis Clustering

Menurut [8], clustering adalah proses membagi-bagi obyek dari suatu set data menjadi beberapa kluster yang homogen. Tujuan utama dari metode kluster adalah pengelompokkan sejumlah data/obyek kedalam kluster (grup) sehingga dalam setiap kluster akan berisi data yang semirip mungkin. Clustering berusaha untuk menempatkan obyek yang mirip atau yang jaraknya berdekatan dalam satu kluster dan membuat jarak antar kluster sejauh mungkin. Ini berarti obyek dalam satu kluster sangat mirip satu sama lain dan berbeda dengan obyek dalam kluster-kluster yang lain.

Secara garis besar, teknik pengklasteran dapat dikelompokkan kedalam empat kategori: (i) metode *partitioning*, (ii) metode hirarkis, (iii) metode *density-based*, dan (iv) metode *grid-based*. Metode *partitioning* melakukan pengklasteran melalui pencarian kluster yang terpisah secara langsung dan merelokasi titik data secara iterative untuk memperbaiki kualitas kluster dari solusi awal. Metode hirarkis membangun kluster secara bertahap, yaitu melalui penggabungan kluster yang kecil menjadi kluster yang lebih besar, atau melalui pemecahan kluster yang lebih besar menjadi kluster yang lebih kecil.

Untuk metode *density-based*, metode ini melakukan identifikasi kluster dengan menghubungkan daerah-daerah yang memiliki titik data yang cukup banyak. Sedangkan pada metode *grid-based*, untuk meningkatkan efisiensi pengklasterannya, metode ini membagi ruang data menjadi sejumlah sel yang membentuk

struktur grid yang rapat (terisi banyak) untuk membentuk klaster.

## 2.2. Algoritma K-Mean

Dalam algoritma K-Mean, obyek atau data yang ada dikelompokkan ke dalam  $k$  kelompok atau klaster. Untuk melakukan klaster ini, nilai  $k$  harus ditentukan terlebih dahulu. Biasanya user atau pengguna sudah mempunyai informasi awal tentang obyek yang sedang dipelajari, termasuk berapa jumlah klaster yang paling tepat. Dalam algoritma K-Mean digunakan ukuran ketidakmiripan untuk mengelompokkan obyek. Ketidakmiripan ini diterjemahkan dalam konsep jarak. Jika jarak dua obyek cukup dekat, maka dua obyek tersebut mirip. Semakin dekat berarti semakin tinggi kemiripannya. Semakin tinggi nilai jarak, semakin tinggi pula ketidakmiripannya [8].

*Clustering* data dengan algoritma K-Mean secara umum dilakukan dengan algoritma sebagai berikut [9]:

1. Tentukan jumlah kelompok.
2. Alokasikan data ke dalam kelompok secara acak

Inisialisasi  $k$  pusat klaster (*centroid*). Ini dapat dilakukan dengan berbagai cara. Yang paling sering dilakukan adalah dengan cara random, yakni pusat-pusat klaster diberi nilai awal dengan angka-angka random.

3. Hitung pusat kelompok (*centroid*) dari data yang ada di masing-masing kelompok
4. Alokasikan masing-masing data ke *centroid* klaster terdekat.

Kedekatan dua obyek ditentukan berdasarkan jarak kedua obyek tersebut. Demikian juga kedekatan suatu data ke klaster tertentu ditentukan berdasarkan jarak antara data tersebut dengan pusat klaster. Data akan dimasukkan ke klaster yang jarak antara data tersebut dengan pusat klasternya paling dekat. Untuk menghitung jarak antara data dengan pusat klaster biasanya digunakan jarak Euclidean, walaupun dapat juga menggunakan jarak Manhattan atau Minkowsky.

Untuk menghitung jarak semua data ke setiap titik pusat *cluster* dengan menggunakan teori jarak *Euclidean* digunakan rumus sebagai berikut :

$$D(ij) = \sqrt{\sum_{i=1}^p |X_{ki} - X_{kj}|^2} \dots\dots\dots(1)$$

Di mana :

$D(ij)$  = jarak data ke ( $i$ ) ke pusat *cluster* ( $j$ )

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

$X_{kj}$  = Titik pusat ( $j$ ) pada atribut ( $k$ )

5. Hitung kembali pusat klaster dengan keanggotaan klaster yang sekarang. Pusat klaster adalah rata-rata dari semua data/objek dalam setiap klasternya.
6. Ulangi lagi langkah (2) dan (3) dengan memakai pusat klaster yang baru. Jika pusat klaster sudah tidak berubah lagi, maka proses pengklasteran selesai atau dinamakan konvergen.

