

## Penerapan Metode Naive Bayes pada Sistem Klasifikasi Kualitas Biji Kopi Robusta

Muhammad Irsyad Indra Fata <sup>1\*</sup>, Donny Avianto <sup>2</sup>

<sup>1,2</sup> Program Studi Informatika, Fakultas Sains & Teknologi, Universitas Teknologi Yogyakarta, Kabupaten Sleman, Provinsi Daerah Istimewa Yogyakarta.

Email: muhammadfata560@gmail.com <sup>1\*</sup>, donny@uty.ac.id <sup>2</sup>

### Histori Artikel:

Dikirim 20 Oktober 2023; Diterima dalam bentuk revisi 15 November 2023; Diterima 1 Desember 2023; Diterbitkan 10 Januari 2024. Semua hak dilindungi oleh Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STMIK Indonesia Banda Aceh.

### Abstrak

Di Indonesia kopi sudah menjadi kebutuhan bagi sebagian masyarakat karena efek yang dihasilkan dapat meningkatkan stamina dan suasana hati sehingga dapat membuat aktivitas seseorang lebih produktif, maka dari itu dibutuhkan biji kopi berkualitas agar minuman kopi dapat menghasilkan efek yang maksimal serta bisa dinikmati oleh khalayak. Hal tersebut menjadi landasan peneliti untuk melakukan penelitian klasifikasi biji kopi berkualitas lebih unggul dengan akurasi yang akurat. Peneliti menggunakan metode naïve bayes classifier karena algoritma tersebut dinilai lebih efektif untuk kasus pengklasifikasian objek dibanding dengan klasifikasi yang dilakukan secara manual. Peneliti menggunakan 220 data primer dengan konfigurasi parameter yaitu, test size 0.2, random state 15, dan model gaussian. Hasil akurasi yang didapatkan setelah dilakukan pengujian yaitu sebesar 0.91 data training dan 0.86 data testing. Disimpulkan bahwa metode naïve bayes terbukti cukup efektif untuk mengklasifikasikan biji kopi yang diteliti.

**Kata Kunci:** Indonesia; Klasifikasi; Kopi; Naive Bayes Classifier.

### Abstract

In Indonesia, coffee has become a necessity for some people because the effects can increase stamina and mood so that it can make a person's activities more productive, therefore quality coffee beans are needed so that coffee drinks can produce maximum effects and can be enjoyed by the public This is the basis for researchers to conduct research on the classification of superior quality coffee beans with accurate accuracy. Researchers use the naïve Bayes classifier method because this algorithm is considered more effective in cases of object classification compared to classification carried out manually. Researchers used 220 primary data with parameter configurations, namely, test size 0.2, random state 15, and Gaussian model. The accuracy results obtained after testing were 0.91 for training data and 0.86 for testing data. It was concluded that the Naive Bayes method proved to be quite effective for classifying the coffee beans studied.

**Keyword:** Indonesia; Classification; Coffee; Naive Bayes Classifier.

## 1. Pendahuluan

Tanaman kopi merupakan salah satu komoditi yang sangat populer di Indonesia dan di banyak negara lain[1][2]. Produksi buah tanaman kopi membuatnya sangat populer karena setiap petani kopi dapat mengolah buahnya dengan berbagai cara. Dengan demikian, pengolahan ini menghasilkan produk biji kopi baru yang mempunyai berbagai macam kriteria. Seorang roaster akan mengolah biji kopi yang dihasilkan hingga siap untuk dikonsumsi sebagai minuman kopi [3]. Selain itu minuman kopi yang dihasilkan sampai saat ini masih dikonsumsi oleh berbagai masyarakat di Indonesia dan di luar negeri.

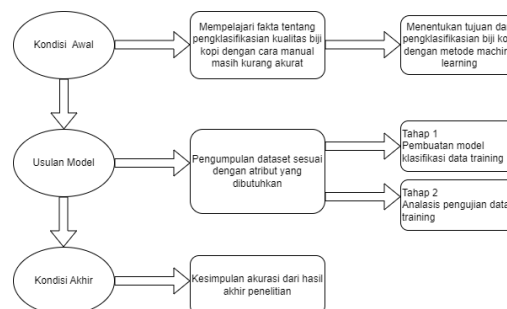
Ada beberapa jenis kopi di Indonesia, tetapi varietas yang populer ditanam oleh masyarakat Indonesia yaitu biji arabika dan robusta[4]. Kedua jenis kopi ini juga dapat diolah dengan berbagai cara atau teknik seperti wash, semi wash, natural, honey, dan masih banyak yang lain[4]. Di Indonesia juga terdapat banyak kedai kopi, mulai dari kedai kopi rumahan maupun café mewah. Selain itu, orang-orang yang menjalankan bisnis kopi bukan hanya mereka yang mahir dalam dunia tentang kopi tetapi ada juga orang awam yang hanya menyukai kopi sebagai minuman dan belum mahir untuk memilah biji kopi yang berkualitas bagus[5]. Oleh karena itu, dapat dikatakan sulit bagi orang awam untuk mensortir kualitas biji kopi yang akan digunakan sebagai bahan baku bisnis mereka, karena pemilahan biji kopi yang berkualitas bisa dijadikan sebagai syarat kopi itu enak dan memiliki nilai jual yang tinggi.

