

Klasifikasi Data Penerimaan Zakat dengan Algoritma K-Nearest Neighbor

Alfin Hernandes, Siska Kurnia Gusti*, Fadhilah Syafria, Lestari Handayani, Siti Ramadhani

Fakultas Sains dan Teknologi, Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ¹11950111669@students.uin-suska.ac.id, ^{2,*}siskakurniagusti@uin-suska.ac.id, ³fadhilah.syafria@uin-suska.ac.id,

⁴lestari.handayani@uin-suska.ac.id, ⁵siti.ramadhani@uin-suska.ac.id

Email Penulis Korespondensi: siskakurniagusti@uin-suska.ac.id

Abstrak—Badan Amil Zakat Nasional (BAZNAS) adalah sebuah lembaga pengelola zakat yang didirikan oleh pemerintah. BAZNAS memiliki keberadaan di setiap kabupaten atau kota, dan salah satu dari mereka adalah BAZNAS kota Pekanbaru. BAZNAS kota Pekanbaru bertanggung jawab untuk menyalurkan zakat kepada berbagai program pendayagunaan, salah satunya adalah program Pekanbaru Peduli. Saat ini, BAZNAS kota Pekanbaru menghadapi permasalahan terkait metode penyaluran zakat, di mana proses penentuan kriteria penerima zakat masih dilakukan secara manual oleh panitia BAZNAS Kota Pekanbaru. Kondisi ini dianggap kurang efektif, dan menjadi salah satu tantangan yang perlu diatasi. Untuk mengatasi kendala yang telah disebutkan, diperlukan langkah-langkah guna meningkatkan efektivitas dan efisiensi pendataan calon penerima zakat. Salah satu solusinya adalah menerapkan sistem klasifikasi untuk mempermudah proses pendataan, dengan menggunakan metode *K-Nearest Neighbour (KNN)*. Pendekatan ini berfungsi sebagai alat bantu dalam mengelompokkan data calon mustahik. Penelitian ini bertujuan untuk melakukan klasifikasi data dan mengukur tingkat akurasi dalam menilai kelayakan calon penerima zakat berdasarkan kriteria yang telah ditetapkan, dengan memanfaatkan algoritma *K-Nearest Neighbor (K-NN)*. Sejumlah 602 data BAZNAS Kota Pekanbaru peduli digunakan dalam penelitian ini, dengan pembagian data latih dan uji yaitu 90:10,80:20, dan 70:30. Hasil evaluasi dari *confusion matrix* dari $k=3, k=5, k=7, k=9$, dan $k=11$, menunjukkan bahwa akurasi tertinggi tercapai pada nilai $k=5$ pada pembagian 80:20, dengan tingkat akurasi mencapai 89,3%. Selain itu, presisi sebesar 87,3% dan recall sebesar 91,4% juga dapat dicapai melalui pendekatan ini.

Kata Kunci: BAZNAS; Data Mining; Klasifikasi; K-Nearest Neighbour

Abstract—National Amil Zakat Agency (BAZNAS) is an institution responsible for managing zakat established by the government. BAZNAS has a presence in every district or city, and one of them is the BAZNAS in the city of Pekanbaru. BAZNAS in Pekanbaru city is responsible for distributing zakat to various empowerment programs, one of which is the Pekanbaru Cares program. Currently, BAZNAS in Pekanbaru city is facing issues related to the method of distributing zakat, where the process of determining the criteria for zakat recipients is still being done manually by the committee of BAZNAS in the city of Pekanbaru. This condition is considered inefficient and poses one of the challenges that need to be addressed. To overcome the mentioned constraints, steps are needed to improve the effectiveness and efficiency of data collection for potential zakat recipients. One of the solutions is to implement a classification system to facilitate the data collection process, using the K-Nearest Neighbor (KNN) method. This approach functions as a tool to classify data for potential beneficiaries. This research aims to classify data and measure the accuracy in assessing the eligibility of zakat recipients based on predetermined criteria, utilizing the K-Nearest Neighbor (K-NN) algorithm. A total of 602 data from BAZNAS in the city of Pekanbaru were used in this study, by dividing the training and test data, namely divided 90:10, 80:20, and 70:30 splits. The evaluation results from the confusion matrix of $k=3, k=5, k=7, k=9$, and $k=11$ show that the highest accuracy is achieved at $k=5$ with an 80:20 split, with an accuracy rate of 89.3%. Furthermore, a precision of 87.3% and a recall of 91.4% can also be attained through this approach.

Keywords: BAZNAS; Data Mining; Classification; K-Nearest Neighbour

1. PENDAHULUAN

Zakat adalah pensucian harta, yaitu dengan menyisihkan separuh harta pada saat waktunya tiba dan jumlahnya diberikan kepada yang berhak menerimanya. Zakat juga memiliki arti sebagai ibadah yang berkaitan dengan harta yang memiliki fungsi sosial. Zakat termasuk dalam rukun Islam tiga yang memiliki tujuan membersihkan harta setiap muslim[1]. Dalam ekonomi Islam, zakat adalah alat lain yang dapat mempengaruhi perilaku seorang Muslim dan memajukan perekonomian[2].

