

Klasifikasi Tingkat Kesejahteraan RW 006 Kelurahan Kalideres Jakarta Barat dengan Metode K-Nearest Neighbor

Veri Arinal¹, Edwin Sentosa²

^{1,2} Teknik informatika Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, Indonesia
Email: veriarinal@gmail.com¹, edwin.keepcalm@gmail.com²

Abstrak

Kesejahteraan adalah suatu keadaan seseorang dapat mampu memenuhi seluruh kebutuhan serta mampu melakukan hubungan baik dengan lingkungan sekitar. Dengan adanya pembangunan diharapkan taraf hidup masyarakat menjadi lebih sejahtera sehingga kualitas sumber daya manusia menjadi lebih baik. Akan tetapi, pada RW 006 Kelurahan Kalideres Jakarta Barat memiliki kualitas sumber daya manusia yang bervariasi. Hal tersebut dikarenakan terdapat tidak meratanya pembangunan. Pada penelitian ini metode yang digunakan adalah K-Nearest Neighbor yang bertujuan untuk mendapatkan akurasi klasifikasi RW 006 Kelurahan Kalideres Jakarta Barat dengan memanfaatkan confusion matriks, didapatkan nilai precision, recall, dan skor F-1. Hasil Nilai presisi terbaik dari $k = 10$ adalah 0,8. Nilai precision yang didapat adalah 0,76 yang menunjukkan proporsi prakiraan benar positif dibandingkan dengan hasil antisipasi positif umum. Sedangkan recall nilai 0.75 dengan skor F-1 0.75 serta nilai kurva ROC dan AUC dari metode K-Nearest Neighbor (k-NN) didapatkan nilai 0.827.

Kata Kunci: Klasifikasi, Kesejahteraan, *K-Nearest Neighbor*.

Abstract

Government assistance is a condition where an individual can satisfy all requirements and can have great relations with the general climate. With the development, it is trusted that the way of life of the local area will be more prosperous so the nature of HR will be better. Be that as it may, at RW 006, Kalideres Town, West Jakarta, the nature of HR shifts. This is on the grounds that there is inconsistent turn of events. In this study, the method used is K-Nearest Neighbor which aims to obtain the classification accuracy of RW 006 Kalideres Village, West Jakarta by utilizing the confusion matrix and the Roc Curve, obtaining precision, recall, and F-1 scores. Results The best precision value of $k = 10$ is 0.8. The precision value obtained is 0.76, which indicates the proportion of true positive predictions compared to the general positive anticipation results. While the recall value is 0.75 with an F-1 score of 0.75 and the value of the ROC and AUC curves from the K-Nearest Neighbor (k-NN) method is 0.827

Keywords: *Classification, Welfare, K-Nearest Neighbor*.

PENDAHULUAN

Kampung Kojan merupakan salah satu di antara sekian banyak kelurahan yang berada di kota DKI Jakarta. Lebih tepatnya di RW 006 Kelurahan Kalideres Jakarta Barat.

DKI Jakarta melalui berbagai program pengembangan praktis terus berupaya menggarap bantuan pemerintah perorangan. Strategi peningkatan yang berbeda ditujukan untuk bekerja pada keadaan keuangan daerah setempat. Hal ini sesuai dengan visinya yang tergambar dalam "Jakarta

adalah kota yang tercipta, wajar, dan berbudaya yang penduduknya diasosiasikan dengan pengakuan kehadiran, pemerataan, dan kemakmuran bagi semua”.

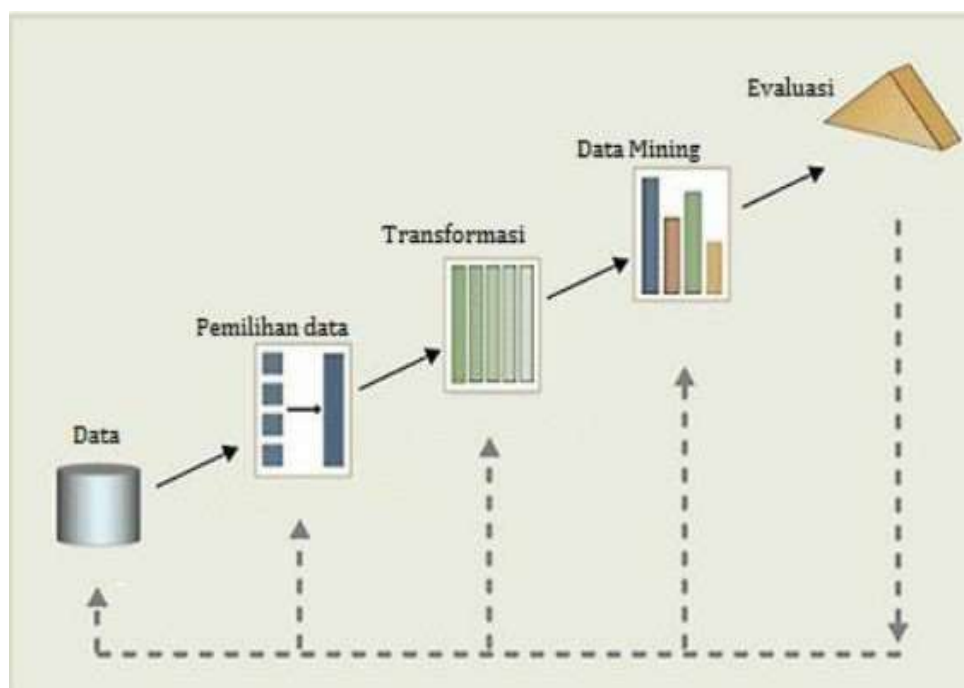
Sebagai ibu kota negara, Jakarta berkali-kali menjadi kesan kemajuan Indonesia sebagai semacam perspektif kemajuan di berbagai kabupaten. Dengan cara ini, penting untuk menilai pameran kemajuan yang berbeda, sehingga pendekatan yang akan diambil ke depan dapat mencapai tujuan dan memiliki opsi untuk bekerja pada bantuan pemerintah daerah secara keseluruhan serta menjadi model untuk perbaikan di berbagai kabupaten.

Jakarta Barat sebagai bagian dari Provinsi DKI Jakarta, merupakan salah satu sentra pertumbuhan ekonomi di Indonesia, dengan andil PDRB sebesar 469.9 Triliun Rupiah pada tahun 2020. Memiliki Luas wilayah sebesar 129.54 km², Jakarta Barat merupakan tempat tinggal bagi 2.62 juta penduduk pada tahun 2020.