Pemilihan kualitas biji kopi dapat dilakukan dengan cara penelitian klasifikasi. Pengklasifikasian ini dapat dilakukan dengan metode machine learning. Penelitian serupa mengenai klasifikasi biji kopi telah dilakukan dengan menggunakan algoritma knn dan dengan menggunakan fitur warna dan tekstur. Algoritma knn merupakan suatu algoritma sederhana machine learning untuk kasus klasifikasi dengan mengikuti strategi “bird of a feather” yang mana data baru akan diletakkan pada tetangga atau data yang mirip dengan jarak yang terdekat[6]. Pada penelitian algoritma knn dihasilkan akurasi 66.67% dengan nilai k yang telah ditentukan yaitu 1, 5, dan 7, sedangkan untuk nilai k 3 dihasilkan akurasi 60%, dan hasil akurasi keseluruhan 65%[7]. Berdasarkan data penelitian sebelumnya maka peneliti ingin melakukan penelitian dengan memaksimalkan akurasi menggunakan metode lain dan variabel data primer yang telah didapatkan[7].

Oleh karena itu dilakukanlah sebuah penelitian untuk pengklasifikasian kualitas biji kopi terutama pada varietas robusta menggunakan metode Naive Bayes Classifier karena algoritma tersebut bisa dikatakan cukup efektif untuk menangani kasus klasifikasi. Dalam penelitian ini peneliti memberikan 2 label yaitu biji kopi berkualitas baik dan baik kurang berkualitas. Selain itu data primer yang digunakan peneliti sangat tepat terhadap algoritma Naive Bayes Classifier. Penelitian ini berfokus untuk memaksimalkan hasil dari keakuratan klasifikasi kualitas biji kopi dari data yang diteliti yang bertujuan agar menghasilkan sistem yang dapat membantu orang awam dalam pemilahan kualitas biji kopi.

## 2. Metode Penelitian

### 2.1 Kerangka Penelitian



Gambar 1. Alur Penelitian

Langkah pertama dalam penelitian dimulai dari kondisi awal yaitu mempelajari tentang suatu definisi masalah dalam kasus pengklasifikasian kualitas biji kopi yang masih dilakukan secara manual dengan cara mengamati objek secara visual yang menghasilkan akurasi kurang akurat. Setelah itu peneliti mengembangkan suatu model sistem dengan *algoritma machine learning* yaitu naive bayes classifier yang bertujuan agar hasil dari klasifikasi kualitas biji kopi memiliki akurasi yang lebih akurat. Dalam penelitian peneliti melakukan pengumpulan data primer sesuai dengan *attribute* yang dibutuhkan kemudian melakukan pemodelan terhadap data training dan menganalisis pengujian data testing. Peneliti melakukan beberapa pengujian sampai mendapatkan hasil akurasi terbaik.

## 2.2 Arsitektur Model



Gambar 2. Usulan Sistem

Pada bagian arsitektur model peneliti memberikan usulan sistem yang akan dikembangkan untuk pengklasifikasi kualitas biji kopi robusta. Seperti yang terlihat pada Gambar diatas merupakan sistem yang diusulkan oleh peneliti. Langkah awal petani melakukan panen pada tanaman kopi. Kemudian petani kopi memberikan data tentang biji kopi sesuai yang dibutuhkan oleh konsumen. Konsumen melakukan klasifikasi kualitas biji kopi dengan sistem yang telah ada.

## 2.3 Pengumpulan Data

Tahap awal yang dilakukan yaitu pengumpulan data dengan cara melakukan wawancara terhadap petani kopi robusta di daerah Temanggung. Peneliti melakukan pengumpulan data secara kualitatif yaitu dengan mengajukan berbagai pertanyaan tentang biji kopi kepada petani kopi yang bertujuan mendapatkan informasi secara real yang kemudian data yang didapatkan bersifat primer[8][9]. Kemudian data primer yang didapatkan akan digunakan dalam penelitian. Adapun data yang didapat berjumlah 220 data biji kopi dengan beberapa atribut yang tersedia. Contoh data yang digunakan dapat terlihat seperti dibawah.

