

# PENERAPAN ALGORITMA K-MEANS UNTUK CLUSTERING SANTRI PRA-SEJAHTERA DI YAYASAN BANTUAN SOSIAL (YBS) AZ-ZAINIYYAH PONDOK PESANTREN NURUL JADID

Sudriyanto <sup>1)</sup>, Ahmad Khairi <sup>2)</sup>, Atoillah Shohibul Hikam <sup>3)</sup>

<sup>1,2,3)</sup>Teknik Informatika, Universitas Nurul Jadid

Pondok Pesantren Nurul Jadid Karanganyar Paiton Probolinggo 67291

e-mail: sudriyanto@unuja.ac.id <sup>1)</sup>, ahmadkhairi@unuja.ac.id <sup>2)</sup>, atoillahaja@email.com <sup>3)</sup>

## ABSTRAK

Permasalahan kemiskinan merupakan persoalan sosial yang memiliki dampak cukup besar diberbagai negara, tidak hanya dalam lingkup besar dalam skala kecilpun kemiskinan menjadi persoalan yang perlu diperhatikan secara khusus salah satu contoh dalam lingkup Pesantren. Pada penelitian ini dataset diperoleh dari Pedatren, yaitu data santri SLTP dan SLTA Pondok Pesantren Nurul Jadid tahun 2021. Algoritma yang digunakan dalam penelitian ini ialah metode K-Means Clustering. Metode K-Means Clustering adalah teknik Unsupervised, serta metode pengelompokan data dalam beberapa kelompok, sesuai karakteristik satu sama lain. Kelebihan dari algoritma K-Means adalah Relatif sederhana dan mudah untuk diterapkan, dapat diskalakan untuk dataset dalam jumlah besar, mudah beradaptasi dengan contoh baru, umum diimplementasikan ke cluster dengan bentuk dan ukuran yang berbeda. Jumlah data yang digunakan sebanyak 749 data santri SLTP dan SLTA, dibagi kedalam 4 cluster yakni (C1) Keluarga Pra-Sejahtera (C2) Keluarga Sejahtera 1 Keluarga Sejahtera 2 (C3) dan Keluarga Sejahtera Plus (C4). Adapun tahapan yang dilakukan dalam penelitian ini yaitu identifikasi masalah, studi literatur, pengumpulan data, preprocessing, implementasi algoritma K-Means Clustering, dan yang terakhir evaluasi hasil. Dari hasil pengelompokan cluster didapatkan data C1 = 163 (Sejahtera 1), C2 = 215 (Sejahtera 2), C3 = 246 (Sejahtera Plus) dan C4 = 125 (Keluarga Pra-Sejahtera). Evaluasi akhir menggunakan matriks Davies Bouldin Index memperoleh akurasi nilai 0,90.

**Kata Kunci:** Davies Bouldin Index, K-Means Clustering, Pra-Sejahtera, Santri..

## ABSTRACT

The problem of poverty is a social problem that has a sizeable impact in various countries, not only in a large scope, but even on a small scale, poverty is a problem that needs special attention, one example is within the scope of Islamic boarding schools. In this study, the dataset was obtained from Pedatren, namely data from SLTP and SLTA Nurul Jadid Islamic Boarding School students in 2021. The algorithm used in this research is the K-Means Clustering method. The K-Means Clustering method is an Unsupervised technique, as well as a method of grouping data into several groups, according to each other's characteristics. The advantages of the K-Means algorithm are relatively simple and easy to implement, scalable for large datasets, easily adaptable to new examples, commonly implemented to clusters of different shapes and sizes. The amount of data used is 749 junior and senior high school students data, divided into 4 clusters namely (C1) Pre-Prosperous Families (C2) Prosperous Families 1 Prosperous Families 2 (C3) and Prosperous Plus Families (C4). As for the phase that done in this research such as identifying a problem, a literature review, data collection, preprocessing, the implementation of k-means algorithm clustering and the last evaluasi the results of. From the results of the grouping of clusters of data or C1 = 163, C2 = 215, C3 = 246 and C4 = 125. The end of use evaluasi matrix davies bouldin mem-peroleh akurasi 0,90 value index.

**Keywords:** Davies Bouldin Index, K-Means Clustering, Pre-Prosperous, Santri.

## I. PENDAHULUAN

**P**ROBLEM kependudukan di negara indonesia merupakan ancaman hebat kepada pemerintah dari masa ke masa. Kependudukan merupakan suatu hal yang begitu penting di setiap negara yang berkaitan dengan jumlah penduduk, mortalitas, natalitas, imigrasi, dan emigrasi, serta kondisi kesejahteraan ekonomi, politik agama budaya dan

sosial yang sudah terlampir sesuai UU (No. 52 2009) pasal 1 ayat 2, yang memanipulasi atas perkembangan kependudukan dan pembangunan ialah Kesejahtaan ekonomi [1][2]. Pada tanggal 15 Juli 2021 Badan Pusat Statistik (BPS) menyampaikan secara resmi laporan bahwa pada bulan Maret 2021 sebesar 10,14% dari total penduduk indonesia atau sebanyak 27,54 juta penduduk masih berstatus miskin. Jika dibandingkan jumlah penduduk miskin pada tahun sebelumnya,

kesmiskinan terendah yang pernah dicapai Indonesia ialah pada bulan September 2019 [3]. Kemiskinan ialah sebagian faktor penghalang bagi setiap penduduk agar mendapatkan kehidupan sejahtera [4]. Pemerintah juga mempunyai peran sangat penting didalam mensejahterakan setiap warga negaranya dari berbagai faktor penyebab krisis kemiskinan. Dalam menangani krisis kemiskinan ini pemerintah sudah melakukan beberapa program upaya mengatasi faktor krisis kemiskinan, diantaranya dengan adanya program bantuan social meliputi program keluarga harapan atau yang lebih di kenal dengan istilah Program Keluarga Harapan (PKH) dan bantuan langsung tunai atau (BLT) serta beberapa program bantuan lainnya [5]. Akan tetapi didalam pelaksanaannya masih banyak terdapat penerima bantuan yang salah sasaran sehingga terdapat masyarakat dengan kategori sejahtera masih mendapatkan bantuan, begitupun sebaliknya masyarakat tidak sejahtera tidak mendapatkan bantuan [6]. Lembaga Bantuan Sosial dibawah naungan Pondok Pesantren Nurul Jadid lebih tepatnya Yayasan Bantuan Sosial (YBS) Az-Zainiyyah lembaga yang menyalurkan bantuan Sosial dan Pendidikan bagi masyarakat serta santri yang kurang mampu. Namun dalam pelaksanaan program bantuan ini, petugas yayasan masih cukup kesulitan dalam menjalankan program tersebut, dikarenakan yayasan hanya memiliki data santri yang sudah mendaftar secara pribadi sebelumnya dan tidak memiliki dataset secara keseluruhan santri yang masuk dalam kategori Pra-Sejahtera tiap tahunnya. Serta saat ini pendataan pengajuan bantuan masih dilakukan secara manual dengan menyamakan data diri yang tersimpan di Pedatren.