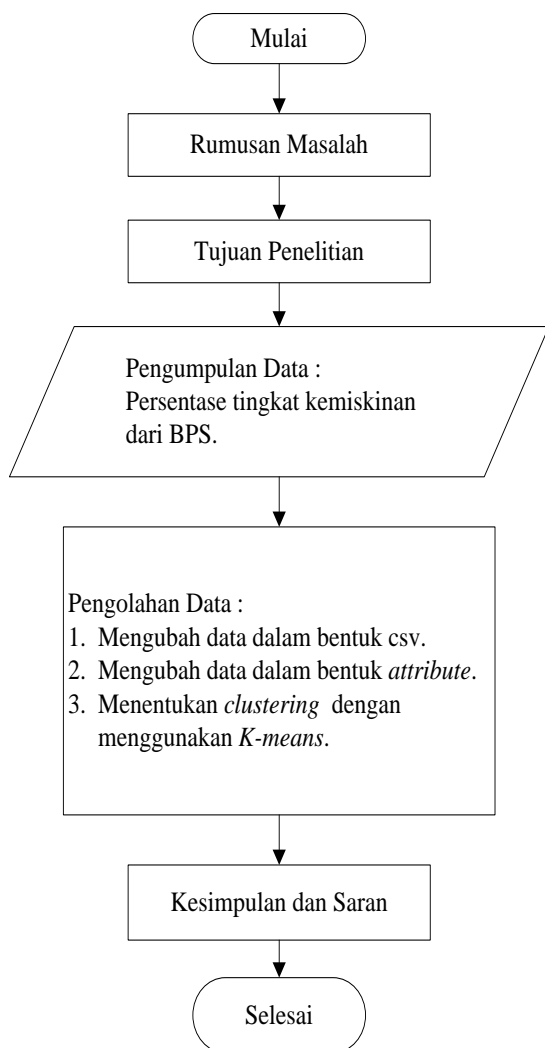
## 3. Metodologi Penelitian

Metode yang digunakan adalah Algoritma *K-means* yang merupakan salah satu algoritma *clustering* (pengelompokan). *K-means clustering* merupakan metode *clustering* non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih *cluster* [10].

Software yang digunakan dalam penelitian ini adalah Weka version 3.8.3. Software ini digunakan untuk membandingkan hasil dengan perhitungan secara teoritis dengan hasil yang didapatkan dengan proses di Weka Interface. Weka Interface, seperti tampak pada Gambar 2 adalah aplikasi open source berbasis bahasa pemrograman Java yang dapat digunakan untuk data mining. Aplikasi ini dikembangkan pertama kali di Selandia Baru. oleh Universitas Waikato Analisis *clustering* dengan algoritma K-Means adalah salah satu fitur data mining yang ada di Weka selain algoritma yang lainnya [7].



Gambar 2. WEKA Interface



**Gambar 1.** Flowchart Metodologi Penelitian

Data yang diperoleh dari BPS memiliki format *excel*, sebelum dapat mengolah data yang kita miliki menggunakan *software* Weka interface 3.8.3, data diharuskan memiliki format tertentu yakni *.Arff*, sehingga kita perlu mengkonversi data tersebut kedalam bentuk *.Arff*. Data yang digunakan untuk pengujian dapat dilihat pada Tabel 1.