## 2.4 Preprocessing

Pada tahap ini dilakukan preprocessing yaitu proses pengolahan dari data asli atau mentah yang telah didapatkan menjadi data yang lebih efektif dan sesuai dengan yang dibutuhkan dalam penelitian[10]. Pada tahap preprocessing dilakukan data cleansing atau pembersihan data digunakan untuk menghilangkan missing value (suatu data yang tidak memiliki nilai atau kosong)[11]. Kemudian didalam pembersihan data juga dilakukan penyesuaian format data yang tidak sesuai dan memlakukan perbaikan data yang mengalami typo/salah penulisan. Langkah berikutnya yaitu melakukan penghilangan beberapa attribute yang bertujuan untuk memaksimalkan hasil dari penelitian. Langkah selanjutnya yaitu melakukan pelabelan data yang mana berfungsi untuk hasil akhir dari klasifikasi yaitu kopi berkualitas baik atau kopi berkualitas kurang baik.

## 2.5 Klasifikasi Data

Selanjutnya dilakukan tahap klasifikasi dengan metode naive bayes classifier, pada tahap ini dilakukan pengklasifikasian data dengan bantuan aplikasi jupyter dan bahasa pemrograman yang digunakan yaitu python. Klasifikasi oleh naive bayes digunakan untuk memprediksi suatu hasil keputusan sesuai dengan probabilitas kelas tersebut[12]. Pada tahap klasifikasi dilakukan pembuatan model dengan gaussian naive bayes setelah itu dilakukan pengujian terhadap data testing. Berikut merupakan rumus perhitungan pada naive bayes[13].

$$P(X | Y) = \frac{P(X|Y)P(X)}{P(Y)}$$

Keterangan :

- Y : Jumlah data yang belum terkategori label.
- X : Hasil dari hipotesis.
- $P(X | Y)$  : Kemungkinan terjadi pada hipotesis X berdasarkan Y.
- $P(Y | X)$  : Kemungkinan pada hipotesis.
- $P(X)$  : Probabilitas hipotesis A.
- $P(Y)$  : Probabilitas B.

### 2.6 Prediksi Pada Matrix

Selanjutnya adalah tahap confusion matrix yaitu dilakukan evaluasi matrix pada model yang telah dibuat dengan tujuan untuk mengukur seberapa jauh hasil akurasi yang didapatkan dari penelitian ini[14]. Selain itu untuk mempertimbangkan nilai prediksi yang benar terhadap nilai aktual. Untuk tabel dan perhitungan pada matrix dapat dilihat sebagai berikut[15].

Tabel 1. Informasi Matrix

Confusion Matrix		Prediksi	
		Yes	No
Kondisi Nyata	Yes	TP	FN
	No	FP	TN

Informasi Tabel 2 :

- TP (*True Positive*) = hasil data Yes yang tepat dikategorikan pada label Yes
- TN (*True Negative*) = hasil data No yang tepat dikategorikan pada label No
- FP (*False Positive*) = hasil data No tetapi dikategorikan kedalam label yang salah
- FN (*False Negative*) = hasil data Yes tetapi dikategorikan kedalam label yang salah

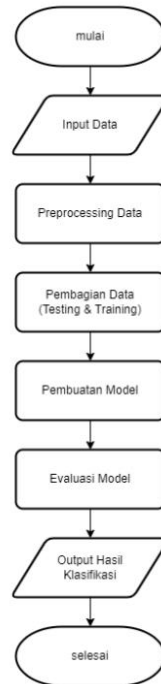
Adapun perhitungan akurasi pada confusion matrix seperti berikut.

$$\text{akurasi} = \frac{TP+TN}{\text{Jumlah Data}} \times 100\%$$

### 2.7 Pembuatan Prototipe

Setelah dilakukan beberapa tahap, langkah terakhir yaitu pembuatan prototipe yaitu untuk mempermudah dalam implementasi sistem. Pembuatan prototipe ini menggunakan aplikasi jupyter dan dengan bahasa pemrograman pyhton. Didalam prototipe ini akan dibuat tampilan menu sederhana yang mana untuk sementara dapat digunakan untuk melakukan klasifikasi data biji kopi baru dan hasil akhir klasifikasi berupa keterangan bahwa kopi tersebut termasuk kedalam biji kopi bagus berkualitas atau biji kopi kurang berkualitas.

## 2.8 Flowchart sistem



Gambar 3. *Flowchart*

Pada bagian ini dijelaskan mengenai alur jalannya sistem digambarkan pada flowchart. Dimulai dari input data biji kopi robusta. Kemudian dilakukan preprocessing data. Setelah data bersih dilakukan pembagian data dengan menggunakan train/test split. Langkah berikutnya yaitu dilakukan pembuatan model dengan gaussian naïve bayes. Kemudian dilakukan pengujian dan dilakukan evaluasi terhadap hasil model.