Badan Amil Zakat Nasional (BAZNAS) adalah lembaga yang dibentuk oleh pemerintah daerah di setiap kabupaten dan kota untuk bertanggung jawab dalam pengawasan dan pengelolaan zakat. Nama Badan Amil Zakat Kota Pekanbaru bernama BAZNAS Kota Pekanbaru[3]. Saat ini panitia BAZNAS Kota Pekanbaru menghitung secara manual kebutuhan penerima zakat untuk dapat melaksanakan metode penyaluran zakat. Hal ini memungkinkan terjadinya perbedaan dan ketidaksinambungan kelayakan dan ketidaklayakan calon penerima zakat serta efisiensi penggunaan waktu yang kurang optimal. Selain itu, penghitungan kelayakan akan memakan waktu yang cukup lama jika orang yang menerima zakat terlalu banyak yang mendaftar untuk mengajukan zakat sekaligus membutuhkan waktu dalam proses seleksi dan pendayagunaan zakat harus mampu menghasilkan produk yang akurat dan tepat sasaran hasil operasi kelayakan. Mengingat permasalahan di atas, diperlukan suatu metode untuk meningkatkan efektivitas dan efisiensi pendataan para penerima zakat. Dengan menggunakan algoritma klasifikasi *Data mining* terhadap data yang ada akan dijelaskan dengan proses analisis karakteristik sampel yang digunakan untuk mengambil keputusan, membuat proses pemilihan lebih mudah, konsisten, dan mampu mengidentifikasi penyimpangan subjektif dari data yang ada[4].

Data mining adalah suatu teknologi yang mengolah volume data yang besar dengan menggabungkan algoritma canggih bersama teknik analisis konvensional. Istilah ini merujuk pada upaya untuk menemukan informasi yang tersembunyi di dalam basis data[5]. selain itu, *data mining* juga merupakan suatu prosedur yang mengekstrak informasi berharga dari dataset yang sangat besar dengan menerapkan pendekatan dua dimensi dan berbagai aspek. Pendekatan ini

mengintegrasikan teknik analisis statistik, visualisasi data, penemuan pengetahuan, pengenalan pola, dan manajemen, dengan memanfaatkan ide-ide interdisipliner yang relatif baru terkait analisis data dan penemuan pengetahuan dari *database*[6]. Saat ini ada lima kategori *Data mining* yang dapat dibagi, *Estimation*, *Forecasting*, *Classification*, *Clustering*, dan *Association*[7].

Salah satu teknik *data mining* adalah klasifikasi. Proses pengklasifikasian data melibatkan identifikasi kumpulan model dan fungsi yang dapat digunakan untuk mendeskripsikan dan membedakannya ke dalam kelompok yang berbeda. Proses mengidentifikasi model yang mengkarakterisasi dan membedakan kelompok ide data dikenal sebagai klasifikasi[8]. Dengan tujuan menerapkan model Klasifikasi ini untuk mengidentifikasi kelas suatu objek ketika kelasnya tidak diketahui. Klasifikasi melibatkan dua proses, yaitu Proses training Melakukan pembangunan model menggunakan data training. Proses testing Melakukan tes terhadap data testing menggunakan model yang telah diperoleh dari proses training[9].

Metode penyelesaian masalah klasifikasi yang populer adalah *K-Nearest Neighbor (K-NN)*. Algoritma *K-Nearest Neighbor (K-NN)* merupakan salah satu teknik paling sederhana dalam mengatasi masalah klasifikasi, namun sering memberikan hasil yang kompetitif dan menarik perhatian[6].*K-Nearest Neighbor* banyak digunakan oleh para peneliti karena kelebihanannya, yaitu akurasi yang tinggi [10]. Pada penelitian [11] membandingkan algoritma KNN dengan algoritma *Naïve Bayes* dan *Decision Tree* pada kualitas air, dengan hasil metode KNN memiliki tingkat akurasi paling tinggi sebesar 86.88% dibandingkan dengan *Decision Tree* sebesar 80.84% dan *Naïve Bayes* sebesar 63.60% , sedangkan dalam penelitian [12] membandingkan algoritma KNN dengan SVM dalam kasus memprediksi struktur sekunder protein, dalam penelitian tersebut tingkat akurasi KNN lebih tinggi dibandingkan dengan metode lainnya, menggunakan nilai $K = 5$, $K = 10$, dan $K = 15$ serta *sliding window* sebesar 15. Akurasi yang dihasilkan yaitu 49.59% menjadi akurasi terbaik, sedangkan metode SVM menggunakan nilai $C = 1$ dan $\gamma = 0.1$ menghasilkan akurasi terbaik sebesar 0.62. Metode umum ini digunakan untuk klasifikasi data. Dalam penerapannya, metode KNN ini memungkinkan klasifikasi suatu set data baru dengan mempertimbangkan sejumlah tetangga terdekat, di mana penentuan ini didasarkan pada jarak terpendek antara data latih dan data uji [13]. KNN beroperasi menggunakan algoritma *supervised learning*, di mana hasil dari *instance* baru yang diajukan diklasifikasikan berdasarkan mayoritas kategori yang dimiliki oleh tetangga terdekatnya. Kelas yang paling sering muncul di antara tetangga tersebut akan menjadi kelas yang digunakan untuk hasil klasifikasi [14].

