

PENERAPAN DATA MINING MENENTUKAN KELOMPOK PRIORITAS PENERIMA BANTUAN BERAS RASTRA DENGAN *CLUSTERING K-MEANS*

Tia Noviana¹, Jasmir², Yudi Novianto³

Program Studi Teknik Informatika, STIKOM Dinamika Bangsa, Jambi

Jl. Jendral Sudirman Thehok – Jambi

E-mail: ¹ tianoviana09@gmail.com, ²ijay_jasmir@yahoo.com, ³yudinovianto@stikom-db.ac.id

Abstract

Data mining which is a process of using statistical techniques, mathematics, artificial intelligence, and machine learning to extract and identify useful information and related knowledge from various large databases. Every subdistrict in Jambi Province has the opportunity to manage the assistance of the Prosperous Rice (rastra). In addition to the level of accuracy of the data, it also requires efficiency in processing data from beneficiaries. There is a complexity in processing data so far, namely determining the population which is the top priority for getting assistance amid the many population data in Muara Papalik District. In carrying out the analysis, the author uses population data from 2017 to 2018 then is presented in the Arff format. In carrying out the analysis, the authors used the WEKA tools with Clustering K-means method with 5 attributes and 3 clusters namely with the results of the first cluster of 46 population data with a percentage of 21%, the second cluster with 143 population data and a 66% percentage result and the third cluster with the results of 28 population data with a percentage of 13%.

Keywords: Data Mining, Rice Rastra, Priority Group, Clustering, K-Means

Abstrak

Data Mining yang merupakan proses menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Setiap Kecamatan di Provinsi Jambi mendapatkan kesempatan untuk mengelola bantuan Beras Sejahtera (rastra). Selain tingkat akurasi data juga dibutuhkan efisiensi waktu pengolahan data penerima bantuan. Terdapat kerumitan dalam pengolahan data selama ini, yaitu menentukan penduduk yang menjadi prioritas utama untuk mendapatkan bantuan ditengah banyaknya data penduduk di Kecamatan Muara Papalik. Dalam melakukan analisis, penulis menggunakan data penduduk dari tahun 2017 sampai tahun 2018 kemudian disajikan kedalam format Arff. Dalam melakukan analisis, penulis menggunakan alat bantu tools WEKA dengan metode *Clustering K-means* dengan 5 atribut dan 3 *cluster* yaitu dengan hasil cluster pertama sebanyak 46 data penduduk dengan hasil presentase 21%, *Cluster* kedua dengan hasil 143 data penduduk dan hasil presentase 66%, dan *Cluster* ketiga dengan hasil 28 data penduduk dengan hasil presentase 13%.

Kata kunci: Data Mining, Beras Rastra, Kelompok Prioritas, *Clustering, K-Means*

1. Pendahuluan

Seiring dengan perkembangan teknologi yang semakin pesat, kehidupan di dunia dipermudah dengan adanya penerapan teknologi tersebut. Salah satu contoh dari penerapan teknologi tersebut adalah penerapan sistem terkomputerisasi. Dengan adanya sistem tersebut, suatu pekerjaan dapat dilakukan dengan cepat, efisien dan efektif sehingga dapat meminimalisir kesalahan yang terjadi.

Dengan perkembangan teknologi ini, dikembangkan pula suatu teknologi yang mampu mengadopsi proses dan cara berpikir manusia yaitu teknologi *Data Mining* yang merupakan proses menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Tujuan utama *data*

mining adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki. Sedangkan *K-Means* merupakan salah satu metode pengelompokan data non hierarki yang berusaha untuk mempartisi data yang ada kedalam bentuk dua atau lebih kelompok, sehingga data berkarakteristik sama dimasukkan kedalam satu kelompok yang sama dan data yang berkarakteristik lain dimasukkan ke dalam kelompok lain.

Program subsidi beras bagi masyarakat berpendapatan rendah adalah program nasional lintas sektoral baik horizontal maupun vertikal, yang bertujuan untuk membantu mencukupi kebutuhan beras masyarakat rendah. Saat ini proses penentuan penerimaan bantuan beras sejahtera (rastra) di Kecamatan Muara Papalik masih dilakukan secara manual yang memungkinkan terjadi kerumitan dalam pengolahan data, yaitu menentukan penduduk miskin yang menjadi prioritas utama untuk mendapatkan bantuan ditengah banyaknya data penduduk miskin di Kecamatan Muara Papalik. Kerumitan tersebut muncul karena banyaknya jumlah penduduk, maka dari itu harus dipertimbangkan dengan beberapa kriteria diantaranya Jumlah Tanggungan, Jumlah Individu Dalam Sebuah Keluarga, Kepala Rumah Tangga, Kondisi Rumah, Jumlah Penghasilan, dan Status Pemilik. Nilai dari setiap kriteria tersebut menjadi patokan untuk menyeleksi penduduk yang menjadi prioritas utama untuk mendapatkan bantuan.

Dengan permasalahan diatas metode *clustering k-mean* sangat tepat digunakan untuk menghasilkan *knowledge* diantara ratusan penduduk miskin di Kecamatan Muara Papalik.

Hal inilah yang melatar belakangi penulis untuk melakukan penelitian guna memberi solusi terhadap masalah yang terjadi dengan mengangkat judul “Penerepan Data Mining Untuk Menentukan Kelompok Prioritas Penerima Bantuan Beras Rastra Menggunakan Metode *Clustering K-Means* Pada Kecamatan Muara Papalik”.

1.1 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dibahas sebelumnya maka dapat dirumuskan permasalahan yang akan diteliti yaitu bagaimana penerapan data mining dalam menentukan penduduk penerima bantuan beras sejahtera (rastra) dengan metode *clustering k-means* ?

1.2 Batasan Masalah

Agar pembahasan ini tidak menyimpang dari apa yang telah dirumuskan, maka dibutuhkan batasan-batasan. Batasan-batasan dalam penelitian ini adalah sebagai berikut:

1. Analisis menggunakan teknik *data mining* dengan metode *clustering k-means*.
2. Atribut yang digunakan yaitu jumlah tanggungan, jumlah individu dalam keluarga, kepala rumah tangga, kondisi rumah, jumlah penghasilan dan status pemilik.
3. Data *training* menggunakan 100% data penduduk.
4. Informasi yang dihasilkan berupa data penerima bantuan beras sejahtera (rastra) dengan dua kategori yaitu ya atau tidak.
5. Pengujian hasil analisis menggunakan *tools* WEKA.

1.3 Tujuan Dan Manfaat Penelitian

1.3.1 Tujuan

Adapun tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Menganalisa data-data penduduk di Kecamatan Muara Papalik dengan metode *clustering k-means* sehingga dapat menentukan penduduk yang mendapatkan bantuan diperiode berikutnya.
2. Mendapatkan akurasi yang baik dari nilai dan kriteria penduduk.

1.3.2 Manfaat Penelitian

Adapun manfaat yang dapat diperoleh dalam penelitian ini adalah:

1. Dapat mengetahui penerima bantuan beras sejahtera (rastra).
2. Dapat menggunakan hasil analisis untuk menentukan strategi dalam meningkatkan kualitas dan kuantitas penerima bantuan unruk periode selanjutnya.
3. Penulis dapat menambah ilmu dan wawasan baru mengenai analisis penerima bantuan.
4. Dapat digunakan sebagai acuan untuk melakukan penelitian berikutnya.

