

Klasifikasi kualitas mutu susu pasteurisasi menggunakan metode klasifikasi *k*-Nearest Neighbor

Nur Wakhidah ¹⁾, Shafira Nur Rochmah ²⁾

^{1,2)}Fakultas Teknologi Informasi dan Komunikasi, Universitas Semarang
Jl. Soekarno-Hatta, Tlogosari, Semarang 50916
Email: ida@usm.ac.id

	Riwayat Artikel	
Diterima:	Direvisi:	Disetujui:
23-09-2023	01-02-2024	01-03-2024

Abstract

*Milk is a product that is often consumed because it has many good benefits for the body. One of the benefits of milk is that it contains calcium, which is useful for the growth of bones and teeth. However, many consumers still choose dairy products based only on their appearance, even though milk quality is not based on its appearance but is found during milk processing. Pasteurization is the process of heating whole milk at a specific temperature and for a certain period to increase the milk's shelf life and maintain the milk's quality. Two types of pasteurization processes are Long Temperature Long Time and High Temperature Short Time. *k*-Nearest Neighbor or *k*-NN is a data mining method used to classify objects from the data, using distance calculations or Euclidean Distance to look for similarities between neighbors. So, based on this problem, milk quality classification is carried out. The research was carried out by classifying milk quality using the *K*-NN method. The public data used to classify is 1,059 pasteurized milk data, with seven regular or ordinary attributes and one special or class attribute. The *K* values that will be used in this research are $k=3$, $k=5$, and $k=7$. From the analysis of calculation results using the *k*-NN method, the accuracy values obtained are $k=3$ is 97%, $k=5$ is 89%, and $k=7$ is 87%.*

Keywords: Data Mining, *k*-Nearest Neighbor, Pasteurized Milk

Abstrak

Susu merupakan salah satu produk yang sering dikonsumsi karena memiliki banyak manfaat yang baik bagi tubuh. Salah satu manfaat susu yaitu terdapat kalsium yang berguna untuk masa pertumbuhan tulang dan gigi. Namun, masih banyak para konsumen yang memilih produk susu hanya berdasarkan tampilannya saja padahal kualitas susu bukan berdasarkan tampilannya melainkan terdapat pada saat proses pengolahan susu tersebut. Pasteurisasi merupakan suatu proses pemanasan susu murni pada suhu tertentu dan jangka waktu tertentu untuk meningkatkan daya tahan susu dan menjaga kualitas dari susu tersebut. Proses pasteurisasi dibagi menjadi dua cara, yaitu *Long Temperature Long Time* (LTLT) dan *High Temperature Short Time* (HTST). *k*-Nearest Neighbor atau *k*-NN adalah suatu metode dalam data mining yang digunakan untuk melakukan klasifikasi terhadap objek dari data tersebut, dengan menggunakan perhitungan jarak atau *Euclidean Distance* untuk mencari kemiripan atau *similarity* dari ketetanggaannya. Dari permasalahan tersebut maka dilakukan pengklasifikasian kualitas susu. Penelitian yang

dilakukan yaitu mengklasifikasikan kualitas susu menggunakan metode K-NN. Data publik yang digunakan untuk melakukan klasifikasi adalah data susu pasteurisasi sebanyak 1.059 data, dan terdapat tujuh atribut reguler atau atribut biasa dan satu atribut spesial atau *class*. Nilai k yang akan digunakan pada penelitian ini yaitu $k=3$, $k=5$, dan $k=7$. Dari analisis hasil perhitungan menggunakan metode k -NN, nilai akurasi yang didapatkan sebesar $k=3$ sebesar 97%, $k=5$ sebesar 89%, dan $k=7$ sebesar 87%.

Kata kunci: *Data Mining*, *k-Nearest Neighbor*, Susu Pasteurisasi

Pendahuluan

Susu merupakan salah satu bahan pangan yang diperlukan untuk kelengkapan gizi di dalam tubuh, susu yang baik juga dapat terlihat dari warna dan bau dari susu tersebut [1]. Namun, dikarenakan susu sapi tidak dapat disimpan terlalu lama dan juga banyak bakteri atau mikroorganisme yang ada pada susu yang dapat membahayakan tubuh, maka dari itu dilakukanlah pasteurisasi untuk meningkatkan daya tahan susu dan juga membunuh mikroorganisme yang membahayakan seperti bakteri, protozoa dan khamir [2]. Selain dari segi daya tahan susu yang tidak bertahan lama, dalam proses pasteurisasi ini masih banyak menggunakan pemanasan secara manual dengan menggunakan kompor dan panci sebagai wadahnya [3]. Minimnya pengetahuan dalam proses pengolahan susu juga menjadi salah satu masalah yang ada, masih banyak sebagian masyarakat berpikir bahwa susu yang telah dipanaskan itu sudah siap untuk dikonsumsi, namun pada kenyataannya proses pemanasan susu tidak boleh sembarang dilakukan, pada suhu tertentu dipanaskan tidak membunuh bakteri baik yang terdapat di dalam susu dan juga masih banyak yang memilih susu berdasarkan dari warnanya saja, padahal warna pada susu juga menjadi salah satu faktor yang mempengaruhi kualitas mutu susu [4]. Proses pasteurisasi atau proses pemanasan pada susu digunakan untuk meningkatkan daya tahan susu dan juga membunuh bakteri-bakteri yang membahayakan tubuh [5], sedangkan susu pasteurisasi adalah suatu proses pemanasan susu sapi pada suhu tertentu yaitu pada suhu 63°C - 66°C pada kisaran waktu kurang lebih selama 30 menit [6].