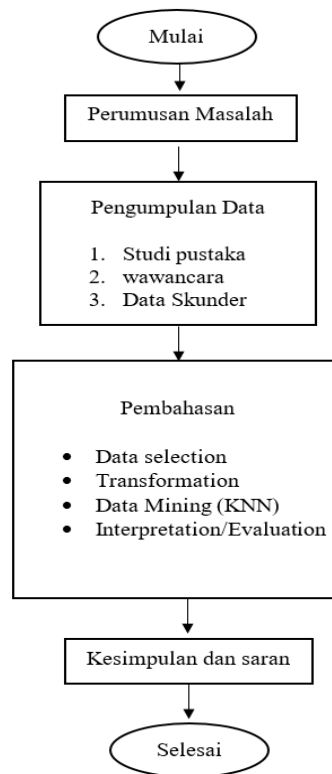
Penelitian ini bertujuan untuk menerapkan algoritma *K-Nearest Neighbor (K-NN)* dalam melakukan klasifikasi data untuk menentukan calon penerima zakat. Tujuan utamanya adalah mengurangi ketidakakuratan yang bisa menghambat proses penentuan penerima zakat, dengan memastikan bahwa klasifikasi yang dilakukan sesuai dengan kriteria yang telah ditetapkan sebelumnya. Hal ini dilakukan dikarenakan mencakup permasalahan diatas maka dari itu kami melakukan penelitian ini.

Beberapa penelitian telah menunjukkan bahwa *K-Nearest Neighbor (KNN)* memiliki tingkat akurasi yang tinggi dalam proses klasifikasi. Misalnya, penelitian[15] Hasil penelitian menunjukkan bahwa pada nilai $K=15$ dan $K=30$, data baru (D160) diklasifikasikan sebagai kategori "Tidak Layak" dengan tingkat akurasi mencapai 100%. Sementara itu, pada nilai $K=45$, $K=60$, dan $K=75$, data baru (D160) diklasifikasikan sebagai kategori "Layak" dengan tingkat akurasi sebesar 81,25%. Studi lain[16] memprediksi Aroma Alkohol Menggunakan Metode KNN, Pengujian terhadap pengaruh nilai K menunjukkan hasil akurasi optimal sebesar 100% pada nilai $K=3$ dan 100% pada nilai $K=4$. Studi [17] Dalam klasifikasi Pencemaran Udara di Kota Jakarta dengan mayoritas $k=7$, hasil menunjukkan bahwa kualitas udara pada bulan berikutnya dikategorikan sebagai "SEDANG", dan penelitian ini mencapai tingkat akurasi sebesar 95,78%. Studi [18] Dalam prediksi penjualan produk Unilever menggunakan metode *K-Nearest Neighbor*, diperoleh tingkat akurasi tertinggi dalam klasifikasi penjualan produk sebesar 86,66%. Studi [19] Dalam analisis penjualan produk paket kuota internet menggunakan metode *K-Nearest Neighbor*, diperoleh prediksi penjualan paket kuota internet yang terdiri dari SP CL1, SP CL2, SP CL4, dan SP CL8 dengan tingkat akurasi sebesar 71,43%.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini melibatkan beberapa tahap yang diimplementasikan melalui serangkaian langkah-langkah, yang dapat ditemukan dalam ilustrasi pada Gambar 1.



Gambar 1. Alur Tahapan Penelitian

Dari gambar 1 tersebut dapat dijelaskan, bahwa:

a. Perumusan Masalah

Ini adalah langkah awal dalam metodologi penelitian ketika masalah penelitian dipahami dan dikaji. Selanjutnya dilakukan upaya mencari solusi untuk mengatasi masalah tersebut. Dalam bab ini juga dibahas kerangka kerja dan topik utama penelitian[5]. Rumusan masalahnya adalah bagaimana penerapan metode *K-Nearest Neighbor (K-NN)* untuk mengklasifikasikan layak tidaknya dalam penerimaan zakat dengan menggunakan algoritma *K-Nearest Neighbor (K-NN)*.

b. Pengumpulan Data

1. Studi Pustaka

Metode ini akan dilakukan dengan menggunakan teknik mencari referensi yang relevan dari berbagai buku, jurnal akademik, dan jurnal yang dianggap penting dan terpercaya.

2. Wawancara

Menganalisis, merancang dan membangun sistem klasifikasi memerlukan pendataan penerima program zakat Pekanbaru Peduli, pada tahap ini. Data kriteria yang diperlukan untuk penelitian penerima manfaat program Pekanbaru peduli dirangkum, pada tahap pengumpulan data peneliti melakukan sesi tanya jawab dengan pengelola BAZNAS yang terkait dengan penelitian.

3. Data Sekunder

Data sekunder mengacu pada informasi yang diperoleh dari sumber selain responden yang dijadikan dasar analisis[20]. Misalnya saja melalui orang lain atau melalui sebuah dokumen. Penelitian ini merupakan sampel pendataan penerima manfaat program zakat Pekanbaru Peduli.

c. Pembahasan

Tahap ini dilakukan setelah pengumpulan data dari metodologi penelitian. Menganalisis metode tertentu yang digunakan untuk menganalisis suatu masalah tertentu. Pada tahap analisis ini, peneliti akan mengkaji permasalahan proses data mining berikut ini. Ini menjelaskan proses analisis data mining.

1. Data Selection

Langkah ini melibatkan pemilihan beberapa atribut yang akan digunakan untuk mengklasifikasikan data dari kumpulan data yang besar. Data yang digunakan dalam penelitian ini berjumlah 602 record. Pada penelitian ini digunakan adalah umur, tanggungan, pekerjaan, jumlah usulan, penghasilan, dan kondisi rumah sebagai atribut dan layak tidak layak sebagai label.

