

# KLASIFIKASI JENIS KACA BERDASARKAN PRODUK KOMPOSISINYA MENGGUNAKAN ALGORITMA KNN

*(Glass Classification Using K-Nearest Neighbor Method)*

Agil Trieanto<sup>[1]</sup>, Cani Ananta Savitri<sup>[2]</sup>, Lalu Farid Auliya Rahman<sup>[3]</sup>

<sup>[1]</sup>Dept Informatics Engineering, Mataram University

Jl. Majapahit 62, Mataram, Lombok NTB, INDONESIA

Email: [dedenogie@gmail.com](mailto:dedenogie@gmail.com), [cani.ananta@gmail.com](mailto:cani.ananta@gmail.com), [lalufarid18@gmail.com](mailto:lalufarid18@gmail.com)

## Abstract

*One of the basic materials that are often used in every company, especially industrial companies, is glass. These types of glass include float glass, building windows, lights, and other dining equipment. In the manufacture of these glasses have the same material but have different production materials among each type of glass. This study was conducted to determine the differences between these production materials using the K-Nearest Neighbor method. The data used in this research is the Glass classification dataset which has 214 data with 10 variables. In this study, the highest accuracy was 69.5% with a K value of 2.*

**Keywords:** K-Nearest Neighbor, Data mining, Classification, Artificial intelligence

## 1. PENDAHULUAN

Kaca adalah material padat yang merupakan zat cair yang sangat dingin, karena molekul-molekunya tersusun seperti air namun kohesinya membuat bentuknya menjadi stabil dan ini terjadi karena proses pendinginan atau perpadatan dari proses kimia dan fisika yang sangat cepat. Hal inilah yang membuatnya menjadi transparan atau tembus[1].

Kaca mempunyai nilai positif dan negatif, nilai positifnya kaca dapat dimanfaatkan untuk menunjang kebutuhan masyarakat dan berbagai kebutuhan dunia industri seperti industri mobil, pesawat, perabot rumah tangga dan industri lainnya, sedangkan nilai negatifnya yaitu kaca banyak menghasilkan limbah yang jika tidak ditangani secara serius akan memberikan efek negatif terhadap lingkungan[2]. Namun hal lainnya yang belum kita ketahui ternyata limbah kaca bisa dimanfaatkan untuk didaur ulang menjadi kaca dan berbagai produk kaca lainnya yang lebih kreatif oleh orang-orang seni lainnya

Data yang diperoleh dari UCI Machine Learning Repository yang dikelola oleh Home Office Forensic Science Service dan Diagnostic Products Corporation terlihat bahwa produksi jenis kaca memiliki bahan yang sama namun yang membedakannya adalah komposisi produksi[3].

Adanya perbedaan antara komposisi pembuatan kaca tersebut kami ingin melakukan pengklasifikasian antara tiap-tiap kaca tersebut. Dalam melakukan pengklasifikasi ini kami menggunakan algoritma klasifikasi yang paling dasar yaitu K-Nearest Neighbor

atau biasa disebut algoritma KNN. Algoritma K-Nearest Neighbor (KNN) adalah sebuah metode klasifikasi terhadap sekumpulan data berdasarkan pembelajaran data yang sudah terklasifikasi sebelumnya. Tujuan dari algoritma ini adalah untuk mengklasifikasikan obyek baru berdasarkan atribut dan sampel-sampel dari data training.

Beberapa penelitian yang membahas tentang klasifikasi seperti dalam percobaan Klasifikasi kaca Menggunakan Metode K-Nearest Neighbor. Pada hasil penelitian tersebut menunjukkan bahwa algoritma K-Nearest Neighbor mendapatkan nilai akurasi terbaik yaitu 69%.

Berdasarkan penjelasan sebelumnya, maka batasan masalah yang akan dibahas dalam penelitian ini peneliti menggunakan data penyakit jantung yang menjadi objek penelitian dimana untuk klasifikasi berdasarkan hasil penelitian-penelitian sebelumnya menunjukkan bahwa algoritma K-Nearest Neighbor dapat digunakan sebagai sistem klasifikasi.

## 2. TINJAUAN PUSTAKA

K-Nearest Neighbor (k-NN atau KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran (neighbor) yang jaraknya paling dekat dengan objek tersebut. Dekat atau jauhnya neighbor biasanya dihitung berdasarkan jarak Euclidean[3]. Algoritma K-Nearest Neighbor atau yang biasa disebut KNN ini pertama kali dikenalkan pada awal tahun 1950-an tetapi mulai terkenal di tahun 1960-an[4]. Sehingga sampai saat ini banyak

penelitian yang memanfaatkan metode K-Nearest Neighbor.