Evaluasi terhadap hasil pembangunan salah satunya dapat dilakukan dengan melihat perkembangan indikator kesejahteraan rakyat, sebagai sasaran dari berbagai program pembangunan. Apabila indikator yang ada menunjukkan kesejahteraan rakyat meningkat, maka dapat diasumsikan bahwa kinerja pembangunan relatif sudah mencapai sasaran yang dituju, demikian pula sebaliknya. Indikator kesejahteraan rakyat yang dimaksud antara lain dapat dilihat dari tingkat pendidikan masyarakat secara umum, tingkat kesehatan, tingkat pendapatan masyarakat, tingkat pengangguran, kondisi perumahan, dan sebagainya.

Pengertian Data Mining

Penambangan informasi adalah siklus yang menggunakan setidaknya satu prosedur pembelajaran PC (AI) untuk memecah dan memusatkan informasi secara konsekuen otomatis (M. A. K. Neighbor, 2017) Mengingat tugas, penggalian informasi dirakit menjadi penggambaran, penilaian, perkiraan, urutan, pengelompokan dan afiliasi (Studi et al., n.d.). Penambangan informasi terdiri dari pembelajaran tanpa bantuan dan pembelajaran terarah (Ayu et al., 2019) Informasi keterbukaan informasi dapat menangani semua informasi yang tidak remeh untuk menemukan desain dalam informasi, di mana contoh-contoh yang ditemukan substansial dan lugas (Yandi Saputra & Primadasa, 2018).



Gambar 2.2 Knowledge Discovery Data

Tahapan KDD adalah:

1. Informasi Membuat indeks informasi yang objektif, memutuskan indeks informasi dan memusatkan perhatian pada subset faktor atau pengujian informasi, di mana pemeriksaan akan dilakukan.
2. Pemilihan Informasi Langkah awal penanganan informasi dan pembersihan informasi adalah kegiatan penting seperti evakuasi huru-hara. Sebelum melakukan proses information mining, diperlukan suatu siklus pembersihan pada informasi yang menjadi konsentrasi di KDD.
3. Perubahan Tahap ini merupakan tahap strategi inovatif dan sangat bergantung pada contoh data yang akan dilihat dalam kumpulan data.
4. Penambangan Informasi Dalam pemilihan perhitungan penggalian informasi untuk mencari proses penambangan informasi antara lain strategi, teknik atau perhitungan dalam pergeseran penambangan informasi secara umum. Kepastian teknik atau perhitungan yang sesuai bergantung pada target dan proses KDD pada umumnya.
5. Penilaian Tahap ini merupakan tahap penilaian, apakah contoh yang ditemukan bertentangan dengan kenyataan saat ini atau spekulasi yang sudah ada.

Kesejahteraan

Kesejahteraan yang berasal dari kata 'Sejahtera' merupakan kata yang berasal dari bahasa sanksekerta dan memiliki arti 'payung'. Kesejahteraan adalah suatu keadaan seseorang dapat mampu memenuhi seluruh kebutuhan serta mampu melakukan hubungan baik dengan lingkungan sekitar. Orang yang hidupnya sejahtera merupakan orang yang terbebas dari kemiskinan, kebodohan, ketakutan, atau kekhawatiran sehingga hidupnya aman dan tentram baik lahir maupun batin.

Klasifikasi

Pengelompokan adalah metode yang terlibat dengan menemukan model yang menggambarkan dan mengenali kelas informasi, atau pendekatan untuk mengkarakterisasi informasi menjadi setidaknya satu kelas yang telah ditentukan. Strategi karakterisasi yang banyak digunakan antara lain Brain, Unpleasant set, K-dekat tetangga, pengklasifikasi Bayesian, organisasi, dan lain-lain. Proses penyusunan informasi terdiri dari 2 tahap, yaitu tahap memahami (tahap persiapan) dan urutan. Pengalaman yang berkembang dibuat untuk memecah informasi persiapan dan setelah itu ditangani sebagai aturan pengelompokan. Sedangkan siklus karakterisasi, dimana informasi pengujian digunakan untuk menilai ketepatan aturan pengaturan. Model ini bekerja dengan menyelidiki kumpulan data tuple. Setiap tuple dianggap sebagai kelas yang telah ditentukan yang dicirikan oleh suatu sifat yang disebut karakteristik nama kelas. Itu cenderung diuraikan dalam gambar di bawah ini:



Gambar 2.3 Class label atribut Proses klasifikasi didasarkan pada empat komponent :

1. Kelas
Variabel dependen adalah kelas yang membahas "tanda" yang terkandung dalam item.
2. Indikator
Variabel otonom ditangani oleh kualitas informasi.
3. Menyiapkan kumpulan data
Indeks informasi yang memiliki kelebihan dari dua bagian yang digunakan untuk menentukan kelas

yang masuk akal berdasarkan indikator.

4. Menguji kumpulan data.

Sebagai informasi baru yang dikelompokkan berdasarkan model informasi yang telah dibuat dan dinilai ketepatan penyusunannya.

Set preparasi dikarakterisasi, dan fokus K yang berdampingan (tersusun) yang terdekat dengan jarak titik uji ditentukan dengan menggunakan teknik estimasi jarak dalam KNN. Tes tidak ditetapkan sebagai titik hubung K yang paling dominan dan memiliki tempat dengan kualitas kelas (Z. Zhang et al., 2020).

K-Nearest Neighbor

Investigasi kontekstual ini diambil dari pemeriksaan masa lalu. Strategi pengelompokan kNN yang menggabungkan Conviction Component (kNN-CF) dapat memasukkan faktor kepastian dalam teknik kNN tradisional dengan tujuan cenderung diterapkan pada awal urutan kNN untuk memenuhi kebutuhan kemajuan yang tidak merata (S. Zhang et al., 2017). Memasukkan ukuran elemen keyakinan untuk mengelompokkan kNN dengan k tetap untuk semua contoh (S. Zhang et al., 2018). Salah satu perhitungan yang dapat digunakan untuk pengelompokan adalah K-Nearest Neighbor (KNN) (K. Neighbor et al., 2020).

Perhitungan dengan kondensasi KNN (K-Nearest Neighbor) adalah suatu teknik untuk mengelompokkan objek berdasarkan penyiapan informasi yang memiliki jarak terdekat dari item tersebut (M. A. K. Neighbor, 2021). Perhitungan K-Nearest Neighbor (K-NN) merupakan strategi pengelompokan objek dilihat dari informasi pembelajaran yang paling dekat dengan item tersebut. K-NN menggabungkan perhitungan pembelajaran yang diatur dimana konsekuensi dari kasus inkuiri baru dicirikan berdasarkan klasifikasi K-NN, maka kelas yang tampaknya paling banyak akan menjadi kelas hasil karakterisasi (Wang et al., 2022).