2. Transformation

Selama prosedur, data akan diubah menggunakan metode pelabelan. Metode *K-Nearest Neighbor (K-NN)* merupakan salah satu metode yang dapat digunakan apabila datanya berupa angka, supaya data tadi dapat diolah memakai algoritma *K-Nearest Neighbor (K-NN)*.

Tabel 1. Transformation Tanggungan

No	Tanggungan	Angka
1	Sangat Sedikit	1
2	Sedikit	2
3	Biasa	3
4	Banyak	4

Tabel 2. transformation pekerjaan

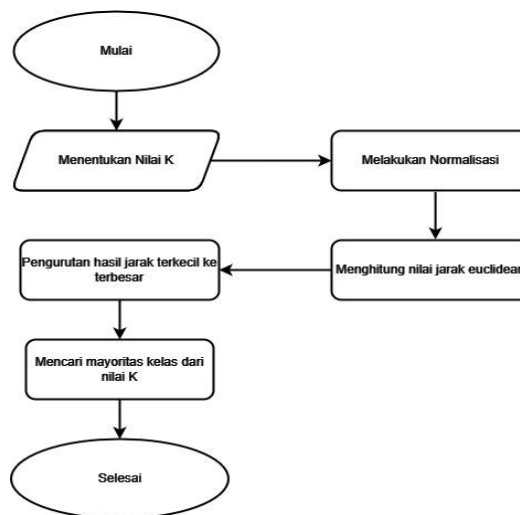
No	pekerjaan	Angka
1	Tidak berkerja	1
2	wiraswasta	2
3	wirusaha	3

Tabel 3. kondisi rumah

No	Kondisi Rumah	Angka
1	Mengontrak	1
2	menumpang	2
3	Rumah sendiri	3

3. Data mining K-Nearest Neighbor (K- NN)

Setelah data melalui tahap *pre-processing* dan tahap tranformation maka data telah bisa diolah menggunakan metode *K-Nearest Neighbor (K- NN)*[21].



Gambar 2. Data mining *K-Nearest Neighbor (K- NN)*

Di bawah ini adalah prosedur penghitungan menggunakan algoritma *K-Nearest Neighbor (KNN)*, seperti terlihat pada Gambar 2.

- a) Pada tahap pertama ditentukan nilai k.
- b) Melakukan normalisasi *min-max* yang memproses data menjadi rentang 0 dan 1
Perhitungan normalisasi dengan metode *min-max*

$$\text{normalized} = \frac{\text{Data}_x - \text{Data}_{\min}}{\text{Data}_{\max} - \text{Data}_{\min}} \tag{1}$$

Data x adalah data yang akan dihitung normalisasinya yang diambil berdasarkan kolom datanya
Data *min* adalah data terkecil dikolom yang sama
Data *max* adalah data terbesar dikolom yang sama dengan data yang akan di normalisasikan

- c) Menghitung jarak *Euclidean* data latih dan data uji.
- d) Mengurutkan hasil jarak data terkecil ke terbesar.
- e) Mencari mayoritas kelas dari nilai K tetangga dan dijadikan hasil prediksi.
Untuk menghitung *Euclidean*, gunakan rumus pada persamaan 2.

$$d_{\text{euclidean}}(x, y) = \sqrt{\sum_i (x_i - y_i)^2} \tag{2}$$

Keterangan :

$d_{\text{euclidean}}$ = Jarak euclidean

- i = Banyak set data
- x = Jumlah set data uji
- y = Jumlah set data latih

4. Interpretation atau Evaluation

Proses menghasilkan informasi atau wawasan baru dari sampel data yang disegmentasi menggunakan algoritma *K-Nearest Neighbor (K-NN)*. Langkah evaluasi ini adalah untuk melihat tingkat kinerja algoritma berdasarkan hasil polling. Salah satu parameter yang digunakan untuk evaluasi algoritma adalah *confusion matrix* yang dapat dilihat pada Tabel 4 dengan melihat nilai akurasinya[10].

Table 4. *confusion matrix*

Actual	Predicted	
	Positive	Negative
	Positive	TP
Negative	FN	TN

Keterangan:

TP (*True Positive*) mengacu pada jumlah poin data ya yang dapat diidentifikasi secara benar.

TN (*True Negative*) mengacu pada jumlah titik data berlabel yang berhasil diidentifikasi sebagai salah.

FP (*False Positive*) mengacu pada jumlah titik data yang dapat diidentifikasi secara meyakinkan sebagai salah.

FN (*False Negative*) mengacu pada jumlah titik data dengan label yang dapat diidentifikasi secara positif.

Untuk mencari nilai *Accuracy, Precision, Recall* dapat menggunakan rumus sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

Accuracy mengukur sejauh mana model dapat secara tepat mengklasifikasikan semua kelas.

Precision mengukur sejauh mana model dapat memprediksi kelas positif dengan benar.