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan pembelajaran mesin (*machine learning*) mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* yang terkait [7]. Dalam pembelajaran mesin terdapat beberapa jenis metode untuk melakukan klasifikasi. *k-Nearest neighbor* atau k -NN adalah metode yang bekerja dengan mencari sejumlah k objek data atau pola yang paling dekat dengan pola masukan, kemudian memilih *class* dengan jumlah terbanyak di antara k pola tersebut. Metode k -NN adalah salah satu metode klasifikasi pada data mining, dimana klasifikasi tersebut dilakukan untuk pengelompokan pada kelas-kelas yang sudah ditentukan sebelumnya [8].

Penelitian sebelumnya yang pernah dilakukan menggunakan objek susu sapi adalah identifikasi kualitas susu sapi menggunakan metode *Gray Level Co-*

occurrence Matrix (GLCM) dan *Local Binary Pattern* (LBP) kemudian menggunakan klasifikasi k -NN dan hasil yang diperoleh berbeda. Akurasi yang didapatkan dari penelitian tersebut sebesar 100% dan sebesar 97,5% [9].

Berdasarkan permasalahan tersebut, maka penelitian ini bertujuan mengklasifikasikan kualitas susu pasteurisasi untuk mengetahui hasil susu yang sudah diolah seperti apa saja yang layak untuk dikonsumsi. Penggunaan metode k -NN pada penelitian ini dikarenakan metode tersebut dapat digunakan dalam proses pengklasifikasian suatu data dan memiliki tingkat akurasi yang tinggi berdasarkan kemiripan (*similarity*) yang dihitung menggunakan perhitungan *Euclidean Distance*.

Susu pasteurisasi adalah salah satu proses pemanasan susu sapi murni pada suhu tertentu dan dalam jangka waktu tertentu untuk membunuh virus, bakteri dan jamur [2]. Dari hasil pasteurisasi tersebut menghasilkan banyak produk olahan susu, seperti susu UHT, susu *full cream*, susu *low fat* dan masih banyak lagi. Metode k -NN digunakan dalam penelitian tersebut karena pelatihan sederhana, cepat, mudah dimengerti, dan efektif apabila ukuran data pelatihan besar [10].

Pada penelitian lain, metode k -NN digunakan untuk penelitian pengaruh nilai k pada identifikasi kerusakan jalan. Dari hasil penelitian tersebut, nilai k yang berbeda dapat mempengaruhi nilai akurasi metode k -NN. Metode klasifikasi pada kerusakan jalan menggunakan Naive bayes memiliki hasil 59% untuk kerusakan ringan dan 8% untuk kerusakan berat.[11] Pada $k=1$ akurasi retaknya adalah 98% dan retak kulit buaya adalah 84%, $k=8$ akurasi retaknya adalah 96% dan retak kulit buaya adalah 8%, sedangkan nilai $k=15$ nilai akurasi retaknya adalah 0% dan retak kulit buaya adalah 98% [12]. Tingkat akurasi metode k -NN dengan $k=7$ sebesar 80% pada penelitian mutu telur berdasarkan warna [13].

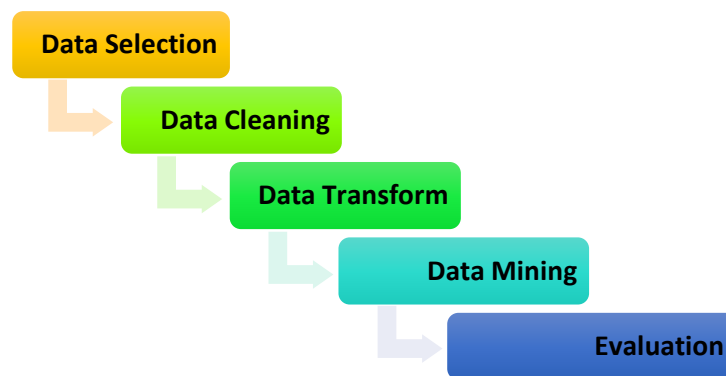
Penggunaan metode k -NN juga diterapkan pada penelitian suatu kinerja satpam menggunakan sistem berbasis web, pada penelitian tersebut metode k -NN digunakan untuk menghitung suatu akurasi dengan data yang digunakan sebanyak 130 data dan 10 atribut dan didapatkan akurasi sebesar 66,67% dengan nilai k yang digunakan adalah 3 [14].

Pada penelitian prediksi suatu produk dari Unilever juga menggunakan metode k -NN. Dari penelitian yang dilakukan menggunakan dataset data penjualan produk Unilever didapatkan nilai akurasi tertinggi yaitu sebesar 86,66% dan akurasi terendah yaitu sebesar 40%. Perhitungan tersebut berdasarkan dari data permintaan konsumen selama kurang lebih tiga tahun [15], sedangkan pada penelitian identifikasi jenis-jenis kaca didapatkan akurasi tertinggi yaitu pada nilai $k=3$ sebesar 64% dengan penggunaan dataset sebanyak 1000 produksi kaca [16]. Penggunaan metode *data mining* k -NN juga digunakan pada analisis sentimen ulasan pada suatu villa, dimana k -NN mendapatkan nilai tertinggi dibandingkan dengan metode *Naïve Bayes* dan metode *Decision Tree* [17]. Selain sentimen

ulasan vila, metode k -NN juga digunakan dalam perhitungan sentimen isu minyak goreng [18].