Tetangga terdekat adalah perhitungan urutan berdasarkan kesamaan. Bandingkan informasi pengujian dan informasi persiapan yang dekat dengan informasi pengujian (Pratama, 2018). Nilai K dalam K-NN menyiratkan informasi K terdekat dari informasi pengujian (Sugiarto et al., 2022). Perhitungan K-Closest Neighbor (K-NN) adalah salah satu perhitungan yang paling mudah untuk menangani masalah pengaturan dan sering memberikan hasil yang kejam dan kritis (Deva Gautama, 2019). K-Nearest neighbor memudahkan untuk melacak karakterisasi atau memutuskan ekspektasi dari suatu variabel. Variabel yang mempengaruhi daya pancar KNN adalah nilai K, jarak Euclidean dan standarisasi batas (Nikmatun & Waspada, 2019)(Lena Oktavianis, 2021) (K. Neighbor et al., 2020). Faktor-faktor yang mempengaruhi performansi KNN adalah nilai K, jarak Euclidean dan normalisasi parameter (Taunk, 2019).

Biasanya, nilai k yang tinggi akan mengurangi dampak keributan pada pertemuan, namun membuat batas antara setiap penggambaran lebih kabur. Nilai berguna dari k dapat diambil dengan meletakkan batas. Ketepatan estimasi K-NN sangat dipengaruhi oleh ada atau tidaknya bagian yang berlebihan atau sebaliknya jika bahan bagian tersebut tidak sesuai dengan kepentingannya untuk pemilihan. Pemeriksaan komputasi ini sebagian besar adalah tentang memilih dan menumpuk fitur, yang semuanya ditujukan untuk pengembangan lebih lanjut eksekusi pengiriman. K-Nearest Neighbor menghargai dua manfaat dan kelemahan. Di mana kelebihan K-NN, khususnya metode pengelompokannya adalah cukup sederhana untuk dijalankan, cukup dapat dikenali (misalnya, kelas tidak perlu membuang waktu dengan segera memisahkan), itu kuat untuk menangani pembatasan informasi. skala, dan memiliki tempat konvergensi yang berbeda untuk referensi (metrik dan k jarak). Sementara ketidaknyamanan K-NN adalah sangat penting untuk memilih nilai k terbaik untuk memberikan jumlah tetangga terdekat yang lebih mudah dan biaya komputasi yang sangat tinggi

mengingat penilaian jarak harus diselesaikan pada setiap masalah model.

Data penelitian

Pada penelitian ini, dataset yang digunakan diambil dari dataset private. Dataset memiliki Kepala Keluarga di RW 006 Kelurahan Kalideres Jakarta Barat sebanyak 2730 KK jumlah individu sebesar 9164 orang. Dataset diambil per tanggal 21 Juni 2022 pada Ketua Dawis pada RW 006 Kelurahan Kalideres Jakarta Barat. Dengan data sampel yang digunakan data 273 data sebagai data latih dan data uji variabel diambil beberapa variabel

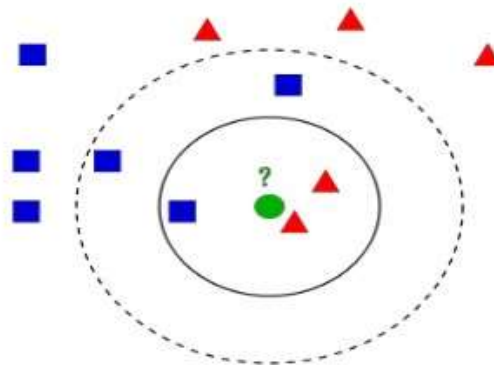
METODE

Euclidean Distance

Jarak pendek atau signifikan umumnya ditentukan berdasarkan jarak Euclidean. Jarak Euclidean diselesaikan menggunakan karakterisasi teks dan kemiripan kosinus. Semakin tinggi harga jarak Euclidean, semakin komparatif informasi tes dengan catatan persiapan. Kemudian lagi, semakin sederhana harga jarak Euclidean, semakin komparatif informasi pengujian dengan laporan persiapan (Kilat, 2019). Semakin sederhana jarak Euclidean menyiratkan jarak antara informasi semakin dekat (Hasanah et al., 2019). Perhitungan dengan kependekan KNN (K-Nearest Neighbor) merupakan teknik karakterisasi objek dalam rangka menyiapkan informasi yang memiliki jarak terdekat dari item tersebut [19]. Perhitungan ini dilakukan dalam tiga tahap sebagai berikut:

1. Pastikan pemisahan dari semua informasi persiapan.
2. Informasi persiapan koordinat berdasarkan jarak dan penentuan K tetangga terdekat.
3. Memanfaatkan kelas bagian yang lebih besar di antara tetangga K terdekat.

Ilustrasi seperti yang terlihat pada gambar di bawah ini:



Gambar 3.2 Klasifikasi KNN

Tes (bintik hijau) harus dipisahkan antara biru dan merah. Jika $k = 3$ (lingkaran garis penuh), dicentang merah karena ada dua merah yang lebih dari satu biru. Jika $k = 5$ (lingkaran berbintik), warnanya biru karena ada tiga biru tambahan daripada dua merah (Kim et al., 2021). Petunjuk langkah demi langkah untuk memainkan siklus penggambaran memanfaatkan estimasi k-NN adalah dengan memanfaatkan kondisi/kondisi jarak Euclidian berikut:

$$d(x, y) = \sqrt{\sum_i^n (x_i - y_i)^2} \dots \dots \dots (1)$$

Data:

$d(x,y)$: jarak Euclidian

x_i : harga informasi persiapan pertama

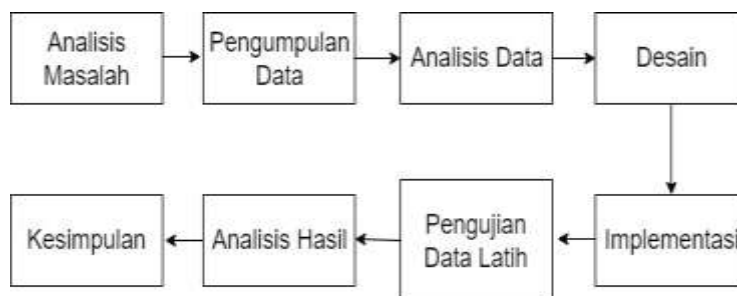
y_i : harga informasi tes pertama

l : Menyebutkan nilai sifat (1,2,3... .n)

Bergerak menuju pekerjaan menggunakan perhitungan KNN mencakup:

1. Putuskan sejauh mungkin.
2. Cari tahu data uji dengan semua data susunan.
3. Urutkan setiap pemisahan yang diuraikan dari asosiasi penilaian.
4. Tentukan jarak terdekat dengan K gathering.
5. Cocokkan kelas yang tepat.
6. Ikuti jumlah kelas dari tetangga terdekat dan turunkan kelas sebagai data yang akan dievaluasi.