## 3. Hasil dan Pembahasan

### 3.1 Pengumpulan Data

Peneliti melakukan pengumpulan data secara langsung dengan mewawancarai petani kopi di daerah Temanggung yaitu biji kopi robusta yang masih dalam bentuk green beans. Peneliti mengajukan beberapa pertanyaan mengenai biji kopi robusta kepada petani kopi kemudian melakukan pencatatan data informasi yang diterima. Data yang diperoleh peneliti bersifat kategorikal dan berjumlah 220 dengan 2 label yaitu 110 data biji kopi berkualitas bagus dan 110 data biji kopi berkualitas kurang bagus. Adapun variabel atribut yang digunakan pada data tersebut diantaranya, no, kopi, bentuk, warna, ukuran, jenis cacat 1, jenis cacat 2, kadar air, aroma, fisik 1, fisik 2, label. Berikut merupakan contoh data yang didapatkan oleh peneliti.

No	Kopi	Bentuk	Warna	Ukuran	Jenis Cacat 1	Jenis Cacat 2	Kadar Air	Aroma	Fisik 1	Fisik 2	Label
1	RBT_1	Normal	Kuning	Large	Jemur Lantai	Normal	14	Bau Kapang	Normal	Normal	Bagus
2	RBT_2	Piberi	Kuning	Large	Cacat Pohon	Normal	14	Bau Kapang	Kurang Sempurna	Normal	Bagus
3	RBT_3	Normal	Kuning	Large	Jemur Lantai	Normal	14	Bau Kapang	Pecah Pinggir	Normal	Bagus
4	RBT_4	Piberi	Kuning	Large	Cacat Mesin Huler	Normal	14	Bau Kapang	Kurang Sempurna	Normal	Kurang Bagus
5	RBT_5	Piberi	Kuning	Large	Petik Hijau	Normal	14	Bau Kapang	Kisut	Normal	Kurang Bagus
6	RBT_6	Normal	Kuning	Large	Hama	Normal	14	Bau Kapang	Berlubang Dua	Normal	Kurang Bagus
7	RBT_7	Normal	Corak Hitam	Semi Larg	Cacat Mesin Pulper	Normal	14	Bau Kapang	Hitam Rusak Setengah	Normal	Kurang Bagus
8	RBT_8	Piksel	Hitam	Mini	Cacat Mesin Huler	Normal	14	Bau Kapang	Hitam Rusak Setengah	Normal	Kurang Bagus
9	RBT_9	Normal	Hijau Hitam	Semi Large	Jemur Lantai	Cacat Mesin Huler	14	Bau Kapang	Kisut	Normal	Kurang Bagus
10	RBT_10	Piberi	Kuning	Large	Jemur Lantai	Normal	14	Bau Kapang	Pecah Pinggir	Normal	Bagus

Gambar 4. *Data Biji Kopi*

### 3.2 Preprocessing

Pada tahap *preprocessing* dilakukanlah proses *cleansing* data dengan tujuan data yang digunakan sesuai dengan sistem dan mendapatkan hasil akurasi yang maksimal serta efektif. Langkah pertama peneliti melakukan pengecekan secara manual pada data biji kopi untuk memastikan tidak adanya missing value atau suatu data yang tidak memiliki nilai pada atribut tertentu. Langkah berikutnya peneliti melakukan penghapusan pada kolom atribut data tertentu yang bertujuan untuk memaksimalkan hasil dari penelitian. Selain itu peneliti melakukan penghapusan kolom atribut tertentu bertujuan agar data lebih efektif. Peneliti menggunakan fungsi `df.drop` untuk melakukan penghapusan kolom, atribut yang dihapus yaitu `no`, `kopi`, dan `aroma`, dalam hal ini peneliti menggunakan aplikasi jupyter dengan bahasa pemrograman python.

	Bentuk	Warna	Ukuran	Jenis Cacat 1	Jenis Cacat 2	Kadar Air	Fisik 1	Fisik 2	Label
0	Normal	Kuning	Large	Jemur Lantai	Normal	14	Normal	Normal	Bagus
1	Piberi	Kuning	Large	Cacat Pohon	Normal	14	Kurang Sempurna	Normal	Bagus
2	Normal	Kuning	Large	Jemur Lantai	Normal	14	Pecah Pinggir	Normal	Bagus
3	Piberi	Kuning	Large	Cacat Mesin Huler	Normal	14	Kurang Sempurna	Normal	Kurang Bagus
4	Piberi	Kuning	Large	Petik Hijau	Normal	14	Kisut	Normal	Kurang Bagus

Gambar 5. Data Setelah Penghapusan Atribut

Kemudian peneliti melakukan transformasi data yaitu mengkonversi data yang bersifat kategorikal atau string menjadi data numeric atau berbentuk angka. Peneliti melakukan tranformasi data yang bertujuan agar dalam pembuatan model data lebih mudah dipahami/dikenali oleh sistem. Pada proses ini peneliti menggunakan aplikasi jupyter dan menggunakan label encoder yaitu library dari python.

