

## BAB II

### LANDASAN TEORI

#### 2.1 Tinjauan Pustaka

Terdapat beberapa penelitian yang telah dilakukan dalam penerapan citra digital dalam hal ini menggunakan metode *K-Nearest Neighbor* dan *Naïve Bayes Classifier*. Sehingga, dalam penelitian ini diperlukannya tinjauan Pustaka sebagai alat dalam penerapan metode ini, untuk menghindari pembuatan ulang, mengidentifikasi kesenjangan, mengetahui penelitian sebelumnya. Perbedaan yang terdapat pada penelitian ini dengan penelitian terdahulu adalah peneliti menggunakan data motif Batik Lampung dan menggunakan metode *K-Nearest Neighbor* dan *Naïve Bayes Classifier* dalam melakukan klasifikasi batik Lampung motif siger, dan motif sembagi. Hal ini bertujuan untuk mengetahui identifikasi motif batik dan nilai akurasi dari metode yang digunakan.

*Tabel 2. 1* Daftar Literatur

Peneliti	Judul	Variabel Terkait	Metode	Perbedaan Penelitian
Budi Soepriyanto (2021)	Comparative Anlysis of K-NN and Naïve Bayes Methods to Predict Stock Prices	Prediksi Saham, Stok emas	KNN Naïve Bayes	▪ Objek yang diteliti adalah prediksi saham

Budi Pangestu (2021)	Compatibility Of Selection Of Student Departments Using K-Nearest Neighbor And Naïve Bayes Classifier In Informatics Private Vocation School, Serang City	Siswa Pemilihan Jurusan SMKS Informatika	KNN NBC	▪ Objek yang diteliti
Rahmat Robi Waliyansyah, Citra Fitriyah (2019)	Perbandingan Akurasi Klasifikasi Citra Kayu Jati Menggunakan Metode Naïve Bayes dan K-Nearest Neighbor	Kayu Jati	Ekstraksi GLCM K-NN NBC	▪ Objek yang diteliti

Riri Nada Devita, Heru Wahyu Herwanto, Aji Prasetya Wibawa (2017)	Perbandingan Kinerja Metode Naïve Bayes dan K- Nearest Neighbor Untuk Klasifikasi Artikel Berbahasa Indonesia	Dokumen	Text preprocessing KNN NBC	<ul style="list-style-type: none"> <li>▪ Objek yang dipilih adalah berupa dokumen</li> </ul>
Mikael Markus Maga, Arini Aha Pekuwali (2021)	Comparison Of Naïve Bayes Classifier Methods With K-Nearest Neighbor In Classifying Apples	Buah Apel	Ekstraksi GLCM KNN NBC	<ul style="list-style-type: none"> <li>▪ Objek yang digunakan</li> </ul>
Calvin Kurniawan, Hafiz Irsyad (2019)	Perbandingan Metode K- Nearest Neighbor dan	Mata Gender	Ekstraksi HSV, dan HOG	<ul style="list-style-type: none"> <li>▪ Objek yang digunakan.</li> <li>▪ Memakai metode</li> </ul>

	Naïve Bayes			ekstraksi
	Untuk			HSV dan
	Klasifikasi			HOG
	Gender			
	Berdasarkan			
	Mata			

### 2.1.1 Tinjauan Literatur 1

Pada literatur 1 ini meneliti tentang Comparative Analysis of K-NN and Naïve Bayes Methods to Predict Stock Prices dalam penelitian ini penulis bertujuan untuk menganalisis dan memberikan informasi prediksi harga saham menggunakan metode K-Nearest Neighbor dan Naïve Bayes. Jual beli saham merupakan transaksi yang banyak dilakukan pada saat ini, untuk melakukan jual beli saham diperlukan kemampuan atau pengetahuan agar pembelian dan penjualan tersebut menguntungkan. Pada penelitian ini memprediksi data harga saham yang digunakan selama satu bulan dalam level menit sebanyak 39.065 data. Pengujian dilakukan menggunakan aplikasi WEKA, dengan menggunakan 4 skenario, setiap pengujian dilakukan menggunakan 10 *Fold Cross Validation*. Pengujian pertama menggunakan metode KNN pada nilai  $K = 1$  menghasilkan akurasi 62,8%. Pada pengujian kedua menggunakan metode KNN pada nilai  $K = 3$  menghasilkan akurasi sebanyak 65,9% . Pada pengujian ketiga menggunakan metode KNN pada  $K = 5$  menghasilkan akurasi sebanyak 67,25%. Pada

pengujian keempat menggunakan metode naïve bayes memperoleh hasil sebanyak 69,38%. Dapat disimpulkan bahwa metode naïve bayes lebih tinggi persentase prediksinya. (Selatan, 2021)