Rancangan Pengujian



Gambar 3.3 Tahap rancangan pengujian



Gambar 3.4 Rancangan Data Uji

Analisis Masalah

Level ini menggambarkan level kunci dari survei dengan alasan bahwa pemeriksaan masalah sangat mendasar dalam penelitian, terutama dalam menentukan ketepatan dengan metode K-Nearest Neighbor di RW 006 Kelurahan Kalideres Jakarta Barat, untuk mengetahui hasil akurasi yang dapat seragam mengedarkan perbaikan daerah.

Pengumpulan data

Ada 3 prosedur yang digunakan dalam pengumpulan informasi, yaitu:

a. Wawancara

Wawancara yang dipimpin oleh pencipta sebagai survei terhadap kumpul-kumpul terkait dengan bantuan sosial.

b. Menulis survei

Studi menulis diselesaikan dengan berkonsentrasi pada buku dan buku harian yang berhubungan dengan penelitian.

c. Dokumentasi

Pendokumentasian dilakukan untuk memperoleh dan selanjutnya mengumpulkan bahan-bahan yang diharapkan untuk memimpin penelitian.

Analisis Data

Pada tahap ini, setelah memperoleh informasi, kemudian dipecah untuk menentukan variabel tingkat kesejahteraan di RW 006, Kelurahan Kalideres, Jakarta Barat

Desain

Pada tahap ini pelaksanaan penetapan urutan tingkat kesejahteraan di RW 006 Kelurahan Kalideres, Jakarta Barat dibangun dengan menggunakan perhitungan K-Nearest Neighbor. Perhitungan ini adalah teknik karakterisasi untuk memiliki pilihan untuk pergi dengan pilihan berdasarkan jarak Euclidian. Rumusnya seperti dibawah ini :

$$D(X1, Y2) = \sqrt{\sum_{i=1}^n (X1 - Y2)^2} \dots \dots \dots (2)$$

Data:

D : Jarak terdekat.

X1 : Menguji informasi/menyiapkan informasi.

X2 : Informasi pengujian

n : Jumlah sifat untuk setiap kasus.

l : Individu menganggap dari 1 sampai n.

Urutan yang menggunakan perhitungan K-NN adalah sebagai berikut:

sebuah. Menerapkan fokus kelas b. Data diinput dan diolah menggunakan software seperti Google Colab

1. Bagilah informasi menjadi dua bagian, menjadi informasi persiapan khusus dan informasi pengujian menggunakan persetujuan terpisah.
2. Menyelenggarakan K-Closest Neighbor (k-NN) dengan pengelompokan sebagai berikut:
 - a. Tentukan jumlah k kualitas untuk (k-NN)
 - b. Menghitung Confusion matriks untuk melihat akurasi klasifikasi dan kurva Roc/nilai AUC.

Pengujian

Pengetesan/pengujian diberlakukan melalui yakni dengan menguji hasil klasifikasi jumlah hasil k untuk K-Nearest Neighbor (k-NN). Pengujian dilakukan pada menu untuk mengetahui ramalan dalam memutuskan kekacauan memastikan jaringan untuk melihat ketepatan pengaturan. Memanfaatkan

bahasa pemrograman Python dan Google Colab.

Analisis Hasil

Ditahap ini penulis menghitung *confusion matrix* untuk melihat akurasi klasifikasi memanfaatkan algoritma K-Nearest Neighbor. Prosedur pengaturan data mining harus dimungkinkan dengan menilai dan menyetujui, khususnya confusion matrix dengan kurva ROC/nilai AUC.

Konfusi matriks adalah prosedur yang biasanya digunakan untuk perhitungan presisi di bidang data mining. Jaringan kecacauan memberikan keputusan yang diperoleh dalam persiapan dan pengujian yang memberikan evaluasi pelaksanaan pengaturan item secara tepat atau salah

Prediksi	Actual	
	True	False
True	True Positif	False Negatif
False	False Positif	True Negatif

Tabel 3. Bentuk Confusion Matrix

Keterangan:

True Positif (TP) = nilai benar positif

True Negatif (TN) = nilai benar negatif

False Positif (FP) = nilai salah positif

False Negatif (FN) = nilai salah negatif

Tabel *Confusion matrix* digunakan untuk mempermudah menghitung nilai akurasi, presisi, recall, dan F-measure.

1. Akurasi : perbandingan kasus yang diidentifikasi benar dengan jumlah seluruh kasus.
2. Presisi : proporsi kasus dengan hasil positif yang diidentifikasi benar.
3. Recall: proporsi kasus positif yang diidentifikasi dengan benar.
4. F-measure : kombinasi pengukuran dari presisi dan recall menjadi suatu metrik.

Rumus yang akan digunakan:

Akurasi : Jumlah nilai benar : Jumlah seluruh data tes X 100%.....(1)

Kurva ROC (Receiver Operating Characteristic) adalah metode yang dapat membayangkan, memilah, dan memilih pengklasifikasi berdasarkan pameran mereka. ROC adalah konsekuensi dari pengelompokan estimasi dalam dua aspek, di mana garis level menunjukkan nilai positif palsu dan garis ke atas menunjukkan nilai positif asli. Untuk mengetahui himpunan mana yang lebih baik, prosedur yang bekerja di dekat tikungan ROC disebut AUC (Are Under the ROC) yang digambarkan sebagai kemungkinan. Skor AUC memprediksi eksekusi yang tidak beralasan dengan mengevaluasi kemungkinan bahwa hasil eksperimen yang dipilih secara diskriminatif dari masyarakat adalah positif atau negatif. Semakin besar nilai AUC, semakin membumi pengumpulan yang digunakan, karena nilai AUC signifikan untuk wilayah unit., nilainya akan terus berada pada kualitas 0 dan 1.

HASIL DAN PEMBAHASAN

Implementasi dan Pengujian

Perancangan sistem untuk menentukan pengujian yang dilakukan menggunakan bahasa pemrograman python dan google colab, perancangannya adalah sebagai berikut:

Import library yang dibutuhkan dan unggah data.


```

%matplotlib inline
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn import neighbors
from matplotlib.colors import ListedColorMap, BoundaryNorm
import matplotlib.patches as mpatches

[79]: data = pd.read_table('/root/.data/Klasifikasi-RW 006 Kelurahan Kalideres.txt')
      data.head()

```

	Nama	Marga RT	Status_Pekerjaan	Penghasilan	Tanggungans	Pendidikan	Asuransi	Tabungan	status_sejahtera
0	Sugrah	3	Karyawan	2.500.000	4	2	1	0	0
1	SiB	3	Wiraswasta	4.453.935	1	3	1	0	1
2	Melyanah	4	Karyawan	2.500.000	2	3	1	0	0
3	Els Marlenah	14	Pegawai swasta	2.500.000	4	3	1	0	0
4	HENDRY EFFENDY	5	Wiraswasta	2.500.000	2	3	1	0	0

Gambar 4.1 Import Library dan unggah data

1. NumPy

Pustaka pertama ini mewakili Python Matematika. Kemampuan modul ini adalah untuk membantu peneliti informasi dalam mengelola masalah matematika. Menekankan pada perhitungan logis, NumPy dapat membongkai objek cluster berlapis-N seperti MATLAB. NumPy memudahkan klien dalam tugas matematika Polinomial Langsung seperti Vektor dan Kisi.