Metode Penelitian

Dalam proses pengumpulan data, data yang didapatkan berasal dari *website* Kaggle. Pada penelitian ini dilakukan proses klasifikasi pada data susu pasteurisasi dengan menggunakan metode k -NN dengan menggunakan tahapan alur KDD yang tampak pada Gambar 1. KDD atau *Knowledge Discovery in Database* adalah seluruh proses ekstraksi atau identifikasi pada pola, pengetahuan dan informasi potensial dari sekumpulan data besar [19].



Gambar 1. Tahapan KDD

Langkah yang dilakukan pada tahapan KDD seperti yang terlihat pada Gambar 1 dan dapat dijelaskan sebagai berikut.

Data Selection. *Data selection* adalah suatu proses pemilihan data dari sekumpulan data operasional sebelum dilakukan pencarian informasi. Data yang diperoleh berasal dari *Kaggle* sebanyak 1.059 data dengan tujuh atribut reguler dan satu atribut spesial atau *class*. Atribut yang didapatkan adalah (1) pH, yang merupakan atribut yang berisi kadar asam atau basa pada susu pasteurisasi. pH adalah jumlah konsentrasi keasaman dan kebasaan yang dimiliki suatu produk [20]. Dengan kadar yang digunakan pada penelitian ini yaitu kadar pH dari range 3 hingga 9,5; (2) *Temperature*, yang merupakan atribut yang berfungsi untuk mengetahui berapa besaran suhu yang digunakan dalam proses pemanasan susu, dengan range *temperature* antara 34°C hingga 90°C; (3) *Taste*, yang merupakan atribut yang berfungsi untuk mengetahui apakah dalam proses pemanasan susu dapat mengubah rasa dari susu tersebut. Dengan *value* yang digunakan dalam *taste* yaitu nilai 0 untuk rasa tidak enak dan nilai 1 untuk rasa enak; (4) *Odor*, yang merupakan atribut yang berfungsi untuk mengetahui apakah susu tersebut memiliki bau yang buruk atau tidak. Dengan *value* yang digunakan dalam *odor* yaitu 0 untuk berbau tidak sedap dan 1 untuk tidak berbau; (5) *Fat*, yang merupakan atribut yang berfungsi untuk mengetahui apakah dalam susu tersebut mengandung lemak yang tinggi atau rendah. Dengan *value* yang digunakan nilai 0 untuk lemak susu rendah dan nilai 1 untuk lemak susu tinggi; (6) *Turbidity*, yang

merupakan atribut yang berfungsi untuk mengetahui apakah dalam susu tersebut memiliki tingkat kekeruhan yang tinggi atau tidak. Dengan *value* yang digunakan nilai 0 untuk keruh rendah dan nilai 1 untuk keruh tinggi; (7) *Colour*, yang merupakan atribut yang berfungsi untuk mengetahui apakah warna dalam susu tersebut. Dengan menggunakan kode *Red Green Blue* (RGB) sebagai *value*-nya dengan *range* 240 hingga 255; (8) *Grade*, yang merupakan atribut yang digunakan sebagai *attribute special* atau *class* pada penelitian ini. *Attribute grade* dibagi menjadi tiga *class* yaitu: *Low*, *Medium* dan *High*.

Data Cleaning. Pada tahapan *data cleaning* dilakukan untuk membersihkan data, dengan kata lain pada tahapan ini dilakukan untuk mengurangi data yang berduplikat atau *noise* dan data derau atau *missing value* pada suatu data. Dengan menggunakan *tools* dari Microsoft Excel yaitu *Remove Duplicates* untuk menghilangkan data yang berduplikat. Dari hasil pembersihan tersebut didapatkan 288 data yang akan digunakan. Data sejumlah 288 tersebut dibagi menggunakan *tools* dari Rapidminer yaitu *Split Data* menggunakan teknik samplingnya yaitu *Stratified Sampling* dengan rasio pembagiannya adalah 80:20. Hasil dari proses tersebut diperoleh 231 data yang digunakan sebagai *data training* dan 57 data yang digunakan sebagai *data testing*.

Data Transform. Pada tahapan *data transform*, data yang didapatkan akan diubah terlebih dahulu untuk menyesuaikan dengan penggunaan algoritma atau metode yang digunakan, sehingga pola pada data dapat ditemukan dengan cepat [21]. Pada tahapan ini, teknik *transform* yang digunakan adalah *Min-Max Normalization*, dengan rumus normalisasi yang terlihat pada Persamaan (1).

$$z = \frac{x - \min(x)}{[\max(x) - \min(x)]} \quad (1)$$

dimana z adalah hasil normalisasi, x adalah nilai x asli, $\min(x)$ adalah nilai minimal untuk variabel x , dan $\max(x)$ adalah nilai maksimal untuk variabel x .

Data Mining. Pada tahapan *Data Mining* ini, dilakukan untuk pencarian informasi yang bermanfaat dari data tersebut dengan menggunakan metode k -NN. Pada perhitungan metode k -NN memiliki beberapa tahapan, antara lain sebagai berikut.

1. Menentukan nilai k . Dalam tahapan ini, penentuan nilai parameter k ditentukan secara manual. Nilai parameter k yang akan digunakan yaitu 3, 5, dan 7.
2. Menghitung jarak antara *data training* dan *data testing*. Dalam tahapan ini perhitungan jarak antara *data training* dan *data testing* menggunakan rumus *Euclidean Distance* yang terlihat pada Persamaan (2).

$$euc = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (2)$$

dimana p_i adalah *data training*, q_i adalah *data testing*, i adalah variabel data, dan n adalah dimensi data.