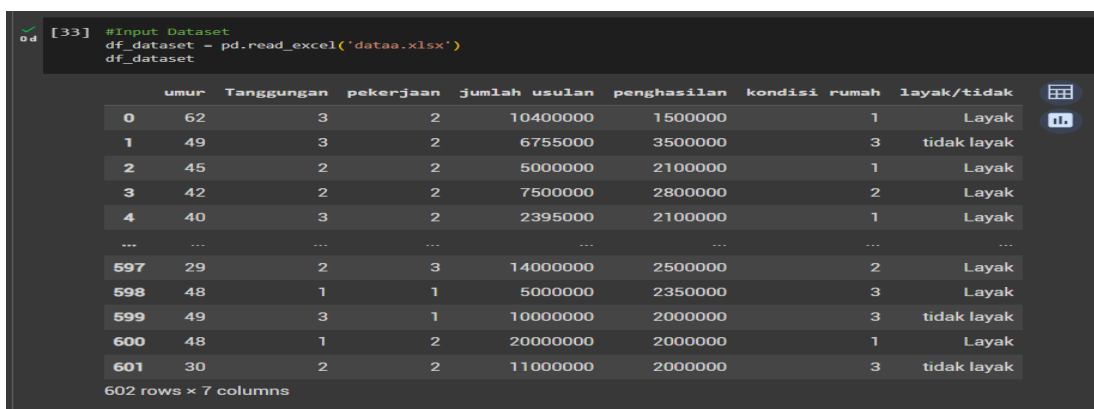
Recall mengukur sejauh mana model dapat menangkap atau mendeteksi kelas yang sebenarnya positif.

2.4 Kesimpulan Dan Saran

Dari hasil pengujian, dapat ditarik kesimpulan yang relevan dengan masalah dan tujuan penelitian. Kesimpulan ini mencakup evaluasi terhadap *Accuracy, Precision,* dan *recall* yang dihasilkan dari perhitungan algoritma *K-Nearest Neighbors*[21].

3. HASIL DAN PEMBAHASAN

Analisis dilakukan dengan memanfaatkan data dari BAZNAS Kota Pekanbaru Peduli. Pengujian data pada penelitian ini dilakukan melalui implementasi pemrograman *Python* dengan menerapkan model algoritma *K-Nearest Neighbors*. Rincian data gabungan dapat diidentifikasi pada Gambar 3.



Gambar 3. Mengimport Dataset

Langkah pertama adalah menginput atau memanggil dataset excel yang akan digunakan kedalam google colab atau *python* untuk analisis data dengan cara *pd.read_excel('dataa.xlsx')*, seperti terlihat pada Gambar 3. Pada Gambar 3 dataset sudah berhasil diinput kedalam *python*, kemudian dataset yang akan digunakan memiliki tujuh atribut yaitu Umur, tanggungan, pekerjaan, jumlah usulan, pendapatan, dan kondisi tempat tinggal dan layak atau tidak sebagai label. Langkah selanjutnya adalah proses pembagian atribut data tersebut menjadi x dan y dapat dilihat pada Gambar 4.

```
[12] x = dataset.iloc[:, [0, 1, 2, 3, 4, 5]].values
     y = dataset.iloc[:, -1].values

[13] print(x)

[[ 62      3      2 10400000 1500000      1]
 [ 49      3      2  6750000 3500000      3]
 [ 45      2      2  5000000 2100000      1]
 ...
 [ 49      3      1 10000000 2000000      3]
 [ 48      1      2 20000000 2000000      1]
 [ 30      2      2 11000000 2000000      3]]

print(y)

['Layak' 'tidak layak' 'Layak' 'Layak' 'Layak' 'Layak' 'Layak' 'Layak'
 'Layak' 'Layak' 'Layak' 'Layak' 'Layak' 'tidak layak' 'Layak'
 'tidak layak' 'Layak' 'Layak' 'Layak' 'Layak' 'Layak' 'Layak'
 'tidak layak' 'Layak' 'Layak' 'Layak' 'Layak' 'Layak' 'Layak' 'Layak'
 'tidak layak' 'Layak' 'tidak layak' 'Layak' 'Layak' 'Layak' 'Layak'
 'Layak' 'Layak' 'Layak' 'Layak' 'Layak' 'Layak' 'tidak layak'
 'tidak layak' 'tidak layak' 'Layak' 'Layak' 'Layak' 'Layak' 'Layak'
 'Layak' 'tidak layak' 'Layak' 'Layak' 'Layak' 'Layak' 'Layak' 'Layak']
```

Gambar 4. Membagi Atribut Data Menjadi X dan Y

Pada Gambar 4, ini merupakan tahap pembagian atribut data, dimana dataset tersebut dibagi menjadi dua atribut yaitu y dan x, variabel x memiliki atribut umur, tanggungan, pekerjaan, jumlah usulan, pendapatan dan kondisi tempat tinggal, dan variabel y memiliki label atribut. Kemudian proses selanjutnya yaitu pembagian data seperti terlihat pada Gambar 5.

```
[183] #Membagi Dataset ke Data Training dan Data Testing
      from sklearn.model_selection import train_test_split
      x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.20, random_state=123)

print(x_train)

[[ 62      2      1 3848000 1500000      3]
 [ 33      1      2 30000000 1350000      3]
 [ 46      3      2 12000000 2000000      1]
 ...
 [ 35      3      2 10000000 1000000      1]
 [ 46      1      2  8000000 2500000      3]
 [ 40      2      2 12000000 1000000      2]]

[185] len(x_train)
481

[186] len(x)
602

[187] len(x_test)
121
```

Gambar 5. Pembagian Data Dalam Komposisi 80:20

Gambar 5 menunjukkan proses pemisahan data untuk pelatihan dan pengujian. Mengimpor fungsi `train_test_split` dari modul `model_selection` dalam pustaka *Scikit-Learn* (`sklearn`) untuk membagi dataset menjadi dua bagian: satu untuk pelatihan model (*train*) dan satu untuk pengujian model (*test*). Setelah pernyataan ini dieksekusi, kita akan memiliki empat variabel, `x_train` (fitur untuk pelatihan), `x_test` (fitur untuk pengujian), `y_train` (target untuk pelatihan), dan `y_test` (target untuk pengujian). Dengan fungsi *train_test_split* dengan perbandingan 80:20. Jumlah data latih sebanyak 481 dan uji 121. Setelah itu, algoritma *K-Nearest Neighbor* dimodelkan menggunakan fungsi klasifikasi *K-Neighbor* dengan rumus jarak *Euclidean*. Nilai k yang digunakan pada penelitian ini adalah 3, 5, 7, 9 dan 11. Sebelum melakukan pencarian nilai K, dataset dinormalisasi terlebih dahulu, di mana perhitungan jarak sangat mempengaruhi hasil, normalisasi dapat membantu meningkatkan kinerja model. Dapat kita lihat proses normalisasi pada tabel 5 dan 6.