2. Tinjauan Pustaka

2.1 Data Mining

Data Mining adalah satu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database, *Data Mining* adalah teknik memanfaatkan data dalam jumlah yang besar untuk memperoleh

informasi berharga yang sebelumnya tidak diketahui dan dapat dimanfaatkan untuk pengambilan keputusan penting. [1]

2.2 Clustering

Clustering atau klasterisasi adalah metode pengelompokan data. *Clustering* adalah sebuah proses untuk mengelompokkan data ke dalam beberapa *cluster* atau kelompok sehingga data dalam satu cluster memiliki tingkat kemiripan yang maksimum dan data antar *cluster* memiliki kemiripan yang minimum [2].

2.3 K-Means

Algoritma K-Means merupakan salah satu algoritma dengan partitional, karena K-Means didasarkan pada penentuan jumlah awal kelompok dengan mendefinisikan nilai centroid awalnya. Algoritma K-Means menggunakan proses secara berulang-ulang untuk mendapatkan basis data cluster. Dibutuhkan jumlah cluster awal yang diinginkan sebagai masukan dan menghasilkan jumlah cluster akhir sebagai output. Jika algoritma diperlukan untuk menghasilkan cluster K maka akan ada K awal dan K akhir. Metode K-Means akan memilih pola k sebagai titik awal centroid secara acak. Jumlah iterasi untuk mencapai cluster centroid akan dipengaruhi oleh calon cluster centroid awal secara random dimana jika posisi centroid baru tidak berubah. Nilai K yang dipilih menjadi pusat awal, akan dihitung dengan menggunakan rumus Euclidean Distance yaitu mencari jarak terdekat antara titik centroid dengan data/objek. Data yang memiliki jarak pendek atau terdekat dengan centroid akan membentuk sebuah cluster. [3]

Dasar algoritma *k-means* adalah sebagai berikut:

1. Tentukan K sebagai jumlah *cluster* yang akan dibentuk
2. Tentukan k Centroid (titik pusat cluster) awal secara random/acak.

$$v = \frac{\sum_{i=1}^n x_i}{n} \quad ; i = 1, 2, 3, \dots, n \quad (1)$$

Dimana; v : centroid pada cluster

x_i : objek ke-i

n : banyaknya objek/jumlah objek yang menjadi anggota cluster.

3. Hitung jarak setiap objek ke masing-masing centroid dari masing-masing cluster. Untuk menghitung jarak antara objek dengan centroid dapat menggunakan Euclidian Distance.

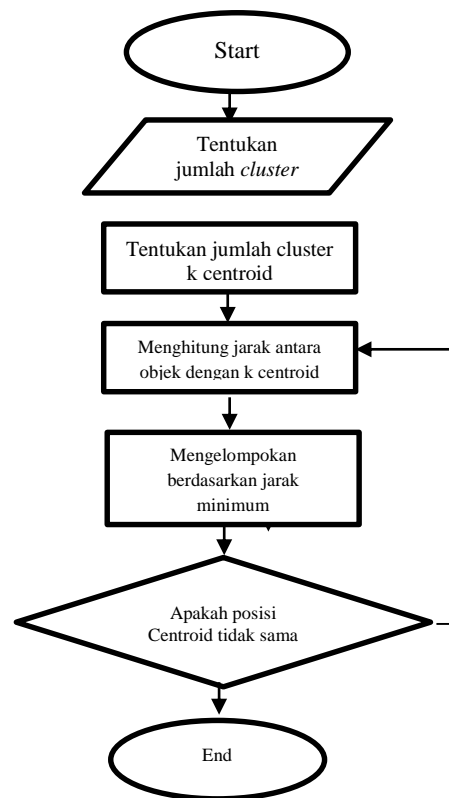
$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad ; i = 1, 2, 3, \dots, n \quad (2)$$

Dimana; x_i : objek x ke-i

y_i : daya y ke-i

n : banyaknya objek

4. Alokasikan masing-masing objek ke dalam centroid yang paling dekat.
5. Lakukan iterasi, kemudian tentukan posisi centroid baru dengan menggunakan persamaan.
6. Ulangi langkah 3 jika posisi centroid baru tidak sama.



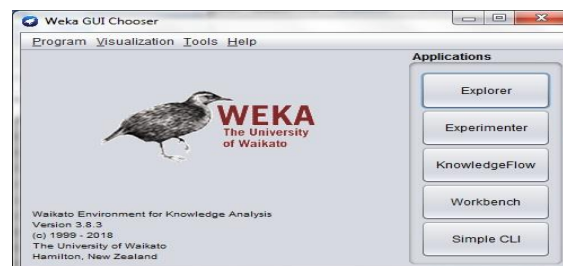
Gambar 1. Flowcart Algoritma K-means

2.4 Beras Sejahtera (RASTRA)

Beras Rastrea merupakan subsidi pangan pokok dalam bentuk beras yang diperuntukan bagi keluarga miskin sebagai upaya untuk meningkatkan ketahanan pangan dari memberikan perlindungan pada keluarga miskin [4].

2.5 WEKA

Waikato Environment for Knowledge Analysis (WEKA) merupakan perangkat lunak pembelajaran mesin yang populer yang ditulis dalam bahasa pemrograman java. WEKA dikembangkan di Universitas Waikato, Selandia Baru. WEKA berisikan kumpulan algoritma beserta visualisasinya untuk analisis data dan pemodelan prediktif. Algoritma-algoritma pembelajaran mesin pada WEKA dapat dimanfaatkan untuk pemecahan masalah dibidang *data mining*. WEKA versi asli awalnya dirancang untuk menganalisis data dari domain pertanian, tetapi WEKA versi lengkap berbasis java (versi 3), yang mulai dibangun pada tahun 1997, yang sekarang dapat digunakan untuk menganalisis data dari berbagai domain, khususnya untuk pendidikan dan penelitian. WEKA memiliki implementasi semua teknik pembelajaran untuk klasifikasi dan regresi, yaitu *decision trees*, *rules set*, pengklasifikasian *teorema bayes*, *Support Vector Machines (SVM)*, logistik dan linier, *multi layers perceptrons* dan metode *nearest neighbor*. [5].

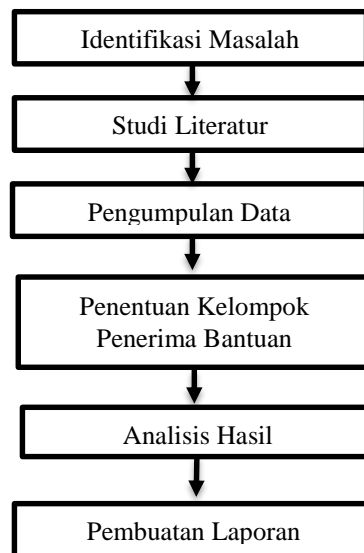


Gambar 2. Tampilan Utama WEKA

3. Metodologi

3.1 Kerangka Kerja Penelitian

Kerangka kerja penelitian merupakan tahapan-tahapan yang dilakukan selama mengerjakan penelitian. Kerangka kerja penelitian dibuat agar mempermudah pencapaian hasil penelitian, dapat menyelesaikan penelitian tepat waktu dan penelitian dapat berjalan sesuai dengan yang diharapkan. Adapun kerangka kerja penelitian yang digunakan adalah sebagai berikut:

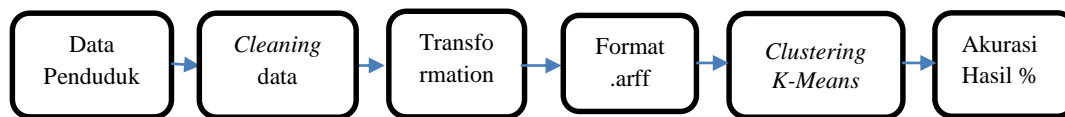


Gambar 3. Kerangka Kerja Penelitian

Keterangan:

1. Identifikasi Masalah: Pada tahap ini penulis mengidentifikasi hal apa saja yang menjadi permasalahan *data mining* mengenai penentuan kelompok prioritas penerima bantuan beras sejahtera (rastra).
2. Studi Literatur: Pada tahap ini penulis melakukan pencarian terhadap landasan-landasan teori yang diperoleh dari berbagai sumber seperti buku, jurnal ilmiah dan juga referensi lainnya.
3. Pengumpulan data: Pada tahap ini penulis melakukan pengumpulan data dan informasi yang dibutuhkan dalam penelitian. Penelitian ini dilakukan dengan beberapa metode yaitu pengamatan, wawancara, dan penelitian kepustakaan (*Library Research*).
 - a. Pengamatan (*Observasi*)
Metode ini dilakukan dengan pengamatan secara langsung pada objek yang akan diteliti. Dalam hal ini penulis mengamati bagaimana pengelolaan pada data-data penduduk, sehingga penulis mendapatkan *softcopy* data-data penduduk tahun 2017 hingga 2018. Data penduduk ini diberikan oleh Staf Kantor Kecamatan Muara Papalik. Adapun atribut data penduduk yang diperoleh Jumlah Tanggungan, Kondisi Rumah, Status Pemilik, Jumlah Penghasilan, dan Status Kesejahteraan.
 - b. Wawancara
Pada metode ini penulis melakukan wawancara singkat secara langsung dengan kepala Staf Tata Usaha untuk memperoleh data dan informasi akurat mengenai data penduduk yang penulis butuhkan.
 - c. Penelitian Kepustakaan (*Library Research*)
Untuk menunjang penelitian yang akan dilakukan, penulis melakukan *library research* yakni dengan mencari data-data dari buku maupun jurnal penelitian sejenis yang berhubungan dengan metode yang penulis gunakan dalam penelitian ini yaitu metode *Clustering K-Means*.
2. Penentuan kelompok prioritas penerima bantuan beras sejahtera (rastra): Pengklasifikasian data penduduk dilakukan dengan beberapa tahapan yaitu Data Penduduk, *Cleaning* data, *Transformation*, Format arff, Training, *Clustering K-means*, dan Akurasi Hasil.

Adapun alur proses penentuan penerima bantuan dapat dilihat gambar dibawah ini:



Gambar 4. Alur Proses Clustering K-Means

Berikut ini penjelasan mengenai tahapan-tahapan pada alur proses penentuan penerima bantuan beras sejahtera.

A. Data Penduduk

Data Penduduk merupakan data-data yang dimiliki oleh Penduduk pada tahun 2017 sampai 2018. Dari sekian banyak atribut data mahasiswa, atribut data yang digunakan yaitu Jumlah Tanggungan, Jumlah Individu Dalam Sebuah Keluarga, Kepala Rumah Tangga, Kondisi Rumah, Jumlah Penghasilan, dan Status Pemilik.

B. *Cleaning Data*

Cleaning data merupakan *preproses* dimana pada tahap ini dilakukan pembersihan atau pemisahan data dari *noise* data dan data yang tidak konsisten seperti Penduduk yang memiliki atribut tidak lengkap. Kemudian hasil dari *cleaning data* ini disusun dalam bentuk format arff pada tools Notepad++.

C. *Transformation*

Transformation merupakan metode yang digunakan dalam *cleaning data* pada tahap ini data yang telah dipisah dan dipilih lalu diubah kedalam *numerical*.

D. Format arff

WEKA mendukung beberapa format file dalam input-nya diantaranya adalah format arff. *Attribute-Relation File Format* (ARFF) merupakan tipe *file text* yang berisi berbagai *instance* data yang berhubungan dengan suatu set atribut data. Kemudian data-data ini disusun dengan menggunakan tanda koma (,) sebagai pemisah data per-atributnya.

E. *Clustering K-means*

Pada tahapan ini, dilakukan proses penentuan data menggunakan tools WEKA terhadap data penduduk yang sudah *testing* menggunakan model *classifier* yang sudah dibentuk.

F. Akurasi Hasil

Akurasi Hasil menunjukkan kedekatan hasil pengukuran dengan nilai sesungguhnya.

3. Analisis Hasil : Dari hasil analisis akan dinilai keakuratan data model yang sudah di dapatkan dengan perhitungan *clustering k-means*.

4. Pembuatan Laporan: Setelah semua tahapan penelitian dilakukan, maka akan dibuat laporan sebagai dokumentasi penelitian agar dapat dimanfaatkan pada waktu yang akan datang. Baik oleh peneliti sendiri maupun peneliti lainnya.

4. Hasil dan Pembahasan

4.1 Gambaran Umum Kantor Kecamatan Muara Papalik

Kecamatan Muara Papalik merupakan pecahan atau pemekaran dari Kecamatan Merlung. Adapun batas – batas wilayah kecamatan Muara Papalik adalah sebelah Utara berbatasan dengan Kecamatan Tungkal Ulu, sebelah Timur berbatasan dengan Kecamatan Merlung, sebelah Selatan berbatasan dengan Kecamatan Batang Hari dan sebelah Barat berbatasan dengan Kabupaten Muaro Jambi. **Luas wilayah Kecamatan Muara Papalik kurang lebih 38.870 Ha atau 388,7 Km²**. Kecamatan Muara Papalik merupakan salah satu dari 13 Kecamatan yang ada dalam Kabupaten Tanjung Jabung Barat. Kecamatan Muara Papalik terdiri dari 1 Kelurahan, 4 desa, 19 dusun. Pusat pemeritahan Kecamatan Muara Papalik terletak di Kelurahan Rantau Badak. Jarak pusat pemerintahan Kecamatan Muara Papalik ke Ibukota Kabupaten kurang lebih 148 Km (Jalan Darat).

4.2 Analisis Sistem yang Sedang Berjalan

Melakukan analisis terlebih dahulu terhadap sistem yang sedang berjalan memang sangat penting sebagai dasar untuk merancang suatu sistem yang tepat guna lebih baik. Hal ini diperlukan sebagai perbandingan antar sistem yang sedang berjalan dengan sistem yang akan dirancang sebagai sistem pembantu dalam training nantinya.

4.3 Solusi Pemecahan Masalah

Berdasarkan analisa sistem yang berjalan, maka penulis melakukan analisis penentuan kelompok prioritas penerima bantuan Beras Sejahtera (Rastra) terhadap data-data penduduk di Kecamatan Muara Papalik pada tahun 2017-2018 agar dapat mengetahui mana penduduk yang menjadi prioritas utama untuk mendapatkan bantuan. Hasil dari penelitian ini diharapkan dapat membantu pihak kecamatan dalam pengambilan keputusan untuk meningkatkan kualitas bantuan dengan tepat dan sebagainya.

4.4 Analisis Data Mining Menggunakan Cluster K-Means

4.4.1 Respresentasi Data

Hasil dari wawancara yang dilakukan maka dapat hasil data dari Kecamatan Muara Papalik yang terdapat 1 Kelurahan dan 9 Desa yaitu Kelurahan Rantau Badak, Desa Dusun Mudo, Desa Intan Jaya/SP.7, Desa Bukit Indah/SP.8, Desa Kemang Manis/SP.9, Desa Rantau Badak Lamo, Desa Sebontang, Desa Sungai Papauh, Desa Sungai Muluk, dan Desa Pematang Balam. Maka didapatkan beberapa data penduduk yang telah dipilih. Atribut yang akan digunakan pada seluruh data penduduk yaitu Jumlah Tanggungan, Kondisi Rumah, Status Pemilik, Jumlah Penghasilan, dan Status Kesejahteraan. Berikut adalah tabel *data training* yang akan digunakan untuk perhitungan *clustering k-means*.