3. Mengurutkan data berdasarkan jarak terkecil. Dalam tahapan ini adalah untuk mengurutkan jarak atau hasil dari perhitungan menggunakan *Eulidean Distance*, dimana jarak atau hasil tersebut diurutkan berdasarkan jarak terdekat atau nilai yang paling kecil.
4. Menetapkan *class*. Pada tahapan menetapkan *class*, berfungsi untuk mengetahui *data testing* tersebut termasuk ke dalam *class* mana. Dimana penetapan *class* diambil berdasarkan mayoritas hasil yang ada.

Evaluation. Dalam tahapan *evaluation* berfungsi untuk mengetahui tingkat akurasi pada metode yang digunakan. Pada penelitian ini, metode yang digunakan adalah metode *k*-NN dan pengukuran akurasi metode *k*-NN menggunakan *confusion matrix*. Persamaan (3) adalah rumus akurasi.

$$Akurasi = \frac{TP + TN}{TP + FP + FN + TN} \quad (3)$$

dimana TP adalah *True Positif* atau jumlah data yang diprediksi benar dan kenyataannya benar, TN adalah *True Negatif* atau jumlah data yang diprediksi salah dan kenyataannya salah, FP adalah *False Positif* atau jumlah data yang diprediksi benar dan kenyataannya salah, dan FN adalah *False Negatif* atau jumlah data yang diprediksi salah dan kenyataannya benar.

Hasil dan Pembahasan

Penelitian ini menggunakan dataset susu pasteurisasi, dengan 57 *data testing* yang akan digunakan sebagai data untuk menguji metode *k*-NN. Tabel 1 memperlihatkan *data testing* yang digunakan dan dari data tersebut digunakan untuk mengetahui akurasi dari metode yang digunakan.

Tabel 1 *Data testing* susu pasteurisasi

No	pH	Temperature	Taste	Odor	Fat	Turbidity	Colour
1	6,6	36	0	1	0	1	253
2	6,6	37	1	1	1	1	255
3	6,8	38	0	0	0	0	255
4	7,4	90	1	0	1	1	255
5	6,5	38	1	0	0	0	255
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
53	6,8	55	0	0	0	1	255
54	6,5	50	1	0	1	0	246
55	6,6	40	0	0	0	0	255
56	6,5	40	1	0	0	0	250
57	6,6	43	0	0	1	0	250

Pada bagian ini dibahas tentang perhitungan dan hasil dari penggunaan metode *k*-NN menggunakan dataset susu pasteurisasi. Perhitungan tersebut antara lain sebagai berikut.

1. *Data transform* menggunakan teknik *min-max normalization*

Pada penggunaan teknik *Min-Max Normalization* digunakan pada tahapan *data transform*. *Data transform* dilakukan untuk mengubah data berdasarkan penggunaan model metode yang akan digunakan. Hasil perhitungan menggunakan Persamaan (1).

Perhitungan data 1

$$z = \frac{6,6-3}{9-3} = 0,60, z = \frac{36-34}{90-34} = 0,04, z = \frac{0-0}{1-0} = 0, z = \frac{1-0}{1-0} = 1, z = \frac{0-0}{1-0} = 0, \\ z = \frac{1-0}{1-0} = 1, z = \frac{253-240}{255-240} = 0,87.$$

Perhitungan data 2

$$z = \frac{6,6-3}{9-3} = 0,60, z = \frac{37-34}{90-34} = 0,05, z = \frac{1-0}{1-0} = 1, z = \frac{1-0}{1-0} = 1, z = \frac{1-0}{1-0} = 1, \\ z = \frac{1-0}{1-0} = 1, z = \frac{255-240}{255-240} = 1.$$

Perhitungan data 3

$$z = \frac{6,8-3}{9-3} = 0,63, z = \frac{38-34}{90-34} = 0,07, z = \frac{0-0}{1-0} = 0, z = \frac{0-0}{1-0} = 0, z = \frac{0-0}{1-0} = 0, \\ z = \frac{0-0}{1-0} = 0, z = \frac{255-240}{255-240} = 1.$$

Perhitungan data 4

$$z = \frac{7,4-3}{9-3} = 0,73, z = \frac{90-34}{90-34} = 1, z = \frac{1-0}{1-0} = 1, z = \frac{0-0}{1-0} = 0, z = \frac{1-0}{1-0} = 1, \\ z = \frac{1-0}{1-0} = 1, z = \frac{255-240}{255-240} = 1.$$

Perhitungan data 5

$$z = \frac{6,5-3}{9-3} = 0,58, z = \frac{38-34}{90-34} = 0,07, z = \frac{1-0}{1-0} = 1, z = \frac{0-0}{1-0} = 0, z = \frac{0-0}{1-0} = 0, \\ z = \frac{0-0}{1-0} = 0, z = \frac{255-240}{255-240} = 1.$$

Untuk hasil seluruh perhitungan atribut tampak pada Tabel 2.

