**KLASIFIKASI SUSU MENGGUNAKAN METODE *K-NEAREST NEIGHBOR* DAN *NAIVE BAYES***

**Laporan Ujian Akhir Pembelajaran Mesin**

Digunakan Sebagai Syarat Ujian Akhir Semester

Politeknik Negeri Malang

**Oleh:**

**ARDHA NUR AZIZAH NIM. 2041720038**

**RAHMA NURWAKHIDATUL H. NIM. 2041720223**

**ROSIS HUDAYA PUTRA NIM. 2041720053**



**PROGRAM STUDI TEKNIK INFORMATIKA**

**JURUSAN TEKNOLOGI INFORMASI**

**POLITEKNIK NEGERI MALANG**

**2022**

**ABSTRAK**

Susu merupakan salah satu bahan makanan yang paling kompleks, karena di dalam kandungan susu ada banyak kandungan yang dibutuhkan bagi tubuh manusia. Manfaat yang dihasilkan dari susu sapi sangat banyak, karena mengandung protein, vitamin, dan juga mineral yang dibutuhkan dalam tubuh. Oleh karena itu, pentingnya pemantauan untuk menguji kualitas susu sangat diperlukan. Pada penelitian ini berguna untuk membangun alat yang mampu menentukan kualitas susu berdasarkan *pH, temperature, taste, odor, fat, turbidity, colour,* dan *grade*. Untuk hasil klasifikasi menggunakan perhitungan Metode *Naive Bayes.* Pemilihan penggunaan Klasifikasi *Naive Bayes* tersebut karena metode tersebut dapat digunakan untuk mengolah data yang bias serta hasil perhitungan yang akurat. Berdasarkan hasil pengujian, didapatkan akurasi perhitungan *Naive Bayes* sebesar 90%. Sehingga dapat menghasilkan produk olahan susu berkualitas tinggi. Kemudian Metode K-Nearest Neighbor dapat digunakan dalam proses pengklasifikasian dan mendapatkan tingkat akurasi yang cukup tinggi. Berdasarkan hasil pengujian, diperoleh rata – rata nilai akurasi pengujian nilai K sebesar 90%, kemudian rata – rata nilai akurasi pengujian pengaruh jumlah data latih sebesar 90%, Maka dapat disimpulkan bahwa metode *Modified K-Nearest Neighbor* dapat diimplementasikan dan diuji kedalam sistem klasifikasi kualitas susu.

**BAB I**

**PENDAHULUAN**

* 1. **Latar Belakang**

Susu berupa cairan putih yang dihasilkan dari sekresi kelenjar ambing hewan ternak mamalia, yang diperoleh dengan cara pemerahan yang benar (Suryadi, 2017). Susu merupakan minuman yang mengandung banyak zat yang sangat bermanfaat untuk tubuh. Sumber zat kalsium yang tinggi, sehingga dapat dikonsumsi untuk pembentukan tulang dan gigi. Selain itu susu juga mengandung protein, lemak yang kaya akan asam lemak omega-3 dan omega-6, karbohidrat, vitamin dan mineral. Kadar protein susu segar mencapai 3,5% dengan kandungan lemak 3,0 hingga 3,8% bahkan ada yang mencapai 5%. Karena kandungan nutrisinya yang tinggi maka susu segar menjadi salah satu medium yang baik untuk pertumbuhan mikroorganisme, sehingga setelah diperah akan mudah mengalami kerusakan apabila proses pengolahannya tidak dilakukan secara benar (Sari, Lailil Muflikah, & Randy Cahya Wihan, 2018). Kandungan tersebut menjadikan susu sebagai makanan ideal yang  dapat diminum oleh semua kalangan dari anak kecil sampai orang dewasa. Dalam aktivitas kehidupan sehari-hari, susu sering dikonsumsi hampir di setiap pagi hari untuk menu sarapan. Susu sapi juga dimanfaatkan sebagai bahan olahan makanan seperti roti, keju dan yogurt.