2. Pandas

Pandas adalah perpustakaan yang harus dipelajari bagi pemula untuk belajar Python. Mengingat kerangka dataframe, modul ini dapat menumpuk catatan ke dalam tabel virtual seperti lembar perhitungan. Pandas juga berfungsi untuk menangani informasi seperti bergabung, khusus, sekelompok, pengumpulan, dan metode lain seperti SQL. Yang penting, ini sudah selesai di atas meja. Manfaat perpustakaan ini adalah juga dapat membaca catatan dari organisasi yang berbeda, misalnya, .txt, .csv, dan .tsv.

3. Matplotlib

Informasi yang telah ditangani harus ditampilkan dengan jelas dan sebanding dengan yang dapat dibayangkan. Pustaka Matplotlib ini akan membantu dalam menampilkan ujian membawa jenis desain variasi dengan cara yang lebih apik dan memikat. Ada dua macam plot dalam menampilkan informasi, yaitu dalam 2D dan 3D. Jadi informasi dapat ditampilkan tergantung pada situasinya. Matplotlib adalah perpustakaan yang paling sering digunakan oleh ilmu informasi dalam memperkenalkan informasi dengan cara visual yang lebih memikat.

Menampilkan 10 Dataset RW 006 lalu membaca data yang sudah di upload.

```

%print(data.shape)
data.head(10)

```

(773, 9)

	Nama	Marga RT	Status_Pekerjaan	Penghasilan	Tanggungans	Pendidikan	Asuransi	Tabungan	status_sejahtera
0	Sugrah	3	Karyawan	2.500.000	4	2	1	0	0
1	SiB	3	Wiraswasta	4.453.935	1	3	1	0	1
2	Melyanah	4	Karyawan	2.500.000	2	3	1	0	0
3	Els Marlenah	14	Pegawai swasta	2.500.000	4	3	1	0	0
4	HENDRY EFFENDY	5	Wiraswasta	2.500.000	2	3	1	0	0
5	Andri	12	Buruh Pabrik	2.500.000	2	3	1	0	0
6	Aska	4	Wiraswasta	4.453.935	1	3	1	1	1
7	Jamal	3	Pegawai swasta	4.453.935	2	3	1	0	1
8	Khalidul Fikron	1	Karyawan TU	2.500.000	1	3	1	0	0
9	Fen Ardiansyah	1	Karyawan Bank	4.453.935	1	4	1	1	1

Gambar 4.2 Menampilkan 10 Dataset RW 006

Berdasarkan gambar 4.2 Menjelaskan berdasarkan Data Klasifikasi Rw 006 Dataset dapat diketahui bahwa terdapat 273 data yang menjelaskan tentang data penduduk RW 006. Dalam data ini terdapat variabel yaitu Status_Pekerjaan , Penghasilan , Tanggungan , Pendidikan , Asuransi , Tabungan, status sejahtera. Klasifikasi Data yaitu Tanggungan : 1 = 1 orang 2 = 2 orang 3 = 3 orang 4 = lebih dari 3 ; Pendidikan : 1 = SD 2 = SMP 3 = 4 = S1 ; Asuransi : 1 = Memiliki 2 = Tidak Memiliki ; Tabungan 0 = Tidak Memiliki 1 = Memiliki, status_sejahtera ; 0 = tidak 1=sejahtera

Membagi data latih dan data uji untuk memprediksi klasifikasi yang dihasilkan sebagai berikut:

```
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.neighbors import KNeighborsClassifier

[87] X = data[["Asuransi", "Tabungan"]]
     Y = data["status_sejahtera"]
     X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.20, random_state=45)

[88] error_rates = []
     import numpy as np

[89] for i in np.arange(1, 101):
     new_model = KNeighborsClassifier(n_neighbors = i)
     new_model.fit(X, Y)
     new_predictions = new_model.predict(X_test)
     error_rates.append(np.mean(new_predictions != y_test))
```

Gambar 4.3 Membagi data latih dan uji

Setelah data ditangani, informasi tersebut dipartisi menjadi 2 bagian, yaitu informasi persiapan khusus dan informasi pengujian. Siklus persiapan selesai menggunakan persetujuan split. Bagian persetujuan adalah kerangka kerja administratif yang mengisolasi informasi menjadi dua bagian secara serampangan, di mana bagian dari informasi persiapan dan bagian lainnya sebagai informasi pengujian. Dengan menggunakan persetujuan split, pengaturan yang mendasari akan menyelesaikan proses merenungkan proporsi split yang baru-baru ini dicirikan. Kemudian, pada saat itu, kelebihan pemisahan persetujuan dari informasi ketersediaan akan dianggap sebagai informasi pengujian. Mempersiapkan informasi adalah informasi yang akan digunakan dalam pembelajaran sedangkan informasi tes adalah informasi yang belum pernah digunakan untuk pembelajaran dan akan digunakan sebagai informasi untuk menguji kenyataan atau ketepatan hasil belajar..


```
✓ [104] knn = KNeighborsClassifier(n_neighbors = 5,  
                                weights='uniform',  
                                algorithm='auto',  
                                metric='euclidean')  
  
knn.fit(X_train, y_train)  
  
↳ KNeighborsClassifier(metric='euclidean')  
  
✓ [106] knn.score(X_test, y_test)  
  
0.76
```

Gambar 4.7 KNN score k=5

```
✓ [124] knn = KNeighborsClassifier(n_neighbors = 10,  
                                weights='uniform',  
                                algorithm='auto',  
                                metric='euclidean')  
  
knn.fit(X_train, y_train)  
  
KNeighborsClassifier(metric='euclidean', n_neighbors=10)  
  
✓ [125] knn.score(X_test, y_test)  
  
0.8
```