Tabel 2 Hasil *transform* menggunakan *min-max normalization data testing*

No	pH	Temperature	Taste	Odor	Fat	Turbidity	Colour
1	0,60	0,04	0,00	1,00	0,00	1,00	0,87
2	0,60	0,05	1,00	1,00	1,00	1,00	1,00
3	0,63	0,07	0,00	0,00	0,00	0,00	1,00
4	0,73	1,00	1,00	0,00	1,00	1,00	1,00
5	0,58	0,07	1,00	0,00	0,00	0,00	1,00
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
53	0,63	0,38	0,00	0,00	0,00	1,00	1,00
54	0,58	0,29	1,00	0,00	1,00	0,00	0,40

55	0,60	0,11	0,00	0,00	0,00	0,00	1,00
56	0,58	0,11	1,00	0,00	0,00	0,00	0,67
57	0,60	0,16	0,00	0,00	1,00	0,00	0,67

2. Data mining menggunakan metode k-NN

Setelah melakukan *data transform* menggunakan teknik *normalization* yaitu *Min-Max Normalization* akan dilakukan perhitungan menggunakan *k-NN* menggunakan rumus *Euclidean Distance* dengan *k* yang digunakan adalah *k=3, k=5, dan k=7*. Perhitungan *Euclidean Distance* menggunakan Persamaan (2), antara lain sebagai berikut:

Perhitungan data 1

$$euc_{(1)} = \sqrt{(0,55-0,6)^2 + (0,02-0,04)^2 + (1-0)^2 + (0-1)^2 + (1-0)^2 + (0-1)^2 + (0,93-0,87)^2} = 2,0$$

$$euc_{(2)} = \sqrt{(0,85-0,6)^2 + (0,64-0,04)^2 + (1-0)^2 + (1-1)^2 + (1-0)^2 + (1-1)^2 + (0,4-0,87)^2} = 1,63$$

$$euc_{(3)} = \sqrt{(1-0,6)^2 + (0-0,04)^2 + (1-0)^2 + (1-1)^2 + (0-0)^2 + (1-1)^2 + (1-0,87)^2} = 1,09$$

$$euc_{(4)} = \sqrt{(0,55-0,6)^2 + (0,05-0,04)^2 + (0-0)^2 + (0-1)^2 + (0-0)^2 + (0-1)^2 + (1-0,87)^2} = 1,42$$

$$euc_{(5)} =$$

$$\sqrt{(0,38-0,6)^2 + (0,20-0,04)^2 + (1-0)^2 + (0-1)^2 + (1-0)^2 + (1-1)^2 + (0,67-0,87)^2} = 1,76$$

$$euc_{(6)} =$$

$$\sqrt{(0,23-0,6)^2 + (0,46-0,04)^2 + (0-0)^2 + (1-1)^2 + (1-0)^2 + (1-1)^2 + (0,67-0,87)^2} = 1,17$$

$$euc_{(7)} = \sqrt{(0,78-0,6)^2 + (0,57-0,04)^2 + (1-0)^2 + (0-1)^2 + (1-0)^2 + (1-1)^2 + (1-0,87)^2} = 1,83$$

$$euc_{(8)} =$$

$$\sqrt{(0,57-0,6)^2 + (0,20-0,04)^2 + (1-0)^2 + (1-1)^2 + (0-0)^2 + (0-1)^2 + (0,47-0,87)^2} = 1,48$$

$$euc_{(9)} =$$

$$\sqrt{(0,57-0,6)^2 + (0,20-0,04)^2 + (1-0)^2 + (1-1)^2 + (1-0)^2 + (0-1)^2 + (0,33-0,87)^2} = 1,82$$

$$euc_{(10)} = \sqrt{(0,4-0,6)^2 + (0,29-0,04)^2 + (0-0)^2 + (1-1)^2 + (1-0)^2 + (1-1)^2 + (1-0,87)^2} = 1,06$$

Perhitungan data 2

$$euc_{(1)} = \sqrt{(0,55-0,6)^2 + (0,02-0,05)^2 + (1-1)^2 + (0-1)^2 + (1-1)^2 + (0-1)^2 + (0,93-1)^2} = 1,42$$

$$euc_{(2)} = \sqrt{(0,85-0,6)^2 + (0,64-0,05)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (0,4-1)^2} = 0,88$$

$$euc_{(3)} = \sqrt{(1-0,6)^2 + (0-0,05)^2 + (1-1)^2 + (1-1)^2 + (0-1)^2 + (1-1)^2 + (1-1)^2} = 1,08$$

$$euc_{(4)} = \sqrt{(0,55-0,6)^2 + (0,05-0,05)^2 + (0-1)^2 + (0-1)^2 + (0-1)^2 + (0-1)^2 + (1-1)^2} = 2,00$$

$$euc_{(5)} = \sqrt{(0,38-0,6)^2 + (0,20-0,05)^2 + (1-1)^2 + (0-1)^2 + (1-1)^2 + (1-1)^2 + (0,67-1)^2} = 1,09$$

$$euc_{(6)} = \sqrt{(0,23-0,6)^2 + (0,46-0,05)^2 + (0-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (0,67-1)^2} = 1,99$$

$$euc_{(7)} = \sqrt{(0,78-0,6)^2 + (0,57-0,05)^2 + (1-1)^2 + (0-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2} = 1,44$$

$$euc_{(8)} = \sqrt{(0,57-0,6)^2 + (0,20-0,05)^2 + (1-1)^2 + (1-1)^2 + (0-1)^2 + (0-1)^2 + (0,47-1)^2} = 1,52$$

$$euc_{(9)} = \sqrt{(0,57-0,6)^2 + (0,20-0,05)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (0-1)^2 + (0,33-1)^2} = 1,21$$

$$euc_{(10)} = \sqrt{(0,4-0,6)^2 + (0,29-0,05)^2 + (0-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2} = 1,05$$