Banyaknya gizi yang ada dalam susu sehingga membuat susu sebagai minuman yg sangat penting untuk dikonsumsi, tetapi susu yang kita pilih untuk dikonsumsi memiliki kemungkinan berkurangnya kualitas karena kemurnian yang ada dalam susu murni tersebut dicampur dengan bahan lain untuk meningkatkan volume dan masa berlaku agar tahan lama dan memberikan keuntungan untuk penjual. Pencampuran bahan lain ke dalam susu menyebabkan penurunan kualitas yang menjadi masalah umum dan serius di beberapa negara, air merupakan substansi yang menyebabkan penurunan kualitas susu yang paling umum digunakan untuk meningkatkan volume dari susu tersebut. Terkadang urea juga ditambahkan untuk meningkatkan masa berlaku susu (masa dimana susu tersebut dapat digunakan atau dapat dikonsumsi) (Ahmad & Hidayat, DE, 2017). Selain dapat diminum secara langsung Salah satu faktor penting dalam pengolahan susu yaitu adanya pasokan bahan baku susu yang berkualitas, sehingga diperlukan pengawasan mutu susu sapi agar menghasilkan produk olahan susu yang berkualitas.

Perkembangan teknologi komputer dapat membantu peran seseorang dalam menentukan kualitas susu yang dihasilkan oleh sapi dengan software maupun perangkat lunak yang ada didalam komputer sehingga dapat diselesaikan dengan lebih cepat dan efisien, salah satu caranya dengan pengklasifikasian. Klasifikasi merupakan proses menemukan sekumpulan model atau fungsi yang membedakan data kedalam kelas-kelas tertentu. Tujuannya menggunakan model tersebut dalam menentukan kelas dari suatu objek yang belum diketahui kelasnya.

Dalam penentuan tingkat kualitas susu ini membutuhkan metode  yang tepat agar menghasilkan klasifikasi yang memiliki akurasi tinggi. Berdasarkan beberapa penelitian, metode naive bayes dan k-nearest neighbor adalah metode yang menghasilkan tingkat akurasi yang cukup tinggi dan cukup efektif. Dengan beberapa latar belakang tersebut maka diharapkan dapat mempermudah penentuan kualitas susu dengan metode yang sesuai

**BAB II**

**METODOLOGI**

* 1. **Metode Penelitian**

Pada bab ini berisi metode dan perancangan algoritma yang akan digunakan dalam penelitian serta perhitungan manual metode *naive bayes* dan metode *k-nearest neighbor.*

* + 1. **Pengambilan Data**

Data yang digunakan diperoleh dari situs “Kaggel”. Dataset ini berjumlah 1059 dengan 7 variabel independen yaitu pH, Temperature, Taste, Odor, Fat, Turbidity, and Color. Umumnya Grade atau Kualitas susu tergantung pada parameter ini yang memainkan peran penting dalam analisis prediktif susu. Dari variabel tersebut nantinya akan menghasilkan grade susu yang memiliki 3 klasifikasi yaitu, low, medium, dan high. Jika taste, odor, fat, dan turbidity terpenuhi dengan kondisi optimal maka mereka akan menetapkan 1 jika tidak 0. Suhu dan ph diberikan nilai aktualnya dalam kumpulan data.