Tabel 5. Normalisasi Dataset Trainig

UMUR	TANGGUNGAN	PEKERJAAN	JUMLAH USULAN	PENGHASILAN	KONDISI RUMAH
0,5614	0,66667	0,5	0,0465013	0,110169492	0
0,33333	0,66667	0,5	0,0299197	0,279661017	1
0,26316	0,33333	0,5	0,0219359	0,161016949	0
0,21053	0,33333	0,5	0,0333088	0,220228983	0,5
0,17544	0,66667	0,5	0,0100854	0,161016949	0
0,29825	0,33333	0	0,0333088	0,13559322	0
0,63158	0,33333	0	0,0061641	0,076271186	0
0,14035	0,66667	0,5	0,0598757	0,088983051	1
0,54386	0,33333	0	0,0968147	0,152542373	0
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
0,26316	0,33333	1	0,3176297	0,06779661	0,5
0,07018	0,33333	0,5	0,0674273	0,088983051	0,5

Tabel 6. Normalisasi Dataset Testing

UMUR	TANGGUNGAN	PEKERJAAN	JUMLAH USULAN	PENGHASILAN	KONDISI RUMAH
0,14286	0,66667	0,5	0,0909091	0,076923077	0,5
0,25	1	0	0,2424242	0,123076923	0,5
0,39286	0,33333	1	0,3030303	0,107692308	0,5
0,75	0,33333	0,5	0,1666667	0,115384615	0
0,53571	0	0,5	0,1212121	0,130769231	0,5
0,32143	0,66667	0,5	0,1515152	0,076923077	0,5
0,53571	0	1	0,1212121	0,107692308	0,5
0,60714	0	0	0,1060606	0,076923077	0
1	0,33333	0,5	0,0909091	0,115384615	0
⋮	⋮	⋮	⋮	⋮	⋮
0,67857	0	0,5	0,2424242	0,230769231	0
0,03571	0,33333	0,5	0,1060606	0,230769231	1

Setelah normalisasi data, kemudian melakukan proses pemodelan algoritma *K-Nearest Neighbour* dengan rumus *Euclidean distance* menggunakan fungsi klasifikasi *K-neighbor* untuk mencari nilai K ditunjukkan pada Gambar 6.

```

#Memanggil Function KNN
from sklearn.neighbors import KNeighborsClassifier
classifier = KNeighborsClassifier(n_neighbors=11, metric="euclidean")
classifier.fit(x_train, y_train)
    
```

KNeighborsClassifier
KNeighborsClassifier(metric='euclidean', n_neighbors=11)

Gambar 6. Pemodelan *K-Nearest Neighbor*

Pada tahapan pemodelan dengan algoritma *K-nearest neighbour (KNN)*, dilakukan pengujian akurasi untuk mengevaluasi kinerja model klasifikasi KNN pada dataset tersebut. Penilaian akurasi didasarkan pada confusion matrix. Hasil dari evaluasi akurasi menggunakan *confusion matrix*.

3.1 Evaluasi

Setelah menyelesaikan tahapan pemodelan dengan algoritma *K-nearest neighbour (KNN)*, dilakukan pengujian akurasi untuk mengevaluasi kinerja model klasifikasi KNN yang telah dibuat sebelumnya. Penilaian akurasi didasarkan pada *confusion matrix*. Dapat dilihat pada gambar 7, dimana pemanggilan *confusion matrix* ada *from sklearn.metrics import confusion_matrix*, Baris pertama mengimpor fungsi *confusion matrix* dari modul *sklearn.metrics*, *cm = confusion_matrix(y_test, y_pred)*, Baris kedua memanggil fungsi *confusion_matrix* dengan parameter *y_test* dan *y_pred*. *y_test* adalah label sebenarnya dari data uji, sementara *y_pred* adalah label yang diprediksi oleh model untuk data uji tersebut. Hasil dari fungsi ini disimpan dalam variabel *cm*, yang akan berisi *confusion matrix*, *print(cm)*, Baris terakhir mencetak *confusion matrix* yang telah dihitung sebelumnya.

```

#Evaluasi dan Validasi
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred)
print(cm)
    
```

[[70 12]
 [1 38]]

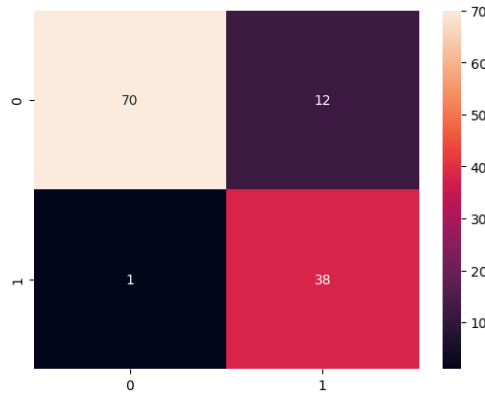
Gambar 7. Pemanggilan *Confusion Matrix 80:20*

Kemudian hasil dari evaluasi akurasi menggunakan confusion matrix dengan variasi nilai k yang diuji tercantum pada Tabel 7 berikut.