*K-Means* adalah metode atau algoritma yang memuat data kedalam jumlah golongan atau kelompok, namun data-data yang mempunyai kesamaan karakteristik berada pada satu golongan, dan terdapat pula sebagian data yang mempunyai ketidak samaan berada pada golongan lainnya [7]. Algoritma *K-Means* yakni teknik perhitungan yang menggunakan kriteria atau tolak ukur input sebanyak ( $k$ ) serta membagi kumpulan data ( $n$ ) bahan ke dalam ( $k$ ) beberapa golongan, agar memperoleh hasil perhitungan dengan susunan analogi setiap komponen disetiap golongan memiliki nilai yang tinggi, meskipun dalam golongan lain tingkat kemiripannya sangatlah rendah [8][9]. Kemiripan anggota tersebut dengan golongan yang dijumlah menggunakan koneksi bahan kepada nilai mean pada setiap golongan. Metode perhitungan *k-means clustering* adalah algoritma penggolongan data ke dalam jumlah golongan sesuai dengan kemiripan karakteristik yang mirip satu dengan

yang lainnya, dan data yang memiliki karakteristik yang tidak mirip berada didalam kelompok lain [10].

Oleh karenanya dalam penerapan Algoritma *K-Means* untuk *clustering* Santri Pra-sejahtera di Yayasan Bantuan Sosial (YBS) Az-Zainiyyah ini diharapkan mampu mempermudah proses pengelompokan data Santri Pra-Sejahtera dan Sejahtera untuk mempersingkat waktu dalam pencarian dan pengelolaan data.

## II. STUDI PUSTAKA

### A. Penelitian Terkait

Penelitian yang pertama oleh [11]. Dalam jurnal yang berjudul “Implementasi Data Mining Dalam Mengelompokkan Jumlah Penduduk Miskin Berdasarkan Provinsi Menggunakan Algoritma K-Means” penelitian ini dilakukan karena jumlah penduduk pra-sejahtera Indonesia pada awal bulan September tahun 2020 menembus angka 27,55 juta penduduk, maka pemerintah harus berupaya menekan jumlah penduduk pra-sejahtera dimasing-masing daerah melalui berbagai program pemerintah berupa bantuan sosial. Dalam penelitian tersebut menggunakan metode clustering dari Algoritma K-Means sebagai teknik untuk mencari jumlah penduduk miskin yang diklasifikasikan menurut setiap daerah. Penelitian ini menggunakan Badan Pusat Statistik (BPS) tahun 2007-2019. Dengan menggunakan algoritma K-Means untuk menjumlah data penduduk miskin secara manual, dapat di simpulkan hasil dari iterasi 2 dan 3 nilainya sama, dengan hasil iterasi  $C1=3$ ,  $C2=9$ , dan  $C3=22$ , posisi kelompok data penduduk pra-sejahtera tetap tidak ada perubahan, sehingga proses perhitungan iterasi berhenti

Hasil perhitungan clustering dari tools Rapid Miner yang menerangkan bahwa ada 3 kelompok yang diawali dari kelompok 0, kelompok 1 dan kelompok 2. Kelompok 0 yakni kelompok tertinggi, kelompok 1 ialah kelompok menengah, dan kelompok 2 merupakan kelompok terkecil. kelompok 0 yaitu Provinsi Banda Aceh Darussalam, Provinsi Sumatera Utara, Provinsi Sumatera Selatan, Provinsi Lampung, Banten, Provinsi Nusa Tenggara Barat (NTB), Provinsi Nusa Tenggara Timur (NTT), Provinsi Sulawesi Selatan dan Provinsi Papua. kelompok 1 yaitu

Provinsi Sumatera Barat, Provinsi Jambi, Provinsi Riau, Provinsi Kepulauan Riau, Provinsi Bengkulu, Provinsi Kepulauan Bangka Belitung, Provinsi DKI Jakarta, Provinsi DI Yogyakarta, Provinsi Bali, Provinsi Kalimantan Barat, Provinsi Kalimantan Tengah, Provinsi Kalimantan Timur, Provinsi Kalimantan Utara,

Provinsi Sulawesi Utara, Provinsi Sulawesi Tengah, Provinsi Sulawesi Tenggara, Provinsi Sulawesi Barat, Provinsi Gorontalo, Provinsi Maluku Utara dan Provinsi Papua Barat. Cluster 2 yaitu Provinsi Jawa Barat, Provinsi Jawa Tengah, Provinsi Jawa Timur.

Penelitian kedua dilakukan oleh saudara [12]. Dalam jurnal yang berjudul “Implementasi K-Nearest Neighbor (KKN) untuk Klasifikasi Masyarakat Pra-Sejahtera di Desa Sapikerep Kecamatan Sukapura” penelitian ini dilakukan karena kemiskinan merupakan masalah umum yang begitu kompleks yang terdapat di berbagai daerah salah satunya di desa Sapikerep Kecamatan Sukapura Kabupaten Probolinggo yang merupakan desa yang saat ini masih banyak terdapat masyarakat yang tergolong kurang mampu dari segi pangan maupun sandangnya. Beberapa kriteria untuk menentukan masyarakat sejahtera yang ditentukan oleh perangkat desa diantaranya aset atau harta benda, bangunan atau tempat tinggal, status lahan atau tanah kepemilikan, profesi dan penghasilan setiap tahunnya. Salah satu upaya untuk mengatasi permasalahan tersebut ialah dengan mengimplementasikan metode K-Nearest Neighbor (KKN) untuk mendapatkan prediksi ataupun klasifikasi dari suatu data berdasarkan nilai  $k$  yang telah ditetapkan yakni  $k=3$ ,  $k=5$ ,  $k=7$ . Hasil dari klasifikasi algoritma K-Nearest Neighbor terdapat data yang menunjukkan nilai akurasi terbesar pada nilai  $k=5$  dan  $k=7$  dengan hasil akurasi 98.68%, sedangkan  $k=3$  memperoleh nilai akurasi 97,35%.

## B. Landasan Teori

### 1. Pra-Sejahtera

Pra-Sejahtera adalah kondisi dimana penduduk atau masyarakat yang tidak mampu mencukupi kebutuhan dasarnya, baik kebutuhan pokok maupun lainnya. Sedangkan sejahtera adalah kondisi dimana penduduk atau masyarakat mampu mencukupi kebutuhan hidupnya, baik pangan maupun sandangnya. Keluarga Pra-sejahtera yakni warga yang belum mampu mencukupi satu dari lima indeks keluarga sejahtera atau indeks kepentingan pokok keluarga. Keluarga sejahtera adalah warga yang mampu memenuhi kelima indikator dan dua indikator tambahan yakni mampu serta sukarela menyalurkan bantuan materil untuk kegiatan sosial, serta aktif didalamnya sebagai anggota pengurus perkumpulan sosial, yayasan atau organisasi masyarakat lainnya [13].