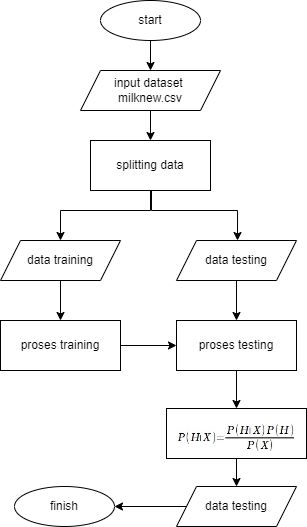
Tabel 2.1 Tabel Dataset

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Temprature | Taste | Odor | Fat | Turbidity | Colour | Grade |
| 37 | 0 | 0 | 0 | 0 | 255 | medium |
| 45 | 1 | 1 | 0 | 0 | 247 | medium |
| 45 | 1 | 1 | 1 | 0 | 245 | medium |
| 50 | 0 | 1 | 1 | 1 | 255 | low |
| 55 | 0 | 1 | 1 | 1 | 255 | low |
| 90 | 1 | 0 | 1 | 1 | 255 | low |
| 45 | 0 | 1 | 1 | 1 | 255 | high |
| 38 | 1 | 0 | 0 | 0 | 255 | medium |
| 38 | 1 | 0 | 1 | 0 | 255 | low |
| 40 | 1 | 1 | 1 | 1 | 255 | low |
| 43 | 1 | 0 | 1 | 1 | 250 | low |
| 40 | 1 | 0 | 1 | 0 | 245 | medium |
| .. | .. | .. | .. | .. | .. | … |
|  |  |  |  |  |  |  |

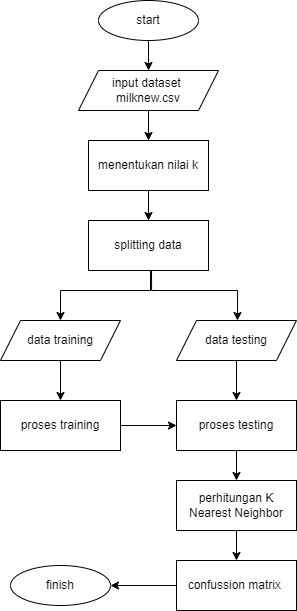
* + 1. **Diagram Alir Pengujian Dataset**

Berikut ini alur dari proses mengklasifikasi susu dengan metode *naïve bayes* dan *k-nearst neighbor :*

1. **Metode *Naïve Bayes***

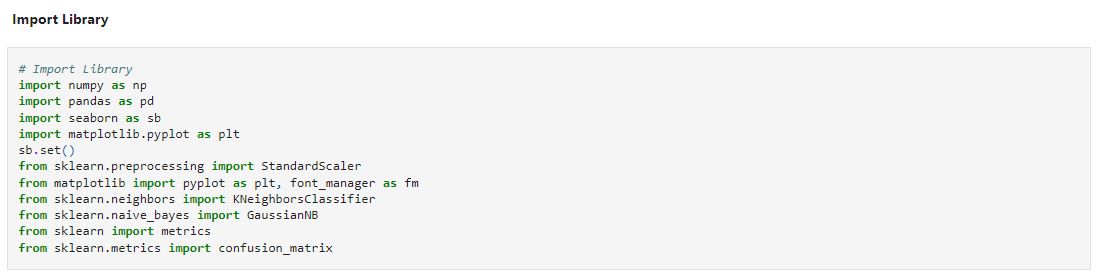


1. **Metode K-Nearst Neighboard**

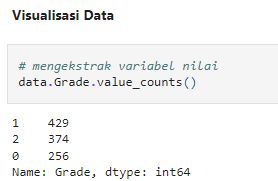


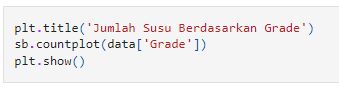
* + 1. **Pengujian Dataset Dengan Phyton**

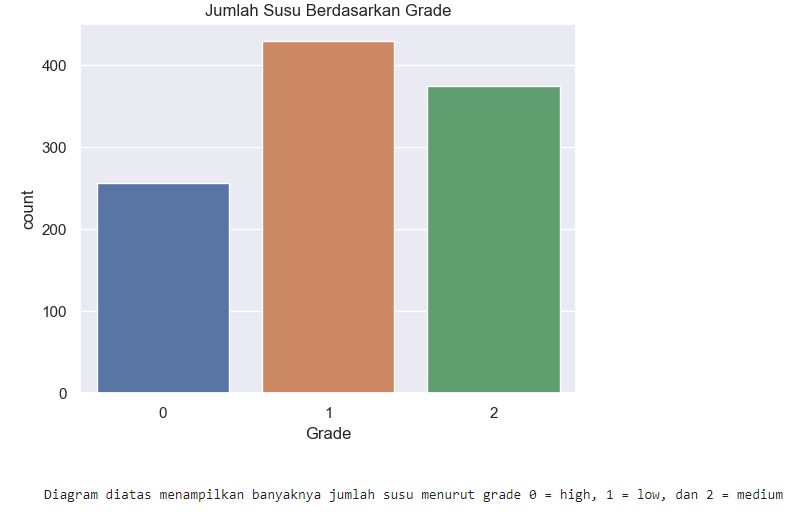
Tahap pertama yang dilakukan adalah dengan import *library* yang akan dibutuhkan. Berikut ini *library* yang akan digunakan :



