

PENGELOMPOKAN PEMBAGIAN ZAKAT DENGAN MENGUNAKAN METODE *CLUSTERING K-MEANS*

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Teknik
Pada Jurusan Teknik Informatika

Oleh:

ALVIN ANZAZ ISLAMI
NIM. 11750115097



FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU

PEKANBARU

2023 M

- Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PERSETUJUAN

**PENGELOMPOKAN PEMBAGIAN ZAKAT DENGAN
MENGUNAKAN METODE *CLUSTERING K-MEANS***

TUGAS AKHIR

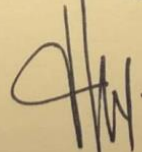
Oleh

ALVIN ANZAZ ISLAMI

NIM. 11750115097

Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir
di Pekanbaru, pada tanggal 15 Juni 2023

Pembimbing I,



Dr. ELIN HAERANI, S.T., M.Kom.

NIP. 19810523 200710 2 003

- Hak Cipta Dilindungi Undang-Undang**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



LEMBAR PENGESAHAN

PENGELOMPOKAN PEMBAGIAN ZAKAT DENGAN MENGUNAKAN METODE *CLUSTERING K-MEANS*

Oleh

ALVIN ANZAZ ISLAMI

NIM. 11750115097

Telah dipertahankan di depan sidang dewan penguji
sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik
pada Universitas Islam Negeri Sultan Syarif Kasim Riau

Pekanbaru, 15 Juni 2023

Mengesahkan,
Ketua Jurusan,



Dr. Hartono, M.Pd.

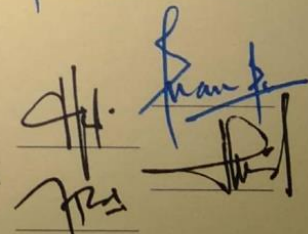
NIP. 19640301 199203 1 003

Iwan Iskandar, M.T.

NIP. 19821216 201503 1 003

DEWAN PENGUJI

Ketua : Iwan Iskandar, M.T.
Pembimbing I : Dr. Elin Haerani, S.T., M.Kom.
Penguji I : Suwanto Sanjaya, S.T., M.Kom.
Penguji II : Novriyanto, S.T., M.Sc.



**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Alvin Anzaz Islami
NIM : 11750115097
Tempat/Tgl. Lahir : Batam, 08 Desember 1998
Fakultas : Sains dan Teknologi
Prodi : Teknik Informatika
Judul Skripsi : Pengelompokan Pembagian Zakat Dengan Menggunakan Metode *Clustering K-Means*

Menyatakan dengan sebenar-benarnya bahwa:

1. Penulisan Skripsi dengan judul sebagaimana tersebut di atas adalah hasil pemikiran dan penelitian saya sendiri.
2. Semua kutipan pada karya tulis saya ini sudah disebutkan sumbernya.
3. Oleh karena itu Skripsi saya ini, saya nyatakan bebas plagiat.
4. Apabila dikemudian hari terbukti terdapat plagiat dalam penulisan Skripsi saya ini, maka saya bersedia menerima sanksi sesuai peraturan perundang-undangan.

Demikian pernyataan ini saya buat dengan penuh kesadaran dan tanpa paksaan dari pihak manapun.

Pekanbaru, 12 Juli 2023

Yang membuat pernyataan



ALVIN ANZAZ ISLAMI
NIM. 11750115097

Pengelompokan pembagian zakat dengan menggunakan metode *clustering k-means*

Anza Islami^{1*}, Elin Haerani², Novriyanto³, Alwis Nazir⁴

Email: ¹1100115097@students.uin-suska.ac.id, ²elin.haerani@uin-suska.ac.id, ³novriyanto@uin-suska.ac.id, ⁴alwis.nazir@uin-suska.ac.id

^{1,2,3,4} Teknik Informatika, Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau

Diterima: 06 April 2023 | Direvisi: 19 April 2023 | Disetujui: 30 April 2023

¹ Program Studi Teknik Informatika Fakultas Ilmu Komputer,
^{2,3,4} Universitas Muhammadiyah Riau, Indonesia