### 2. Data Mining

Data mining merupakan suatu metode pembentukan acuan informasi yang memiliki fungsi untuk pengumpulan data, serta perlu diekstrak untuk menjadi sebuah informasi baru dan menjadi acuan dalam

pengambil keputusan. Sumber informasi bisa berupa *dataset*, *data warehouse*, *data web*, dan *repositoril* dan lainnya.

Hal ini menunjang kelompok agar terfokus pada suatu sumber data ataupun informasi yang sangat fundamental di *data warehouse* mereka. Data mining bisa menjadi alat dalam menganalisa mode era ditahun mendatang, serta dapat membantu organisasi membuat sebuah hasil keputusan yang berbasis pengetahuan proaktif [7].

Berdasarkan fungsinya data mining terbagi kedalam beberapa kelompok, yaitu :

#### a. Deskripsi

Dalam suatu peneliti yang dilakukan dengan metode sederhana untuk dapat menemukan pola untuk mencerminkan serta melihat gaya yang ada dalam kumpulan sumber informasi. Deskripsi dari bentuk dan gaya terkadang meneruskan peluang pengertian dalam bentuk ataupun gaya.

#### b. Estimasi

Estimasi tidak jauh beda dengan klasifikasi, namun apabila variabel tujuan lebih dominan numerik dari pada kategori. Pola dibentuk dari trek record komplit yang menyimpan nilai dari sebuah variabel tujuan sebagai acuan dari nilai prediksi. Berikutnya pada pemantauan, estimasi nilai dari sebuah variabel tujuan dibentuk berdasar nilai variabel prediksi.

#### c. Prediksi

Prediksi mirip dengan klasifikasi dan estimasi, namun pada prediksi nilai yang dihasilkan akan ada di masa mendatang.

#### d. Klasifikasi

Klasifikasi adalah metode penemuan variabel atau fungsi untuk membedakan suatu objek ataupun kelompok sesuai labelnya.

#### e. Pengklusteran

Pengklusteran adalah penggolongan pemantauan, atau menyimak sejumlah data atau objek yang mempunyai persamaan karakteristik [14]. *Cluster* merupakan kumpulan dari data yang mempunyai persamaan karakteristik satu dengan lainnya serta memiliki ketidak samaan dengan data didalam kelompok lain. Pengelompokan atau pengklusteran tidak sama dengan klasifikasi, bedanya terletak pada tidak adanya variabel tujuan dalam pengelompokan. Didalam pengklusteran tidak ada tindakan untuk melakukan klasifikasi, mengukur atau menganalisa nilai dari variabel tujuan, namun dalam metode pengklusteran mencoba agar menjalankan penguraian terhadap keseluruhan objek menjadi kelompok-kelompok sesuai kesamaan karakteristik (*homogen*), yang mana kesamaan data dalam suatu kelompok memiliki nilai tinggi, sedangkan kesamaan dalam kelompok lain akan memiliki nilai yang rendah [15].

Data mining adalah metode pada teknik *Knowledge*

*Discovery in Database* (KDD) yang terbangun dari beberapa penerapan penelitian data dan penemuan algoritma yang bernilai enumerasi bentuk tertentu atau model dari suatu data. Teknik KDD memiliki sifat interaktif dan perulangan, yang banyak melibatkan proses serta berbagai ketentuan yang dibuat [16]. Adapun langkah-langkahnya sebagai berikut :

a. Seleksi Data

Pada proses seleksi data memiliki proses yang cukup singkat, karena hanya mengolah data didalam database yang akan digunakan dan membuang data yang tidak diperlukan, oleh karenanya dibutuhkan penyeleksian data yang akan proses lebih lanjut.

b. Data Integrasi

Data integrasi adalah kumpulan dari suatu dataset ke suatu database yang lain atau baru. Integrasi data diproses melalui atribut dan variable yang mengidentifikasi ciri-ciri yang unik seperti atribut pembelian, jenis produk, nama pelanggan, nomer pelanggan dan lainnya

c. Data Tranformasi

Data Tranformasi merupakan pengelolaan data yang digabungkan ataupun diubah sesuai dengan algoritman yang digunakan dalam data mining. Terdapat beberapa format data mining yang membutuhkan format khusus sehingga dalam prosesnya memerlukan cara yang khusus pula.

d. Proses Mining

Proses mining merupakan pencarian model ataupun informasi unik dari data yang dipilih dengan teknik tertentu. Pemilihan teknik yang tepat sangat berpengaruh kepada proses KDD secara keseluruhan.

e. Evaluasi

Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang sudah ada sebelumnya. Sehingga menghasilkan kesimpulan yang mendekati target ataupun hipotesa untuk langkah berikutnya.

f. Presentasi Pengetahuan

Dalam langkah akhir dari proses data mining ini ialah merancang kesimpulan keputusan dan memberikan informasi untuk dipresentasikan dari proses analisa yang telah laksanakan sebelumnya.

### 3. Algoritma K-Means

Algoritma sendiri memiliki pengertian urutan atau langkah-langkah untuk menyelesaikan suatu problem yang ditulis secara sistematis [17]. *K-Means* adalah teknik penjabaran data ataupun data mining yang menjalankan prosedur pemodelan non pengawas atau supervisi (*Unsupervised*), serta merupakan suatu teknik penggolongan data menggunakan pola

pembatas. Teknik *K-Means* berupaya menggolongkan data kebeberapa golongan sesuai sifat satu sama lain, sehingga data dengan sifat yang sama digolongkan ke dalam golongan yang sama, dan sebaliknya data dengan sifat yang berbeda digolongkan ke dalam golongan yang berbeda pula. Sehingga dalam teknik ini berfungsi meminimalkan perbedaan data yang berada digolongan lain, serta memaksimalkan perbedaan data yang ada digolongan lain [18].

Teknik *K-Means Clustering* bermaksud mengurangi fungsi sasaran yang diatur dalam metode penggolongan dengan cara mengurangi perbedaan antar data disuatu golongan dan menambah perbedaan dengan daya yang ada digolongan lain, serta usaha untuk mencari golongan disebagian data dan variabel  $k$  merupakan jumlah golongan [19]. Algoritma tersebut menampung input berbentuk data non predikat, bertentangan dengan pembelajaran menggunakan pengawas yang memperoleh input data berbentuk vektor  $(x_1, y_1), (x_2, y_2), \dots, (x_a, y_a)$ , yang mna  $x_a$  yakni data dari data penataan serta  $y_a$  adalah kategori dari label  $x_a$ .