Perhitungan data 3

$$euc_{(1)} = \sqrt{(0,55-0,63)^2 + (0,02-0,07)^2 + (1-0)^2 + (0-0)^2 + (1-0)^2 + (0-0)^2 + (0,93-1)^2} = 1,42$$

$$euc_{(2)} = \sqrt{(0,85-0,63)^2 + (0,64-0,07)^2 + (1-0)^2 + (1-0)^2 + (1-0)^2 + (1-0)^2 + (0,4-1)^2} = 2,18$$

$$euc_{(3)} = \sqrt{(1-0,63)^2 + (0-0,07)^2 + (1-0)^2 + (1-0)^2 + (0-0)^2 + (1-0)^2 + (1-1)^2} = 1,78$$

$$euc_{(4)} = \sqrt{(0,55-0,63)^2 + (0,05-0,07)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (1-1)^2} = 0,08$$

$$euc_{(5)} = \sqrt{(0,38-0,63)^2 + (0,20-0,07)^2 + (1-0)^2 + (0-0)^2 + (1-0)^2 + (1-0)^2 + (0,67-1)^2} = 1,79$$

$$euc_{(6)} = \sqrt{(0,23-0,63)^2 + (0,46-0,07)^2 + (0-0)^2 + (1-0)^2 + (1-0)^2 + (1-0)^2 + (0,67-1)^2} = 1,85$$

$$euc_{(7)} = \sqrt{(0,78-0,63)^2 + (0,57-0,07)^2 + (1-0)^2 + (0-0)^2 + (1-0)^2 + (1-0)^2 + (1-1)^2} = 1,81$$

$$euc_{(8)} = \sqrt{(0,57-0,63)^2 + (0,20-0,07)^2 + (1-0)^2 + (1-0)^2 + (0-0)^2 + (0-0)^2 + (0,47-1)^2} = 1,52$$

$$euc_{(9)} = \sqrt{(0,57-0,63)^2 + (0,20-0,07)^2 + (1-0)^2 + (1-0)^2 + (1-0)^2 + (0-0)^2 + (0,33-1)^2} = 1,86$$

$$euc_{(10)} = \sqrt{(0,4-0,63)^2 + (0,29-0,07)^2 + (0-0)^2 + (1-0)^2 + (1-0)^2 + (1-0)^2 + (1-1)^2} = 1,76$$

Hasil perhitungan tersebut digunakan untuk menentukan *class* dari *data testing* susu pasteurisasi. Hasil *class* dari perhitungan tersebut berdasarkan $k=3$, $k=5$, dan $k=7$, masing-masing ditunjukkan pada Tabel 3, 4, dan 5.

Tabel 3 Hasil *class* menggunakan $k=3$

No	<i>Prec. Class</i>	<i>Real Class</i>
1	<i>High</i>	<i>High</i>
2	<i>High</i>	<i>High</i>
3	<i>Medium</i>	<i>Medium</i>
4	<i>Low</i>	<i>Low</i>
5	<i>Medium</i>	<i>Medium</i>
6	<i>Low</i>	<i>Low</i>
7	<i>Medium</i>	<i>Medium</i>
8	<i>Medium</i>	<i>Medium</i>
9	<i>Low</i>	<i>Low</i>
10	<i>Medium</i>	<i>Medium</i>
...
48	<i>Low</i>	<i>Low</i>
49	<i>Medium</i>	<i>Medium</i>
50	<i>High</i>	<i>High</i>
51	<i>High</i>	<i>High</i>
52	<i>High</i>	<i>High</i>
53	<i>Low</i>	<i>Medium</i>
54	<i>Low</i>	<i>Low</i>
55	<i>Medium</i>	<i>Medium</i>
56	<i>Medium</i>	<i>Medium</i>
57	<i>Medium</i>	<i>Medium</i>

Tabel 4 Hasil *class* menggunakan $k=5$

No	<i>Prec. Class</i>	<i>Real Class</i>
1	<i>High</i>	<i>High</i>
2	<i>High</i>	<i>High</i>
3	<i>Medium</i>	<i>Medium</i>
4	<i>Low</i>	<i>Low</i>
5	<i>Medium</i>	<i>Medium</i>
6	<i>Low</i>	<i>Low</i>

No	Prec. Class	Real Class
7	Medium	Medium
8	Medium	Medium
9	Low	Low
10	Medium	Medium
...
48	Low	High
49	Medium	Medium
50	High	High
51	High	High
52	High	High
53	Low	Medium
54	Low	Low
55	Medium	Medium
56	Medium	Medium
57	Medium	Medium

Tabel 5 Hasil *class* menggunakan $k=7$

No	Prec. Class	Real Class
1	High	High
2	High	High
3	Medium	Medium
4	Low	Low
5	Medium	Medium
6	Low	Low
7	Medium	Medium
8	Medium	Medium
9	Low	Low
10	Medium	Medium
...
48	Low	High
49	Medium	Medium
50	High	High
51	High	High
52	High	High
53	Low	Medium
54	Low	Low
55	Medium	Medium
56	Medium	Medium
57	Medium	Medium

3. Evaluation

Pada tahapan *evaluation* berfungsi untuk mengetahui akurasi suatu metode yang digunakan. Akurasi pada metode k -NN yang digunakan akan memiliki perbedaan dikarenakan nilai k yang digunakan berbeda. Nilai k mempengaruhi perbedaan tingkat akurasi karena nilai K adalah nilai atau

hasil dari *class* terdekat dari *data testing*. Namun, terdapat perbedaan antara prediksi *class* dan *class* asli hasil dari perhitungan. Perbedaan tersebut ditunjukkan pada Tabel 6.