Abstrak

Zakat merupakan ibadah yang menyangkut harta benda. Zakat juga termasuk rukun islam yang ke empat yang memiliki tujuan menyekatkan harta bagi setiap muslim dengan cara menysihkan sebagian harta kekayaannya, jika telah mencapai waktu dan besaran jumlahnya diberikan kepada orang yang berhak menerimanya. Pengumpulan dan penyaluran zakat biasanya ditangani oleh Badan Amil Zakat (BAZ) yang ada disetiap wilayah Indonesia, salah satunya di Pekanbaru. Sesuai dengan peraturan ada dua tahap yang dilakukan dalam memberikan bantuan kepada para mustahik yaitu melakukan wawancara dan observasi lapangan, kemudian menentukan nominal bantuan yang diberikan dengan kategori Mustahik penerima bantuan zakat 1, zakat 2, dan zakat 3. Masalah yang sering dijumpai dalam penentuan calon penerima bantuan adalah cara dalam pemilihan Mustahik yang masih menggunakan cara manual, sehingga sering menimbulkan masalah seperti lamanya proses pemilihan dan terjadinya salah hitung sehingga hasil seleksi Mustahik menjadi kurang akurat. Untuk itu, perlu dibuat suatu analisis yang dapat mengolah data menjadi informasi. *Data mining* ialah proses untuk mengolah data menjadi suatu informasi dengan teknik statistik, AI, dan *machine learning*. Ada banyak metode dalam *data mining*. Pada penelitian ini menggunakan algoritma *k-means clustering* dan untuk pengujian menggunakan *Davies Bouldin Index*. berdasarkan pengujian menggunakan *davies bouldin index* (DBI) kluster 4 merupakan kluster terbaik dengan nilai 0.671, dimana jika nilainya semakin rendah maka akan semakin baik kluster tersebut

Kata kunci: *clustering, data mining, k-means, davies bouldin index*

Grouping the distribution of zakat by using the method *clustering k-means*

Abstract

Zakat is a worship that involves property. Zakat is also included in the fourth pillar of Islam which has the aim of purifying the assets of every Muslim by setting aside a portion of his wealth, if it has reached the time and the amount is given to those who are entitled to receive it. The collection and distribution of zakat is usually handled by the Amil Zakat Agency (BAZ) in every region of Indonesia, one of which is in Pekanbaru. In accordance with the regulations, there are two stages in providing assistance to mustahik, namely conducting interviews and field observations, then determining the nominal amount of assistance given to the Mustahik category of recipients of zakat 1, zakat 2, and zakat 3. Problems that are often encountered in determining potential recipients assistance is a way of selecting Mustahik which still uses the manual method, so that it often causes problems such as the length of the selection process and the occurrence of miscalculations so that the results of Mustahik's selection become inaccurate. For that, it is necessary to create analysis that can process data into information. *Data mining* is a process for processing data into information using statistical techniques, AI, and *machine learning*. There are many methods in *data mining*. In this study using the *k-means clustering* and for testing use *Davies Bouldin Index*. based on testing using the *davies bouldin index* (DBI) cluster 4 is the best cluster with a value of 0.671, where the lower the value, the better the cluster.

Keywords: *clustering, data mining, k-means, davies bouldin index*



I. PENDAHULUAN

Zakat dalam agama Islam merupakan pembersih dari harta kekayaan yaitu dengan cara menyisihkan setengah dari harta kekayaan apabila sudah sampai waktu dan jumlahnya diberikan kepada orang yang berhak menerimanya [1]. Zakat juga memiliki arti sebagai ibadah yang berkaitan dengan harta benda yang berfungsi sosial. Zakat termasuk kedalam rukun islam yang keempat yang memiliki tujuan untuk membersihkan harta bagi setiap muslim.

Undang-Undang No. 23 tahun 2011 tentang cara pengolahan zakat, bahwasannya dalam mengolah zakat bisa dilakukan oleh lembaga pemerintah dan badan amil zakat nasional. Maksud dari pernyataan tersebut adalah operasi pengumpulan dan penyaluran zakat pada daerah-daerah diatur dan dikendalikan oleh badan nasional yaitu Badan Amil Zakat (BAZ) yang ada di setiap wilayah Indonesia terutama di kota Pekanbaru. Badan Amil Zakat Nasional (BAZNAS) ialah organisasi pengelola dan pengoperasian zakat yang ditangani oleh pemerintah, yang tersedia di setiap kabupaten maupun kota [2].

Zakat mempunyai potensi untuk mengurangi kesenjangan sosial. Rasio antara penerima zakat dan Muzakki yang ada di provinsi Riau mencapai 0,56 artinya 100 Muzakki dapat menanggung 56 Mustahik. Sesuai dengan prosedur yang ada, para Muzakki melakukan pembayaran zakat di kantor Baznas Pekanbaru. Zakat yang di terima Baznas dari Muzakki akan disalurkan kepada Mustahik. Sesuai dengan peraturan, ada dua tahap yang akan dilakukan oleh Baznas Pekanbaru dalam memberikan bantuan kepada para Mustahik. Pertama, Baznas melakukan wawancara pada setiap calon penerima, kemudian melakukan observasi lapangan untuk menentukan orang yang berhak menjadi Mustahik. Kedua, menentukan nominal bantuan yang diberikan dengan kategori Mustahik penerima bantuan zakat 1, zakat 2, dan zakat 3 [3].