Pada metode ini terdapat dua macam penggolongan data yang kerap kali dipakai dalam penggolongan data yakni *analisis Hierarchinal* dan *analisis Non-hieraechinal*. *Analisis Hierarchinal* merupakan studi penggolongan yang hasil penggolongannya ditampilkan bersusun dari  $n, (n-1)$  sampai diperoleh suatu golongan. Analisis hierarchinal ini dibagi menjadi dua teknik yaitu *agglomerative* dan *divisive*. *Analisis Non-hieraechinal* merupakan teknik penggolongan yang mana bahan hendak digolongkan ke dalam  $k$  golongan yang sudah terpilih diawal.

Rumus yang digunakan untuk menentukan jarak data dari masing-masing *centroid* dapat dilihat pada formula 1:

$$d(K,L)=\sqrt{(\sum_{j=1}^p [(x_{-j}(K)-x_{-j}(L))]^2) \dots(1)}$$

Keterangan :

D = titik *dataset*

K = data *record*

L = data *centroid*

Jarak yang terpendek antara centroid dengan dataset menentukan posisi cluster suatu data. Setelah jarak ditemukan lalu menghitung rumus iterasi dapat dilihat pada formula 2 :

$$C(i)=(x_1+x_2+x_3+\dots+x_n)/(\sum x) \dots (2)$$

Keterangan :

- $X_1$  = Nilai data *record* ke-1
- $X_2$  = Nilai data *record* ke-2
- $\sum x$  = jumlah data *record*

4. Davies Bouldin Index (DBI)

*Davies Bouldin Index* atau yang lebih populer dengan nama DBI ini, merupakan matrik untuk mengevaluasi hasil algoritma *clustering*. Pengukuran hasil DBI adalah memaksimalkan jarak antara *cluster* serta meminimalkan jarak dalam *cluster*. Semakin kecil nilai akurasi yang diperoleh maka semakin optimal skema *cluster* yang di hasil kan, namun nilai berada di antara 0 dan 1 tidak bernilai minus [20].

Sebelum perhitungan *Davies Bouldin Index* (DBI), dihitung lebih dahulu *variance* dari tiap-tiap *cluster* seperti ditampilkan pada formula 3.

$$var(x) = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \quad \dots (3)$$

dimana :

$\bar{x}$  : rata-rata *cluster* x dan N iyalah jumlah keanggotaan *cluster*

Kemudian hitung *Davies- Bouldin Index* (DBI) dengan persamaan seperti ditampilkan pada formula 4, 5 dan 6 :

$$DBI = \frac{1}{k} \sum_{i=1}^k R_i \quad \dots (4)$$

dengan

$$R_i = \max R_{ij_{j=1, \dots, k, i \neq j}} \quad \dots (5)$$

dan

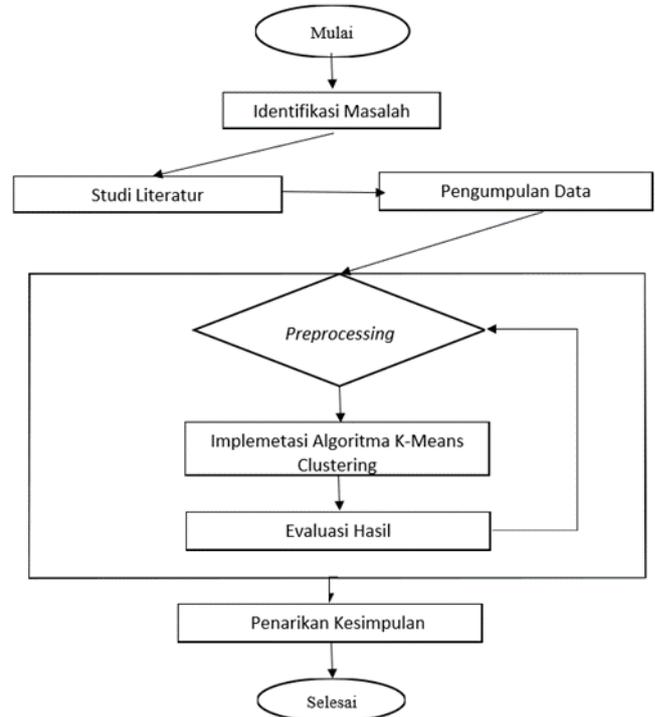
$$R_{ij_{i \neq j}} = \frac{varC_i + varC_j}{\|c_i - c_j\|} \quad \dots (6)$$

dimana :

$C_i$  : *cluster* ke i dan  $c_i$  iyalah *centroid* dari *cluster* ke i.

III. METODE PENELITIAN

Kerangka penelitian ini dibuat untuk mempermudah peneliti menyelesaikan proses penelitian penerapan Algoritma *K-Means* untuk *Clustering* Santri Pra-Sejahtera Di Yayasan Bantuan Sosial (YBS) Az-Zainiyyah Pondok Pesantren Nurul Jadid. Terdapat beberapa tahapan yang akan dilalui. Tahapan kerangka penelitian seperti ditampilkan pada gambar 1:



Gambar 1. Bagan Alur Penelitian

Pada tahap ini diuraikan setiap langkah-langkah pada kerangka rancangan penelitian sebagai berikut :

A. Identifikasi Masalah

Berlandaskan pembahasan diatas, maka dapat diidentifikasi permasalahan yang akan diteliti yaitu, Bagaimana menghasilkan *clustering* Santri Pra-Sejahtera, berdasarkan data Pedatren di Yayasan Bantuan Sosial (YBS) Az-Zainiyyah Pondok Pesantren Nurul Jadid menggunakan penerapan Algoritma *K-Means*.

B. Studi Litelatur

Langkah awal dari penelitian ini, mencari serta mempelajari masalah yang ada pada data santri yang diperoleh dari Pengelolaan Data Pesantren (Pedantren) Pondok Pesantren Nurul Jadid. Data yang digunakan ialah data santri Sekolah Lanjutan Tingkat Pertama (SLTP) sampai Sekolah Lanjutan Tingkat Atas (SLTA). Kemudian menganalisa masalah yang akan ditulis dilatar belakang penelitian, serta membuat tujuan pada penelitian, menyampaikan manfaat penelitian dan membuat batasan pada penelitian.

C. Pengumpulan Data

Data diperoleh dari Pengelolaan Data Pesantren (Pedantren) Pondok Pesantren Nurul Jadid. Kemudian dilakukan pengelolaan dan pengelompokan terhadap data yang telah terkumpul berdasarkan karakteristik dan variable yang telah di tentukan.

#### D. Proprocessing

Sebelum proses pengelolaan data mining dapat dilaksanakan perlu dilakukan tahap *preprocessing*. Pada tahap ini akan dilakukan proses integrasi data, yang mana merupakan proses pengolahan data yang dilakukan dengan menganalisa data yang sudah diperoleh serta membuang data yang tidak dibutuhkan dalam proses penghitungan.