Terdapat visualisasi yang digunakan untuk memperjelas dataset dari jumlah kualitas susu berdasarkan grade



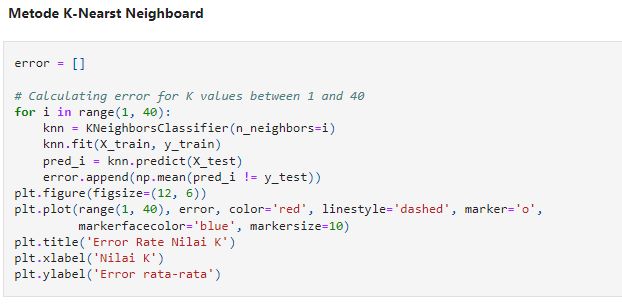


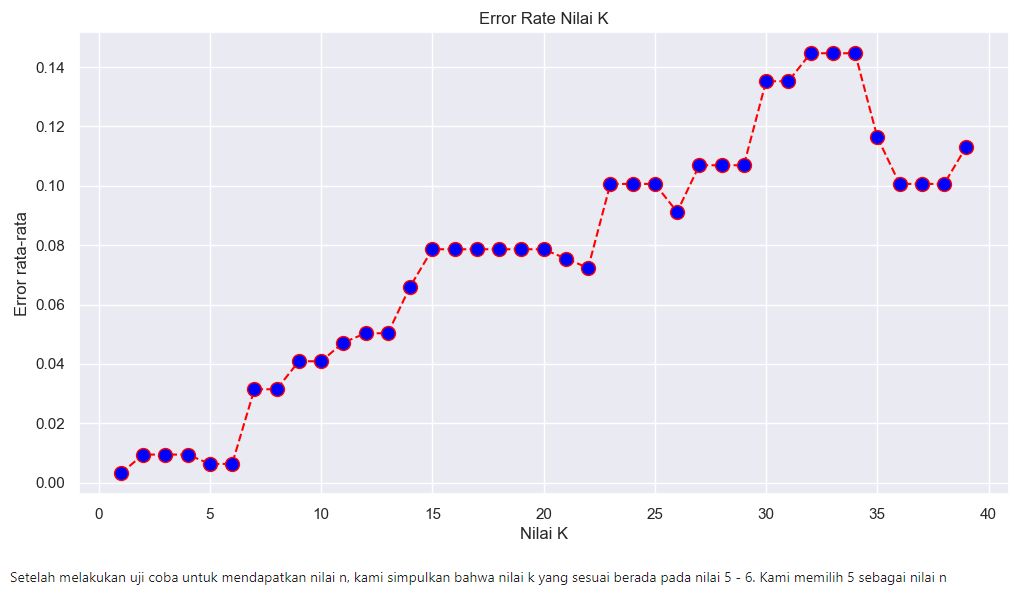


Melakukan splitting data untuk membagi menjadi dua data sebagai data tes dan data train