### **2.1.2 Tinjauan Literatur 2**

Pada literatur 2 ini meneliti tentang Compatibility Of Selection Of Student Departments Using K-Nearest Neighbor And Naïve Bayes Classifier In Informatics Private Vocation School, Serang City dalam penelitian ini tujuan penulis untuk membantu siswa dalam membantu menyesuaikan jurusan yaitu dengan memperhatikan nilai akademik semester 1 dan 2. Penelitian ini menggunakan 997 data yang diambil dari tahun 2014-2019, jurusan yang diminati yaitu : akutansi, administrasi perkantoran, rekayasa perangkat lunak dan computer, rekaya jaringan. Metologi penelitian yang pertama memakai teknis analisis yang digunakan untuk penambangan data yang membandingkan metode naïve bayes dan k-Nearest Neighbor, yang kedua membuat rancangan sistem , menentukan kesesuaian department dalam pemilihan jurusan. Proses yang akan dirancang adalah impor data dari excel yang sudah diproses dan disimpan dalam format .csv atau .xls. Data yang telah diimpor diperiksa untuk kelayakan, dan diproses oleh sistem. Selanjutnya melakukan perbandingan data sesuai kebutuhan, hasil perbandingannya berupa deskripsi sesuai atau tidak sesuai. Didapat kesimpulan bahwa pada penelitian ini menghasilkan akurasi 99%

untuk metode *K-Nearest Neighbor*, dan 98% untuk metode *Naïve Bayes*. (Pangestu, 2021)

### 2.1.3 Tinjauan Literatur 3

Pada literatur 3 ini membahas tentang Perbandingan Akurasi Klasifikasi Citra Kayu Jati Menggunakan Metode *Naïve Bayes* dan *K-Nearest Neighbor*, dalam penelitian ini penulis melakukan klasifikasi pada citra kayu jati yaitu semarangan, blora, dan Sulawesi. Tujuan dari penelitian ini adalah menganalisis suatu tekstur kayu jati agar bisa diklasifikasikan ke dalam kelompok-kelompok tertentu. Pengelompokan jenis kayu jati menggunakan beberapa parameter yaitu tekstur, berat, warna, dan lain sebagainya. Pada penelitian data yang digunakan sebanyak 50 citra, data yang diuji 40 dan data latih sebanyak 10 citra. Penelitian ini menggunakan ekstraksi GLCM yaitu kontras, korelasi, energi, dan *homogeneity*. Hasil yang didapat dari penelitian ini metode K-NN menghasilkan akurasi sebanyak 76% sedangkan memakai metode *naïve bayes* memperoleh akurasi sebanyak 82,7%. Dari penelitian ini memperoleh kesimpulan bahwa metode *Naïve Bayes* lebih baik dari metode *K-Nearest Neighbor*. (Robi et al., n.d.)

### 2.1.4 Tinjauan Literatur 4

Pada literatur 4 ini membahas tentang Perbandingan Kinerja Metode *Naïve Bayes* dan *K-Nearest Neighbor* Untuk Klasifikasi Artikel Berbahasa Indonesia untuk klasifikasi dokumen berbahasa Indonesia. Metode dengan kinerja yang baik dapat membantu

memudahkan mahasiswa dalam memilih tema jurnal yang sesuai untuk artikel yang dimilikinya. Dengan adanya sistem tersebut peluang diterimanya suatu jurnal akan semakin besar sehingga mampu untuk mendapatkan publikasi. Pada penelitian ini metode naïve bayes akan menggunakan dokumen yang belum diketahui kategorinya, sehingga akan mencari kata pada data uji yang sesuai dengan data latih. Pada proses pelatihan metode k-NN, dokumen dikelompokkan secara manual, dan akan melalui tahapan preprocessing yang akan menghasilkan bobot untuk setiap kata yang di semua data latih. Tahapan preprocessing yaitu : tokenizing adalah tahapan pemfilteran membuang karakter tertentu, seperti tanda baca, tahap selanjutnya case folding yang mengubah semua katakte huruf menjadi kapital. Dari hasil penelitain menghasilkan kesimpulan bahwa dari 40 data uji metode Naïve Bayes mampu mengklasifikasikan artikel jurnal Bahasa Indonesia sebanyak 28 dokumen atau memperoleh hasil akurasi sebanyak 70%. Sedangkan untuk metode K-Nearest Neighbor dari 40 data uji dapat mengklasifikasikan sebanyak 16 dokumen atau akurasi sebesar 40%. (Devita et al., 2018)