Tabel 6 Perbedaan hasil *class*

No	Predict Class	Real Class		
		k=3	k=5	k=7
1	High	High	High	High
2	High	High	High	High
3	Medium	Medium	Medium	Medium
4	Low	Low	Low	Low
5	Medium	Medium	Medium	Medium
6	Low	Low	Low	Low
7	Medium	Medium	Medium	Medium
8	Medium	Medium	Medium	Medium
9	Low	Low	Low	Low
10	Medium	Medium	Medium	Medium
...
48	Low	Low	High	High
49	Medium	Medium	Medium	Medium
50	High	High	High	High
51	High	High	High	High
52	High	High	High	High
53	Low	Medium	Medium	Medium
54	Low	Low	Low	Low
55	Medium	Medium	Medium	Medium
56	Medium	Medium	Medium	Medium
57	Medium	Medium	Medium	Medium

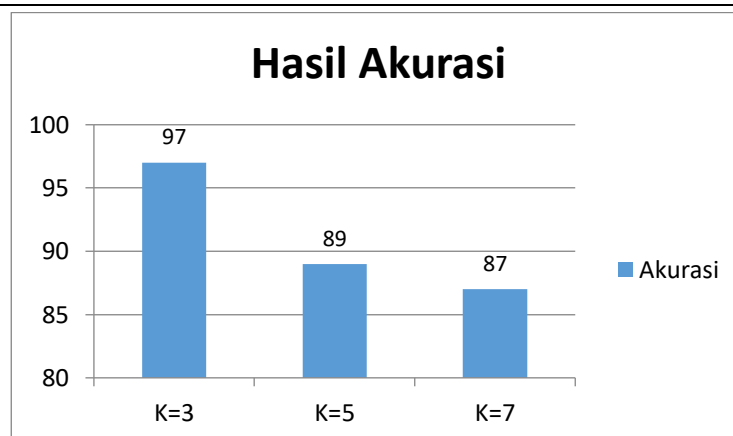
Perhitungan akurasi metode *k*-NN menggunakan Persamaan (3) seperti berikut:

$$\text{Evaluation } k=3, \text{ akurasi} = \frac{16+21+19}{57} = 0,97 = 97\%$$

$$\text{Evaluation } k=5, \text{ akurasi} = \frac{16+15+20}{57} = 0,89 = 89\%$$

$$\text{Evaluation } k=7, \text{ akurasi} = \frac{14+18+14}{57} = 0,87 = 87\%$$

Berdasar perhitungan akurasi dengan beberapa nilai *k* dalam melakukan klasifikasi menggunakan *k*-NN memberikan hasil bahwa nilai akurasi dengan nilai *k*=3 memiliki nilai akurasi paling tinggi, seperti yang ditunjukkan pada Gambar 2.



Gambar 2. Hasil akurasi metode k -NN dengan $k=3, 5$ dan 7

Simpulan

Pada penelitian ini, dapat ditarik kesimpulan bahwa hasil implementasi tahapan KDD menggunakan metode k -NN mampu melakukan klasifikasi *class* mutu susu pasteurisasi dengan baik. Klasifikasi *class* mutu susu pasteurisasi menggunakan dataset susu pasteurisasi sebagai *database* untuk pola pembelajaran pada sistem. Klasifikasi pada penelitian ini menggunakan delapan *attribute*, yaitu pH, *temperature*, *taste*, *odor*, *fat*, *turbidity*, *colour*, dan *grade*. Hasil analisis penggunaan metode k -NN dengan beberapa nilai k menghasilkan hasil akurasi yang berbeda. Penelitian yang dilakukan menunjukkan bahwa nilai $k=3$ memiliki tingkat akurasi tertinggi dengan nilai sebesar 97%, sedangkan nilai $k=5$ memiliki tingkat akurasi sebesar 89%, dan nilai $k=7$ memiliki nilai akurasi sebesar 87%.

Daftar Pustaka

- [1] Z. E. Fitri, L. N. Sahenda, R. S. A. Holili, and D. L. Rukmi, "PERHITUNGAN KOLONI BAKTERI SUSU SEGAR PADA RUANG WARNA YCBCR," vol. 8, no. 2, 2023.
- [2] H. Apriliawan, "IMPLEMENTASI MESIN PASTEURISASI MODERN PEF (PULSED ELECTRIC FIELD) TERHADAP PENGOLAHAN SUSU SAPI PERAH," 2022.
- [3] A. Dany Priyanto, S. Djajati, and R. Yulistiani, "KHADIMUL UMMAH IMPLEMENTASI PULSED ELECTRIC FIELD UNTUK MENINGKATKAN EFEKTIVITAS PASTEURISASI SUSU PADA CV MILKINESIA NUSANTARA DI PONOROGO IMPLEMENTATION OF PULSED ELECTRIC FIELD TO INCREASE THE EFFECTIVENESS OF MILK PASTEURIZATION AT CV MILKINESIA NUSANTARA IN PONOROGO," vol. 4, 2020, doi: 10.21111/ku.v4i1.
- [4] B. Hariono, F. Erawantini, A. Budiprasojo, and T. D. Puspitasari, "Perbedaan nilai gizi susu sapi setelah pasteurisasi non termal dengan HPEF (High Pulsed Electric Field)," *AcTion: Aceh Nutrition Journal*, vol. 6, no. 2, p. 207, Nov. 2021, doi: 10.30867/action.v6i2.531.