	Bentuk	Warna	Ukuran	Jenis Cacat 1	Jenis Cacat 2	Kadar Air	Fisik 1	Fisik 2	Label
0	0	15	1	5	4	14	25	9	0
1	1	15	1	3	4	14	20	9	0
2	0	15	1	5	4	14	29	9	0
3	1	15	1	1	4	14	20	9	1
4	1	15	1	8	4	14	17	9	1

Gambar 6. Data Setelah Tranformasi

### 3.3 Pembuatan Model

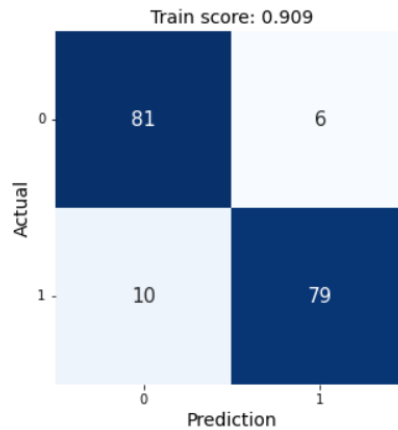
Pembuatan model merupakan tahap yang digunakan untuk melakukan klasifikasi dan prediksi pada data biji kopi yang telah ada. Pada proses pembuatan model dilakukan pembagian data menjadi 2 yaitu data fitur dan label. Data yang digunakan sebagai variabel fitur yaitu atribut bentuk, warna, ukuran, jenis cacat 1, jenis cacat 2, kadar air, fisik 1, fisik 2, sedangkan untuk variabel targer yaitu atribut "label" dijadikan sebagai label klasifikasi biji kopi. Setelah dilakukan tahap preprocessing data, langkah berikutnya yaitu pembagian data yaitu menggunakan `splitting` data membagi data menjadi data `train/test`. Langkah tersebut digunakan untuk mengevaluasi hasil performa pada model gaussian naive bayes. Peneliti melakukan `train/test split` dengan rasio perbandingan 0.2 dan 0.8 yaitu 80% untuk `train` data dan 20% untuk `test` data. Kemudian untuk `random state` yang digunakan sebesar 15, setelah itu dengan parameter tersbut model dapat dilatih menggunakan data latih dan dilakukan pengujian dengan data testing. Berikut merupakan hasil akurasi yang diperoleh dari model :

Tabel 2. Hasil Akurasi

Algoritma	Train Score	Test Score
Gaussian Naive Bayes	0.909	0.863

### 3.4 Confusion Matrix

Pada tahap ini dilakukan evaluasi pada confusion matrix yang bertujuan untuk mengukur hasil kinerja model gaussian naive bayes terhadap pengklasifikasian kualitas biji kopi. Selain itu evaluasi model bertujuan untuk mengetahui seberapa jauh pemahaman model dalam memecahkan kasus klasifikasi kualitas biji kopi, dan melakukan prediksi dengan akurasi yang cukup tinggi. Hasil dari evaluasi model gaussianNB dapat dilihat sebagai berikut.



Gambar 7. Matrix pada data latih

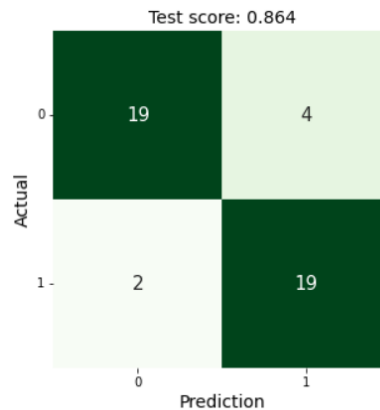
Berdasarkan hasil pada gambar 7 diperoleh informasi sebagai berikut :

- TP (*True Positive*) = hasil data Yes yang tepat dikategorikan pada label Yes berjumlah 81 data.
- TN (*True Negative*) = hasil data No yang tepat dikategorikan pada label No berjumlah 79 data.
- FP (*False Positive*) = hasil data No tetapi dikategorikan kedalam label yang salah berjumlah 10 data.
- FN (*False Negative*) = hasil data Yes tetapi dikategorikan kedalam label yang salah berjumlah 6 data.

Perhitungan akurasi sebagai berikut :

$$\text{accuracy} = \frac{TP+TN}{\text{Jumlah Data}} \times 100\% = \frac{81+79}{176} \times 100\% = 90.9 \%$$

Berdasarkan hasil diatas model yang telah dibuat dengan menggunakan data latih (train data) memiliki akurasi sebesar 90.9% dalam memprediksi kualitas biji kopi.



Gambar 8. Confusion Matrix Test Data

Berdasarkan hasil confusion matrix pada data uji diatas dapat dijelaskan sebagai berikut :

- TP (*True Positive*) = hasil data Yes yang tepat dikategorikan pada label Yes berjumlah 19 data.
- TN (*True Negative*) = hasil data No yang tepat dikategorikan pada label No berjumlah 19 data.
- FP (*False Positive*) = hasil data No tetapi dikategorikan kedalam label yang salah berjumlah 2 data.
- FN (*False Negative*) = hasil data Yes tetapi dikategorikan kedalam label yang salah berjumlah 4 data.