Tabel 1. *Respresentasi Data*

No.	Nama KK	Jumlah Tanggungan	Kondisi Rumah	Status Pemilik	Jumlah Penghasilan	Status Kesejahteraan
1	Marsini	1	1	1	2	2
2	Ahmad Hafiz	1	1	2	1	3
3	Usman Jakfar	1	1	1	1	3
4	Mansyur	6	2	2	3	2
5	Aswin	3	1	1	3	2
6	Sulaiman	2	1	3	2	3
7	Jedan	2	2	1	1	3
8	Eka Kurniawan	2	1	1	3	2
9	Mustam	3	1	1	3	2
10	Pungut	3	1	2	2	2
11	Mirza Erdi	2	1	1	5	1
12	Mustainillah	3	1	1	4	1
13	Junaidi	6	1	1	3	1
14	Arifin Siregar	1	1	1	4	1
15	Herman	2	2	2	1	3

Data Penduduk Tidak Mampu digali dengan disesuaikan kriteria penerima bantuan beras sejahtera. Sehingga *field* yang diolah adalah Jumlah Tanggungan, Kondisi Rumah, Status Pemilik, Jumlah Penghasilan, serta Status Kesejahteraan.

4.4.2 Selection dan Transformation

Penyesuaian format data perlu dilakukan untuk mempermudah pengolahan data berupa nilai sehingga butuh Transformasi data bagi data yang belum dapat diolah kedalam bentuk angka. Dalam hal ini data “keterangan” yang terdapat dalam data tidak mampu harus ditransformasi kedalam bentuk angka, berikut hasil tranformasi teks kedalam angka yang sudah ditentukan pemerintah pada Data Penduduk Miskin Muara Papalik.

Tabel 2. *Transformasi Nilai Untuk Keterangan*

1	JUMLAH TANGGUNGAN			
1	UNTUK 1 ANAK			
2	UNTUK 2 ANAK DAN SETERUSNYA SESUAI DENGAN JUMLAH ANAK			
2	KONDISI RUMAH			
1	BAIK			
2	KURANG BAIK			
3	STATUS PEMILIK			
1	HAK MELIK			
2	NUNDAANG			
3	SEWA			
4	JUMLAH PENGHASILAN			
1	>=1600000			
2	>=2400000			
3	>=2200000			
4	>=4800000			
5	>=6800000			
5	STATUS KESEJAHTERAAN			
1	BAIK			
2	CUKUP			
3	KURANG			

4.4.3 Data Mining

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data. Untuk menghasilkan kelompok penduduk prioritas diantara ratusan data yang ada lebih tepat menggunakan metode Cluster, karena metode ini akan mengelompokkan anggota sejenis dan memisahkan anggota lain tidak sejenis pada kelompok lainnya. Dalam penelitian ini penulis menggunakan algoritma K-Means yang akan diuraikan lebih lanjut pada tahapan dibawah ini.

1. Pembentukan Cluster dan Centroid

Berdasarkan data yang telah digali sesuai dengan kriteria, dapat ditentukan *centroid* sebagai batas *cluster*. Dimana penulis menggunakan 3 *cluster* yang dibentuk, nilai centroid (M) dan cluster tersebut ditentukan secara acak dan dapat dilihat pada gambar berikut:

Tabel 3. Nilai *Centroid* (M) Tahap Awal

CENTROID 1	2	1	1	2	2
CENTROID 2	3	1	1	4	2
CENTROID 3	5	1	1	3	2

No.	Nama KK	Jumlah tanggungan	Kondisi Rumah	Status Pemilik	Jumlah Penghasilan	Status Kesejahteraan	
1	Marsini	1	1	1	2	2	
2	Ahmad Hafiz	1	1	2	1	3	
3	Usman Jakfar	1	1	1	1	3	
4	Mansyur	6	2	2	3	2	
5	Aswin	3	1	1	3	2	
6	Sulaiman	2	1	3	2	3	
7	Jedan	2	2	1	1	3	
8	Eka Kurniaswan	2	1	1	3	2	
9	Mustam	3	1	1	3	2	
10	Pungut	3	1	2	2	2	
11	Mirza Erdi	2	1	1	5	1	
12	Mustainbillah	3	1	1	4	1	
13	Junaidi	6	1	1	3	1	
14	Arifin Siregar	1	1	1	4	1	
15	Herman	2	2	2	1	3	
16	Digos Putra	2	1	1	2	2	
17	M Ridho	3	1	1	3	2	
18	Nurainah	2	1	1	2	2	MI
19	RD. Andi Jefri	2	1	1	3	2	
20	Sanusi	1	1	1	4	1	
21	Ali Murtadi	3	1	1	4	1	
22	Halik	2	2	1	2	3	
23	Fauzi	3	1	1	5	1	
24	Ibrahim	1	1	1	3	2	
25	Miyo	3	1	1	3	2	
26	Abdullah	3	1	1	5	1	
27	Abdullah Aap	3	1	1	4	1	
28	Edison	1	1	1	3	2	
29	Ali Amran	3	1	1	4	1	
30	M. Sejahtera	3	1	1	4	1	
31	Erwan	3	1	1	3	2	
32	Sefizal	3	1	1	3	2	
33	Susanti	1	1	1	2	3	

Gambar 5. Pemilihan Cluster dan Centroid

2. Perhitungan Jarak Pusat Cluster

Untuk mengukur jarak dengan pusat *Cluster* digunakan *Euclidian Distance*, kemudian akan didapatkan matriks jarak yaitu C1, C2, dan C3 sebagai berikut:

Rumus *Euclidian Distance*:

$$d(x, y) = |x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3)$$

Jarak "MARSINI" ke *Centroid* I =

$$\sqrt{(1 - 2)^2 + (1 - 1)^2 + (1 - 1)^2 + (2 - 2)^2 + (2 - 2)^2} = 1.0000$$

Jarak "AHMAD HAFIZ" ke *Centroid 2* =

$$\sqrt{(1-3)^2 + (1-1)^2 + (2-1)^2 + (1-4)^2 + (3-2)^2} = 3.8730$$

Jarak "USMAN JAKFAR" ke *Centroid 3* =

$$\sqrt{(1-5)^2 + (1-1)^2 + (1-1)^2 + (1-3)^2 + (3-2)^2} = 4.5826$$

Dengan perhitungan jarak seluruh data terhadap 3 *centroid* yang sudah ditetapkan, maka dapat dilihat jarak semua nama yang menjadi sampel ke *Centroid 1, 2, dan 3* terdapat pada tabel berikut:

Tabel 4. Tabel Jarak Pusat data ke *Centroid 1, 2, dan 3*

No.	Nama KK	C1	C2	C3
1	Marsini	1,0000	2,8284	4,1231
2	Ahmad Hafiz	2,0000	3,8730	4,6904
3	Usman Jakfar	1,7321	3,7417	4,5826
4	Mansyur	4,3589	3,4641	1,7321
5	Aswin	1,4142	1,0000	2,0000
6	Sulaiman	2,2361	3,1623	3,8730
7	Jedan	1,7321	3,4641	3,8730
8	Eka Kurniawan	1,0000	1,4142	3,0000
9	Mustam	1,4142	1,0000	2,0000
10	Pungut	1,4142	2,2361	2,4495