#### E. Implementasi Algoritma K-Means Clustering

Pada tahap ini dilakukan penyelesaian dari algoritma *K-Means Clustering* menggunakan rumus yang telah ditetapkan. Sebagaimana yang dijelaskan pada Bab Pendahuluan bahwa pada penelitian ini bertujuan mengelompokkan data *clustering* santri Pra-sejahtera menggunakan *K-Means Clustering*. Cara kerjanya yaitu membagi data kedalam bentuk *cluster* melalui proses tersistematis, setelah didapatkan *cluster* kemudian dilakukan analisa terhadap pola pembentukan *cluster* yang dapat menghasilkan informasi yang bermanfaat.

#### F. Evaluasi Hasil

Tahapan evaluasi hasil dari perhitungan algoritma *K-Means Clustering*, baik secara manual ataupun dilakukan secara pengujian *software*. Apakah data yang dihasilkan sesuai atau sebaliknya.

### IV. HASIL DAN PEMBAHASAN

Data diperoleh dari admin Pengelolaan Data Pesantren (Pedantren) Pondok Pesantren Nurul Jadid dengan situs web <https://nuruljadid.app>. Kemudian dilakukan pengelolan dan pengelompokan terhadap data yang telah terkumpul berdasarkan karakteristik dan variable yang telah di tentukan seperti ditampilkan pada tabel 1:

Tabel 1. Data Santri SLTP – SLTA angkatan 2021

No	Nama	Jumlah Saudara	Angkatan	status	Lembaga	Jurusan	Pekerjaan Orangtua	Penghasilan Orangtua
1	Fillah Syarifah Mz	4	2021	santri-pelajar	MA-NJ	Bahasa Reguler	Petani	Rp. 500.000
2	Lady Gifary Ellizka Wibowo	2	2021	santri-pelajar	SMP-NJ	Reguler	Wiraswasta	Rp. 1.999.999
3	Alifia Raziqa Munir	3	2021	santri-pelajar	MTs-NJ	Keagamaan Unggulan	Pedagan g Kecil	Rp. 1.999.999
4	Fikri Iqbal Zam Zam	5	2021	santri-pelajar	SMA-NJ	IPA Unggulan	Pedagan g Kecil	Rp. 999.999

5	Fahmi Akbar Zam Zam	5	2021	santri-pelajar	SMA-NJ	IPA Unggulan	Petani	Rp. 500.000
6	Muhamma d Ali Qudsi	2	2021	santri-pelajar	MTs-NJ	Reguler	Wiraswasta	Rp. 4.999.999
7	Rizky Romadhoni	4	2021	santri-pelajar	MTs-NJ	Reguler	Wiraswasta	Rp. 4.999.999
8	Tegar Andarbeni Karusan Iman	2	2021	santri-pelajar	MTs-NJ	Reguler	Wiraswasta	Rp. 500.000
9	Muhamma d Zaky Royhan	5	2021	santri-pelajar	SMA-NJ	Bahasa Unggulan	Wiraswasta	Rp. 999.999
10	Rifji Amaliatus Soliha	2	2021	santri-pelajar	MA-NJ	IPA Tahfidz	Nelayan	Rp. 500.000
11	Anindia Wardani	3	2021	santri-pelajar	SMP-NJ	LIPS	Lainnya	Rp. 500.000
12	Nadiatus Sya'adah	2	2021	santri-pelajar	SMP-NJ	LIPS	Petani	Rp. 999.999
13	Nuraida Nadya Handayani	3	2021	santri-pelajar	SMP-NJ	LIPS	Petani	Rp. 500.000
14	Shila Maya	3	2021	santri-pelajar	SMP-NJ	LIPS	Wiraswasta	Rp. 999.999
15	Naura Merdhiyani z Prayogo	2	2021	santri-pelajar	SMP-NJ	LIPS	Petani	Rp. 999.999
16	Moh. Nashihul Ibad	0	2021	santri-pelajar	MTs-NJ	Tahfidz	Petani	Rp. 1.999.999
17	Nasihul Ibad	2	2021	santri-pelajar	MTs-NJ	Reguler	Wiraswasta	Rp. 4.999.999
18	Humairo Aulia Zain	2	2021	santri-pelajar	MTs-NJ	Tahfidz	Wiraswasta	Rp. 500.000
19	Aghistni Yusyairoh Romadhani	2	2021	santri-pelajar	MTs-NJ	Tahfidz	Wiraswasta	Rp. 1.999.999
20	Wildaningrum Ayu Putri Maghfiroh	3	2021	santri-pelajar	MTs-NJ	Tahfidz	Karyawan Swasta	Rp. 1.999.999
21	Rohmatika Fathoni Fitri	2	2021	santri-pelajar	MTs-NJ	Tahfidz	Petani	Rp. 999.999
22	Vina Noor Khoiyah	3	2021	santri-pelajar	MTs-NJ	Tahfidz	Pedagan g Besar	Rp. 500.000
23	Hanuna Alva Binarqie Elhas	4	2021	santri-pelajar	MTs-NJ	Tahfidz	Pedagan g Besar	Rp. 1.999.999
24	Hanunah Nafiyah	3	2021	santri-pelajar	MTs-NJ	Tahfidz	Wiraswasta	Rp. 500.000
25	Raia Maulia Hudiyanto	3	2021	santri-pelajar	MTs-NJ	Tahfidz	Wiraswasta	Rp. 1.999.999
26	Safiratul Firdausi	4	2021	santri-pelajar	MTs-NJ	Tahfidz	Petani	Rp. 1.999.999
27	Marisa Muzayyana	3	2021	santri-pelajar	MTs-NJ	Tahfidz	Petani	Rp. 500.000
28	Rofiatul Andawiyah	2	2021	santri-pelajar	MTs-NJ	Tahfidz	Tidak bekerja	Rp. 500.000
29	Zaskia Ulya Az Zahro	1	2021	santri-pelajar	MTs-NJ	Tahfidz	Sudah Meninggal	Rp. 500.000
30	Hawa Kamela	2	2021	santri-pelajar	MTs-NJ	Tahfidz	Pedagan g Kecil	Rp. 500.000

Dataset yang dipakai dalam proses clustering adalah data Santri SLTP sampai dengan SLTA Pondok Pesantren Nurul Jadid tahun masuk 2021 dengan jumlah data sebanyak 749 santri. Pada tahap ini atribut - atribut data yang tidak diperlukan akan di buang.