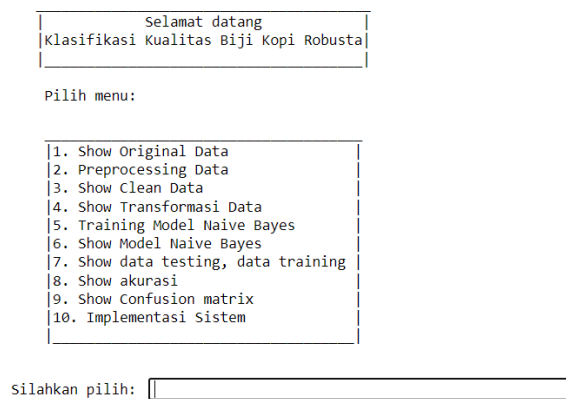
Perhitungan akurasi sebagai berikut :

$$\text{accuracy} = \frac{TP+TN}{\text{Total}} \times 100\% = \frac{19+19}{44} \times 100\% = 86.36 \%$$

Berdasarkan hasil diatas model yang telah dibuat kemudian dilakukan pengujian dengan menggunakan data uji (test data) memiliki akurasi sebesar 86.36 % dalam memprediksi kualitas biji kopi.

### 3.5 Implementasi

Pada bagian ini dijelaskan mengenai implementasi sistem melalui prototipe yang bertujuan agar lebih efektif saat digunakan. Pada pembuatan prototipe peneliti menggunakan aplikasi jupyter dan bahasa pemrograman python. Peneliti membuat 10 menu pada prototipe, untuk tampilan menu dapat dilihat sebagai berikut:



Gambar 9. Tampilan awal *prototipe*

*Prototipe* memiliki 7 menu untuk menampilkan, 2 menu untuk proses, dan 1 menu untuk implementasi. Cara kerja sistem ini ketika dijalankan yaitu user memasukkan suatu inputan nilai atau angka yang tersedia didalam tampilan sistem kemudian menekan enter pada *keyboard* agar diproses ke selanjutnya oleh sistem. Jika *user* memasukkan inputan diluar fitur menu maka akan sistem menampilkan pilihan apakah user akan kembali ke menu atau tidak dengan pilihan y itu iya, t itu tidak dan sistem otomatis selesai, apabila dalam proses ini user memasukkan inputan diluar y dan t maka sistem otomatis kembali masuk ke awal sistem dijanakan yaitu menu.

```
Silahkan pilih: 1
No  Kopi  Bentuk  Warna  Ukuran  Jenis Cacat 1  Jenis Cacat 2  \
0  1  RBT_1  Normal  Kuning  Large  Jemur Lantai  Normal
1  2  RBT_2  Piberi  Kuning  Large  Cacat Pohon  Normal
2  3  RBT_3  Normal  Kuning  Large  Jemur Lantai  Normal
3  4  RBT_4  Piberi  Kuning  Large  Cacat Mesin Huler  Normal
4  5  RBT_5  Piberi  Kuning  Large  Petik Hijau  Normal

Kadar Air  Aroma  Fisik 1  Fisik 2  Label  \
0  14  Bau Kapang  Normal  Normal  Bagus
1  14  Bau Kapang  Kurang Sempurna  Normal  Bagus
2  14  Bau Kapang  Pecah Pinggir  Normal  Bagus
3  14  Bau Kapang  Kurang Sempurna  Normal  Kurang Bagus
4  14  Bau Kapang  Kisut  Normal  Kurang Bagus

Aktivitas Air  Density
0  NaN  NaN
1  NaN  NaN
2  NaN  NaN
3  NaN  NaN
4  NaN  NaN

Kembali Ke Menu ? [y/t] 
```

Gambar 10. Tampilan show original data

Pada menu Show Original Data diperoleh pemanggilan data asli atau data pertama sebelum dilakukan preprocessing. Dalam pemanggilan tersebut didapatkan 5 data teratas dengan masing masing attribute. Pemanggilan data tersebut memanfaatkan library pandas.

```
Silahkan pilih: 2
Preprocessing Data Selesai...

Kembali Ke Menu ? [y/t] 
```