#### **2.1.5 Tinjauan Literatur 5**

Pada literatur 5 ini membahas tentang *Comparison Of Naïve Bayes Classifier Methods With K-Nearest Neighbor In Classifying Apples*. Tujuan dalam penelitian ini adalah membandingkan tingkat akurasi dari metode NBC dan KNN yang memiliki perhitungan

klasifikasi yang berbeda namun menggunakan metode ekstraksi yang sama yaitu ekstraksi GLCM yaitu contrast, energy, dan homogeneity. Data yang digunakan sebanyak 150 foto dari 3 jenis buah apel menggunakan format .jpg. Jumlah data yang digunakan sebanyak 150 data citra yang dibagi menjadi 6 folder untuk data latih dan data uji. Pengujian yang memakai metode Naïve Bayes menghasilkan akurasi sebanyak 100%, sedangkan pengujian yang dilakukan menggunakan metode KNN menghasilkan akurasi sebesar 88,33%. Dapat disimpulkan dalam penelitian ini metode Naïve Bayes lebih akurat dalam mengklasifikasi buah apel. (Maga et al., n.d.)

#### **2.1.6 Tinjauan Literatur 6**

Pada literatur 6 ini membahas tentang Perbandingan Metode K-Nearest Neighbor dan Naïve Bayes Untuk Klasifikasi Gender Berdasarkan Mata dalam penelitian ini penulis melakukan pengklasifikasian gender berdasarkan mata. Penelitian ini bertujuan untuk mengklasifikasi gender berdasarkan mata menggunakan metode K-NN dan *Naïve Bayes*. Data yang digunakan sebanyak 10.541 data uji dan 4.518 data latih. Ekstraksi yang digunakan adalah *Histogram of Oriented Gradient* (HOG) metode untuk mendeteksi objek, dan *Hue Saturation Value* (HSV) untuk ukuran dari jenis warna yang ada, artinya warna menjadi lebih berwarna. Pengujian dilakukan sebanyak 4 kali, pada pengujian pertama menggunakan metode K-NN dengan ekstraksi HSV dan HOG nilai  $k = 3$  tanpa



cropping menghasilkan akurasi sebesar 60,63%. Pengujian kedua menggunakan metode K-NN dengan ekstraksi fitur HSV dan HOG nilai  $k=3$  dengan cropping menghasilkan akurasi sebesar 68,81%. Pengujian ketiga menggunakan metode naïve bayes dengan ekstraksi HSV dan HOG tanpa cropping menghasilkan akurasi sebesar 56,44%. Pada pengujian keempat menggunakan metode naïve bayes dengan ekstraksi fitur HSV dan HOG dengan *cropping* menghasilkan akurasi sebesar 62,58%. Dapat disimpulkan metode K-NN lebih baik daripada metode *Naïve Bayes* untuk klasifikasi gender berdasarkan mata. (Kurniawan & Irsyad, 2022)

## **2.2 Batik Lampung**

Batik sebagai warisan budaya Indonesia tidak bisa dipungkiri memiliki popularitas tersendiri di mata dunia. Batik Lampung mulai berkembang pada tahun 1970-an dan dipelopori oleh Andrean Sangaji (seorang budayawan Lampung). Batik di daerah Lampung adalah salah satu pakaian yang memiliki corak indah dengan memakai motif yang sudah ada sejak zaman dahulu. Motif Lampung memiliki keunikan tersendiri yang sangat berbeda dengan motif daerah lain di Indonesia. (Amaris Trixie, n.d.-b)

*Tabel 2. 2 Motif Batik Lampung*



### 2.2.1 Batik Motif Sembagi

Batik sembagi adalah batik asli Lampung yang memiliki ciri khas yakni terdapat ornament geometris, bunga kaca piring, sependudung, dan lain-lain. Ciri khas batik Lampung yakni terdapat gambar bunga kaca piring, sependudung, dan lain-lain. Motif-motif ini merupakan motif yang diadopsi oleh masyarakat Lampung.

### 2.2.2 Batik Motif Siger

Motif batik siger Lampung sangat populer di masyarakat Lampung. Batik motif ini sangat mencerminkan ciri khas Lampung yaitu Siger.

## 2.3 Citra

Citra atau dikenal dengan istilah gambar dapat diartikan sebagai fungsi dua dimensi dari intensitas cahaya yang dinyatakan dengan  $f(x,y)$ , dimana nilai atau amplitude  $f$  dalam koordinat spasial  $(x,y)$  menyatakan intensitas (kecerahan) dari gambar pada gambar. Menurut kamus besar

Webster, citra adalah representasi, kemiripan atau tiruan dari suatu objek atau benda. Sebuah gambar dinyatakan sebagai fungsi konstan dari intensitas cahaya dalam bidang dua dimensi. Bayangan tampak adalah cahaya yang dipantulkan dari suatu benda. Sebuah sumber cahaya menerangi suatu objek, objek tersebut memantulkan Kembali beberapa berkas cahaya dan cahaya yang dipantulkan ditangkap oleh perangkat optik.