Dalam percobaan Klasifikasi Algoritma K-Nearest Neighbor untuk mengklasifikasikan mutu jambu biji berdasarkan warna dan tekstur didapatkan hasil menggunakan nilai k (ketetanggaan paling dekat) [5].



Gambar 1. Hasil grafik rata-rata akurasi.

Kajian ini dapat diambil kesimpulan bahwa metode klasifikasi KNN yang diusulkan dapat melakukan klasifikasi jambu biji dengan akurasi 91.25% pada nilai ketetanggaan terbaik K=3. Fitur yang diekstraks untuk klasifikasi jambu biji adalah r,g,b, luas cacat, energy, homogeneity, dan contrast.

Dalam percobaan penerapan algoritma K-Nearest Neighbor untuk mengklasifikasi dokumen didapatkan hasil penelitian dengan tingkat akurasi 93,4% [6].

Berdasarkan dari beberapa penelitian yang menjadi tinjauan pustaka tersebut dapat diketahui bahwa klasifikasi menggunakan K-Nearest Neighbor cukup baik untuk mengetahui hasil klasifikasi. Pada penelitian ini, dilakukan suatu penelitian untuk mengelompokkan kaca berdasarkan jenis komposisi pembuatannya menggunakan metode klasifikasi K-Nearest Neighbor.

## 2.1. Data Mining

Data mining adalah proses yang menggunakan statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat [7]. Data Mining didefinisikan sebagai proses penemuan pola dalam data. Berdasarkan tugasnya, data mining dikelompokkan menjadi deskripsi, estimasi, prediksi, klasifikasi, clustering dan asosiasi. Proses dalam tahap data mining terdiri dari tiga langkah Utama, yaitu data Preparation Pada langkah ini, data dipilih, dibersihkan, dan dilakukan preprocessed mengikuti pedoman dan knowledge dari ahli domain yang menangkap dan mengintegrasikan data internal dan eksternal ke dalam tinjauan organisasi secara menyeluruh. Penggunaan algoritma data mining dilakukan untuk

menggali data yang terintegrasi untuk memudahkan identifikasi informasi bernilai. Namun semakin besar data yang diolah maka semakin besar pula waktu prosesnya [7]

## 2.2. K-Nearest Neighbor (KNN)

KNN adalah salah satu metode dimana metode ini melakukan klasifikasi berdasarkan data training atau data pembelajaran dilihat dari jarak yang paling dekat dengan objek berdasarkan nilai k. Metode ini bertujuan untuk mengklasifikasikan objek baru berdasarkan atribut dan training sample. Diberikan suatu titik query, selanjutnya akan ditemukan sejumlah K objek atau titik training yang paling dekat dengan titik query. Nilai prediksi dari query akan ditentukan berdasarkan klasifikasi tetangga [6].

Sebelum melakukan perhitungan dengan metode K-Nearest Neighbor, terlebih dahulu harus menentukan data latih dan data uji. Kemudian akan dilakukan proses perhitungan untuk mencari jarak menggunakan Euclidean. Teknik ini sangat sederhana dan mudah diimplementasikan. Mirip dengan teknik clustering, yaitu mengelompokkan suatu data baru berdasarkan jarak data baru itu ke beberapa data/tetangga terdekat. Pertama sebelum mencari jarak data ke tetangga adalah menentukan nilai K tetangga (neighbor). Lalu, untuk mendefinisikan jarak antara dua titik yaitu titik pada data training dan titik pada data testing, maka digunakan rumus Euclidean.

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2} \quad [1]$$

dengan

$d(x_i, x_j)$  : Jarak Euclidean (Euclidean Distance).

$(x_i), (x_j)$  : record ke-i, record ke-j

$(a_r)$  : data ke-r

i,j : 1,2,3,...,n

n : dimensi objek[3]

## 3. METODE PENELITIAN

### 3.1. Sumber Data

Sumber data yang digunakan dalam penelitian ini diperoleh dari website Kaggle yang menyediakan berbagai macam dataset untuk penelitian. Dalam hal ini peneliti menggunakan dataset *Glass classification*. Dataset ini berisi tentang variabel predicator dari kandungan kimia kaca tersebut. Terdapat data kandungan kimia kaca sebanyak 214 dengan data RI

(*Refractive Index*), Na (*Sodium*), Mg (*Magnesium*), Al (*Aluminium*), Si (*Silicon*), K (*Potassium*), Ca (*Calcium*), Ba (*Barium*), Fe (*Iron*) [8]. Berdasarkan isi dari dataset ini, peneliti akan melakukan proses klasifikasi menggunakan algoritma *K-Nearest Neighbor* (KNN).