**Tabel 1.** Persentase Kemiskinan Setiap Provinsi

Provinsi	Persentase Penduduk Miskin (%)	Garis Kemiskinan (Rp)	P1 (%)	P2 (%)
Aceh	18.58	321 893	3.07	0.83
Sumut	10.41	271 738	1.82	0.5
Sumbar	8	292 052	1.24	0.31
Riau	8.05	310 603	1.13	0.25
Jambi	8.28	273 267	1.37	0.44
Sumsel	13.48	259 668	1.85	0.43
Bengkulu	17.51	283 252	3.05	0.8
Lampung	15.65	263 088	2.53	0.62
Babel	5.37	382 412	0.66	0.14
Riau	6.83	363 450	0.85	0.19
DKI Jakarta	3.7	392 571	0.56	0.15
Jawa Barat	9.89	242 104	1.62	0.42
Jateng	14.98	233 769	2.39	0.57
DIY	15.88	270 110	2.89	0.75
Jatim	13.08	243 783	1.93	0.44
Banten	5.71	251 161	0.95	0.28
Bali	3.95	254 221	0.39	0.07
NTB	18.02	248 758	3.2	0.83
NTT	20.41	222 507	3.47	0.91
Kalbar	7.96	239 162	1.24	0.33
Kalteng	6.19	277 407	1.08	0.27
Kalsel	5.01	269 714	0.76	0.17
Kaltim	6.38	363 887	0.99	0.25
Sulut	7.64	223 883	1.18	0.3
Sulteng	14.94	266 718	2.82	0.82
Sulsel	9.82	195 627	1.68	0.42
Sulawesi Tenggara	13.06	203 333	1.92	0.49
Gorontalo	17.22	212 476	3.21	0.84
Sulawesi Barat	13.01	207 072	1.74	0.4
Maluku	20.76	295 904	4.38	1.31
Maluku Utara	8.06	250 184	0.85	0.14

#### 4. Hasil dan Pembahasan

##### 4.1. Data pengujian

Pada pengujian data input, data yang digunakan adalah berupa table yang memiliki komponen penyusun sebagai berikut:

1. Memiliki 5 atribut yaitu nama provinsi, persentase penduduk miskin, garis kemiskinan, indeks kedalaman kemiskinan (P1) dan indeks keparahan kemiskinan (P2)
2. Jumlah instance adalah 33 provinsi

##### 4.2. Pengolahan data dengan software WEKA

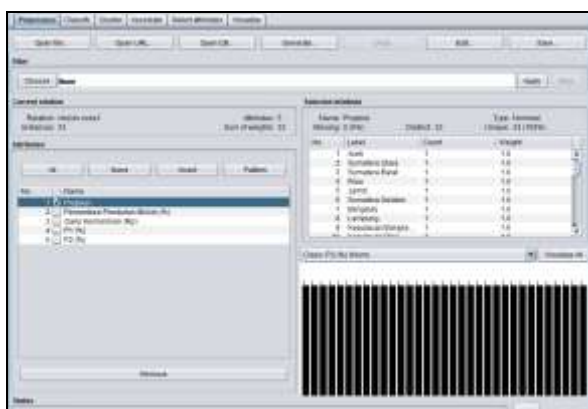
Sebelum data dapat diolah, data terlebih dahulu disiapkan dengan cara mengubah format dari data yang kita miliki, yang kemudian data tersebut akan diolah menggunakan *software* Weka interface 3.8, berikut merupakan langkah - langkah dari preparasi data yang dimiliki.

1. Persiapan Data

Papua Barat	27.04	354 626	5.71	1.71
Papua	30.66	297 502	7.35	2.44

## 2. Mengolah Data Menggunakan Software

Setelah data disesuaikan dengan format .Arff maka langkah selanjutnya ialah data di input kedalam software Weka Interface 3.8.3, pada saat penginputan data, langkah yang dilakukan ialah melakukan *preprocessing* data dengan me-remove satu elemen yang ada di data yang akan kita olah yakni 'provinsi', tujuan menghapus 'provinsi' ialah karena data yang akan dijadikan pertimbangan untuk melakukan *cluster* adalah tidak berdasarkan provinsi melainkan berdasarkan 4 atribut lainnya yaitu persentase penduduk miskin, garis kemiskinan, indeks kedalaman kemiskinan (P1 dan P2). Langkah tersebut dapat dilihat pada Gambar 3.



Gambar 3. Preprocessing Data

Setelah *preprocessing* data dilakukan maka langkah selanjutnya barulah data dapat diolah untuk di *cluster*-kan. Algoritma *cluster* yang dipilih dalam kasus ini algoritma *K-Means*. Kemudian memilih jumlah *cluster* yang diinginkan. Dalam pengkategorian data kemiskinan ini dipilih 3 *cluster* pembagian provinsi yakni *cluster 0* (memiliki tingkat kemiskinan rendah), *cluster 1* (memiliki tingkat kemiskinan sedang), dan terakhir *cluster 3* (memiliki tingkat kemiskinan tinggi). Seperti yang terlihat pada Gambar 4 dibawah ini.