Citra digital adalah citra  $f(x,y)$  yang telah didigitalisasi, baik pada daerah koordinat maupun tingkat kecerahannya. Nilai  $f$  pada koordinat  $(x,y)$  menunjukkan kecerahan atau tingkat keabuan citra pada titik tersebut. Dengan kata lain hanya gambar digital yang telah disimpan atau diubah ke format digital.(Ginantra, 2016)

## **2.4 Pengolahan Citra Digital**

Citra Digital adalah disiplin ilmu yang mempelajari teknik dalam pengolahan citra, citra yang dimaksud adalah gambar diam (foto) atau gambar bergerak (seperti merekam video). Sedangkan arti dari digital adalah pengolahan citra/gambar yang menggunakan komputer secara digital. RGB adalah artinya Red-Green-Blue, adalah tiga warna dasar biasanya digunakan sebagai referensi warna selain RGB dasar, kami mendapatkan ubah warna menjadi kode bilangan yang membentuk warna adalah tampaknya universal. Komputer sudah mengemas informasi warna ke dalam model warna yang sama untuk membuat Pemrosesan warna RGB dimungkinkan dengan mudah. (Ratna, 2020)

## 2.5 Preprocessing

Proses awal dilakukannya perbaikan suatu citra untuk menghilangkan *noise*. Seperti yang dikatan oleh Bahri dan Maliki, preprocessing merupakan suatu proses untuk mengilangkan bagian-bagain yang tidak diperlukan pada gambar input untuk proses selanjutnya.

### 2.5.1 Resize

Resize adalah merubah ukuran besarnya citra kedalam pixel. Perubahan bisa mengecilkan ukuran dari file aslinya atau lebih besar dari ukuran file aslinya. Ukuran berubah menjadi kecil atau besar dilakukan secara proposional baik pada Panjang maupun lebar citra.

### 2.5.2 Grayscale

Grayscale adalah matrisk data yang setiap nilainya mewakili intensitas setiap piksel berkisar antara 0 sampai 255. Setiap piksel membutuhkan 8 bit memori. Untuk memudahkan model citra. Warna yang dihasilkan bukan berasal dari warna hitam putih melainkan mencari nilai rata-rata dari R, G, B, sehingga membuat warna keabuan. (Solihin et al., n.d.)

### 2.5.3 Thresholding

Thresholding citra adalah suatu metode yang digunakan untuk memisahkan antara objek dan backgroundnya. Thresholding adalah teknik yang sederhana tapi efektif untuk segmentasi citra. Proses thresholding sering disebut dengan proses binerisasi. Cara

kerjanya adalah mengkonversi ruang warna citra RGB menjadi *grayscale*. Melakukan operasi komplemen agar objek yang bernilai 1 (warna putih), sedangkan background yang bernilai 0 (berwarna hitam).

### **2.5.3.1 Adaptive Thresholding**

Suatu metode dimana dalam pencarian nilai ambang batas, gambar dipecah menjadi beberapa bagian gambar yang lebih kecil kemudian tiap-tiap bagian tersebut akan dicari nilai ambang batasnya. Nilai ambang adalah jumlah tertimbang nilai lingkungan dimana bobot adalah window gaussian.

## **2.6 Ekstraksi Ciri**

Ekstraksi ciri citra merupakan tahapan mengekstrak ciri/informasi dari objek di dalam citra yang ingin dikenali/dibedakan dengan objek lainnya. Ciri yang diekstrak digunakan sebagai nilai masukan untuk membedakan antara objek satu dengan lainnya pada tahap identifikasi/klasifikasi. Ciri yang dapat digunakan untuk membedakan objek satu dengan yang lainnya diantaranya adalah ciri bentuk, ciri ukuran, ciri geometri, ciri tekstur, dan ciri warna.

## **2.7 Normalisasi**

Normalisasi adalah proses penskalaan nilai atribut dari data sehingga bisa terletak pada rentang tertentu.

a. Min-Max Normalization

Metode normalisasi dengan melakukan transformasi linier terhadap data asli sehingga menghasilkan keseimbangan nilai perbandingan antar data saat sebelum dan sesudah proses.

## 2.8 Klasifikasi Citra

Klasifikasi citra digital merupakan proses pengelompokan piksel kelas kelas tertentu, seperti perbedaan pada nilai warna. Perbedaan warna ini dimulai dari 0- hingga gradasi warna. Semakin banyak klasifikasi citra, maka semakin banyak warna yang digunakan dan semakin bagus pula hasilnya. Pada umumnya klasifikasi citra digital terbagi ke dalam dua jenis, yaitu *supervised* dan *unsupervised*. *Supervised learning* merupakan proses pengelompokan data yang telah diberi label dan akan di klasifikasi berdasarkan labelnya. Sedangkan *unsupervised learning* merupakan proses data yang tidak memiliki label. Sehingga bebas menentukan berapa jumlah kelompok yang akan dibuat. Tujuan klasifikasi citra adalah untuk menghasilkan peta tematik, dimana suatu warna mewakili suatu objek tertentu.