Gambar 11. Tampilan pemilihan menu no 2

Pada menu tersebut maka akan diperoleh pemberitahuan bahwa *Preprocessing* Data Selesai yang artinya bahwa telah dilakukan pembersihan data atau pengolahan data dari data sebelumnya agar bisa digunakan dalam penelitian didalam *prototipe*.

```
Silahkan pilih: 3
Bentuk  Warna  Ukuran  Jenis Cacat 1  Jenis Cacat 2  Kadar Air  \
0  Normal  Kuning  Large  Jemur Lantai  Normal  14
1  Piberi  Kuning  Large  Cacat Pohon  Normal  14
2  Normal  Kuning  Large  Jemur Lantai  Normal  14
3  Piberi  Kuning  Large  Cacat Mesin Huler  Normal  14
4  Piberi  Kuning  Large  Petik Hijau  Normal  14

Fisik 1  Fisik 2  Label
0  Normal  Normal  Bagus
1  Kurang Sempurna  Normal  Bagus
2  Pecah Pinggir  Normal  Bagus
3  Kurang Sempurna  Normal  Kurang Bagus
4  Kisut  Normal  Kurang Bagus

Kembali Ke Menu ? [y/t] 
```

Gambar 12. Tampilan pemilihan menu no 3

Pada menu tersebut akan ditampilkan 5 data teratas dari data yang sudah dilakukannya *preprocessing*. Dalam tahap ini peneliti melakukan beberapa penghilangan beberapa *attribute* yang tidak digunakan.

```
Silahkan pilih: 4
Bentuk  Warna  Ukuran  Jenis Cacat 1  Jenis Cacat 2  Kadar Air  Fisik 1  \
0  0  0  0  2  0  14  2
1  1  0  0  1  0  14  1
2  0  0  0  2  0  14  3
3  1  0  0  0  0  14  1
4  1  0  0  3  0  14  0

Fisik 2  Label
0  0  0
1  0  0
2  0  0
3  0  1
4  0  1

Kembali Ke Menu ? [y/t] 
```

Gambar 13. Tampilan pemilihan menu no 4

Pada menu tersebut akan ditampilkan 5 data teratas dari hasil transform data dari string ke *numeric*. Peneliti melakukan transformasi data supaya data dapat digunakan untuk proses penelitian selanjutnya. Peneliti menggunakan library dari sklearn preprocessing dan mengimportkan fungsi dari Label Encoder.

```
Silahkan pilih: 5
Training Model Selesai...
Kembali Ke Menu ? [y/t] 
```

Gambar 14. Tampilan pemilihan menu no 5

Pada menu diatas maka akan ditampilkan pemberitahuan dari sistem bahwa Training Model Selesai yang artinya mesin telah melakukan pemodelan pada data.

```
Silahkan pilih: 6
model yang digunakan adalah : GaussianNB()
Kembali Ke Menu ? [y/t] 
```

Gambar 15. Tampilan pemilihan menu no 6

Pada menu tersebut ditampilkan model yang digunakan dalam sistem klasifikasi kualitas biji kopi lokal adalah GaussianNB() yang mana model tersebut merupakan model dari algoritma Naïve Bayes Classifier. Peneliti menggunakan library sklearn.naive\_bayes dan mengimportkan GaussianNB.

```
Silahkan pilih: 7
Panjang data x train :176
Panjang data y train :176
Panjang data x test :44
Panjang data y test :44
Kembali Ke Menu ? [y/t] 
```

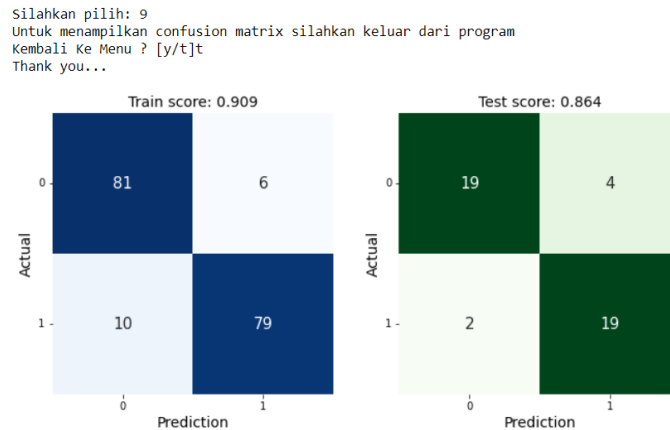
Gambar 16. Tampilan pemilihan menu no 7

Pada menu tersebut ditampilkan total dari data x train dan y train masing-masing sebesar 176 data. Untuk data x test dan y test masing-masing sebesar 44 data. Peneliti menggunakan library dari sklearn untuk pembagian data latih dan uji.

```
Silahkan pilih: 8
Akurasi yang didapatkan dari data training adalah : 0.9090909090909091
Akurasi yang didapatkan dari data testing adalah : 0.8636363636363636
Kembali Ke Menu ? [y/t] 
```

Gambar 17. Tampilan pemilihan menu no 8

Pada menu tersebut akan ditampilkan dari akurasi yang didapatkan dari model yang digunakan yaitu akurasi dari data training sebesar 0.909 dan akurasi dari data testing sebesar 0.863.