Masalah yang sering dijumpai dalam penentuan calon penerima bantuan ialah pada saat penentuan Mustahik yang masih menggunakan cara manual, sehingga sering timbul masalah seperti lamanya proses dalam pemilihan dan terjadinya salah hitung sehingga hasil seleksi Mustahik menjadi kurang akurat. Selain itu tidak tersedianya akses informasi untuk melihat Mustahik yang sudah masuk ke dalam daftar penerima zakat, akibatnya Mustahik yang sudah menerima zakat dapat menerima bantuan dari satu kali dalam setahun. Hal ini akan memungkinkan terjadinya perbedaan dan ketidaklayakan calon penerima zakat pada efisiensi waktu yang kurang maksimal. Selain itu perhitungan kelayakan akan membutuhkan waktu yang lama jika jumlah Mustahik yang mengajukan permohonan zakat dalam waktu bersamaan terlalu banyak, sehingga proses pemilihan dan penyaluran zakat akan membutuhkan waktu dan dituntut agar mampu menghasilkan hasil operasi kelayakan yang akurat dan tepat sasaran. Untuk itu, perlu dibuat suatu analisa yang dapat mengolah data menjadi informasi pada penentuan kelayakan calon penerima zakat serta besaran dana.

Data mining atau yang dikenal dengan *Knowledge Discovery in Database (KDD)* ialah tahapan untuk analisis data dari sudut pandang yang berbeda dan menjadikannya menjadi informasi-informasi yang menggunakan teknik statistik, *machine learning*, *artificial intelegent* dan matematika untuk mengekstraksi dan identifikasi informasi yang berguna dan pengetahuan yang terkait dari berbagai *database* yang besar demi menaikkan keuntungan, mengurangi biaya pengeluaran, atau keduanya [4]. Metode yang digunakan pada penelitian ini ialah algoritma *k-means*. Algoritma *k-means* merupakan salah satu metode klasterisasi [5]. Prinsip utama metode ini yaitu menyusun K buah centroid dari beberapa data [6]. Tujuan klasterisasi yaitu untuk menjadikan data ke dalam bentuk beberapa kalster, sehingga data yang berada di satu klaster akan mempunyai tingkat kesamaan yang tinggi dan data antar klaster akan memiliki tingkat kesamaan yang kecil [4].

Penelitian sebelumnya yang menggunakan metode *k-means* clustering ada banyak sekali, salah satunya ialah penelitian terkait yang dilakukan oleh Sulastri, dkk. tentang mengembangkan sistem untuk pengelompokan penderita *Thalassaemia* dengan metode *K-Means*. Pada penelitian ini kriteria yang digunakan terdiri dari umur, Hb, level, dan jumlah darah dengan tujuan untuk mengelompokkan data kriteria penderita *Thalassaemia*. Hasilnya adalah klaster yang digunakan berjumlah 3 kelompok, yaitu kelompok 1 berjumlah 214 data, kelompok 2 berjumlah 138 data, dan kelompok 3 berjumlah 23 data [7]. Penelitian lain yang dilakukan oleh Azwan tentang *clustering* pada data fasilitas pelayanan kesehatan Kota Pekan baru menggunakan metode *k-means*, hasil penelitiannya adalah menghasilkan klaster yang optimal dengan menggunakan uji validasi *Davies bouldien index* yang menghasilkan 9 klaster [8].

Berdasarkan peneliasan yang telah disampaikan, maksud dari penelitian ini adalah untuk pengelompokan pembagaian zakat menggunakan *clustering k-means*. Penelitian ini menggunakan data yang berasal dari Baznas Kota Pekanbaru, dengan menggunakan data pada tahun 2020 sampai 2021 yang berjumlah 1174 data. Atribut yang digunakan yaitu penghasilan, kondisi rumah, dan kepemilikan harta dengan parameter K yang digunakan adalah $k=2$ sampai dengan $k=5$ yang akan diimplementasikan pada pemrograman *python*. Tujuan dari penelitian ini adalah membentuk *cluster* berdasarkan jenis dan jumlah pembagian zakat menggunakan algoritma *k-means* sehingga bisa menghasilkan suatu informasi untuk pengelompokan pembagian zakat di Baznas dalam menentukan para penerima zakat.