### 3.2. Pengumpulan Data

Dalam proses penelitian, data yang digunakan harus data yang berkualitas. Dalam penelitian ini data yang akan diolah sebanyak 214 data dengan jumlah variabel sebanyak 10.

Berikut adalah variabel datanya :

Tabel 1. Variabel data

Variabel	Keterangan
RI	<i>Refractive Index</i>
Na	<i>Sodium</i>
Mg	<i>Magnesium</i>
Al	<i>Aluminium</i>
Si	<i>Silicon</i>
K	<i>Potassium</i>
Ca	<i>Calcium</i>
Ba	<i>Barium</i>
Fe	<i>Iron</i>

### 3.3. Pengolahan Data

#### a. Data Validation

Tahapan data validation dilakukan untuk proses mengidentifikasi dan menghapus anomali dan inkonsistensi dari data. Anomali dan inkonsistensi data dapat terjadi dikarenakan oleh beberapa masalah, seperti; adanya missing values dan juga karena data yang dipakai mengandung nilai-nilai yang salah atau disebut dengan noise dan outlier.

#### b. Data Integration dan data Transformation

Tahap selanjutnya adalah menerapkan Data Integration dan Data Transformation pada beberapa dataset untuk meningkatkan akurasi dan efisiensi dari model pembelajaran algoritma yang digunakan. Data yang digunakan dalam penelitian ini adalah data numerik sehingga data transformation tidak diperlukan karena data transformation ini mengubah data string menuju data numerik.

#### c. Data Reduction

Tahapan terakhir adalah Data Reduction yaitu suatu teknik yang digunakan untuk mereduksi ukuran data jika dataset yang digunakan cukup besar yang berguna untuk membuat model pembelajaran algoritma menjadi lebih efisien tanpa mengurangi kualitas dari hasil yang diperoleh. Penelitian ini juga tidak menggunakan data reduction karena penelitian ini menggunakan data yang sedikit dan termasuk dalam golongan

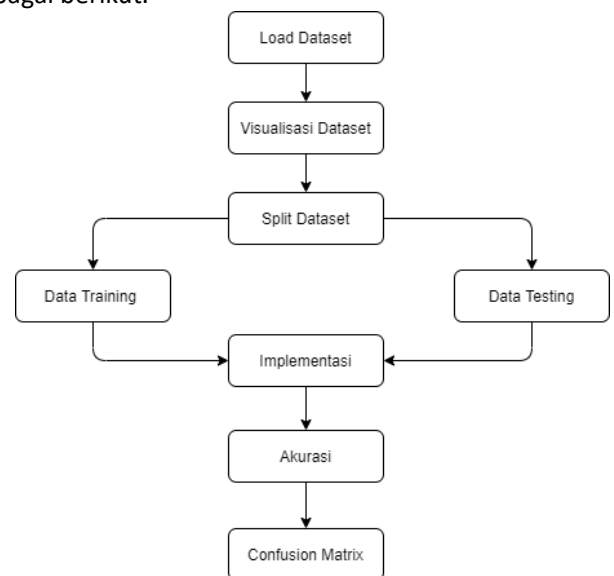
data berukuran kecil dimana memiliki data yang kurang dari 1000 data.

### 3.4. Model yang Diusulkan

Dalam penelitian ini metode yang diusulkan adalah menggunakan metode pengujian dan dalam pengolahan data menggunakan Algoritma K-Nearest Neighbor untuk memprediksi akurasi data.

Output dari proses pengujian akan dijadikan acuan untuk mengambil kesimpulan dari sistem apakah sistem dapat memprediksi nilai akurasi tinggi dengan baik atau tidak.

Model yang diusulkan dalam penelitian ini adalah sebagai berikut.



Gambar 2. Model yang diusulkan

### 3.5. Evaluasi

1. Pada penelitian ini dilakukan evaluasi terhadap model yang ditetapkan untuk mengetahui tingkat keakuratan model.

## 4. HASIL DAN PEMBAHASAN

Hasil dari Penelitian yang dilakukan bertujuan untuk mengklasifikasikan jenis kaca berdasarkan produk komposisinya menggunakan algoritma KNN

### 4.1 Pengujian

Pengujian dilakukan dengan mengambil dataset yang telah didapatkan dari kaggle yang berisi 214 data. Pengujian ini dilakukan dengan jumlah data testing 67 data dan data training 147 data. Algoritma KNN ini bekerja dengan cara yang pertama mencari jarak terdekat antara 2 titik, kemudian mencari tetangga terdekat dari data yang diuji, selanjutnya menentukan akurasi dari data prediksi dan data test