## 2.9 K-Nearest Neighbor (K-NN)

K-NN adalah sebuah metode klasifikasi terhadap sekumpulan data berdasarkan pembelajaran data yang sudah terklasifikasikan sebelumnya. Termasuk dalam supervised learning, dimana hasil query instance yang baru di klasifikasikan berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam K-NN. Hal ini hanya didasarkan pada gagasan bahwa suatu objek yang ‘dekat’ satu sama lain juga akan memiliki karakteristik yang mirip. Ini berarti jika kita mengetahui ciri-ciri dari salah satu objek, maka kita juga dapat

memprediksi objek lain berdasarkan tetangga terdekatnya. Hal ini hanya didasarkan pada gagasan bahwa suatu objek yang ‘dekat’ satu sama lain juga akan memiliki karakteristik yang mirip. (Suguna dan Thanushkodi, 2010). K-nearest neighbor merupakan metode klasifikasi yang memiliki sifat ketetanggaan, ini didapat dari perhitungan nilai persamaan ataupun pertidaksamaan. KNN menggunakan perhitungan nilai pertidaksamaan (*Euclidean, Mahnhattan, Square Euclidian*, dll). KNN akan memilih K tetangga terdekat dengan jumlah kelas yang paling banyak yang akan menjadi kelas klasifikasi. (Putra Pamungkas et al., 2019)

### 2.9.1 Tahapan Metode K-Nearest Neighbor (K-NN)

Langkah-langkah untuk menghitung metode K-Nearest Neighbor antara lain :

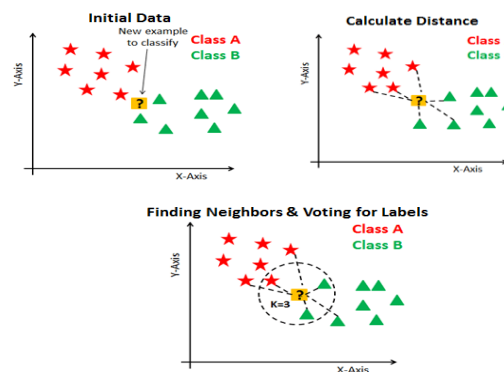
- a) Menentukan bobot dengan melihat jarak terdekat dan terjauh.
- b) Menentukan parameter k sebagai banyaknya jumlah tetangga terdekat dengan objek baru.
- c) Menghitung jarak antara data yang akan diklasifikasi dengan seluruh data pelatihan menggunakan jarak Euclidean.

Dinyatakan dalam persamaan (2.1)

$$Distance = \sqrt{\sum_{i,j=1}^n (x - x_{i,j})^2} \quad (2.1)$$

- d) Mengurutkan hasil perhitungan
- e) Tentukan tetangga terdekat berdasarkan jarak minimum ke K.
- f) Tentukan kategori dari tetangga terdekat dengan objek/data.
- g) Gunakan kategori mayoritas sebagai klasifikasi objek/data baru.

- h) Dengan menggunakan kategori Nearest Neighbor yang paling mayoritas maka dapat diprediksi nilai query instance yang telah dihitung.



Gambar 2. 1 Algoritma K-Nearest Neighbor

Menurut aturan *K-Nearest Neighbor* klasifikasi berdasarkan proses hasil mayoritas. Terlihat pada Gambar 2.1 terdapat bentuk bintang dan segitiga dan mengidentifikasi data baru masuk di kelompok bintang atau segitiga, selanjutnya menentukan nilai  $k$  yang bernilai ganjil, ditentukan nilai  $k = 3$ , selanjutnya pilih jarak terdekat dengan data baru, jarak terdekat dari data baru adalah 1 bintang dan 2 segitiga. Didapat kesimpulan bahwa data baru termasuk kedalam kelompok segitiga berdasarkan mayoritas.

KNN memiliki beberapa kelebihan yaitu ketangguhan terhadap training data yang memiliki noise dan efektif apabila training data-nya besar. Sedangkan kelemahannya adalah KNN perlu menentukan nilai dari  $K$  (jumlah dari tetangga terdekat), training berdasarkan jarak yang kurang jelas dan jarak apa yang harus digunakan untuk mendapatkan hasil terbaik.