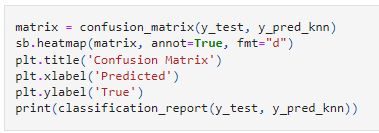


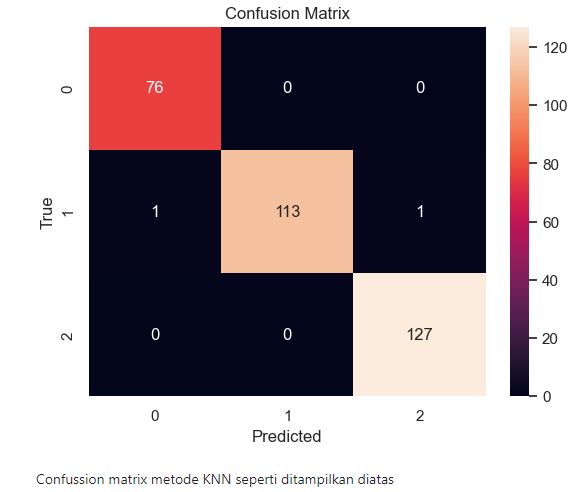
Menentukan nilai k antara hasil 1-40



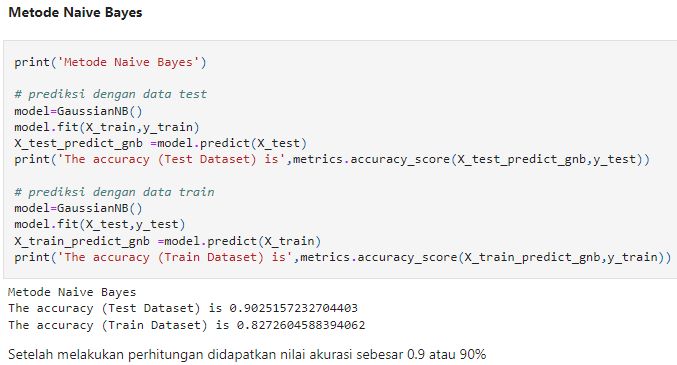


Confusion matrix dapat digunakan untuk mengukur performa dalam permasalahan klasifikasi biner maupun permasalahan klasifikasi multiclass. Berikut ini confusion matrix yang dihasilkan :





Melakukan perhitungan akurasi menggunakan metode naïve bayes dengan menggunakan data x dan y yang didaoat dari hasil splitting data diatas



* + 1. **Perhitungan Manual**

Untuk memperjelas langkah-langkah pada phyton, diperlukan sebuah contoh kasus dengan perhitungan manual. Sebagai contoh kasus, terdapat 100 data yang diperoleh dari dataset seperti pada Tabel 2.2 berikut ini :

Tabel 2.2 Contoh Data Training

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Temprature | Taste | Odor | Fat | Turbidity | Colour | Grade |
| 37 | 0 | 0 | 0 | 0 | 255 | medium |
| 45 | 1 | 1 | 0 | 0 | 247 | medium |
| 45 | 1 | 1 | 1 | 0 | 245 | medium |
| 50 | 0 | 1 | 1 | 1 | 255 | low |
| 55 | 0 | 1 | 1 | 1 | 255 | low |
| 90 | 1 | 0 | 1 | 1 | 255 | low |
| 45 | 0 | 1 | 1 | 1 | 255 | high |
| 38 | 1 | 0 | 0 | 0 | 255 | medium |
| 38 | 1 | 0 | 1 | 0 | 255 | low |
| 40 | 1 | 1 | 1 | 1 | 255 | low |
| 43 | 1 | 0 | 1 | 1 | 250 | low |
| 40 | 1 | 0 | 1 | 0 | 245 | medium |
| 45 | 0 | 1 | 1 | 1 | 250 | high |
| 36 | 0 | 0 | 1 | 0 | 255 | medium |
| 38 | 0 | 1 | 1 | 1 | 255 | low |
| 45 | 1 | 1 | 1 | 1 | 245 | high |
| 35 | 1 | 0 | 1 | 0 | 246 | medium |
| 36 | 0 | 1 | 1 | 0 | 253 | high |
| 70 | 0 | 0 | 0 | 0 | 246 | low |
| 34 | 0 | 0 | 0 | 1 | 240 | medium |
| 37 | 0 | 0 | 0 | 0 | 245 | medium |
| 37 | 1 | 0 | 1 | 0 | 255 | high |
| 45 | 0 | 1 | 0 | 0 | 240 | medium |
| 40 | 0 | 1 | 0 | 1 | 250 | low |
| 42 | 1 | 1 | 1 | 1 | 255 | high |
| 41 | 1 | 0 | 0 | 0 | 247 | medium |
| 50 | 1 | 1 | 1 | 0 | 245 | low |
| 45 | 0 | 1 | 1 | 1 | 255 | high |
| 55 | 0 | 1 | 0 | 0 | 255 | low |
| 65 | 0 | 0 | 0 | 0 | 255 | low |
| 41 | 0 | 0 | 0 | 0 | 255 | medium |
| 38 | 1 | 1 | 1 | 1 | 255 | high |
| .. | .. | .. | .. | .. | .. | .. |