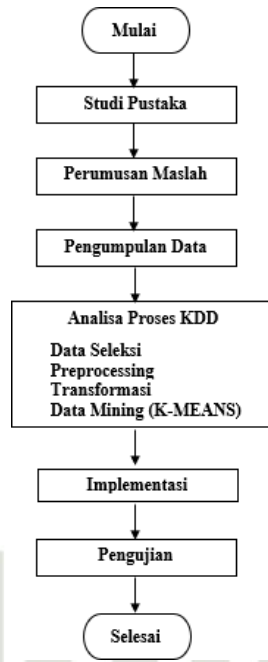
2. METODE PENELITIAN

Metodologi penelitian ialah gambaran tahapan proses yang dijadikan panduan dan dilakukan secara berurutan supaya tujuan dan hasil sesuai dengan yang diminta. Berikut merupakan tahapan-tahapan yang akan dilakukan pada penelitian yang digambarkan seperti dibawah ini:



Hak Cipta Dilindungi Undang-Undang

1. Diarang mengutip sebagian atau seluruh karya tulis ini tanpa mengutip sumbernya.
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan artikel atau surat kabar, atau sebagai sumber informasi umum.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Diarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 1. Metodologi Penelitian

1. Studi Pustaka

Studi Pustaka ini bertujuan untuk mencari informasi-informasi beserta teori yang mendukung dalam pendefinisian masalah yang ada pada penelitian ini. Sumber dari buku-buku, jurnal, penelitian atau referensi lain yang berhubungan dengan penelitian [9].

2. Perumusan Masalah

Perumusan masalah adalah tahapan awal dalam penelitian dengan mengidentifikasi dan memahami permasalahan yang ada dalam penelitian dan menemukan solusi dari masalah tersebut. Rumusan masalah pada penelitian ini yaitu bagaimana menerapkan algoritma *k-means* dalam mengelompokkan data pembagian ke dalam beberapa cluster.

2.3 Pengumpulan Data

Tahapan pengumpulan data ialah proses yang memiliki tujuan untuk mendapatkan data-data dan informasi yang dibutuhkan dalam penelitian [10]. Data yang dipakai pada penelitian ini ialah data permohonan zakat tahun 2020 sampai 2021 yang bersumber dari Baznas Kota Pekanbaru yang berjumlah 1174 data yang atributnya terdiri dari jenis kelamin, nama pemohon, nomor rekening, alamat, kecamatan, kelurahan, telepon, tempat lahir, tanggal lahir, umur, status perkawinan, program, sub program, rencana penggunaan dana, penghasilan, kondisi rumah, kepemilikan harta, jumlah usulan, tanggungan, pekerjaan, dan usulan bantuan. Berikut merupakan dataset dari Baznas kota Pekanbaru dapat dilihat pada tabel berikut:

Tabel 1. Dataset

Jenis Kelamin	Nama	Usulan Bantuan	...	Penghasilan	Kondisi Rumah	Kepemilikan Harta
L	Jasman	Bantuan Biaya Pendidikan	...	Rp.1.200.000	Sewa	< 5 juta
P	Sahyati Harahap	Bantuan Biaya Pendidikan	...	Rp.700.000	Kontrak	< 5 juta
L	Defianto	Bantuan Biaya Pendidikan	...	Rp.2.100.000	Kontrak	< 5 juta
P	Noor Amelia Sumarni	Bantuan Biaya Pendidikan	...	-	Tinggal dengan Ortu	< 10 juta
L	Abdul Muhardi	Bantuan Biaya Pendidikan	...	Rp.1.350.000	Menumpang	< 5 juta
....
1174	L	Jonywat	Bantuan Biaya Pendidikan	...	-	-

2.4 Analisa Proses Knowledge Discovery in Database (KDD)

Proses ini adalah bagian utama dalam penelitian. Ada beberapa tahapan dalam proses *Knowledge Discovery in Database*, diantaranya adalah *data selection, preprocessing, transformation, data mining* [11].

a. Data selection

Data selection ialah proses memilih data yang akan digunakan dari sekumpulan data yang ada. Tahap ini perlu dilakukan sebelum tahap pengkajian informasi dalam KDD. Data dari hasil seleksi ini akan disimpan dalam satu berkas dan terpisah dari basis data operasional [4]. Dari data yang didapatkan tidak semuanya digunakan, atribut yang digunakan hanya penghasilan, kondisi rumah, dan kepemilikan harta.

b. Cleaning



Sebelum masuk ke tahap data mining, data akan melalui tahap *cleaning* terlebih dahulu, tujuannya yaitu untuk menghapus data duplikat, memeriksa data yang tidak konsisten, dan memperbaiki data [12]. Jumlah data yang kosong adalah 41 data, sehingga data berubah menjadi 1133 dari 1174 data. Data yang telah melewati tahap *cleaning* dapat dilihat pada tabel 2:

Tabel 2. Data Hasil *Cleaning*

No	Penghasilan	Kondisi Rumah	Kepemilikan Harta
	Rp.1.200.000	Sewa	< 5 juta
	Rp.700.000	Kontrak	< 5 juta
	Rp.2.100.000	Kontrak	< 5 juta
	Rp.1.000.000	Menumpang	< 10 juta
	Rp.1.350.000	Menumpang	< 5 juta

1133	Rp.2.100.000	Hak milik	> 5 juta

Transformasi

Transformasi adalah tahap mengubah jenis atau pola informasi yang akan digunakan dalam *database*. Pada tugas akhir ini setiap atribut akan memiliki rentang nilai yang berbeda, sehingga dapat menyebabkan tidak berfungsinya atribut yang memiliki nilai jauh lebih kecil dibandingkan dengan atribut lainnya. Maka diperlukan tahap transformasi dengan metode normalisasi yang menjadikan rentang nilai dari setiap atribut dapat disamakan pada rentang nilai tertentu [13]. Berikut merupakan inisialisasi data pada penelitian ini:

Tabel 3. Inisialisasi Penghasilan

Penghasilan	Inisialisai
<= Rp. 500.000	1
> Rp.500.000 s/d Rp. 1 jt	2
> Rp.1 jt s/d Rp. 1,5 jt	3
> Rp. 1,5 jt s/d Rp. 2 jt	4
> Rp.2 jt s/d Rp. 2,5 jt	5
> Rp. 2,5 jt s/d Rp. 3 jt	6
> 3 jt	7

Tabel 4. Inisialisasi Kondisi Rumah

Kondisi Rumah	Inisialisai
Sewa	1
Kontrak	2
Menumpang	3
Tinggal dengan orang tua	4
Hak milik	5

Tabel 5. Inisialisasi Kepemilikan Harta

Kepemilikan Harta	Inisialisai
<= 5 juta	1
> 5 juta sampai 10 juta	2
>10 juta	3

Data Mining

Pada tahap ini, data akan dianalisis menggunakan salah satu metode *data mining* yaitu *k-means*. Metode *k-means* ialah salah satu metode yang digunakan untuk proses pengelompokan data dengan cara partisi yang memisahkan data menjadi kluster-kluster [14]. Data yang akan dikelompokkan ialah data yang sudah melewati proses seleksi, *cleaning*, dan transformasi. Tahapan proses *k-means* dijelaskan dalam *flowchart* seperti gambar di bawah ini:

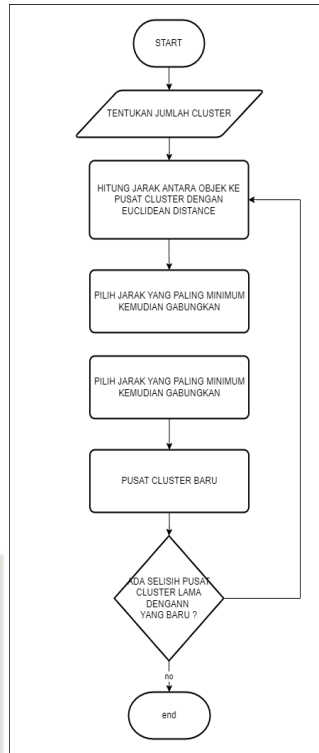


Hak Cipta Dilindungi Undang-undang
 1. Di larang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumbernya.
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Di larang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mengemukakan sumbernya.
- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penerbitan kritik atau penerbitan lain-lain.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 2 . Flowchart K-Means

Dari gambar di atas bisa kita lihat tahapan pertama pada prose *k-means* ialah menentukan jumlah *cluster*, kemudian menentukan *centroid*. Titik pusat pada tahap awal dilakukan secara random kemudian untuk tahap iterasi menggunakan rumus pada persamaan berikut:

$$c_j = \frac{1}{n_j} \sum_{k=0}^{n_j-1} x_{kj} \quad (1)$$

Keterangan:

- c_j = centroid rata-rata cluster ke-*i* untuk variabel ke-*j*
- n_j = jumlah anggota cluster ke-*i*
- x_{kj} = indeks dari cluster
- x_{kj} = indeks dari variabel x_{kj} = nilai data ke-*k* variabel ke-*j* dalam cluster tersebut

Selanjutnya ialah menghitung jarak objek ke titik pusat menggunakan rumus *euclidean distance*, berikut merupakan rumus persamaan dari *euclidean distance*:

$$DE_{ij} = \sqrt{(x_i - s_j)^2 + (y_i - t_j)^2} \quad (2)$$

Keterangan:

- DE_{ij} = Euclidean distance
- i = banyaknya objek
- (x_i, y_i) = koordinat objek
- (s_j, t_j) = koordinat centroid

Langkah selanjutnya yaitu mengelompokkan objek berdasarkan jarak ke centroid terdekat, kemudian ulangi langkah ke-2, lakukan perulangan hingga hasil iterasi sama dengan hasil iterasi sebelumnya.