Gambar 4. Process Clustering

## 3. Hasil Clustering

Setelah data di sesuaikan, kemudian algoritma dan jumlah *cluster* yang diinginkan telah dipilih maka data dapat diolah menggunakan software weka interface 3.8.3. Hasil clustering data setelah di *run* menggunakan software weka interface 3.8.3. dapat dilihat pada Gambar 5 dan 6.

Dapat terlihat setelah data di input dan dijalankan dengan software didapatkan hasil 3 *cluster* seperti yang diinginkan dengan hasil yang diperoleh terdapat 9 provinsi yang dikategorikan sebagai *Cluster 0* (Memiliki tingkat kemiskinan rendah), 21 provinsi dikategorikan sebagai *Cluster 1* (Memiliki tingkat kemiskinan sedang), dan 3 provinsi dikategorikan sebagai *Cluster 2* (Memiliki tingkat kemiskinan tinggi).

## 4. Visualisasi Data

Setelah data diolah kemudian data dapat divisualisasikan menggunakan grafik, tujuan memvisualisasikan data adalah menunjukkan pola penyebaran data yang dimiliki tersebut, Visualisasi hasil clustering dapat dilihat pada Gambar 7.



```
Missing values globally replaced with mean/mode
Final cluster centroids:
Attribute                Full Data      Cluster#
                        (33.0)        0           1           2
=====
Persentase Penduduk Miskin (%) 12.2888    17.0211     8.28      26.1533
Garis Kemiskinan (Rp)         321 893    321 893     271 738   295 904
P1 (%)                        2.1176     2.9589     1.229     5.8133
P2 (%)                        0.5703     0.7744     0.3043     1.82

Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances
0      9 ( 27%)
1     21 ( 64%)
2      3 ( 9%)
```

Gambar 5. Hasil cluster centroids dan clustered Instances dengan Weka

```
Scheme: weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 1.0 -t1 -1.25 -t2 -1.0 -M 3 -A "weka.core.EuclideanDis
Relation: miskin_coba1-weka.filters.unsupervised.attribute.Remove-R1
Instances: 33
Attributes: 4
    Persentase Penduduk Miskin (%)
    Garis Kemiskinan (Rp)
    P1 (%)
    P2 (%)
Test mode: evaluate on training data

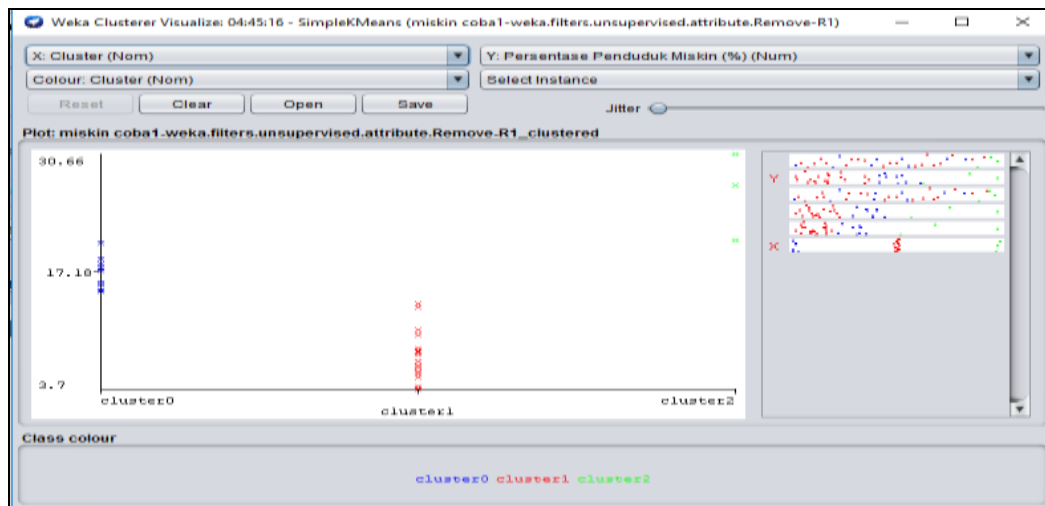
=== Clustering model (full training set) ===

KMeans
=====
Number of iterations: 3
within cluster sum of squared errors: 30.752242452877276

Initial starting points (random):
Cluster 0: 17.22, '212 476', 3.21, 0.84
Cluster 1: 13.08, '243 783', 1.93, 0.44
Cluster 2: 20.76, '295 904', 4.38, 1.31

Missing values globally replaced with mean/mode
```