4.4.4 Pra Proses Data

Tahapan ini melakukan beberapa persiapan proses data. Persiapan proses data tersebut yaitu:

4.4.4.1 Data Mentah

Data mentah ini merupakan semua data penduduk Kecamatan Muara Papalik sebelum dilakukannya penyeleksian data. Adapun data tersebut yaitu:

No	Kode Provinsi	Kode Kabupaten/Kota	Kode Kecamatan	Kode Desa/Kelurahan	Provinsi	Nama Kabupaten/Kota	Nama Kecamatan	Nama Desa/Kelurahan	Nama KK	Jumlah Tanggungan	Kondisi Rumah	Status Pemilik	Jumlah Penghasilan	Status Kesejahteraan
1	15	06	09	1001	JAMBI	TANIUNG JABUNG BARAT	MUARA PAPALIK	RANTAU BADAK	Marsini	1	Baik	Hak Milik	2.000.000	Cukup
2	15	06	09	1001	JAMBI	TANIUNG JABUNG BARAT	MUARA PAPALIK	RANTAU BADAK	Ahmad Hafiz	1	Baik	Numpang	1.200.000	Kurang
3	15	06	09	1001	JAMBI	TANIUNG JABUNG BARAT	MUARA PAPALIK	RANTAU BADAK	Usman Jakfar	1	Baik	Hak Milik	800.000	Kurang
4	15	06	09	1001	JAMBI	TANIUNG JABUNG BARAT	MUARA PAPALIK	RANTAU BADAK	Mansyur	6	Kurang Baik	Numpang	2.500.000	Cukup
5	15	06	09	1001	JAMBI	TANIUNG JABUNG BARAT	MUARA PAPALIK	RANTAU BADAK	Aswin	3	Baik	Hak Milik	2.500.000	Cukup
6	15	06	09	1001	JAMBI	TANIUNG JABUNG BARAT	MUARA PAPALIK	RANTAU BADAK	Sulaiman	2	Baik	Sewa	1.900.000	Kurang
7	15	06	09	1001	JAMBI	TANIUNG JABUNG BARAT	MUARA PAPALIK	RANTAU BADAK	Jedan	2	Kurang Baik	Hak Milik	1.500.000	Kurang
8	15	06	09	1001	JAMBI	TANIUNG JABUNG BARAT	MUARA PAPALIK	RANTAU BADAK	Eka Kurniawan	2	Baik	Hak Milik	2.800.000	Cukup
9	15	06	09	1001	JAMBI	TANIUNG JABUNG BARAT	MUARA PAPALIK	RANTAU BADAK	Mustam	3	Baik	Hak Milik	3.000.000	Cukup
10	15	06	09	1001	JAMBI	TANIUNG JABUNG BARAT	MUARA PAPALIK	RANTAU BADAK	Pungut	3	Baik	Numpang	2.100.000	Cukup
11	15	06	09	1001	JAMBI	TANIUNG JABUNG BARAT	MUARA PAPALIK	RANTAU BADAK	Mirza Erdi	2	Baik	Hak Milik	5.000.000	Baik
12	15	06	09	1001	JAMBI	TANIUNG JABUNG BARAT	MUARA PAPALIK	RANTAU BADAK	Mustainbillah	3	Baik	Hak Milik	4.000.000	Baik
13	15	06	09	1001	JAMBI	TANIUNG JABUNG BARAT	MUARA PAPALIK	RANTAU BADAK	Junaidi	6	Baik	Hak Milik	6.800.000	Baik
14	15	06	09	1001	JAMBI	TANIUNG JABUNG BARAT	MUARA PAPALIK	RANTAU BADAK	Arifin Siregar	1	Baik	Hak Milik	3.800.000	Baik
15	15	06	09	1001	JAMBI	TANIUNG JABUNG BARAT	MUARA PAPALIK	RANTAU BADAK	Herman	2	Kurang Baik	Numpang	1.500.000	Kurang
16	15	06	09	1001	JAMBI	TANIUNG JABUNG BARAT	MUARA PAPALIK	RANTAU BADAK	Digos Putra	2	Baik	Hak Milik	2.000.000	Cukup
17	15	06	09	1001	JAMBI	TANIUNG JABUNG BARAT	MUARA PAPALIK	RANTAU BADAK	M Ridho	3	Baik	Hak Milik	2.600.000	Cukup
18	15	06	09	1001	JAMBI	TANIUNG JABUNG BARAT	MUARA PAPALIK	RANTAU BADAK	Nurainah	2	Baik	Hak Milik	2.100.000	Cukup
19	15	06	09	1001	JAMBI	TANIUNG JABUNG BARAT	MUARA PAPALIK	RANTAU BADAK	RD. Andri Jefri	2	Baik	Hak Milik	2.700.000	Cukup
20	15	06	09	1001	JAMBI	TANIUNG JABUNG BARAT	MUARA PAPALIK	RANTAU BADAK	Sanusi	1	Baik	Hak Milik	4.500.000	Baik
21	15	06	09	1001	JAMBI	TANIUNG JABUNG BARAT	MUARA PAPALIK	RANTAU BADAK	Ali Murtadi	3	Baik	Hak Milik	3.500.000	Baik
22	15	06	09	1001	JAMBI	TANIUNG JABUNG BARAT	MUARA PAPALIK	RANTAU BADAK	Halik	2	Kurang Baik	Hak Milik	1.700.000	Kurang

Gambar 6. Data Mentah Penduduk Kecamatan Muara Papalik

4.4.4.2 Seleksi Data Mentah

Seleksi data mentah ini merupakan proses penyeleksian data dengan memfokuskan pada atribut-atribut yang digunakan untuk perhitungan seperti jumlah tanggungan dalam keluarga, kondisi rumah, status pemilik, jumlah penghasilan dan status kesejahteraan. Adapun data tersebut yaitu:

No.	Nama KK	Jumlah Tanggungan	kondisi rumah	status pemilik	jumlah penghasilan	status kesejahteraan
1	Marsini	1	1	1	2	2
2	Ahmad Hafiz	1	1	2	1	3
3	Usman Jakfar	1	1	1	1	3
4	Mansyur	6	2	2	3	2
5	Aswin	3	1	1	3	2
6	Sulaiman	2	1	3	2	3
7	Jedan	2	2	1	1	3
8	Eka Kurniawan	2	1	1	3	2
9	Mustam	3	1	1	3	2
10	Pungut	3	1	2	2	2
11	Mirza Erdi	2	1	1	5	1
12	Mustainbillah	3	1	1	4	1
13	Junaidi	6	1	1	0	1
14	Arifin Siregar	1	1	1	4	1
15	Herman	2	2	2	1	3
16	Digos Putra	2	1	1	2	2
17	M Ridho	3	1	1	3	2
18	Nurainah	2	1	1	2	2
19	RD. Andri Jefri	2	1	1	3	2
20	Sanusi	1	1	1	4	1
21	Ali Murtadi	3	1	1	4	1
22	Halik	2	2	1	2	3
23	Fauzi	3	1	1	5	1
24	Ibrahim	1	1	1	3	2
25	Miyo	3	1	1	3	2
26	Abdullah	3	1	1	5	1
27	Abdullah Aap	3	1	1	4	1
28	Edison	1	1	1	3	2
29	Ali Amran	3	1	1	4	1
30	M. Sejahtera	3	1	1	4	1
31	Erwan	3	1	1	3	2
32	Sefrizal	3	1	1	3	2
33	Supardi	1	1	1	2	3
34	Patimah	1	1	1	2	3