Data santri yang ada digunakan sebagai data yang akan diproses, kemudian dilakukan penyesuaian format data agar pengelolaan data lebih efisien. Format data disesuaikan sehingga menjadi bentuk nilai atau angka seperti ditampilkan pada tabel 2 dan 3:

Tabel 2. Kategori Penilaian

No	Pekerjaan Orang Tua	Nilai
1	Meninggal/Tidak Bekerja	0
	Buruh	1
	Pedagang Kecil	2
	Guru	3
	Karyawan Swasta	4
	Nelayan	5
	Petani	6
	Peternak	7
	Wiraswasta	8
	Wirausaha	9
	Pesiunan	10
	Lainnya	11
	Pedagang besar	12
PNS/TNI/Polri	13	

Tabel 3. Penghasilan Orang Tua

Penghasilan Orang Tua	Nilai
Rp. 0 - Rp.500.000	1
Rp. 500.100 - Rp.999.999	2
Rp. 1.000.000 -Rp. 1.999.999	3
Rp. 2.000.000-Rp. 4.999.999	4
Rp. 5.000.000-Rp. 20.000.000	5
Lebih dari Rp. 20.000.000	6

Sesudah proses klasifikasi ditentukan, data santri dilakukan tranformasi ke dalam bentuk nilai atau angka. Berikut proses transformasi teks ke dalam bentuk nilai atau angka yang sudah ditetapkan. Seperti ditampilkan pada tabel 4:

Tabel 4. Tranformasi Nilai

Data ke-	X	Y	Z
1	4	1	1
2	2	4	3

3	3	2	3
4	5	2	2
5	5	1	1
6	4	4	4
7	4	4	4
8	2	4	1
9	5	4	2
10	2	1	1

Banyak *cluster* yang dipakai pada data santri ini berjumlah 4 *cluster*. *Cluster* tersebut diantaranya adalah (C1) Keluarga Pra-Sejahtera (C2) Keluarga Sejahtera 1 Keluarga Sejahtera 2 (C3) dan Keluarga Sejahtera Plus (C4).

Penentuan pusat awal *cluster* (*centroid*) ditentukan secara manual atau random yang diambil dari data yang ada dalam *range* seperti ditampilkan pada tabel 5:

Tabel 5. Data *Centroid* Awal

<i>Centroid Random</i>			
Data ke-	X	Y	Z
89	2	4	2
163	4	6	2
217	2	6	2
225	2	8	2

Dalam perhitung jarak tiap titik *centroid* dengan titik antar objek menggunakan *Euclidian Distance* dengan persamaan atau rumus.

Dalam menentukan *cluster* dengan mencari nilai *cluster* berdasarkan nilai minimal dari nilai *cluster* dan diletakkan pada *cluster* yang sesuai dengan nilai minimal pada iterasi awal. Berikut tabel *cluster* pada iterasi awal seperti ditampilkan pada tabel 6:

Tabel 6. Perhitungan Jarak dan *Cluster* Iterasi Awal

Data ke	X	Y	Z	C1	C2	C3	C4	Jarak Terdekat	Keterangan
1	4	6	1	3	1	2,236 068	3	1	Cluster 2
2	2	8	3	4,123 106	3	2,236 068	1	1	Cluster 4
3	3	2	3	2,449 49	4,242 641	4,242 641	6,16 4414	2,44 949	Cluster 1
4	5	2	2	3,605 551	4,123 106	5	6,70 8204	3,60 5551	Cluster 1
5	5	6	1	3,741 657	1,414 214	3,162 278	3,74 1657	1,41 4214	Cluster 2
6	2	8	4	4,472 136	3,464 102	2,828 427	2	2	Cluster 4

7	4	8	4	4,898	2,828	3,464	2,82	2,82	Cluster
				979	427	102	8427	8427	2
8	2	8	1	4,123	3	2,236	1	1	Cluster
				106		068			4
9	5	8	2	5	2,236	3,605	3	2,23	Cluster
					068	551		6068	2
10	2	5	1	1,414	2,449	1,414	3,16	1,41	Cluster
				214	49	214	2278	4214	1
11	3	11	1	7,141	5,196	5,196	3,31	3,31	Cluster
				428	152	152	6625	6625	4
12	2	6	2	2	2	0	2	0	Cluster
									3
13	3	6	1	2,449	1,414	1,414	2,44	1,41	Cluster
				49	214	214	949	4214	2
14	3	8	2	4,123	2,236	2,236	1	1	Cluster
				106	068	068			4
15	2	6	2	2	2	0	2	0	Cluster
									3
16	0	6	3	3	4,123	2,236	3	2,23	Cluster
					106	068		6068	3
17	2	8	4	4,472	3,464	2,828	2	2	Cluster
				136	102	427			4
18	2	8	1	4,123	3	2,236	1	1	Cluster
				106		068			4
19	2	8	3	4,123	3	2,236	1	1	Cluster
				106		068			4
20	3	4	3	1,414	2,449	2,449	4,24	1,41	Cluster
				214	49	49	2641	4214	1
21	2	6	2	2	2	0	2	0	Cluster
									3
22	3	12	1	8,124	6,164	6,164	4,24	4,24	Cluster
				038	414	414	2641	2641	4
23	4	12	3	8,306	6,082	6,403	4,58	4,58	Cluster
				624	763	124	2576	2576	4
24	3	8	1	4,242	2,449	2,449	1,41	1,41	Cluster
				641	49	49	4214	4214	4
25	3	8	3	4,242	2,449	2,449	1,41	1,41	Cluster
				641	49	49	4214	4214	4
26	4	6	3	3	1	2,236	3	1	Cluster
						068			2
27	3	6	1	2,449	1,414	1,414	2,44	1,41	Cluster
				49	214	214	949	4214	2
28	2	0	1	4,123	6,403	6,082	8,06	4,12	Cluster
				106	124	763	2258	3106	1
29	1	0	1	4,242	6,782	6,164	8,12	4,24	Cluster
				641	33	414	4038	2641	1
30	2	2	1	2,236	4,582	4,123	6,08	2,23	Cluster
				068	576	106	2763	6068	1

Hasil dari *cluster* pada iterasi awal seperti ditampilkan pada tabel 7:

Tabel 7. Hasil *Cluster* Iterasi awal

Cluster	Jumlah Data
C1	176
C2	153
C3	159
C4	261

Selanjutnya dalam metode *K-Means*, perhitungan berhenti apabila *cluster* pada iterasi yang dihasilkan sama pada iterasi sebelumnya. Maka selanjutnya mencari *cluster* pada iterasi selanjutnya sampai nilai iterasinya sama. Untuk mencari nilai *centroid* selanjutnya dengan menggunakan *centroid* baru pada iterasi ke-1 dan seterusnya dengan menjumlahkan nilai

sesuai yang tertera pada *cluster* tabel di atas. Adapun *centroid* baru untuk mencari *cluster* selanjutnya adalah dengan menjumlahkan nilai yang terpilih pada *cluster* tersebut kemudian membagikannya sebanyak jumlah nilai. Maka didapat hasil data *centroid* baru pada iterasi terakhir seperti ditampilkan pada tabel 8:

Tabel 8. Data *Centroid* Akhir

<i>Centroid</i> Iterasi ke-7			
<i>Centroid</i>	X	Y	Z
C1	2,552147239	1,889570552	1,803680982
C2	3,451162791	6,772093023	1,725581395
C3	1,792682927	6,589430894	1,756097561
C4	2,512	11,856	2,24

Dengan menggunakan langkah-langkah yang sama seperti sebelumnya untuk menentukan jarak dari *centroid* pada iterasi terakhir dengan menggunakan *centroid* baru, maka didapatkan hasil keseluruhan dari iterasi terakhir seperti ditampilkan pada tabel 9:

Tabel 9. Perhitungan Jarak dan *Cluster* Iterasi Akhir

Data Ke-	X	Y	Z	C1	C2	C3	C4	Jarak Terdekat	Keterangan
1	4	6	1	4,431	1,193	2,406	6,168	1,193	Cluster 2
				457	239	525	021	239	
2	2	8	3	6,250	2,288	1,892	3,963	1,892	Cluster 3
				871	618	083	392	083	
3	3	2	3	1,282	4,959	4,905	9,897	1,282	Cluster 1
				165	895	893	297	165	
4	5	2	2	2,458	5,024	5,604	10,16	2,458	Cluster 1
				194	647	395	801	194	
5	5	6	1	4,851	1,876	3,347	6,482	1,876	Cluster 2
				136	564	536	321	564	
6	2	8	4	6,516	2,964	2,658	4,269	2,658	Cluster 3
				597	221	53	482	53	
7	4	8	4	6,652	2,642	3,449	4,492	2,642	Cluster 2
				627	339	21	269	339	
8	2	8	1	6,187	2,034	1,613	4,082	1,613	Cluster 3
				739	723	806	705	806	
9	5	8	2	6,585	1,995	3,512	4,595	1,995	Cluster 2
				429	484	275	267	484	
10	2	5	1	3,259	2,402	1,772	6,986	1,772	Cluster 3
				684	635	274	02	274	
11	3	1	1	9,156	4,313	4,634	1,583	1,583	Cluster 4
				768	376	913	818	818	
12	2	6	2	4,151	1,666	0,670	5,883	0,670	Cluster 3
				992	525	744	237	744	
13	3	6	1	4,212	1,151	1,541	6,005	1,151	Cluster 2
				138	583	664	704	583	
14	3	8	2	6,129	1,336	1,872	3,894	1,336	Cluster 2
				964	641	647	16	641	
15	2	6	2	4,151	1,666	0,670	5,883	0,670	Cluster 3
				992	525	744	237	744	
16	0	6	3	4,984	3,759	2,260	6,417	2,260	Cluster 3
				001	095	185	202	185	
17	2	8	4	6,516	2,964	2,658	4,269	2,658	Cluster 3
				597	221	53	482	53	
18	2	8	1	6,187	2,034	1,613	4,082	1,613	Cluster 3
				739	723	806	705	806	

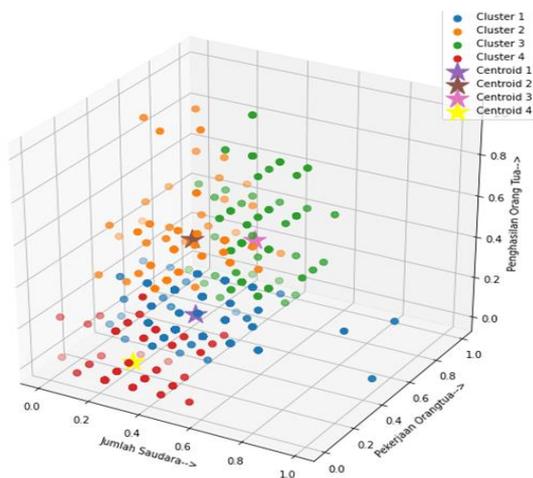
1	2	8	3	6,250	2,288	1,892	3,963	1,892	Cluster
9				871	618	083	392	083	3
2	3	4	3	2,466	3,084	3,116	7,907	2,466	Cluster
0				914	184	097	748	914	1
2	2	6	2	4,151	1,666	0,670	5,883	0,670	Cluster
1				992	525	744	237	744	3
2	3	1	1	10,15	5,297	5,594	1,340	1,340	Cluster
2				22	266	958	328	328	4
2	4	1	3	10,28	5,408	5,974	1,677	1,677	Cluster
3				34	916	429	045	045	4
2	3	8	1	6,179	1,495	2,004	4,079	1,495	Cluster
4				306	918	745	765	918	2
2	3	8	3	6,242	1,826	2,234	3,960	1,826	Cluster
5				523	32	863	364	32	2
2	4	6	3	4,519	1,587	2,601	6,089	1,587	Cluster
6				191	921	34	703	921	2
2	3	6	1	4,212	1,151	1,541	6,005	1,151	Cluster
7				138	583	664	704	583	2
2	2	0	1	2,126	6,963	6,635	11,93	2,126	Cluster
8				322	734	907	166	322	1
2	1	0	1	2,574	7,238	6,679	12,01	2,574	Cluster
9				013	502	867	618	013	1
3	2	2	1	0,981	5,040	4,655	9,946	0,981	Cluster
0				307	358	914	883	307	1

Hasil dari *cluster* pada iterasi akhir seperti ditampilkan pada tabel 10:

Tabel 10. Hasil *Cluster* Iterasi akhir

Cluster	Jumlah Data
C1	163
C2	215
C3	246
C4	125

Perhitungan pada data Santri SLTP sampai dengan SLTA Pondok Pesantren Nurul Jadid tahun masuk 2021, didapatkan iterasi akhir pengelompokan data yang dilakukan didapatkan hasil yang sama pada iterasi sebelumnya. Sehingga posisi *cluster* pada data tersebut tidak mengalami perubahan lagi maka proses iterasi berhenti. Berikut gambar hasil Visualisasi data *Clustering* seperti ditampilkan pada gambar 2 :



Gambar 2. Visualisasi data hasil *Clustering*

Pada tahap akhir ini dilakukan evaluasi hasil dari implemtasi algoritma diatas, menggunakan matrik DBI (*Davies Bouldin Index*) untuk mendapatkan nilai akurasi.

Dari evaluasi hasil *clustering* menggunakan matrik DBI (*Davies Bouldin Index*) memperoleh nilai akurasi sebesar 0,90 (90%) yang mana hasil *clustering* menunjukkan hasil dengan katagori baik, karena nilai akurasi DBI yang baik adalah nilai yang mendekati 1 dan 0 namun tidak minus (-).