### 2.9.2 Contoh Perhitungan *K-Nearest Neighbor* (K-NN)

Setelah citra diekstraksi menjadi 4 fitur yaitu *contrast*, *energy*, *homogeneity*, dan *correlation*. Selanjutnya adalah mengklasifikasikan citra berdasarkan identifikasi dengan metode *K-Nearest Neighbor*.

No	Kontras	Homogeneity	Energy	Corelation	Motif
1	0,043	0,160	0,079	0,936	Sembagi
2	0,002	0,646	0,380	0,999	Siger
3	0,449	0,127	0,048	0,987	Siger
4	0,006	0,372	0,134	0,993	Sembagi
5	0,047	0,284	0,227	0,990	Siger
6	0,299	0,133	0,044	0,891	?

Tabel 2. 3 Contoh Soal *K-NN*

Berikut tahapan dalam penerapan metode K-NN pada citra:

1. Tentukan parameter k. Misalnya ditentukan nilai k= 3
2. Hitung jarak antara data baru dengan semua data training. Disini menggunakan Euclidean Distance. Untuk mendapatkan nilai jarak menggunakan Persamaan (2.1).

*distance1*

$$= \sqrt{(0,299 - 0,043)^2 + (0,133 - 0,160)^2 + (0,044 - 0,079)^2 + (0,891 - 0,936)^2}$$

$$= 0,655 + 0,0007 + 0,00122 + 0,002 = 0,659$$

*distance2*

$$= \sqrt{(0,299 - 0,002)^2 + (0,133 - 0,646)^2 + (0,044 - 0,380)^2 + (0,891 - 0,999)^2}$$

$$= 0,088 + 0,263 + 0,112 + 0,012 = 0,475$$

*distance3*

$$= \sqrt{(0,299 - 0,449)^2 + (0,133 - 0,127)^2 + (0,044 - 0,048)^2 + (0,891 - 0,987)^2}$$

$$= 0,022 + 0,000036 + 0,000016 + 0,0092 = 0,031$$

*distance4*

$$= \sqrt{(0,299 - 0,006)^2 + (0,133 - 0,372)^2 + (0,044 - 0,134)^2 + (0,891 - 0,993)^2}$$

$$= 0,086 + 0,057 + 0,0081 + 0,01 = 0,1611$$

*distance5*

$$= \sqrt{(0,299 - 0,047)^2 + (0,133 - 0,284)^2 + (0,044 - 0,227)^2 + (0,891 - 0,990)^2}$$

$$= 0,064 + 0,228 + 0,033 + 0,009 = 0,334$$

3. Mengurutkan jarak Euclidean secara ascending.

*Tabel 2. 4 Pengurutan secara ascending*

No	Kontras	Homogeneity	Energy	Corelation	Motif
1	0,043	0,160	0,079	0,936	Sembagi
2	0,002	0,646	0,380	0,999	Siger
5	0,047	0,284	0,227	0,990	Siger
4	0,006	0,372	0,134	0,993	Sembagi
3	0,449	0,127	0,048	0,987	Siger

4. Ambil jarak terkecil sejumlah k

$Distance1 = 0,659$  Sembagi

$Distance2 = 0,475$  Siger

$Distance5 = 0,334$  Siger

$Distance4 = 0,161$  Sembagi

$Distance3 = 0,031$  Siger

5. Hasil klasifikasi citra dengan metode k-NN menggunakan mayoritas dari tetangga terdekat tersebut sebagai nilai prediksi data yang baru.

Kelas siger = 3

Kelas sembagi = 2

Dari proses klasifikasi citra dengan metode K-NN diatas, dapat disimpulkan bahwa data baru masuk kedalam kelas siger.

*Tabel 2. 5 Hasil Perhitungan K-NN*

No	Contrast	Homogeneity	Energy	Correlation	Motif
1	0,043	0,160	0,079	0,936	Sembagi
2	0,002	0,646	0,380	0,999	Siger
3	0,449	0,127	0,048	0,987	Siger
4	0,006	0,372	0,134	0,993	Sembagi
5	0,047	0,284	0,227	0,990	Siger
<b>6</b>	<b>0,299</b>	<b>0,133</b>	<b>0,044</b>	<b>0,891</b>	<b>Siger</b>

## 2.9 Naïve Bayes Classifier

*Naïve bayes* merupakan sebuah metode klasifikasi menggunakan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes. Algoritma ini memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Ciri utama dari *Naïve Bayes Classifier* ini adalah mengasumsi yang sangat kuat dari masing-masing kondisi/kejadian. Keuntungan penggunaan metode ini hanya membutuhkan jumlah data pelatihan yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian.