Gambar 4.8 KNN score k=10

Berdasarkan Gambar 4.7 dan 4.8 di atas yang terlihat hasil dari beberapa test pada nilai k di k-NN itulah yang ditunjukkan oleh teknik skor knn terbaik adalah dengan nilai k = 10 dengan akurasi 0,8.

```

import matplotlib.patches as mpatches
import matplotlib.pyplot as plt
from sklearn import neighbors
import numpy

def plot_surga_nampak(X, y, n_neighbors, weights, X_test, y_test):
    X_mat = X
    y_mat = y

    cmap_light = ListedColorMap(['#FFDAB9', '#FFDAB9', '#FFDAB9', '#FFDAB9'])
    cmap_bold = ListedColorMap(['#FFDAB9', '#FFDAB9', '#FFDAB9', '#FFDAB9'])

    clf = neighbors.KNeighborsClassifier(n_neighbors, weights=weights)
    clf.fit(X_mat, y_mat)

    mesh_step_size = .01
    plot_symbol_size = 50

    x_min, x_max = X_mat[:, 0].min() - 1, X_mat[:, 0].max() + 1
    y_min, y_max = X_mat[:, 1].min() - 1, X_mat[:, 1].max() + 1
    xx, yy = numpy.meshgrid(numpy.arange(x_min, x_max, mesh_step_size),
                            numpy.arange(y_min, y_max, mesh_step_size))
    Z = clf.predict(numpy.c_[xx.ravel(), yy.ravel()])

    Z = Z.reshape(xx.shape)
    plt.figure()
    plt.pcolormesh(xx, yy, Z, cmap=cmap_light)

    plt.scatter(X_mat[:, 0], X_mat[:, 1], s=plot_symbol_size, c=y, cmap=cmap_bold, edgecolor='black')
    plt.xlabel('x1')
    plt.ylabel('x2')

    title = "neighbors = {}".format(n_neighbors)
    if (X_test is not None):
        train_score = clf.score(X_mat, y_mat)
        test_score = clf.score(X_test, y_test)
        title = title + "\ntrain score = {:.7f}, test score = {:.7f}".format(train_score, test_score)

    patch0 = mpatches.Patch(color='#FFDAB9', label='1')
    patch1 = mpatches.Patch(color='#FFDAB9', label='1')
    plt.legend(handles=[patch0, patch1])
    plt.xlabel('x1')
    plt.ylabel('x2')
    plt.title(title)
    plt.show()

```

```

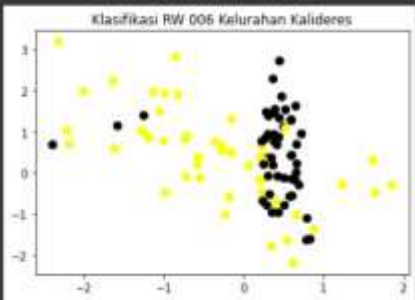
from sklearn.datasets import make_classification, make_blobs
from matplotlib.colors import ListedColorMap

cmap_bold = ListedColorMap(['#FFDAB9', '#FFDAB9', '#FFDAB9', '#FFDAB9'])

plt.figure()
plt.title('Klasifikasi RW 006 Kelurahan Kalideres')
X_C2, y_C2 = make_classification(n_samples=100, n_features=2,
                                n_redundant=0, n_informative=2,
                                n_clusters_per_class=1, flip_y=0.2,
                                class_sep=0.5, random_state=0)

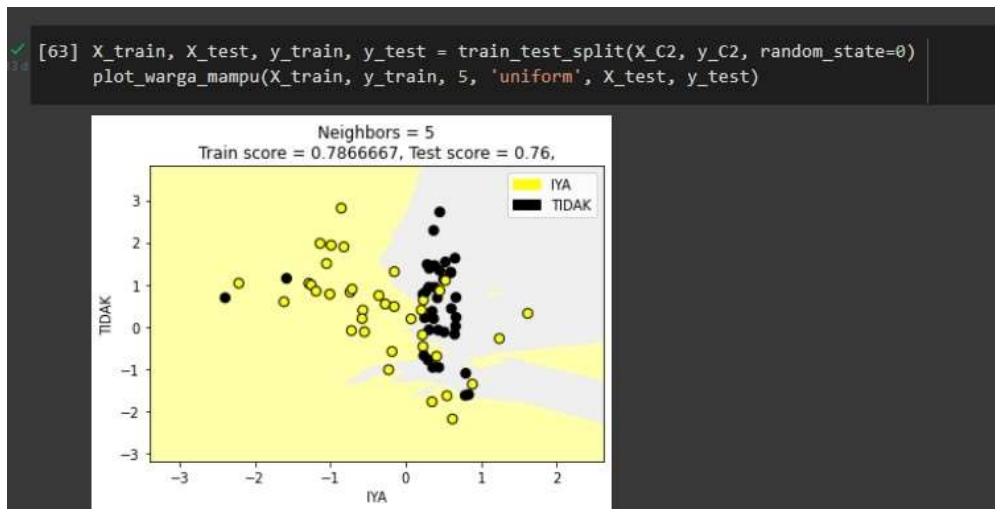
plt.scatter(X_C2[:, 0], X_C2[:, 1], c=y_C2,
            marker='o', s=50, cmap=cmap_bold)
plt.show()

```



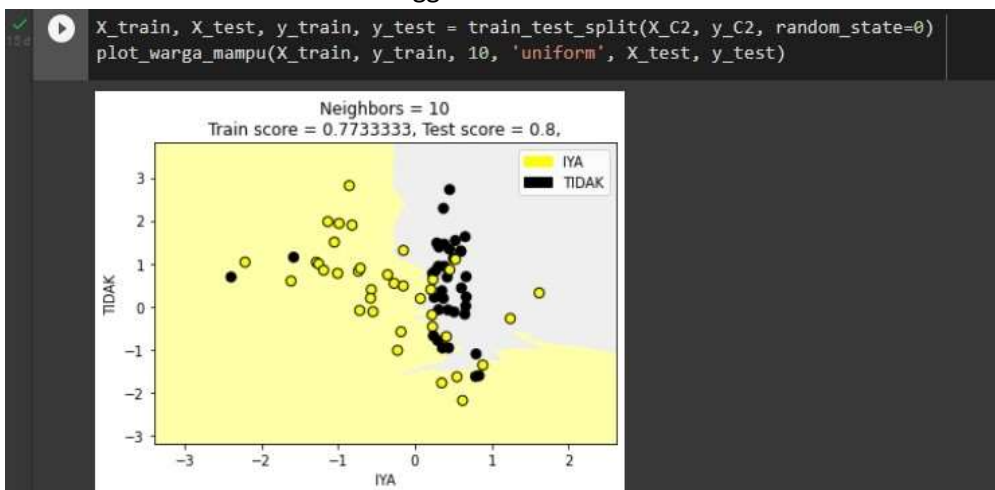
Gambar 4.9 Visualisasi Klasifikasi Algoritma K-Nearest Neighbor

Visualisasi Klasifikasi K-NN Menggunakan Nilai k = 5



Gambar 4.10 Visualisasi dan akurasi Nilai k = 5

Visualisasi Klasifikasi K-NN Menggunakan nilai k = 10



Gambar 4.11 Visualisasi dengan nilai k = 10

Hasil Akurasi Confusion Matrix nilai k = 10 untuk Klasifikasi RW 006 Kelurahan Kalideres

```

✓ [131] #import confusion_matrix
      from sklearn.metrics import confusion_matrix

✓ [133] #import confusion_matrix
      from sklearn.metrics import confusion_matrix

✓ [134] confusion_matrix(y_test,y_pred)

array([[ 7,  4],
       [ 2, 12]])

✓ pd.crosstab(y_test, y_pred, rownames=['True'], colnames=['Predicted'], margins=True)

Predicted  0   1  All
True
0         7   4  11
1         2  12  14
All        9  16  25