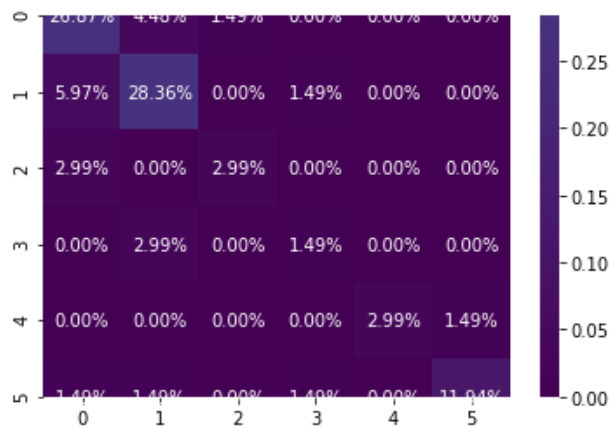
yang diuji untuk mengetahui nilai K terbaik pada klasifikasi kaca.

Tabel 2. Hasil Pengujian

Nilai K	Akurasi
3	71.64%
5	74.62%
7	70.14%
9	62.68%
11	62.68%

Tabel 2. Mendeskripsikan tingkat akurasi nilai K, nilai K yang digunakan dalam jumlah ganjil (3, 5, 7, 9, 11) bertujuan untuk menghindari munculnya jumlah jarak yang sama dalam proses klasifikasi. Jika diperhatikan terdapat nilai K yang memiliki akurasi yang sama yaitu nilai K = 9 dan nilai K = 11. Hal ini disebabkan saat proses klasifikasi dimana proses ini mencari jarak terdekat, kemungkinan jarak terdekat merupakan jenis kaca yang sama. Dari pengujian yang dilakukan didapatkan nilai K = 5 sebagai nilai K terbaik karena memiliki akurasi tertinggi yaitu 74.62%.

Setelah melakukan pengujian dapat dilihat confusion matrix yang didapatkan dari akurasi tertinggi yaitu 74.62%. Confusion matrix sendiri adalah suatu metode yang digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Presisi atau confidence adalah proporsi kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya. Berikut merupakan gambaran dari confusion matriks :



Gambar 3. Confusion matrix

Gambar 5. Menggambarkan confusion matriks dari pengujian yang telah dilakukan dengan mengambil nilai K terbaik yaitu 5 dengan akurasi 74.62% Confusion matrix ini bertujuan untuk mendeskripsikan atau menggambarkan berapa banyak data yang berhasil diprediksi oleh algoritma KNN yang telah dibuat. Sumbu x merupakan sumbu yang menggambarkan nilai dari dataset yang asli dan sumbu y merupakan sumbu yang menggambarkan

nilai yang telah di prediksi. Nilai 0 hingga 5 merupakan representasi dari label yang digambarkan dalam bentuk numerik. Label dari data ini adalah 1 = 0, 2 = 1, 3 = 2, 5 = 3, 6 = 4, dan 7 = 5. Jika dilihat pada Gambar 5. Maka program berhasil memprediksi label 1 memiliki tingkat keberhasilan sebesar 26.87% dan memiliki kegagalan sebesar 10.45%, label 2 memiliki tingkat keberhasilan sebesar 28.36% dan memiliki kegagalan sebesar 8.96%, label 3 memiliki tingkat keberhasilan sebesar 2.99% dan memiliki kegagalan sebesar 1.49%, label 5 memiliki tingkat keberhasilan sebesar 1.49% dan memiliki kegagalan sebesar 2.98%, label 6 memiliki tingkat keberhasilan sebesar 2.99% dan memiliki kegagalan sebesar 0.00%, kemudian label 7 memiliki tingkat keberhasilan sebesar 11.94% dan memiliki kegagalan sebesar 1.49%. Dari confusion matriks tersebut dapat dikatakan bahwa proses pengujian ini menghasilkan tingkat akurasi yang baik hal ini juga dapat dilihat bahwa algoritma KNN ini lebih banyak berhasil dalam memprediksi dibandingkan dari kesalahan prediksi.

## 5. KESIMPULAN DAN SARAN

### 5.1. Kesimpulan

Setelah melakukan analisa penelitian ini, maka dapat disimpulkan bahwa pada penerapan metode K-Nearest Neighbor (KNN) yang digunakan telah mampu digunakan sebagai acuan dalam mengklasifikasikan jenis kaca berdasarkan produk komposisinya sehingga dapat membantu masyarakat dan berbagai kebutuhan dunia industri dalam mengklasifikasi jenis kaca tersebut. Penelitian ini dapat membangun dan merancang sebuah sistem pendukung keputusan yang digunakan untuk membantu proses penyeleksian kaca.