Tahapan dari proses algoritma *naïve bayes* adalah:

1. Menentukan nilai probabilitas hipotesis masing-masing target atau kelas  $P(c_i)$ .
2. Menghitung nilai *Gaussian mean* ( $\mu$ ) dan standar deviasi ( $\sigma$ ) didefinisikan dalam Persamaan (2.2) dan Persamaan (2.3).

$$\mu = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n} \quad (2.2)$$

$$\sigma^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{n} \quad (2.3)$$

Sedangkan fungsi distribusi *Gaussian* dinyatakan dalam Persamaan (2.4).

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}e^{\frac{-(x-\mu)^2}{2(\sigma^2)}}} \quad (2.4)$$

Keterangan :

$\mu$  : Hasil perhitungan tiap atribut data latih berdasarkan kelas yang telah dilakukan sebelumnya

$\sigma$  : Hasil perhitungan  $\sigma$  tiap atribut data latih berdasarkan kelas yang telah dilakukan sebelumnya

$\Pi$  : 3.14

e : 2,718282

x : nilai atribut pada data uji yang digunakan

3. Hitung nilai probabilitas masing-masing kelas. Untuk mendapatkan nilai probabilitas dinyatakan dalam Persamaan (2.5).

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \quad (2.5)$$

Keterangan:

$X$  : Data dengan kelas yang belum diketahui.

$C$  : Hipotesis data  $X$  merupakan suatu kelas spesifik.

$P(C|X)$  : Probabilitas hipotesis  $C$  berdasarkan kondisi  $X$ .

$P(C)$  : Probabilitas hipotesis  $C$ .

$P\{X|C\}$  : Probabilitas berdasarkan kondisi pada hipotesis  $C$ .

$P(X)$  : Probabilitas data dari  $X$ .

4. Mencari nilai tertinggi *posterior probability* dari setiap kelas. Dimana probabilitas posterior,  $P(C|X)$  (*probabilitas* dari nilai atribut  $X$  milik kelas  $C$ ), dihitung menggunakan *Class Prior Probability*  $P(C)$

(probabilitas kelas), *Predictor Prior Probability*  $P(X)$  (probabilitas nilai atribut) dan *Likelihood*  $P(X/C)$  (probabilitas nilai atribut  $X$  yang diberikan kelas  $C$ ).

Berikut contoh perhitungan dalam penerapan metode *naïve bayes* pada citra :

- a. Menentukan nilai probabilitas hipotesis masing-masing target atau kelas  $P(c_i)$ .

$$P(\text{target} = \text{"Sembagi"}) = 2/6 = 0,33$$

$$P(\text{target} = \text{"Siger"}) = 4/6 = 0,66$$

- b. Menghitung nilai *Gaussian mean* ( $\mu$ ) dan standar deviasi ( $\sigma$ ) untuk setiap atribut pada masing-masing kelas. Untuk mendapatkan *Gaussian mean* ( $\mu$ ) dan standar deviasi ( $\sigma$ ) menggunakan Persamaan (2.2) dan Persamaan (2.3). Tabel 2.5 menunjukkan hasil *Gaussian mean* ( $\mu$ ) dan standar deviasi ( $\sigma$ ).

Tabel 2. 6 Kelas Sembagi

No	Contrast	Homogeneity	Energy	Correlation	Motif
1	0,043	0,160	0,079	0,936	Sembagi
4	0,006	0,372	0,134	0,993	Sembagi
$\mu$	0,0245	0,266	0,1065	0,9645	
$\Sigma$	0,026163	0,149907	0,038891	0,040305	

Tabel 2. 7 Kelas Siger

No	Contrast	Homogeneity	Energy	Correlation	Motif
2	0,002	0,646	0,380	0,999	Siger
3	0,449	0,127	0,048	0,987	Siger
5	0,047	0,284	0,227	0,990	Siger
6	0,299	0,133	0,044	0,891	Siger
$\mu$	0,19925	0,2975	0,17475	0,96675	
$\Sigma$	0,21167	0,243423	0,161265	0,050757	

- c. Hitung nilai probabilitas masing-masing kelas. Untuk mendapatkan nilai probabilitas . Untuk mendapatkan nilai probabilitas untuk kelas sembagi menggunakan Persamaan (2.4) .