-
- [5] F. Ardiansyah, D. Syauqy, and G. E. Setyawan, "Klasifikasi Tingkat Kematangan Susu Kefir dengan Metode K-Nearest Neighbor (KNN) menggunakan Sensor Cahaya dan Sensor Warna," vol. 3, no. 9, pp. 8608–8613, 2019.
- [6] M. H. Abdurrahman, E. Suhartono, and E. Wulandari, "Deteksi Kualitas Kemurnian Susu Sapi Melalui Pengolahan Citra Digital Menggunakan Metode Scale Invariant Feature Transform Dengan Klasifikasi K-nearest Neighbor," *eProceedings of Engineering*, vol. 6, no. 2, pp. 3845–3852, 2019.
- [7] N. Karolina, "Data Mining Pengelompokan Pasien Rawat Inap Peserta BPJS Menggunakan Metode Clustering (Studi Kasus : RSUD.Bangkalan)."
- [8] C. Marshela, N. Ibrahim, and E. Wulandari, "Deteksi Kualitas Kemurnian Susu Sapi Melalui Pengolahan Citra Digital Menggunakan Metode Adaptive Region Growing Dan Klasifikasi Learning Vector Quantization Quality Detection of Cow'S Milk Purity Based on Digital Image Processing Using Adaptive Region Gr," *Techno*, vol. 6, no. 2, pp. 3822–3829, 2019.
- [9] H. Gustiawidi, B. Hidayat, and ..., "Identifikasi Kualitas Kesegaran Susu Sapi Melalui Pengolahan Citra Digital Menggunakan Metode Gray Level Co-occurrence Matrix Dan Local Binary Pattern ...," *eProceedings ...*, vol. 5, no. 3, pp. 4843–4850, 2018.
- [10] D. Cahyanti, A. Rahmayani, and S. Ainy Husniar, "Indonesian Journal of Data and Science Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara," vol. 1, no. 2, pp. 39–43, 2020.
- [11] S. Asmiatun, N. Wakhidah, and A. N. Putri, "Implementing The Augmented Reality Map To Classify The Damaged Roads," in *Engineering, Information and Agricultural Technologi in The Global Digital Revolution*, Jul. 2019.
- [12] I. A. Angreni, S. A. Adisasmita, M. I. Ramli, and S. Hamid, "PENGARUH NILAI K PADA METODE K-NEAREST NEIGHBOR (KNN) TERHADAP TINGKAT AKURASI IDENTIFIKASI KERUSAKAN JALAN," *Rekayasa Sipil*, vol. 7, no. 2, p. 63, Jan. 2019, doi: 10.22441/jrs.2018.v07.i2.01.
- [13] M. Sholihin and M. Ghofar Rohman, "KLASIFIKASI MUTU TELUR BERDASARKAN FITUR WARNA DENGAN MENGGUNAKAN METODE K-NEAREST NEIGHBOR," *Seminar Nasional Sistem Informasi*, vol. 9, 2018.
- [14] M. R. Alghifari and A. P. Wibowo, "Knn 14," *Jurnal Teknologi & Manajemen Informatika*, vol. 5, no. 1, 2019.
- [15] A. Alfani W.P.R., F. Rozi, and F. Sukmana, "Prediksi Penjualan Produk Unilever Menggunakan Metode K-Nearest Neighbor," *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 6, no. 1, pp. 155–160, 2021, doi: 10.29100/jipi.v6i1.1910.

-
- [16] M. M. Baharuddin, H. Azis, and T. Hasanuddin, “Analisis Performa Metode K-Nearest Neighbor Untuk Identifikasi Jenis Kaca,” *ILKOM Jurnal Ilmiah*, vol. 11, no. 3, pp. 269–274, 2019, doi: 10.33096/ilkom.v11i3.489.269-274.
 - [17] N. L. W. S. R. Ginantra, C. P. Yanti, G. D. Prasetya, I. B. G. Sarasvananda, and I. K. A. G. Wiguna, “Analisis Sentimen Ulasan Villa di Ubud Menggunakan Metode Naive Bayes, Decision Tree, dan K-NN,” *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 11, no. 3, pp. 205–215, 2022, doi: 10.23887/janapati.v11i3.49450.
 - [18] C. P. Yanti, N. W. E. Agustini, and ..., “Perbandingan Metode K-NN Dan Metode Random Forest Untuk Analisis Sentimen pada Tweet Isu Minyak Goreng di Indonesia,” *Jurnal Media ...*, vol. 7, no. April, pp. 756–765, 2023, doi: 10.30865/mib.v7i2.5900.
 - [19] F. Ginting, E. Buulolo, and E. R. Siagian, “IMPLEMENTASI ALGORITMA REGRESI LINEAR SEDERHANA DALAM MEMPREDIKSI BESARAN PENDAPATAN DAERAH (STUDI KASUS: DINAS PENDAPATAN KAB. DELI SERDANG),” *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 3, no. 1, Nov. 2019, doi: 10.30865/komik.v3i1.1602.
 - [20] M. Jurnal and Vol, “RANCANG BANGUN ALAT UKUR PH DAN SUHU BERBASIS SHORT MESSAGE SERVICE (SMS) GATEWAY,” vol. 1, no. 1, 2014.
 - [21] L. S. Ummah, Karunia Vio Nita Rusyatul., Sutedjo., Rifadil, Moch. Machmud., Mahendra, “Emitor: jurnal teknik elektro,” *Emitor: Jurnal Teknik Elektro*, vol. 22, no. 2, pp. 141–147, 2022.