Dari percobaan tersebut didapatkan nilai akurasi untuk k = 3 adalah 71.64%, k = 5 adalah 74.62%, k = 7 adalah 70.14%, k = 9 adalah 62.68%, k = 11 adalah 62.68%. Keseluruhan nilai akurasi dari k = 3, k = 5, k = 7, k = 9, dan k = 11 menunjukkan bahwa akurasi terbaik didapatkan pada nilai k = 5 yaitu 74.62%. Dan nilai K sangat mempengaruhi hasil proses dari klasifikasi.

### 5.2. Saran

Saran yang dapat diberikan untuk penelitian selanjutnya adalah diharapkan menggunakan dataset yang memiliki jumlah data yang lebih banyak supaya mendapatkan hasil yang lebih baik dari penelitian sebelumnya dan juga menggunakan metode klasifikasi lain yang memungkinkan mendapatkan akurasi yang lebih bagus juga.

## 6. UCAPAN TERIMA KASIH

Puji syukur kehadiran Tuhan Yang Maha Kuasa atas segala limpahan rahmat, inayah, taufik, dan hidayahnya sehingga peneliti dapat menyelesaikan penyusunan paper ini.

Dalam proses penelitian dan penyusunan paper ini tentu banyak pihak yang membantu. Peneliti ingin mengucapkan Terima Kasih untuk pihak yang telah membantu penelitian dan penyusunan paper ini, diantaranya Ucapan Terima Kasih untuk:

1. Bapak Ramaditia Dwiyanaputra S.T., M.Eng sebagai Dosen yang baik dan sabar dalam membimbing peneliti dalam proses penyusunan paper ini.
2. Ibu dan Ayah peneliti sebagai orang tua dan memberi semangat.
3. Teman-teman yang membantu dan memberi semangat dalam penyelesaian penyusunan paper ini.

## 7. DAFTAR PUSTAKA

- [1] A. Avansyah, "(99+) (DOC) Pengertian kaca | andri avsyah - Academia.edu," *Acedemia*, 2015.  
[https://www.academia.edu/35924677/Pengertian\\_kaca](https://www.academia.edu/35924677/Pengertian_kaca) (accessed Dec. 24, 2020).
- [2] P. Meilita, A. Dhiyafakhri, R. G. Lina, and I. A. Resti, "Industri Kaca," 2019, doi: 10.31227/osf.io/ed963.
- [3] M. M. Baharuddin, H. Azis, and T. Hasanuddin, "Analisis Performa Metode K-Nearest Neighbor Untuk Identifikasi Jenis Kaca," *Ilk. J. Ilm.*, vol. 11, no. 3, pp. 269–274, 2019, doi: 10.33096/ilkom.v11i3.489.269-274.
- [4] T. R. A. Arjo, "Pengenalan Aktivitas Berkendara Berbasis Sensor Accelerometer Pada Smartphone Dengan Fitur Auto Feedback Menggunakan Algoritma k-Nearest Neighbour | Journal of Applied Informatics and Computing (JAIC)," *polibatam*, 2018.  
<https://jurnal.polibatam.ac.id/index.php/JAIC/article/view/735> (accessed Dec. 24, 2020).
- [5] T. Y. Prahudaya and A. Harjoko, "Metode Klasifikasi Mutu Jambu Biji Menggunakan Knn Berdasarkan Fitur Warna Dan Tekstur," *J. Teknosains*, vol. 6, no. 2, p. 113, 2017, doi: 10.22146/teknosains.26972.
- [6] A. Nugraha, pratama dwi,, Said al faraby, "Klasifikasi Dokumen Menggunakan Metode Knn Dengan Information Gain," *eProceedings Eng.*, vol. 5, no. 1, pp. 1541–1550, 2018.
- [7] N. Yadi, "Tugas Akhir Data Mining," *Iterative Dichotomiser*, vol. 3, pp. 1–62, 2015.
- [8] Q. Lina, "Klasifikasi pada Glass Identification Data Set dengan Menggunakan K-Nearest Neighbour (KNN) | by QOLBIYATUL LINA | Medium," *Medium*, 2019.  
<https://medium.com/@16611110/klasifikasi-pada-glass-identification-data-set-dengan-menggunakan-k-nearest-neighbour-knn-25589a80158c> (accessed Dec. 24, 2020).