Gambar 6. Hasil Clustering model



Gambar 7. Visualisasi Hasil Clustering

ARFF-Viewer - C:\Users\Hp\Documents\Kuliah\hasil.csv

File Edit View

hasil.csv

Relation: miskin coba1-weka.filters.unsupervised.attribute.Remove-R1\_clustered

No.	1: Instance_number	2: Persentase Penduduk Miskin (%)	3: Garis Kemiskinan (Rp)	4: P1 (%)	5: P2 (%)	6: Cluster
	Numeric	Numeric	Nominal	Numeric	Numeric	Nominal
1	0.0	18.58	321 893	3.07	0.83	cluster0
2	1.0	10.41	271 738	1.82	0.5	cluster1
3	2.0	8.0	292 052	1.24	0.31	cluster1
4	3.0	8.05	310 603	1.13	0.25	cluster1
5	4.0	8.28	273 267	1.37	0.44	cluster1
6	5.0	13.48	259 668	1.85	0.43	cluster1
7	6.0	17.51	283 252	3.05	0.8	cluster0
8	7.0	15.65	263 088	2.53	0.62	cluster0
9	8.0	5.37	382 412	0.66	0.14	cluster1
10	9.0	6.83	363 450	0.85	0.19	cluster1
11	10.0	3.7	392 571	0.56	0.15	cluster1
12	11.0	9.89	242 104	1.62	0.42	cluster1
13	12.0	14.98	233 769	2.39	0.57	cluster0
14	13.0	15.88	270 110	2.89	0.75	cluster0
15	14.0	13.08	243 783	1.93	0.44	cluster1
16	15.0	5.71	251 161	0.95	0.28	cluster1
17	16.0	3.95	254 221	0.39	0.07	cluster1
18	17.0	18.02	248 758	3.2	0.83	cluster0
19	18.0	20.41	222 507	3.47	0.91	cluster0
20	19.0	7.96	239 162	1.24	0.33	cluster1
21	20.0	6.19	277 407	1.08	0.27	cluster1
22	21.0	5.01	269 714	0.76	0.17	cluster1
23	22.0	6.38	363 887	0.99	0.25	cluster1
24	23.0	7.64	223 883	1.18	0.3	cluster1
25	24.0	14.94	266 718	2.82	0.82	cluster0
26	25.0	9.82	195 627	1.68	0.42	cluster1
27	26.0	13.06	203 333	1.92	0.49	cluster1
28	27.0	17.22	212 476	3.21	0.84	cluster0
29	28.0	13.01	207 072	1.74	0.4	cluster1
30	29.0	20.76	295 904	4.38	1.31	cluster2
31	30.0	8.06	250 184	0.85	0.14	cluster1
32	31.0	27.04	354 626	5.71	1.71	cluster2
33	32.0	30.66	297 502	7.35	2.44	cluster2

Gambar 8. Penyimpanan dan Pengelompokan Data Hasil Clustering

## 5. Penyimpanan Data

Setelah data selesai diolah maka langkah berikutnya ialah menyimpan hasil *clustering* tersebut kedalam format .Arff terlebih dahulu, yang kemudian disimpan kembali dalam format .csv seperti yang dapat dilihat pada Gambar 8.

Tujuan dilakukannya penyimpanan hasil dari *clustering* yang telah kita lakukan ialah agar data yang telah ter-*cluster* tersebut dapat di bandingkan kembali hasil pengelompokan kedalam data .csv yang kita miliki sehingga akan terlihat provinsi manakah yang masuk kedalam *cluster 0*, *cluster 1*, ataupun *cluster 2*.