```

Gambar 4.12 Hasil Confusion Matrix

```

✓ [136] #import classification_report
      from sklearn.metrics import classification_report

✓ print(classification_report(y_test,y_pred))

              precision    recall  f1-score   support

0               0.78        0.64        0.70         11
1               0.75        0.86        0.80         14

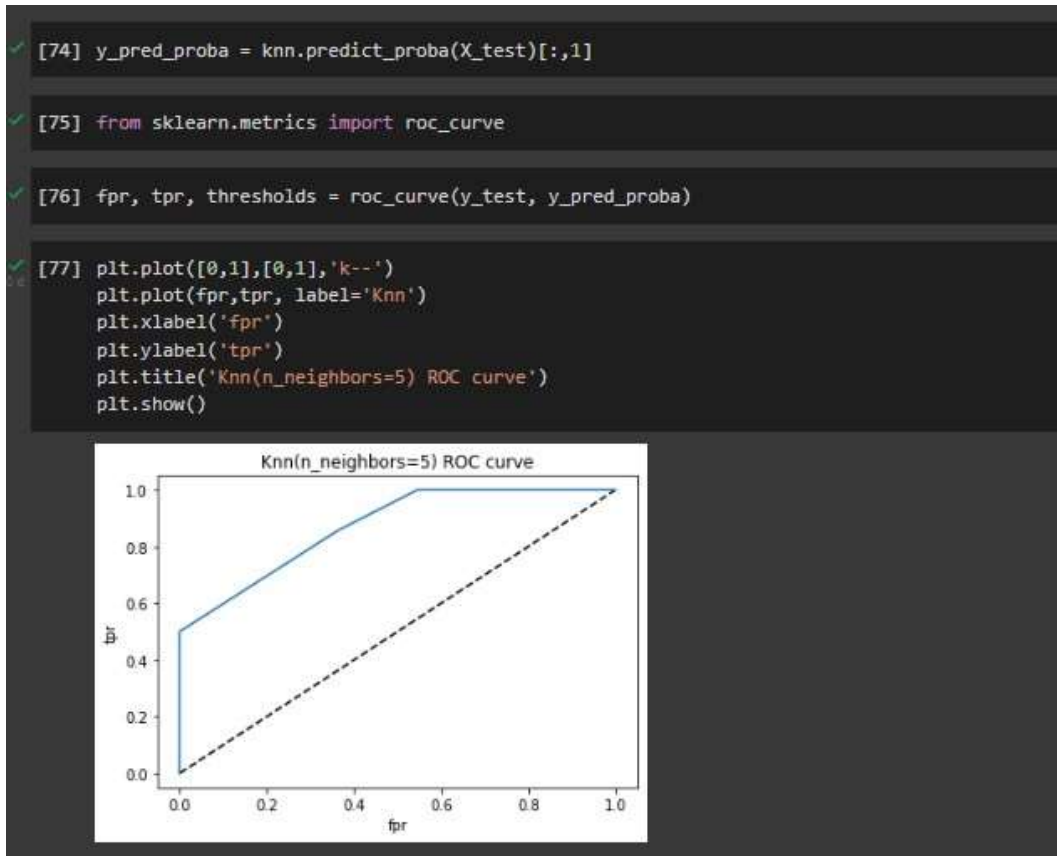
 accuracy          0.76
 macro avg         0.76        0.75        0.75         25
 weighted avg     0.76        0.76        0.76         25

```

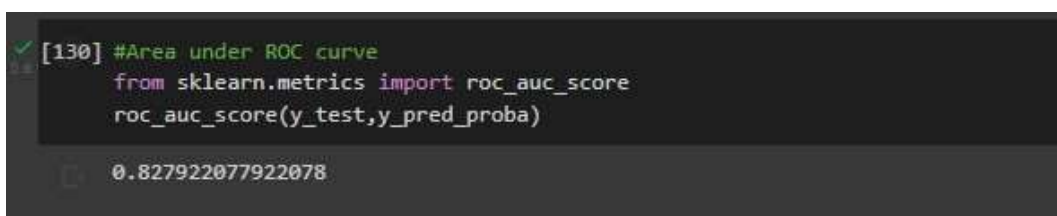
Gambar 4.13 Hasil Precision, Recall, dan F1-score

Confusion matrix adalah teknik penilaian yang memanfaatkan gambar matriks seperti yang ditampilkan pada Gambar 4.12 di atas yang menunjukkan hasil *confusion matrix* menggunakan metode k-NN. Berdasarkan Gambar 4.12, terlihat dari hasil *confusion matrix* pada k = 10 terdapat 7 data yang benar diprediksi dan sebanyak 4 data yang salah diprediksi pada data klasifikasi Rw 006 Kelurahan Kalideres yang tergolong rendah dan ada 12 data yang benar diprediksi dan 2 data yang salah diprediksi pada data klasifikasi Rw 006 Kelurahan Kalideres yang tergolong tinggi. Dengan memanfaatkan confusion matriks diperoleh presisi, recall, dan skor F-1 skor. Hasil akurasi paling tinggi pada nilai k = 10. Nilai presisi yang didapat adalah 0,76 yang menunjukkan rasio prediksi benar positif dibandingkan dengan hasil keseluruhan yang diprediksi positif. Sedangkan recall 0,75 dengan skor F-1 0,75.

Hasil Akurasi Kurva ROC



Gambar 4.13 Kurva ROC nilai k = 10



Gambar 4.14 Hasil Akurasi Kurva ROC nilai k = 10

Curve ROC menunjukkan presisi dan memeriksa classifier secara eksternal. ROC adalah grafik dua lapis dengan sisi atas yang menyestatkan sebagai garis level dan sisi atas asli sebagai garis ke atas. Sedangkan AUC (Region Under Bend) esteem adalah wilayah di bawah tikungan ROC, sebuah tikungan yang menggambarkan kemungkinan/probabilitas dengan faktor responsif dan partikularitas. AUC adalah salah satu jenis estimasi presisi untuk kemungkinan/kemungkinan dalam penilaian. Dilihat dari Gambar 4.13 di atas, cenderung terlihat bahwa efek samping dari ROC bend dan nilai AUC dari strategi K-NN. Pada prosedur k-NN dengan nilai k = 10, nilai AUC adalah 0,827. Untuk karakterisasi information mining, nilai AUC dapat diisolasi menjadi beberapa himpunan.