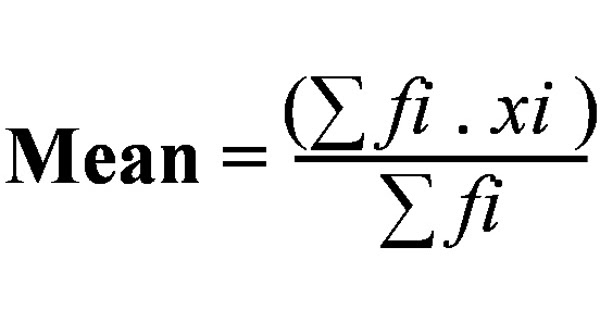
Diketahui pula data testing seperti pada Tabel 2.3 dengan parameter nilai k yang digunakan adalah .

Tabel 2.3 Contoh Kasus Data Testing

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Temprature | Taste | Odor | Fat | Turbidity | Colour | Grade |
| 45 | 1 | 0 | 1 | 1 | 250 | **?** |

1. **Metode *Naïve Bayes***

Langkah pertama yang dilakukan dengan mencari nilai mean di masih-masing variabel. Dengan menggunakan rumus berikut ini :



Gambar 2.1 Rumus Mean

Berikut ini hasil dari perhitungan mean berdasarkan masing-masing variabel :

Tabel 2.4 Hasil Perhitungan Mean

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Grade | Temprature | Taste | Odor | Fat | Turbidity | Colour |
| Low | 44,03370787 | 0,5056179775 | 0,4606741573 | 0,6179775281 | 0,4943820225 | 251,6179775 |
| Medium | 44,04444444 | 0,5 | 0,4666666667 | 0,6 | 0,4666666667 | 251,5111111 |
| High | 43,28235294 | 0,5058823529 | 0,4588235294 | 0,6 | 0,4705882353 | 251,5176471 |

Setelah melakukan perhitungan mean, lalu menghitung standard devisiasi. Perhitungan tersebut juga dilakukan masing-masing variabel.

Tabel 2.5 Hasil Perhitungan Standard Devisiasi

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Grade | Temprature | Taste | Odor | Fat | Turbidity | Colour |
| Low | 10,62057515 | 0,4999684373 | 0,4984510789 | 0,4858819845 | 0,4999684373 | 4,550603412 |
| Medium | 10,56189283 | 0,5 | 0,4988876516 | 0,4898979486 | 0,4988876516 | 4,600107352 |
| High | 9,530893131 | 0,4999653967 | 0,4983016138 | 0,4898979486 | 0,4991341985 | 4,608462203 |

Langkah ketiga menghitung probabilitas berdasarkan grade kualitas susu

Tabel 2.6 Hasil Perhitungan Probabilitas

|  |  |
| --- | --- |
| Grade | Nilai |
| Low | 0,402173913 |
| Medium | 0,3695652174 |
| High | 0,2282608696 |

Hasil probabilitas di atas jika dijumlahkan akan menghasilkan nilai 1. Langkah selanjutnya menghitung data tes yang telah ditentukan sebelumnya.