Tabel 7. Hasil Semua Akurasi

Nilai K	Hasil Akurasi (90:10)	Hasil Akurasi (80:20)	Hasil Akurasi (70:30)
K 3	85	88	87
K 5	86	89	86
K 7	85	85	85
K 9	88	85	86
K 11	83	85	85

Pengujian dilakukan dengan *confusion matrix* dimana banyaknya tetangga terdekat adalah K = 3,5,7,9 dan 11, kemudian digunakan perbandingan 90:10,80:20,70:30. Dari hasil pengujian yang tercatat dalam Tabel 7, ditemukan bahwa akurasi tertinggi mencapai 89% dengan rasio data latih dan uji sebesar 80:20, dan nilai K = 5. Detail *confusion matrix* untuk tingkat akurasi tertinggi dapat diperiksa pada Gambar 8.



Gambar 8. Tampilan Hasil *Confusion Matrix* 80:20

Gambar diatas menjelaskan hasil pengujian model klasifikasi yang dilakukan, angka pada gambar menunjukkan seberapa baik model dapat menentukan kelas sebenarnya terhadap kelas prediksi dengan perbandingan data 80:20 ketika K = 5. Kemudian laporan klasifikasi hasil *Classification Report*, ditunjukkan pada Gambar 9.

```
[ ] from sklearn.metrics import classification_report
akurasi = classification_report(y_test,y_pred)
print(akurasi)

              precision    recall  f1-score   support

   Layak              0.99      0.85      0.92         82
  tidak layak         0.76      0.97      0.85         39

   accuracy              0.89         121
  macro avg              0.87      0.91      0.88         121
 weighted avg           0.91      0.89      0.90         121

[ ] from sklearn.metrics import accuracy_score
akurasi = accuracy_score(y_test,y_pred)
print("Tingkat Akurasi :%d persen"%(akurasi*100))

Tingkat Akurasi :89 persen
```

Gambar 9. Tampilan Hasil *Classification Report*

Dari gambar 9 terlihat terdapat laporan *Classification Report*, yang diperoleh dari perhitungan pengujian *confusion matrix*. Hasil perhitungan yang diperoleh adalah nilai *Precision* 87%, *recall* 91% ,dan *accuracy* 89%. Hasil tersebut menunjukkan bahwa model klasifikasi dapat bekerja dengan baik.

4. KESIMPULAN

Berdasarkan hasil penelitian dapat disimpulkan bahwa metode yang digunakan efektif dalam mengklasifikasikan data BAZNAS Kota Pekanbaru peduli. Data yang digunakan sebanyak 602 dataset, yang kemudian dibagi menjadi tiga skenario perbandingan: 90:10, 80:20, dan 70:30. Hasil pengujian menggunakan Google Colab *Python* menunjukkan performa hasil terbaik terjadi pada pengujian dengan K=5 dan perbandingan data latihan dan uji 80:20 dengan *accuracy* 89%. Hasil tersebut juga diikuti dengan nilai *precision* sebesar 87%, *recall* sebesar 91%. Untuk penyelidikan lebih lanjut, disarankan untuk meningkatkan nilai akurasi dengan mencoba kumpulan data lain yang berbentuk tipe data numerik (belum dikonversi).

REFERENCES

- [1] Islami Alvin Anzaz, "Jurnal Computer Science and Information Technology (CoSciTech) menggunakan algoritma lexrakn," vol. 4, no. 1, pp. 154–163, 2023.
- [2] Suryani Dyah and Fitriani Laitul, "Fungsi Zakat dalam Mengatasi Kemiskinan," *Al-Iqtishad J. Islam. Econ.*, vol. 10, no. 1, pp. 43–62, 2022, [Online]. Available: <https://jurnal.stai-alazharmenganti.ac.id/index.php/AlIqtishod/article/view/307/176>
- [3] Haerani Elin and Ramdaril Ramdaril, "Sistem Pendukung Keputusan Pendistribusian Zakat Pada Baznas Kota Pekanbaru Menggunakan Fuzzy Multiple Attribute Decision Making (FMADM) Dan Simple Additive Weighting (SAW)," *J. Tek. Inform.*, vol. 3, no. 2, pp. 15–20, 2019, doi: 10.15408/jti.v10i2.6994.
- [4] Riyyan Muhamad and Firdau Hafiz, "Perbandingan Algoritma Naive Bayes Dan KNN Terhadap Data Penerimaan Beasiswa (Studi Kasus Lembaga Beasiswa Baznas Jabar)," *J. Inform. dan Rekayasa Elektron.*, vol. 5, no. 1, pp. 1–10, 2022, doi: 10.36595/jire.v5i1.547.
- [5] Kurnia Fitra, Kurniawan Jhoni, Fahmi Ichsan, and Monalisa Siti, "Klasifikasi Keluarga Miskin Menggunakan Metode K-Nearest Neighbor Berbasis Euclidean Distance," *Semin. Nas. Teknol. Inf. Komun. dan Ind.*, pp. 230–239, 2019, [Online]. Available:

- <https://ejournal.uin-suska.ac.id/index.php/SNTIKI/article/view/8089>
- [6] Nikmatun Inna Alvi and Indra Waspada, "Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *J. SIMETRIS*, vol. 10, no. 2, pp. 421–432, 2019.
 - [7] Julianto Indri Tri, Kurniadi Dede, Nashrulloh Muhammad Rizka, and Mulyani Asri, "Comparison of Classification Algorithm and Feature Selection in Perbandingan Algoritma Klasifikasi Dan Feature Selection," *Jutif*, vol. 3, no. 3, pp. 739–744, 2022.
 - [8] Ula Mutammimul, Zulhusna Ria, Putra Fhonna Rizki, and Pratama Angga, "Penerapan Model Klasifikasi K-Nearest Neighbor Dalam Pencarian Kesesuaian Pekerjaan," *Metik J.*, vol. 6, no. 1, pp. 18–23, 2022, doi: 10.47002/metik.v6i1.343.
 - [9] Raysyah Siti, Arinal Veri, and Mulyana Dadang Iskandar, "Klasifikasi Tingkat Kematangan Buah Kopi Berdasarkan Deteksi Warna Menggunakan Metode Knn Dan Pca," *JSil (Jurnal Sist. Informasi)*, vol. 8, no. 2, pp. 88–95, 2021, doi: 10.30656/jsii.v8i2.3638.
 - [10] Akhmad Muhammad Rhosyid and Siswa Taghfirul Azhima Yoga, "Implementasi K-Nearest Neighbor Dalam Memprediksi Keterlambatan Pembayaran Biaya Kuliah Di Perguruan Tinggi," *Progresif J. Ilm. Komput.*, vol. 18, no. 2, pp. 185–192, 2022, doi: 10.35889/progresif.v18i2.921.
 - [11] Tangkelayuk Aldi, "The Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes, dan Decision Tree," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 1109–1119, 2022, doi: 10.35957/jatisi.v9i2.2048.
 - [12] Anggi Tasari, Dinata Tarigan Dewan, Nia Erika, Br Devina, and S Kana Saputra, "Perbandingan Algoritma Support Vector Machine dan KNN dalam Memprediksi Struktur Sekunder Protein," vol. 9, no. 2, pp. 172–179, 2022.
 - [13] Iswanto, Tulus, and Sihombing Poltak, "Comparison of Feature Selection To Performance Improvement of K-Nearest Neighbor Algorithm in Data Classification," *J. Tek. Inform.*, vol. 3, no. 6, pp. 1709–1716, 2022, doi: 10.20884/1.jutif.2022.3.6.471.
 - [14] M. Syukri Mustafa and I. Wayan Simpen, "Implementation of the K-Nearest Neighbor (KNN) Algorithm to Predict Patients Affected by Diabetes at the Manyampa Health Center, Bulukumba Regency," *Pros. Semin. Ilm. Sist. Indormasi dan Teknol. Inf.*, vol. VIII, no. 1, pp. 1–10, 2019.
 - [15] Hasanah Riyan Latifahul, Hasan Muhamad, Pangesti Witriana Endah, Wati Fanny Fatma, and Gata Windu, "Klasifikasi Penerima Dana Bantuan Desa Menggunakan Metode Knn (K-Nearest Neighbor)," *J. Techno Nusa Mandiri*, vol. 16, no. 1, pp. 1–6, 2019, doi: 10.33480/techno.v16i1.25.
 - [16] Tangguh Admojo Fadhila, "Indonesian Journal of Data and Science Klasifikasi Aroma Alkohol Menggunakan Metode KNN," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 34–38, 2020.
 - [17] Nurjanah Siti, Siregar Amril Mutoi, and Kusumaningrum Dwi Sulistya, "Penerapan Algoritma K – Nearest Neighbor (KNN) Untuk Klasifikasi Pencemaran Udara Di Kota Jakarta," *Sci. Student J. Information, Technol. Sci.*, vol. 1, no. 2, pp. 71–76, 2020, [Online]. Available: <http://journal.ubpkarawang.ac.id/mahasiswa/index.php/ssj/article/view/14>
 - [18] Aisha Alfani W.P.R., Rozi Fahrur, and Sukmana Farid, "Prediksi Penjualan Produk Unilever Menggunakan Metode K-Nearest Neighbor," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 6, no. 1, pp. 155–160, 2021, doi: 10.29100/jupi.v6i1.1910.
 - [19] Handoko Dedi, Tambunan Heru Satria, and Tata Hardinata Jaya, "Analisis Penjualan Produk Paket Kuota Internet Dengan Metode K-Nearest Neighbor," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.)*, vol. 6, no. 1, pp. 111–119, 2021, doi: 10.30645/jurasik.v6i1.275.
 - [20] Ali Jamaluddin and Faroji Ridwan, "Pengaruh Profitabilitas terhadap Nilai Perusahaan," *J. Neraca Perad.*, vol. 1, no. 2, pp. 128–135, 2021, doi: 10.55182/jnp.v1i2.36.
 - [21] Basuki Beni, Alwis Nazir, Gusti Siska Kurnia, Handayani Lestari, and Iskandar Iwan, "Klasifikasi Tingkat Keberhasilan Produksi Ayam Broiler di Riau Menggunakan Algoritma K-Nearest Neighbor," *J. Sist. Komput. dan Inform.*, vol. 4, no. 3, pp. 510–516, 2023, doi: 10.30865/json.v4i3.5665.