- 0,90-1,00 = pengelompokan yang sangat baik
- 0,80-0,90 = pengelompokan yang baik
- 0,70-0,80 = pengelompokan yang cukup baik
- 0,60-0,70 = pengelompokan yang buruk
- 0,50-0,60 = pengelompokan yang salah

Mengingat kelompok karakterisasi berakhir, strategi k-NN dengan nilai $k = 10$ diingat untuk kelompok pesanan besar. Maka dengan melihat akibat dari presisi dan curve ROC serta harga AUC, ditemukan bahwa strategi pengelompokan K-NN baik untuk memesan informasi pada Klasifikasi Rw 006 Kelurahan Kalideres, Jakarta Barat.

SIMPULAN

Pada penelitian ini dilakukan menggunakan algoritma k-nearest neighbor dengan menggunakan data yang diolah berdasarkan tahapan Knowledge Discovery in Database (KDD). Berdasarkan hasil perhitungan data mining menggunakan teknik klasifikasi dan algoritma k-nearest neighbor, didapatkan hasil akurasi dengan menggunakan konfusi matriks dan kurva ROC dengan data klasifikasi tingkat kesejahteraan RW 006 Kelurahan Kalideres. Hasil Akurasi terbaik pada nilai $k = 10$. Nilai presisi yang didapat adalah 0,76 yang menunjukkan rasio prediksi benar positif dibandingkan dengan hasil keseluruhan yang diprediksi positif. Sedangkan recall 0,75 dengan skor F-1 0,75 serta hasil kurva ROC dan nilai AUC dari metode k-NN dengan nilai pada $k=10$ diperoleh nilai AUC sebesar 0,827 dapat disimpulkan untuk tingkat kesejahteraan pada RW 006 adalah baik.

DAFTAR PUSTAKA

- Ayu, I., Sita, A., & Wahyu, A. (2019). Klasifikasi Indeks Pembangunan Manusia dengan Metode K-Nearest Neighbor dan Support Vector Machine di Pulau Jawa. 15, 8–21.
- Deva Gautama, A. (2019). Analisis Kualitas Hasil Panen Tembakau Menggunakan K – Nearest Neighbor Berbasis Web. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 3(2), 1–8. <https://doi.org/10.36040/jati.v3i2.846>
- Hasanah, R. L., Hasan, M., Pangesti, W. E., Wati, F. F., & Gata, W. (2019). Klasifikasi Penerima Dana Bantuan Desa Menggunakan Metode Knn (K-Nearest Neighbor). *Jurnal Techno Nusa Mandiri*, 16(1), 1–6. <https://doi.org/10.33480/techno.v16i1.25>
- Kilat, J. (2019). KLASIFIKASI SENTIMENT ANALYSIS PADA KOMENTAR PESERTA DIKLAT MENGGUNAKAN METODE K-NEAREST NEIGHBOR. 8(1), 81–92.
- Kim, J. H., Choi, J. H., Park, Y. H., Leung, C. K. S., & Nasridinov, A. (2021). Knn-sc: Novel spectral clustering algorithm using k-nearest neighbors. *IEEE Access*, 9, 152616–152627. <https://doi.org/10.1109/ACCESS.2021.3126854>
- Lena Oktavianis. (2021). Klasifikasi Kelayakan Peminjaman Kredit Bank Menggunakan Algoritma Probabilistic Neural Network Dan K-Nearest Neighbor Proposal Tugas Akhir.
- Neighbor, K., Kurniawan, Y. I., Barokah, T. I., Soedirman, U. J., Surakarta, U. M., Prof, J., & Boenjamin, H. R. (2020). Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan. 22(1), 73–82.
- Neighbor, M. A. K. (2017). Kata Kunci : resiko kredit, algoritma k-nearest neighbor, data mining, PHP. 1. 1(1), 32–39.
- Neighbor, M. A. K. (2021). Penentuan masyarakat miskin penerima zakat menggunakan algoritma k-nearest neighbor. 2(2), 191–204.
- Nikmatun, I. A., & Waspada, I. (2019). Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor. *Jurnal SIMETRIS*, 10(2), 421–432.
- Pratama, T. C. (2018). Penerapan Metode K-Nearest Neighbour Dalam Menentukan Kelayakan Calon Nasabah Yang Layak Untuk Kredit Mobil (Studi Kasus : Pt . Astra International , Tbk-Toyota). 5(4), 402–408.
- Studi, P., Informatika, T., Teknik, F., Pancasila, U., & Risk, M. (n.d.). PENERAPAN ALGORITMA NEAREST NEIGHBOR UNTUK PENENTUAN RISIKO PEMINJAMAN UANG. 171–176.
- Sugiarto, E., Fahmi, A., Muslih, M., & Hendriyanto, N. (2022). Penerapan K-Nearest Neighbors (KNN) untuk Klasifikasi Aset dalam Upaya Menentukan Aset Wakaf Produktif. *Jurnal Transformatika*, 19(2), 114. <https://doi.org/10.26623/transformatika.v19i2.3356>
- Taunk, K. (2019). 2019 International Conference on Intelligent Computing and Control Systems, ICCS 2019. In 2019 International Conference on Intelligent Computing and Control Systems, ICCS 2019 (Issue Iccics, pp. 1255–1260). IEEE.

- Wang, Y. U., Cao, X., & Li, Y. (2022). Unsupervised Outlier Detection for Mixed-Valued Dataset Based on the Adaptive k-Nearest Neighbor Global Network. *IEEE Access*, 10, 32093–32103. <https://doi.org/10.1109/ACCESS.2022.3161481>
- Yandi Saputra, A., & Primadasa, Y. (2018). Penerapan Teknik Klasifikasi Untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbour Implementation of Classification Method to Predict Student Graduation Using K-Nearest Neighbor Algorithm. 17(4), 395–403.
- Zhang, S., Li, X., Zong, M., Zhu, X., & Cheng, D. (2017). Learning k for kNN Classification. *ACM Transactions on Intelligent Systems and Technology*, 8(3). <https://doi.org/10.1145/2990508>
- Zhang, S., Li, X., Zong, M., Zhu, X., & Wang, R. (2018). Efficient kNN classification with different numbers of nearest neighbors. *IEEE Transactions on Neural Networks and Learning Systems*, 29(5), 1774–1785. <https://doi.org/10.1109/TNNLS.2017.2673241>
- Zhang, W. E. I., Chen, X., Liu, Y., & Xi, Q. (2020). A Distributed Storage and Computation k-Nearest Neighbor Algorithm Based Cloud-Edge Computing for Cyber-Physical-Social Systems. *IEEE Access*, 8, 50118–50130. <https://doi.org/10.1109/ACCESS.2020.2974764>
- Zhang, Z., Hui, J., Shi, Z., Gao, S., Zhang, X., & Fan, H. (2020). Cycle Condition Identification of Loader Based on Optimized KNN Algorithm. *IEEE Access*, 8, 69532–69542. <https://doi.org/10.1109/ACCESS.2020.2985052>