Tabel 2.7 Perhitungan Data Tes

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Temprature | Taste | Odor | Fat | Turbidity | Colour | Grade |
|  | 45 | 1 | 0 | 1 | 1 | 250 | ? |
| Low | 0,122 | 0,346 | 0,369 | 0,420 | 0,338 | 0,176 | 0,0001563284092 |
| Medium | 0,122 | 0,342 | 0,365 | 0,409 | 0,319 | 0,176 | 0,0001296352359 |
| High | 0,127 | 0,346 | 0,369 | 0,409 | 0,322 | 0,176 | 0,00008609381331 |

Dengan hasil perhitungan grade, jika di amati maka nilai maksimum berada pada grade low. Sehingga dapat disimpulkan bahwa data tes tersebut memiliki grade low.

1. **Metode K-Nearest Neighbor**

Langkah pertama yang dilakukan dalam perhitungan manual menggunakan Metode K-Nearest Neighbor yaitu, menentukan jarak atau *Euclidean Distance* dengan rumus berikut ini :



Gambar 2.2 Rumus Jarak

Tabel 2.8 Perhitungan Jarak

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Temprature | Taste | Odor | Fat | Turbidity | Colour | Grade | Jarak |
| 37 | 0 | 0 | 0 | 0 | 255 | medium | 9,591663047 |
| 45 | 1 | 1 | 0 | 0 | 247 | medium | 251,0697114 |
| 45 | 1 | 1 | 1 | 0 | 245 | medium | 249,1043958 |
| 50 | 0 | 1 | 1 | 1 | 255 | low | 259,8615016 |
| 55 | 0 | 1 | 1 | 1 | 255 | low | 260,8696993 |
| 90 | 1 | 0 | 1 | 1 | 255 | low | 270,4218926 |
| 45 | 0 | 1 | 1 | 1 | 255 | high | 258,9459403 |
| 38 | 1 | 0 | 0 | 0 | 255 | medium | 257,8177651 |
| 38 | 1 | 0 | 1 | 0 | 255 | low | 257,8197044 |
| 40 | 1 | 1 | 1 | 1 | 255 | low | 258,1259383 |
| 43 | 1 | 0 | 1 | 1 | 250 | low | 253,6769599 |
| 40 | 1 | 0 | 1 | 0 | 245 | medium | 248,24786 |
| 45 | 0 | 1 | 1 | 1 | 250 | high | 254,0236209 |
| 36 | 0 | 0 | 1 | 0 | 255 | medium | 257,5305807 |
| 38 | 0 | 1 | 1 | 1 | 255 | low | 257,8216438 |
| 45 | 1 | 1 | 1 | 1 | 245 | high | 249,106403 |
| 35 | 1 | 0 | 1 | 0 | 246 | medium | 248,4813876 |
| 36 | 0 | 1 | 1 | 0 | 253 | high | 255,552343 |
| .. | . | . | . | . | . | . | .. |