## 5. Kesimpulan dan saran

Berdasarkan pada hasil penelitian dengan bantuan *software* weka interface 3.8.3 menggunakan algoritma K-Means didapatkan 3 cluster provinsi di Indonesia berdasarkan tingkat kemiskinannya yaitu *Cluster 0* (provinsi dengan tingkat kemiskinan rendah), *Cluster 1* (provinsi tingkat kemiskinan sedang), dan *Cluster 2* (provinsi dengan tingkat kemiskinan tinggi). Terdapat 9 provinsi yang termasuk dalam kategori provinsi dengan tingkat kemiskinan rendah (*Cluster 0*) yaitu Aceh, Bengkulu, Lampung, Jawa Tengah, DI Yogyakarta, NTB, NTT, Sulawesi Tengah, dan Gorontalo. Sedangkan provinsi yang termasuk dalam kategori provinsi dengan tingkat kemiskinan sedang (*Cluster 1*) terdapat 21 provinsi dan provinsi yang termasuk dalam kategori provinsi dengan tingkat kemiskinan tinggi (*Cluster 2*) terdapat 3 provinsi yaitu Maluku, Papua Barat dan Papua.

Penelitian ini diharapkan menjadi dasar bagi pemerintah dalam memberikan kebijakan atau pendekatan yang tepat dan cepat untuk mengatasi kemiskinan yang terjadi dengan diketahuinya karakteristik di setiap provinsi. Untuk penelitian berikutnya akan lebih baik apabila ditambahkan metode lain untuk menghitung dan meningkatkan kevalidan dari hasil analisis cluster.

## Daftar Pustaka:

- [1] N. I. Febianto and N. Palasara, "Analisa Clustering K-Means Pada Data Informasi Kemiskinan Di Jawa Barat Tahun 2018," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 8, no. 2, p. 130, 2019, doi: 10.32736/sisfokom.v8i2.653.
- [2] R. Harun, K. C. Pelangi, and Y. Lasena, "PENERAPAN DATA MINING UNTUK MENENTUKAN POTENSI HUJAN HARIAN DENGAN MENGGUNAKAN ALGORITMA K NEAREST NEIGHBOR (KNN)," *J. Manaj. Inform. dan Sist. Inf.*, vol. 3, no. 1, pp. 8–15, 2020, [Online]. Available: <http://e-journal.stmiklombok.ac.id/index.php/misi/article/download/125/84>.
- [3] H. Garcia-Molina, *Database systems: the complete book*. Pearson Education India, 2008.
- [4] K. L. Du, "Clustering: A neural network approach," *Neural Networks*, vol. 23, no. 1, pp. 89–107, 2010, doi: 10.1016/j.neunet.2009.08.007.
- [5] F. Fajrianti, M. N. Bustan, and M. A. Tiro, "Penggunaan Analisis Cluster K-Means dan Analisis Diskriminan Dalam Pengelompokan Desa Miskin di Kabupaten Pangkep," *VARIANSI J. Stat. Its Appl. Teach. Res.*, vol. 1, no. 2, 2019.
- [6] G. F. Mandias, G. A. Sandag, S. Susanti, and H. R. Musak, "Penerapan Algoritma K-Means Untuk Analisis Prestasi Akademik Mahasiswa Fakultas Ilmu Komputer Universitas Klabat," *CogITo Smart J.*, vol. 3, no. 2, p. 230, 2017, doi: 10.31154/cogito.v3i2.72.230-239.
- [7] R. A. Asroni, "Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang," *Ilm. Semesta Tek.*, vol. 18, no. 1, pp. 76–82, 2015, doi: 10.1038/hdy.2009.180.
- [8] B. Santosa and A. Umam, *Data Mining dan Big Data Analytics: Teori dan Implementasi Menggunakan Python & Apache Spark*. Penebar Media Pustaka, 2018.
- [9] E. Prasetyo, "Data mining konsep dan aplikasi menggunakan matlab," *Yogyakarta Andi*, 2012.
- [10] J. Jaroji, D. Danuri, and F. P. Putra, "K-Means Untuk Menentukan Calon Penerima Beasiswa Bidik Misi Di Polbeng," *INOVTEK Polbeng - Seri Inform.*, vol. 1, no. 1, p. 87, 2016, doi: 10.35314/isi.v1i1.129.
- [11] L. Mutawalli, M. T. A. Zaen, and I. F. Suhriani, "Sistem Identifikasi Persebaran Pecemaran Air Oleh Limbah di Indonesia Menggunakan Average Linkage Dan K-Mean Cluster," *MISI (Jurnal Manaj. Inform. Sist. Informasi)*, vol. 1, no. 2, pp. 36–42, 2018.