1	JUMLAH TANGGUNGAN		
1	UNTUK 1 ANAK		
2	UNTUK 2 ANAK		
	DAN SETERUSNYA SESUAI		
	DENGAN JUMLAH ANAK		
2	KONDISI RUMAH		
1	BAIK		
2	KURANG BAIK		
3	STATUS PEMILIK		
1	HAK MILIK		
2	NUMPANG		
3	SEWA		
4	JUMLAH PENGHASILAN		
1	<1600000		
2	<2400000		
3	<3200000		
4	<4800000		
5	>6800000		
5	STATUS KESEJAHTERAAN		
1	BAIK		
2	CUKUP		
3	KURANG		

Gambar 7. Seleksi Data Mentah

4.4.4.3 Konversi Hasil Data

Hasil dari data tersebut, kemudian dipindahkan ke notepad dengan format .arff. Adapun bentuk data tersebut yaitu:

```

@data
1,Marsini,1,1,1,2,2
2,'Ahmad Hafiz',1,1,2,1,3
3,'Usman Jakfar',1,1,1,1,3
4,Mansyur,6,2,2,3,2
5,Aswin,3,1,1,3,2
6,Sulaiman,2,1,3,2,3
7,Jedan,2,2,1,1,3
8,'Eka Kurniawan',2,1,1,3,2
9,Mustam,3,1,1,3,2
10,Pungut,3,1,2,2,2
11,'Mirza Erdi',2,1,1,5,1
12,Mustainbillah,3,1,1,4,1
13,Junaidi,6,1,1,0,1
14,'Arifin Siregar',1,1,1,4,1
15,Herman,2,2,2,1,3
16,'Digos Putra',2,1,1,2,2
17,'M Ridho',3,1,1,3,2
18,Nurainah,2,1,1,2,2
19,'RD. Andri Jefri',2,1,1,3,2
20,Sanusi,1,1,1,4,1
21,'Ali Murtadi',3,1,1,4,1
22,Halik,2,2,1,2,3
23,Fauzi,3,1,1,5,1
24,Ibrahim,1,1,1,3,2
25,Miyo,3,1,1,3,2
26,Abdullah,3,1,1,5,1
27,Abdullah/Asp,3,1,1,4,1
28,Edison,1,1,1,3,2
29,'Ali Amran',3,1,1,4,1
30,'M. Sejahtera',3,1,1,4,1
31,Erwan,3,1,1,3,2
32,Sefrizal,3,1,1,3,2
33,Supardi,1,1,1,2,3
34,Patimah,1,1,1,2,3
35,Efendi,3,1,1,2,2

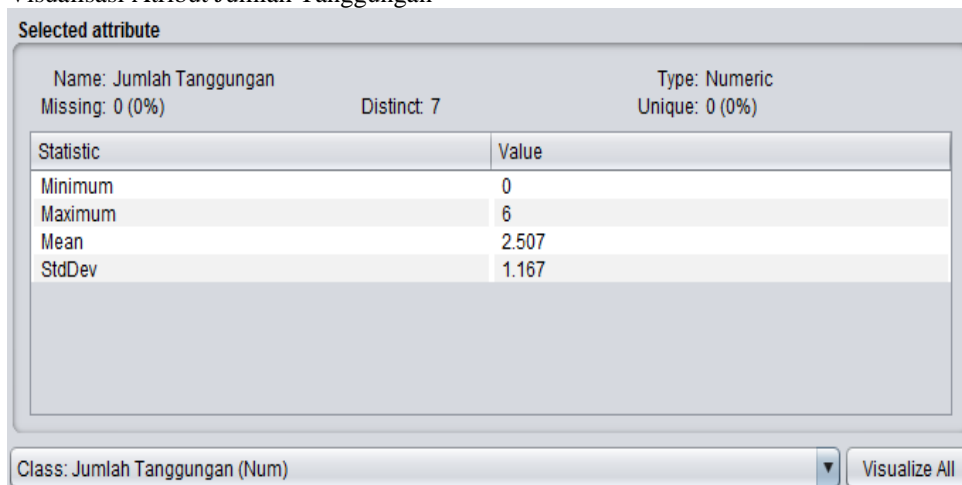
```

Gambar 8. Konversi Hasil Data

4.4.4.4 Hasil Visualisasi Data Setiap Atribut Dengan Menggunakan Weka

Berikut merupakan bentuk visualisasi atribut beras dengan menggunakan tools WEKA, yaitu:

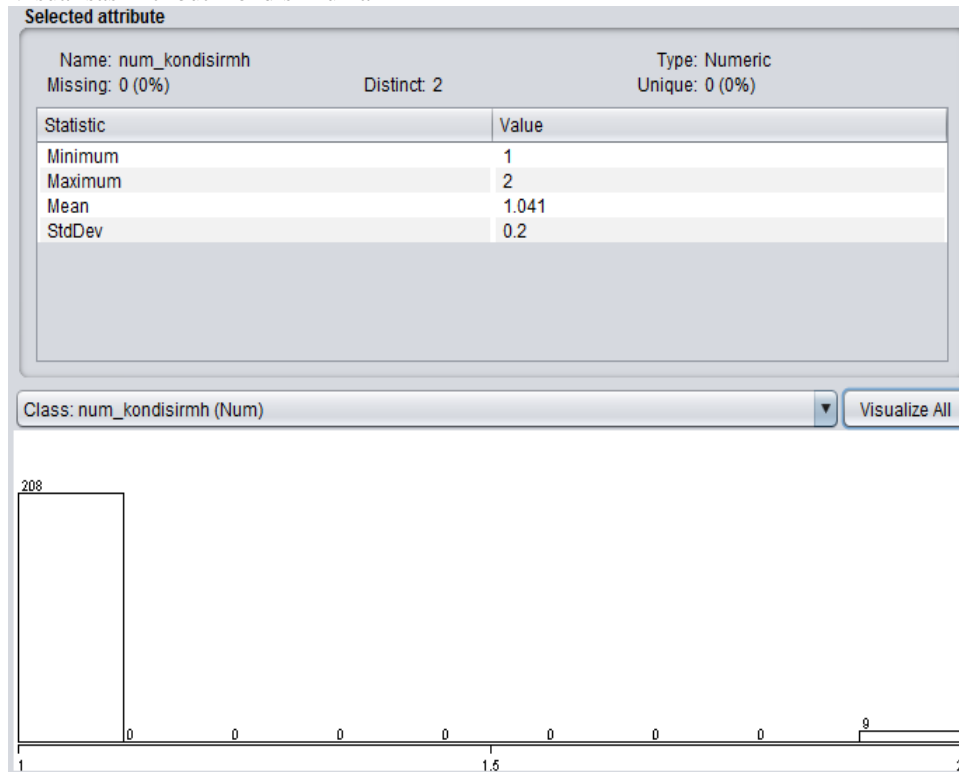
1. Visualisasi Atribut Jumlah Tanggungan



Gambar 9. Visualisasi Atribut Jumlah Tanggungan

Gambar 9 adalah visualisasi dari Atribut jumlah tanggungan dalam keluarga. Diketahui bahwa dari 217 kepala keluarga terdapat 10 data penduduk yang tidak memiliki tanggungan (*minimum*), kemudian 3 kepala keluarga yang memiliki tanggungan sebanyak 6 orang (*maximum*). Nilai rata – rata (*mean*) dari 217 data penduduk adalah 2.507 dan standar deviasi terdapat nilai 1,167.