Setelah perhitungan jarak selesai, selanjutnya menghitung nilai K

Tabel 2.9 Perhitungan K

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Temprature | Taste | Odor | Fat | Turbidity | Colour | Grade | K1 | K2 | K3 | K4 | K5 | K17 |
| 37 | 0 | 0 | 0 | 0 | 255 | medium | medium | medium | medium | medium | medium | medium |
| 45 | 1 | 1 | 0 | 0 | 247 | medium |  |  |  |  |  |  |
| 45 | 1 | 1 | 1 | 0 | 245 | medium |  |  |  |  |  | medium |
| 50 | 0 | 1 | 1 | 1 | 255 | low |  |  |  |  |  |  |
| 55 | 0 | 1 | 1 | 1 | 255 | low |  |  |  |  |  |  |
| 90 | 1 | 0 | 1 | 1 | 255 | low |  |  |  |  |  |  |
| 45 | 0 | 1 | 1 | 1 | 255 | high |  |  |  |  |  |  |
| 38 | 1 | 0 | 0 | 0 | 255 | medium |  |  |  |  |  |  |
| 38 | 1 | 0 | 1 | 0 | 255 | low |  |  |  |  |  |  |
| 40 | 1 | 1 | 1 | 1 | 255 | low |  |  |  |  |  |  |
| 43 | 1 | 0 | 1 | 1 | 250 | low |  |  |  |  |  |  |
| 40 | 1 | 0 | 1 | 0 | 245 | medium |  |  |  |  |  | medium |
| 45 | 0 | 1 | 1 | 1 | 250 | high |  |  |  |  |  |  |
| 36 | 0 | 0 | 1 | 0 | 255 | medium |  |  |  |  |  |  |
| 38 | 0 | 1 | 1 | 1 | 255 | low |  |  |  |  |  |  |
| 45 | 1 | 1 | 1 | 1 | 245 | high |  |  |  |  |  | high |
| 35 | 1 | 0 | 1 | 0 | 246 | medium |  |  |  |  |  | medium |
| 36 | 0 | 1 | 1 | 0 | 253 | high |  |  |  |  |  |  |
| 70 | 0 | 0 | 0 | 0 | 246 | low |  |  |  |  |  |  |
| 34 | 0 | 0 | 0 | 1 | 240 | medium |  | medium | medium | medium | medium | medium |
| 37 | 0 | 0 | 0 | 0 | 245 | medium |  |  |  |  |  | medium |
| 37 | 1 | 0 | 1 | 0 | 255 | high |  |  |  |  |  |  |
| 45 | 0 | 1 | 0 | 0 | 240 | medium |  |  |  | medium | medium | medium |

**BAB III**

**HASIL**

Hasil klasifikasi dari dataset klasifikasi susu menggunakan Metode *Naive Bayes* dan *K-Nearest Neighbor* menghasilkan akurasi yang lumayan baik yakni masing masing senilai 90%. Performa klasifikasi dari hasil pengujian kepada 10% dataset menghasilkan presentase sebesar 100%. Akurasi klasifikasi semakin baik jika data latih terus ditambah. Hasil dari penelitian ini juga menunjukkan bahwa algoritma N*aive Bayes* dan *K-Nearest Neighbor* handal dalam menangani kesalahan inputan dengan memperbaharui bobot.

**BAB IV**

**KESIMPULAN DAN SARAN**

* 1. **Kesimpulan**

Berdasarkan perancangan dan pengujian dari langkah-langkah yang telah ditentukan, maka dapat disimpulkan :

1. Algoritma *k-nearest neighbor* dan *naïve bayes* dapat diterapkan dalam klasifikasi susu. Dengan menggunakan data yang telah diperoleh dari situs “Kaggle”. Langkah selanjutnya adalah :
2. Untuk Metode *K-Nearest Neighbor*, langkah dilanjutkan dengan perhitungan jarak, di urutkan berdasarkan jarak terkecil, kemudian diambil data terdekat sejumlah nilai k.
3. Untuk Metode *Naïve Bayes*, langkah dilanjutkan dengan perhitungan mean, standard devisiasi, lalu probabilitas.
4. Dalam penelitian ini kedua metode memiliki kinerja yang sangat baik, dimana akurasi untuk Metode *K-Nearest Neighbor* sebesar 0,99 sedangkan Metode *Naïve Bayes* sebesar 0,90

# **DAFTAR PUSTAKA**

Ahmad, H. M., & Dr. Ir.Bambang Hidayat, DE. (2017). IDENTIFIKASI DAN KLASIFIKASI KEMURNIAN SUSU SAPI BERDASARKAN PEMROSESAN SINYAL VIDEO MENGGUNAKAN METODE GABOR WAVELET DAN SUPPORT VECTOR MACHINE. 3649.

Sari, P., Lailil Muflikah, & Randy Cahya Wihan. (2018). Klasifikasi Kualitas Susu Sapi Menggunakan Metode Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1204-1211.

Suryadi. (2017). Perbandingan Metode Naive Bayes dan K-Nearest Neighbor Untuk Klasifikasi Mutu Susu Sapi.