## V. KESIMPULAN

Berdasarkan hasil dari pengamatan dan penelitian yang telah dilakukan pada data santri maka dapat diambil kesimpulan. Dari total 749 data santri SLTA dan SLTP tahun 2021 diatas, mendapatkan hasil perhitungan terakhir yaitu pada Iterasi ke-7 dan data hasil *Clustering* menunjukan bahwa C1 = 163 (Sejahtera 1), C2 = 215 (Sejahtera 2), C3 = 246 (Sejahtera Plus) dan C4 = 125 (Keluarga Pra-Sejahtera). Evaluasi hasil akhir dari clustering menggunakan matriks DBI (*Davies Bouldin Index*) memperoleh akurasi nilai sebesar 0,90 (90%) yang menunjukan pengclustering menggunakan algoritma *k-means* sudah sangat baik. Analisa perhitungan data santri diatas masih sangat jauh dari kata sempurna, sehingga perlu untuk dilakukan perbaikan dan pengembangan, maka ada beberapa saran dari penelitian ini untuk kedepannya. *Clustering* data santri pra-sejahtera menggunakan metode algoritma *k-means* ini, dapat dikembangkan lebih lanjut menggunakan metode yanag lain dalam menganalisis data dan mengimplementasikan dalam bentuk aplikasi berbasis *web* dan *android* sehingga bisa menghasilkan sistem aplikasi yang efisien serta siap pakai, sehingga data hasil pengolahan bisa menjadi acuan di YBS dalam proses pemilihan santri kurang mampu.

## REFERENCES

- [1] Presiden Republik Indonesia, "UU nomor 52 Tahun 2009 ten-tang Perkembangan Kependudukan dan Pembangunan Keluar-ga," J Am Chem Soc, vol. 123, no. 10, 2009.
- [2] F. Maulidah, "Proyeksi Penduduk Dan Bonus Demografi Kota Surabaya Tahun 2020-2045 Dengan Menerapkan Aplikasi Spec-trum," Jurnal Penelitian Kesehatan (Jpk), vol. 18, no. 1, 2020.
- [3] D. Fitriani, T. N. Padilah, and B. N. Sari, "Penerapan Algoritma K-Means Dalam Pengelompokan Kesejahteraan Rakyat Berdasar-kan Kecamatan di Kabupaten Karawang," Progresif: Jurnal Ilmiah Komputer, vol. 17, no. 2, 2021, doi: 10.35889/progresif.v17i2.649.
- [4] P. Magal, G. M. V. Kawung, and M. TH. B. Maramis, "Pengaruh Alokasi Dana Desa terhadap Kemiskinan dan Peningkatan Kesejahteraan Masyarakat di Desa Solimandungan Baru Kecamatan Bolaang Kabupaten Bolaang Mongondow," Jurnal EMBA, vol. 9, no. 1, 2021.
- [5] S. C. Azhari, E. Mulyanie, and S. I. Saputri, "Kegiatan Verifikasi Dan Validasi Data Program Keluarga Harapan (Pkh) Dan Bpnt Melalui Program Pejuang Muda Di Kecamatan Banjaranyar Kabupaten

- Ciamis,” *Jurnal Pengabdian Masyarakat Ilmu Keguruan dan Pendidikan (JPM-IKP)*, vol. 5, no. 1, 2022.
- [6] W. Gie and D. Jollyta, “Perbandingan Euclidean dan Manhattan Untuk Optimasi Cluster Menggunakan Davies Bouldin Index: Status Covid-19 Wilayah Riau,” *Prosiding Seminar Nasional Riset Dan Information Science (SENARIS)*, vol. 2, 2020.
- [7] D. Triyansyah and D. Fitriana, “Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing,” *Jurnal Telekomunikasi dan Komput-er*, vol. 8, no. 3, 2018, doi: 10.22441/incomtech.v8i3.4174.
- [8] K. P. Sinaga and M. S. Yang, “Unsupervised K-means clustering algorithm,” *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.2988796.
- [9] M. Ahmed, R. Seraj, and S. M. S. Islam, “The k-means algorithm: A comprehensive survey and performance evaluation,” *Electronics (Switzerland)*, vol. 9, no. 8. 2020. doi: 10.3390/electronics9081295.
- [10] R. Al Izzati, “Situasi Kemiskinan Selama Pandemi,” *The SMERU Research Institute*, 2021.
- [11] Yuni Radana Sembiring, Saifullah, and Riki Winanjaya, “Implementasi Data Mining Dalam Mengelompokkan Jumlah Penduduk Miskin Berdasarkan Provinsi Menggunakan Algoritma,” *KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen) Vol. 2, No. 2, vol. 2, no. 2, 2021.*
- [12] A. Khairi, A. F. Ghozali, and A. D. N. Hidayah, “Implementasi K-Nearest Neighbor (KNN) untuk Mengklasifikasi Masyarakat Pra-Sejahtera Desa Sapikerep Kecamatan Sukapura,” *TRILOGI: Jurnal Ilmu Teknologi, Kesehatan, dan Humaniora*, vol. 2, no. 3, 2021, doi: 10.33650/trilogi.v2i3.2878.
- [13] Y. Fitriadi, W. Novita, and D. Endriani, “Ekonomi Kreatif Sebagai Solusi Bagi Keluarga Pra Sejahtera Untuk Bertahan Dalam Era New Normal,” *To Maega : Jurnal Pengabdian Masyarakat*, vol. 4, no. 1, 2021, doi: 10.35914/tomaega.v4i1.473.
- [14] R. A. Farissa, R. Mayasari, and Y. Umaidah, “Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokkan Data Obat dengan Silhouette Coefficient di Puskesmas Ka-rangsambung,” *Journal of Applied Informatics and Computing*, vol. 5, no. 2, 2021, doi: 10.30871/jaic.v5i1.3237.
- [15] S. R. Durugkar, R. Raja, K. K. Nagwanshi, and S. Kumar, “Introduction to data mining,” *Data Mining and Machine Learning Applications. 2022.* doi: 10.1002/9781119792529.ch1.
- [16] M. M. Ghazal and A. Hammad, “Application of knowledge discovery in database (KDD) techniques in cost overrun of construction projects,” *International Journal of Construction Management*, vol. 22, no. 9, 2022, doi: 10.1080/15623599.2020.1738205.
- [17] D. Abdullah, S. Susilo, A. S. Ahmar, R. Rusli, and R. Hidayat, “The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data,” *Qual Quant*, vol. 56, no. 3, 2022, doi: 10.1007/s11135-021-01176-w.
- [18] A. Nur Khormarudin, “Teknik Data Mining: Algoritma K-Means Clustering,” *Jurnal Ilmu Komputer*, 2016.
- [19] Y. Niagara, Ernawati, and E. P. Purwandari, “Pemanfaatan Citra Penginderaan Jauh Untuk Pemetaan Klasifikasi Tutupan Lahan Menggunakan Metode Unsupervised K-Means Berbasis Web Gis (Studi ...,” *Rekursif: Jurnal Informatika*, vol. 8, no. 1, 2020.
- [20] N. Zulfa, R. I. Auliya, and A. Zaenal, “Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means,” *Jurnal Teknologi dan Sistem Informasi (JTSI)*, vol. 2, no. 2, 2021.