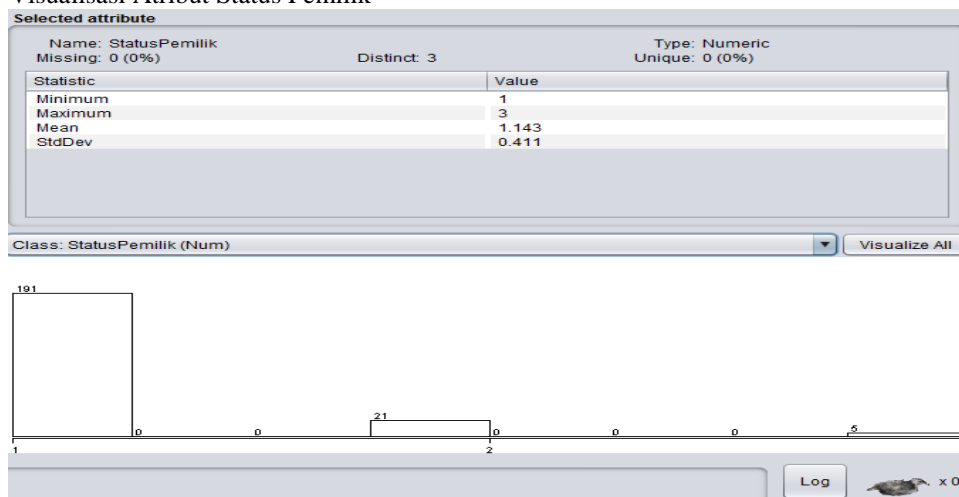
2. Visualisasi Atribut Kondisi Rumah



Gambar 10. Visualisasi Atribut Kondisi Rumah

Gambar 10 adalah visualisasi dari Atribut kondisi rumah. Diketahui bahwa dari 217 data penduduk dapat dijelaskan bahwa sebanyak 208 yang kondisi rumah nya baik sedangkan sebanyak 9 orang memiliki kondisi rumah kurang baik. Nilai rata – rata (*mean*) yang didapat adalah 1,041 da standar deviasi adalah 0,2.

3. Visualisasi Atribut Status Pemilik

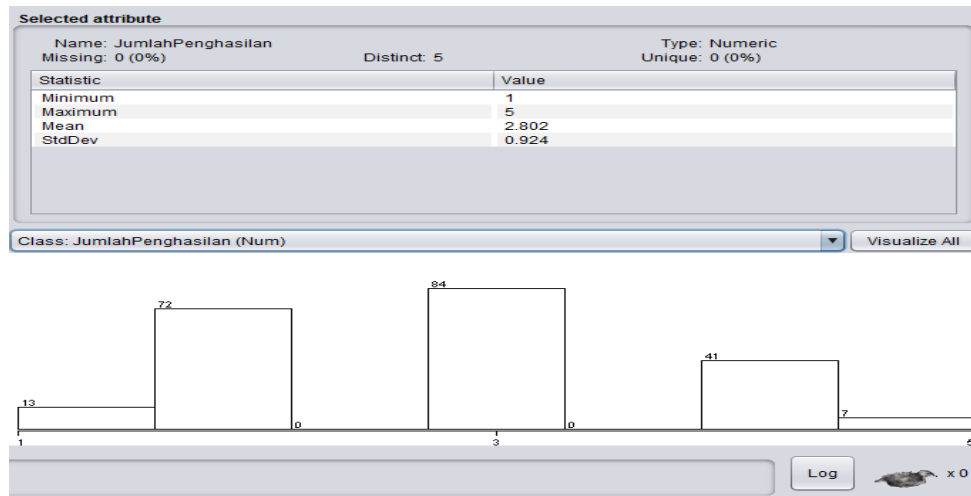


Gambar 11. Visualisasi Atribut Status Pemilik

Gambar 11 adalah visualisasi dari Atribut status pemilik. Diketahui bahwa dari 217 data penduduk dapat dijelaskan bahwa sebanyak 191 penduduk yang tempat tinggalnya hak milik atau milik

sendiri, 21 penduduk yang numpang, dan 3 penduduk sewa. Nilai rata-rata (*mean*) adalah 1,143 dan standar deviasi 0,411.

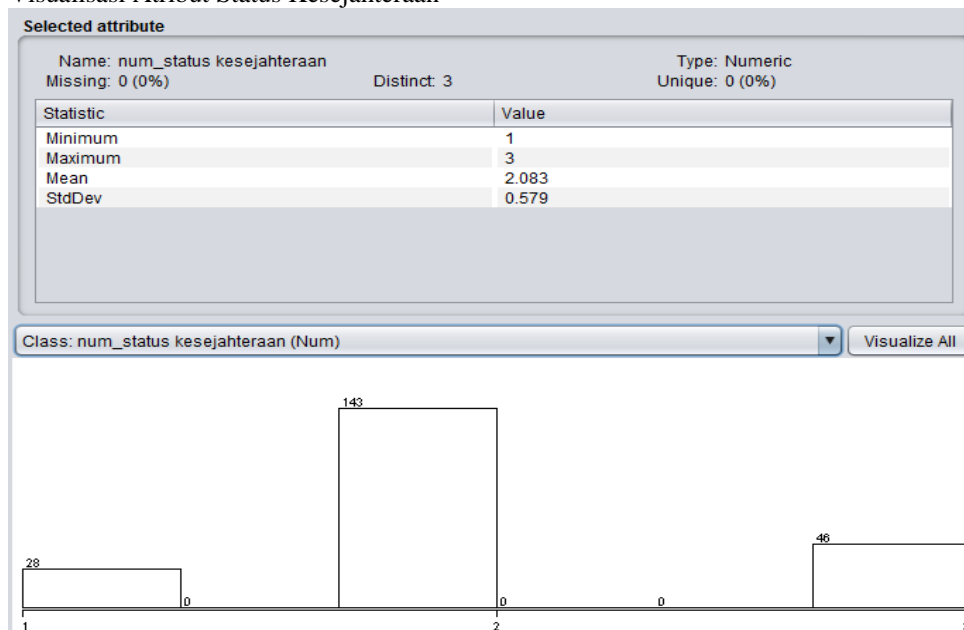
4. Visualisasi Atribut Jumlah Penghasilan



Gambar 12. Visualisasi Atribut Jumlah Penghasilan

Gambar 12 adalah visualisasi dari Atribut kondisi rumah. Diketahui bahwa dari 217 data penduduk dapat dijelaskan bahwa sebanyak 83 penduduk yang status jumlah penghasilannya sama dengan atau 3200000 sampai 4700000 perbulan. Nilai rata-rata yang didapat adalah 2,788 dan standar deviasi adalah 0,943.

5. Visualisasi Atribut Status Kesejahteraan



Gambar 13. Visualisasi Atribut Status Kesejahteraan

Gambar 13 adalah visualisasi dari Atribut status kesejahteraan. Diketahui bahwa dari 217 data penduduk dapat dijelaskan bahwa sebanyak 28 penduduk status kesejahteraannya baik, 143 penduduk yang status kesejahteraannya cukup dan 46 penduduk yang status kesejahteraannya kurang.