2.5 Implementasi

Selanjutnya dengan *k-means* selesai, tahap selanjutnya adalah proses implementasi. Pada penelitian ini akan diimplementasi menggunakan pemrograman *python* yang dijalankan pada *google colaboratory*.



2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Pengujian dilakukan menggunakan metode *Davies Bouldin Index* (DBI). DBI ialah salah satu metode untuk menentukan kluster terbaik berdasarkan nilai kohesi dan separasi. Kohesi ialah jumlah kedekatan data terhadap titik pusat dan kluster yang diikuti, sedangkan separasi ialah didasarkan pada jarak antara titik pusat dari klusternya [15]. Semakin rendah nilai DBI yang dihasilkan maka akan semakin bagus kualitas kluster tersebut [16]. Berikut rumus-rumus yang digunakan untuk menghitung nilai DBI. Untuk menghitung jarak antar kluster bisa menggunakan *Sum of Square Within Cluster* (SSW) dengan rumus sebagai berikut:

$$SSW_i = \frac{1}{n_i} \sum_{j=1}^{n_i} (x_{ij} - \mu_i)^2 \quad (3)$$

Keeterangan:
 ini = jumlah data yang berada pada cluster ke-i
 μ_i = centroid cluster ke-i
 x_{ij} = data yang berada pada cluster ke-i

selanjutnya menghitung jarak antar kluster dengan menggunakan *Sum of Square Between Cluster* (SSB) seperti persamaan berikut ini:

$$SSB_{ij} = (n_i - n_j) \quad (4)$$

Kelebihan melakukan pengukuran rasio seberapa baik perbandingan antara kluster ke-I dan kluster k-j menggunakan persamaan sebagai berikut:

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}} \quad (5)$$

Untuk menghitung nilai DBI dengan menggunakan rumus sebagai berikut:

$$DBI = \frac{1}{K} \sum_{i=1}^K R_{ij} \neq R_{ij} \quad (6)$$

Dimana K adalah jumlah kluster yang digunakan.

HASIL DAN PEMBAHASAN

3.1.1. Dataset

Data yang digunakan ialah data permohonan bantuan zakat yang diambil dari Baznas Kota Pekanbaru, dengan kriteria yang digunakan ialah penghasilan, kondisi rumah, dan kepemilikan harta. Dataset yang digunakan telah melalui proses seleksi, *cleaning*, dan transformasi. Data ini akan digunakan dalam proses *clustering*. Dataset ini akan dipanggil didalam *google colab* untuk di proses. Pada tahap ini yang dilakukan pertama kali ialah dengan menginstall *library python* yang akan digunakan pada *google colab*. Berikut merupakan dataset yang akan digunakan dalam proses *clustering* yang terdiri dari 1133 data.

Tabel 6. Dataset

No	Penghasilan	Kondisi Rumah	Kepemilikan Harta
1	3	1	1
2	2	2	1
3	5	2	1
4	3	3	1
5	2	3	1
6	2	3	2
7	4	2	1
8	5	2	1
9	5	2	1
10	3	5	1
11	5	2	1
12	3	3	1
13	4	3	1
....
1133	4	2	1



Berikut merupakan *library-library python* yang akan digunakan:

```
[1] import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
from scipy.spatial.distance import cdist
import time
import itertools
```

Gambar 3. Import Library

```
memanggil data
#panggil data
from google.colab import files
file = files.upload()
```

Gambar 4. Proses Memanggil Data Menggunakan Python

```
[3] #read data
import pandas
data01 = pandas.read_excel('DATA BAZNAS FIX.xlsx')
data01.head(1133)
```

Gambar 5. Proses Membaca Data

1. Diarangi mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumpukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

selanjutnya yang akan dilakukan ialah memanggil data yang akan digunakan pada proses *import library*, selanjutnya yang akan dilakukan ialah memanggil data yang akan digunakan pada proses *import library*.

selanjutnya ialah membaca data yang sudah dipanggil sebelumnya:

Implementasi Algoritma *K-Means*

- a. Menentukan jumlah cluster yang akan digunakan. Pada penelitian ini jumlah klaster yang digunakan ialah $k=4$.
- b. Menentukan titik pusat. Sebelum dilakukannya tahap pengelompokan data menggunakan *k-means clustering*, yang perlu dilakukan adalah menentukan titik pusat. Dalam menentukan titik pusat, maka *python* akan otomatis menentukan nilai-nilai dari titik pusat tersebut. Berikut merupakan hasil dari *centroid* yang digunakan:

```
[23] kmeans.cluster_centers_
array([[4.88174274, 1.98340249, 1.24273859],
       [2.625      , 4.87096774, 1.42741935],
       [6.14529915, 4.58974359, 1.93162393],
       [2.43006993, 2.03146853, 1.29370629]])
```

Gambar 6. Kode Program Penentuan Centroid

Clustering dengan algoritma *k-means*

Langkah awal yang dilakukan pada tahap klastering ini ialah dengan menginstall *library k-means* pada *python* yang dijalankan pada *google colaboratory*. Berikut merupakan tampilan hasil dari proses klastering menggunakan *k-means* pada *python*:

```
[22] kmeans = KMeans (n_clusters=4, random_state=30)
#kmeans.fit(df)
y_kmeans = kmeans.fit_predict(df)
```

Gambar 8. Kode Program Clustering K-Means pada Python



Berikut merupakan tabel hasil dari klastering menggunakan algoritma *k-means* yang dijalankan pada *python*:

Tabel 7. Hasil Klastering dengan *K-Means* menggunakan *python*

No	Penghasilan	Kondisi Rumah	Kepemilikan Harta	Clusters
1	3	1	1	3
2	2	2	1	3
3	5	2	1	0
4	3	3	1	3
5	2	3	1	3
6	2	3	2	3
7	4	2	1	0
8	5	2	1	0
9	5	2	1	0
10	3	5	1	1
11	5	2	1	0
12	3	3	1	3
13	4	3	1	0
....
1133	4	2	1	0

3. Interpretasi

Interpretasi ialah tahap yang dilakukan setelah proses data mining, yang memiliki tujuan untuk menghasilkan pola informasi yang sudah dipahami. Berikut merupakan penjelasan hasil dari pengelompokan menggunakan 1133 data dengan 3 atribut:

Tabel 8. Hasil Interpretasi

Cluster	Kesimpulan
0	<p>Jumlah data 482</p> <ul style="list-style-type: none"> Rata-rata pemohon berpenghasilan mulai dari > 1.5 juta sampai dengan > 3 juta. Pemohon yang berpenghasilan > 1.5 s/d 2 juta berjumlah 184 orang, > 2 juta s/d 2.5 juta berjumlah 203 orang, > 2.5 juta s/d 3 berjumlah 65 orang, dan > 3 juta berjumlah 32 orang. Pemohon didominasi oleh orang yang kondisi rumahnya masih berstatus kontrak. Pemohon didominasi oleh orang yang kepemilikan hartanya <= 5 juta.
1	<p>Jumlah data 248</p> <ul style="list-style-type: none"> Rata-rata pemohon berpenghasilan mulai dari <= 500 ribu sampai dengan 2 juta. Pemohon yang berpenghasilan <=500 ribu berjumlah 29 orang, > 500 ribu s/d 1 juta berjumlah 87 orang, >1 juta s/d 1.5 juta berjumlah 80 orang, > 1.5 juta s/d 2 juta berjumlah 52 orang. Pemohon didominasi oleh orang yang kondisi rumahnya berstatus hak milik sendiri. Pemohon didominasi oleh orang yang kepemilikan hartanya <= 5 juta.
2	<p>Jumlah data 117</p> <ul style="list-style-type: none"> Rata-rata pemohon berpenghasilan > 2 juta sampai dengan > 3 juta. Pemohon yang berpenghasilan > 2 juta s/d 2.5 juta berjumlah 41 orang, >2.5 juta s/d 3 juta berjumlah 31 orang, dan > 3 juta berjumlah 45 orang. Pemohon didominasi oleh orang yang kondisi rumahnya berstatus hak milik sendiri. Pemohon didominasi oleh orang yang kepemilikan hartanya <= 5 juta
3	<p>Jumlah data 286</p> <ul style="list-style-type: none"> Rata-rata pemohon berpenghasilan mulai dari <=500 ribu sampai dengan 1.5 juta. Pemohon yang berpenghasilan <= 500 ribu berjumlah 28 orang, > 500 ribu s/d 1 juta berjumlah 101 orang, dan >1 juta s/d 1.5 juta berjumlah 157 orang. Pemohon di dominasi oleh orang yang kondisi rumahnya masih menumpang. Pemohon didominasi oleh orang yang kepemilikan hartanya <=5 juta.