Perhitungan probabilitas untuk kelas sembagi adalah sebagai berikut:

$$P(X1 = 0,043|Sembagi) = \frac{1}{0,026163\sqrt{2(3.14)}} 2.718282 \frac{(0,043-0,0245)^2}{2(0,026163^2)}$$

$$= 1,921335$$

$$P(X2 = 0,160|Sembagi) = \frac{1}{0,149907\sqrt{2(3.14)}} 2.718282 \frac{(0,160-0,266)^2}{2(0,14997^2)}$$

$$= 0,802668$$

$$P(X3 = 0,079|Sembagi) = \frac{1}{0,038891\sqrt{2(3.14)}} 2.718282 \frac{(0,079-0,1065)^2}{2(0,038891^2)}$$

$$= 1,575878$$

$$P(X4 = 0,936|Sembagi) = \frac{1}{0,149907\sqrt{2(3.14)}} 2.718282 \frac{(0,936-0,9645)^2}{2(0,040305^2)}$$

$$= 1,547985$$

d. Perhitungan probabilitas untuk kelas siger sebagai berikut :

$$P(X1 = 0,002|Siger) = \frac{1}{0,21167\sqrt{2(3.14)}} 2.718282 \frac{(0,043-0,19925)^2}{2(0,21167^2)}$$

$$= 0,721549343$$

$$P(X2 = 0,646|Siger) = \frac{1}{0,243423\sqrt{2(3.14)}} 2.718282 \frac{(0,646-0,2975)^2}{2(0,243423^2)}$$

$$= 0,026941$$

$$P(X3 = 0,380|Siger) = \frac{1}{0,161265\sqrt{2(3.14)}} 2.718282 \frac{(0,380-0,17475)^2}{2(0,161265^2)}$$

$$= 0,004505$$

$$P(X4 = 0,999|Siger) = \frac{1}{0,96675\sqrt{2(3.14)}} 2.718282 \frac{(0,999-0,96675)^2}{2(0,96675^2)}$$

$$= 0,422726$$

e. Menghitung *posterior probability*  $P(C_i|X)$ . Untuk mendapatkan nilai *posterior probability* masing-masing kelas menggunakan Persamaan (2.3). Berikut persamaan nilai *posterior probability* masing-masing kelas menggunakan Persamaan (2.5).

$$P(\text{Sembagi} | X \text{ uji}) = (1,9213 * 0,8026 * 1,5758 * 1,5479 * 0,33)$$

$$= 1,24123$$

$$P(\text{Siger} | X \text{ uji}) = 0,7215 * 0,0269 * 0,0045 * 0,4227 * 0,66$$

$$= 2,43656E-05$$

f. Cari nilai tertinggi *posterior probability*  $P(C_i|X)$  dari setiap kelas. Karena  $P(\text{Sembagi} | X_{uji}) > P(\text{Siger} | X_{uji})$ , maka hasil tersebut menunjukkan bahwa citra uji diklasifikasikan ke dalam kelas/target “Siger”.



## 2.10 Google Collaboratory

*Google collaboratory* atau disingkat dengan *Google Collab* adalah satu tools yang dikembangkan oleh Google untuk data science. *Google Collab* persis seperti jupyter notebook. Kelebihannya adalah bahwa tools ini memiliki kumpulan *built-in-library machine learning* paling populer yang dapat dimuat dengan mudah dengan mudah dalam notebook.



Gambar 2. 2 Google Collaboratory

## 2.11 Confusion Matrix

Berisi informasi tentang hasil presiksi klasifikasi dan data actual yang dilakukan oleh sistem klasifikasi. Kinerja sistem klasifikasi umumnya dihitung menggunakan data dalam matriks ini. Adapun tabel *confusion matrix* sebagai berikut:

		Nilai sebenarnya	
		TRUE	FALSE
Nilai Prediksi	TRUE	<b>TP</b> (True Positive) Correct result	<b>FP</b> (False Positive) Unexpexted result
	FALSE	<b>FN</b> (False Negatif) Missing result	<b>TN</b> (True Negative) Correct absence of result

Tabel 2. 8 Confusion Matrix

Matriks tersebut memiliki empat nilai yang dijadikan acuan dalam perhitungan, dimana:

- *True Positive (TP)* = ketika kelas yang diprediksi positif dan faktanya positif
- *True Negatif (TN)* = ketika kelas yang diprediksi negatif dan faktanya negatif
- *False Positive (FP)* = ketika kelas yang diprediksi positif dan faktanya negatif
- *False Negative (FN)* = ketika kelas yang diprediksi negatif dan faktanya positif

*Berdasarkan nilai, TP, TN, FP, dan FN dapat diperoleh nilai akurasi. Nilai akurasi menggambarkan seberapa akurat system dapat mengklasifikasi data secara benar.*

Nilai akurasi menggambarkan seberapa akurat sistem dapat mengklasifikasi data secara benar. Untuk mencari nilai presisi, recall, dan akurasi dinyatakan dalam Persamaan (2.4), (2.5), dan (2.6).

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (2.4)$$

$$\text{Recall} = \frac{TP+TN}{TP+FN} \quad (2.5)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.6)$$

- ✓ Presisi (*precision*) adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem.
- ✓ Recall adalah tingkat keberhasilan sistem dalam menemukan Kembali sebuah informasi.
- ✓ Akurasi (*accuracy*) didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai actual.