6. Pengambilan Pusat *Cluster*

```
Clusterer output
=====
num_statuspmlk
num_jlhpnghsln
num_status kesejahteraan
Test mode: evaluate on training data

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

kMeans
=====
Number of iterations: 8
Within cluster sum of squared errors: 235.76496460499897

Initial starting points (random):

Cluster 0: Pungut,3,1,2,2,2
Cluster 1: Safrizal,3,1,1,4,2
Cluster 2: Mahyudin,5,1,1,4,2

Missing values globally replaced with mean/mode

Final cluster centroids:
Attribute          Full Data          Cluster#
                   (217.0)          (48.0)          (140.0)          (29.0)
```

Gambar 14. Pengambilan cluster

Pada weka pengambilan pusat *cluster* diambil secara random dan di sebutkan nilainya dari 5 atribut, pada weka dijelaskan jumlah iterasi yang di lakukan untuk mendapatkan sebanyak 3 *cluster* yaitu sebanyak 3 iterasi. Kemudian setiap atribut mendapatkan nilai masing-masing dapat dilihat pada gambar 15 Sebagai berikut:

```
Clusterer output

Missing values globally replaced with mean/mode

Final cluster centroids:
Attribute          Full Data          Cluster#
                   (217.0)          (48.0)          (140.0)          (29.0)
=====
Nama KK            Junaidi            Deni            Burhan            Mansyur
Jumlah Tanggungan  2.5069            1.75            2.5857            3.3793
num_kondisirmh     1.0415            1.125           1.0143            1.0345
num_statuspmlk     1.129            1.2292          1.0929            1.1379
num_jlhpnghsln     2.788            1.75            2.8786            4.069
num_status_kesejahteraan  2.0829          2.9583            2            1.0345

Time taken to build model (full training data) : 0.03 seconds

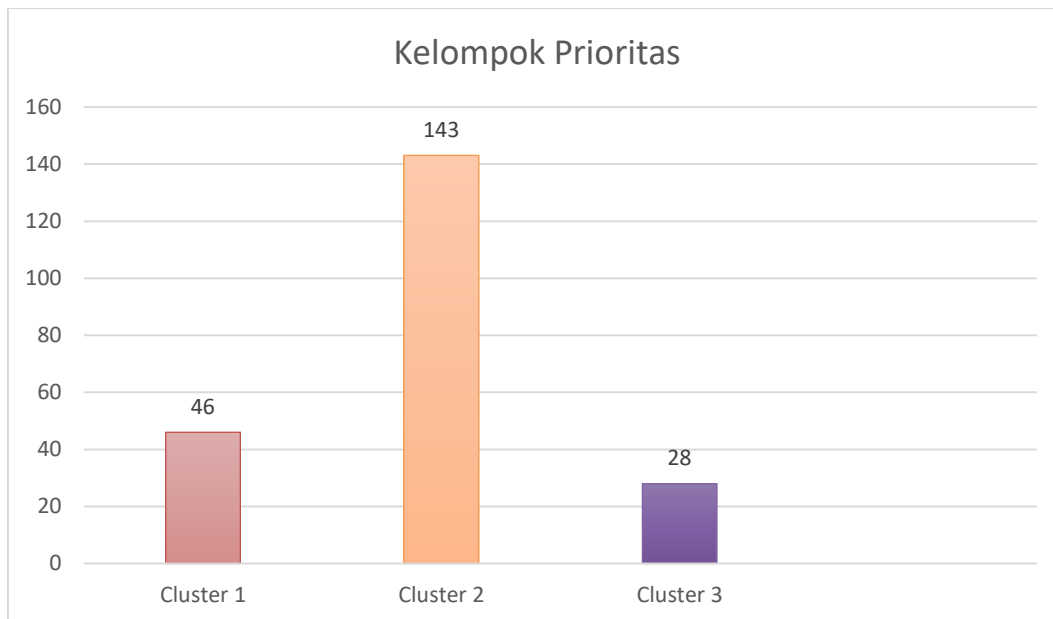
=== Model and evaluation on training set ===

Clustered Instances

0          48 ( 22%)
1          140 ( 65%)
2           29 ( 13%)
```

Gambar 15. Centroid cluster

Kemudian didapatlah jumlah dalam setiap *clusternya*
Cluster 0 sebanyak 46 data atau sebesar 21%
Cluster 1 sebanyak 143 data atau sebesar 66%
Cluster 2 sebanyak 28 data atau sebesar 13%



Gambar 16. Cluster Kelompok Prioritas

Dapat disimpulkan dari hasil cluster dengan tools *WEKA* adalah sebanyak 217 data penduduk di Kecamatan Muara Papalik yang terbagi menjadi 3 *cluster* dimana dari masing-masing cluster nantinya akan dikelompokkan kedalam 3 prioritas, pada *cluster 1* terdapat 46 data penduduk, kemudia *cluster 2* terdapat 143 data penduduk, dan *cluster 3* terdapat 28 data penduduk.

5 Kesimpulan

5.1 Simpulan

Dari penelitian yang telah dilakukan, dapat diambil kesimpulan-kesimpulan sebagai berikut:

1. Penelitian ini menggunakan data penduduk tahun 2017 dan 2018 dengan jumlah data 217.
2. Sistem Penilaian prioritas penduduk tidak mampu, yang menggunakan *Datamining* metode *Clustering K-Means* dapat dipergunakan sebagai salah satu alat Sistem Penunjang Keputusan didalam menentukan Kelompok Penduduk Prioritas yang mendapatkan bantuan beras sejahtera (rastra).
3. Persentasi hasil akurasi klasifikasi *clustering* dengan menghasilkan 3 kelompok prioritas penerima bantuan beras sejahtera (rastra).
4. Hasil penentuan kelompok prioritas penerima bantuan baeras sejahtera mendapatkan hasil akurasi dari 3 *cluster*, *cluster* pertama terdapat 46 data penduduk dengan hasil persentase 21%, *cluster* kedua terdapat 143 data penduduk dengan hasil presentase 66%, dan *cluster* ketiga terdapat 28 data penduduk dan hasil presentase 13%.
5. Hasil dari data penduduk menggunakan algoritma *clustering k-mean* menghasilkan bahwa penduduk yang paling berpengaruh terhadap bantuan penerima beras adalah yang terdapat pada prioritas pertama.

5.2 Saran

Adapun saran-saran yang berguna untuk perkembangan lebih lanjut dari sistem yang diusulkan adalah:

1. Dalam acuan penerapan aplikasi agar bisa efektif, maka diperlukan aturan dan perundangan serta dukungan anggaran dari pemerintah daerah untuk mewujudkan terbentuknya *e-government* yang dapat meningkatkan kepercayaan masyarakat terhadap pemerintah dalam mengelola dana bantuan beras sejahtera.
2. Diperlukan adanya kegiatan pengumpulan data evaluasi secara berkala sebagai bahan input dari sistem penilaian Prioritas Penduduk Tidak Mampu Penerima Bantuan Beras Sejahtera (Rastra).

6. Daftar Rujukan

- [1] Jamaris, M. (2017). Implementasi Metode Rough Set Untuk Menentukan Kelayakan Bantuan Dana Hibah Fasilitas Rumah Ibadah, 2(2).
- [2] M. Rixco Setiawan, “Analisa Data Mining Untuk menentukan Kelompok Prioritas Penerima Bantuan Bedah rumah di desa Semen Menggunakan Metode Clustering K-Means”, Universitas Nusantara PGRI Kediri. (2018).
- [3] Aranda, J., Astari, W., & Natasya, G. (2016). Penerapan Metode K-Means Cluster Analysis Pada Sistem Pendukung Keputusan Pemilihan Konsentrasi Untuk, 6–7.
- [4] Chairul Fadlan, Selfia Ningsih, Agus Perdana Windarto (2018). Penerapan Metode Naive Bayes Dalam Klasifikasi Kelayakan Penerima Beras Rastra, 1-8.
- [5] Sartika, D., Sensuse, D. I., Indo, U., Mandiri, G., & Komputer, F. I. (2017). Perbandingan Algoritma Klasifikasi Naive Bayes , Nearest Neighbour , Dan Decision Tree Pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian, 1(2), 151–161.