Berdasarkan tabel di atas dapat kita berikan definisi dari masing-masing klaster. Pada klaster 0 dan 2, sebagian pemohon bisa dimasukkan ke dalam kelompok data yang layak menjadi sebagai penerima yaitu yang penghasilannya tidak lebih dari 2,5 juta, dan selebihnya dinyatakan tidak layak sebagai penerima. Sedangkan pada klaster 1 dan 3 semua pemohon bisa disebut sebagai kelompok data yang layak menjadi penerima.



of Zakat in BAZNAS Pekanbaru,” *Proc. - 2018 2nd Int. Conf. Electr. Eng. Informatics Towar. Most Effic. W. Mak. Deal. with Futur. Electr. Power Syst. Big Data Anal. ICon EEI 2018*, no. October, pp. 12–17, 2018, doi: 10.1109/ICon-EEI.2018.8784142.

K. Fatmawati and A. P. Windarto, “Data Mining: Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (Dd) Berdasarkan Provinsi,” *Comput. Eng. Sci. Syst. J.*, vol. 3, no. 2, p. 173, 2018, doi: 10.24114/cess.v3i2.9661.

H. Annur, “Penerapan Data Mining Menentukan Strategi Penjualan Variasi Mobil Menggunakan Metode K-Means Clustering (Studi Kasus Toko Motor Variasi Gorontalo),” vol. 5, no. 1, 2019.

M. L. Sibua and A. Safta, “Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering,” *Jurteks*, vol. 4, no. 1, pp. 85–92, 2017, doi: 10.33330/jurteks.v4i1.28.

H. Mulaistri and A. I. Gufroni, “Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 3, no. 2, pp. 299–305, 2017, doi: 10.25077/teknosi.v3i2.2017.299-305.

R. Lurmiawan, P. Pizaini, and F. Insani, “Penerapan Algoritma K-Means Clustering dan Correlation Matrix Untuk Menganalisis Risiko Penyebaran Demam Berdarah di Kota Pekanbaru,” *JIMP (Jurnal Inform.)*, vol. 6, no. 3, pp. 1–6, 2021, [Online]. Available: <http://ejournal.unperas.ac.id/index.php/informatika/article/view/353>.

C. Neural, N. Cnn, R. Firdaus, and J. Satria, “Jurnal Computer Science and Information Technology (CoSciTech) Klasifikasi Jenis Kelamin Berdasarkan Gambar Mata Dengan Menggunakan Algoritma,” vol. 3, no. 3, pp. 267–273, 2022.

E. L. Susanto, Paminto Agung Christianto, Mohammad Reza Maulana, and Satriedi Wahyu Binabar, “Analisis Kinerja Algoritma Naïve Bayes Pada Dataset Sentimen Masyarakat Aplikasi NEWSAKPOLE Samsat Jawa Tengah,” *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 3, no. 3, pp. 234–241, 2022, doi: 10.37859/coscitech.v3i3.4343.

R. Anggraini, E. Haerani, and I. Afrianty, “Pengelompokan Penyakit Pasien Menggunakan Algoritma K-Means,” vol. 9, no. 6, pp. 1840–1849, 2022, doi: 10.30865/jurikom.v9i6.5145.

R. G. Gunawan, Erik Suanda Handika, and Edi Ismanto, “Pendekatan Machine Learning Dengan Menggunakan Algoritma Xgboost (Extreme Gradient Boosting) Untuk Peningkatan Kinerja Klasifikasi Serangan Syn,” *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 3, no. 3, pp. 453–463, 2022, doi: 10.37859/coscitech.v3i3.4356.

D. A. Nasution, H. H. Khotimah, and N. Chamidah, “Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN,” *Comput. Eng. Sci. Syst. J.*, vol. 4, no. 1, p. 78, 2019, doi: 10.24114/cess.v4i1.11458.

R. W. Sari, A. Wanto, and A. P. Windarto, “Implementasi Rapidminer Dengan Metode K-Means (Study Kasus: Imunisasi Campak Pada Balita Berdasarkan Provinsi),” *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 224–230, 2018, doi: 10.30865/komik.v2i1.930.

R. Hablum, A. Khairan, and R. Rosihan, “Clustering Hasil Tangkap Ikan Di Pelabuhan Perikanan Nusantara (Ppn) Ternate Menggunakan Algoritma K-Means,” *JIKO (Jurnal Inform. dan Komputer)*, vol. 2, no. 1, pp. 26–33, 2019, doi: 10.33387/jiko.v2i1.1053.

F. Indriyani and E. Irfiani, “Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means,” *JUITA J. Inform.*, vol. 7, no. 2, p. 109, 2019, doi: 10.30595/juita.v7i2.5529.

1. Diarahkan mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.