Gambar 18. Tampilan pemilihan menu no 9

Pada menu tersebut akan ditampilkan sebuah *confusion matrix* yang dihasilkan dari proses pemodelan sebelumnya dan didapatkan akurasi dari masing-masing data train dan test. Selain itu terdapat beberapa salah prediksi dan benar prediksi dari sistem seperti yang tertera pada Gambar 4.10. Untuk menampilkan *confusion matrix* user harus memilih nomor 9 lalu keluar dari menu dengan input t/T supaya confusion matrix tampil.

```
Silahkan pilih: 10
[[0. 0. 0. 0. 0. 0. 0.]]
Masukan Nilai Bentuk :
masukan : 1
Masukan Nilai Warna :
masukan : 1
Masukan Nilai Ukuran :
masukan : 2
Masukan Nilai Jenis Cacat 1 :
masukan : 1
Masukan Nilai Jenis Cacat 2 :
masukan : 1
Masukan Nilai Kadar Air :
masukan : 14
Masukan Nilai Fisik 1 :
masukan : 11
Masukan Nilai Fisik 2 :
masukan : 2
Data yang Ingin Diprediksi : [[ 1.  1.  2.  1.  1. 14. 11.  2.]]
Hasil Klasifikasi Biji Kopi Kurang Bagus
Kembali Ke Menu ? [y/t] 
```

Gambar 19. Tampilan pemilihan menu no 10

Pada menu tersebut sistem akan menampilkan suatu inputan dan user harus menginputkan nilai dari tiap-tiap *attribute* yang disediakan dan sesuai dari nilai yang disediakan. Setelah itu akan muncul hasil prediksi dari inputan data biji kopi yang diinputkan user bahwa biji tersebut Bagus atau Kurang Bagus.

#### 4. Kesimpulan

Berdasarkan hasil dari penelitian klasifikasi kualitas biji kopi dengan menggunakan Metode Naive Bayes dan model gaussian, disimpulkan cukup efektif dengan hasil akurasi training sebesar 0.91% dan testing 0.86%. Dengan total data 220 yang terdiri dari 176 data latih dan 44 data uji serta random state yang digunakan sebesar 15. Sehingga dengan penelitian ini algoritma naive bayes dapat menghasilkan performa yang cukup baik dalam mengklasifikasikan kualitas biji kopi robusta dengan atribut data penelitian yang digunakan seperti bentuk, warna, ukuran, jenis cacat 1, jenis cacat 2, kadar air, fisik 1, dan fisik 2.

Dalam rangka meningkatkan hasil penelitian terdapat beberapa saran yang diusulkan. Saran pertama yaitu meningkatkan jumlah data dan menambahkan beberapa atribut yang sesuai dengan klasifikasi kualitas biji kopi agar akurasi dari model yang dibuat lebih akurat. Saran kedua yaitu dapat

mengembangkan model dengan beberapa parameter seperti k fold validation, menambahkan nilai recall agar dapat mengetahui seberapa jauh model dapat memprediksi. Saran berikutnya yaitu hasil penelitian dapat dibuat dalam bentuk web agar lebih memudahkan dalam implementasi sistem.

## 5. Daftar Pustaka

- [1] Karyani, T. (2019). PENINGKATAN KEMAMPUAN PETANI DALAM MENGELOLA RISIKO USAHATANI KOPI (coffea arabica). *Dharmakarya: Jurnal Aplikasi Ipteks Untuk Masyarakat*, 8(4), 268-273. DOI: <https://doi.org/10.24198/dharmakarya.v8i4.24915>.
- [2] Jamil, A. S. (2019). Daya Saing Perdagangan Kopi Indonesia di Pasar Global. *Agriekonomika*, 8(1), 26-35. DOI: <https://doi.org/10.21107/agriekonomika.v8i1.4924>.
- [3] Karyani, T. (2018). Peningkatan kemampuan manajemen pemasaran coffee shop dan nilai tambah pangan berbahan baku kopi di koperasi produsen kopi margamulya (KPKM). *Dharmakarya: Jurnal Aplikasi Ipteks Untuk Masyarakat*, 7(3), 157-162. DOI : <https://doi.org/10.24198/dharmakarya.v7i3.19496>.
- [4] Wibowo, Y., & Palupi, C. B. (2022). Analisis Nilai Tambah Pengolahan Biji Kopi Arabika (Studi Kasus: Rumah Kopi Banjarsengon, Jember). *Jurnal Agroteknologi*, 16(01), 37-48. DOI: <https://doi.org/10.19184/j-agt.v16i01.28209>.
- [5] Hasibuan, S. A. F., Harahap, R. H., & Purwoko, A. (2021). Peran pemuda dalam pengembangan usaha kopi di kawasan wisata Simarjarunjung. *Perspektif*, 10(2), 644-655. DOI: [10.31289/perspektif.v10i2.4970](https://doi.org/10.31289/perspektif.v10i2.4970).
- [6] Cahyanti, D., Rahmayani, A., & Husniar, S. A. (2020). Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara. *Indonesian Journal of Data and Science*, 1(2), 39-43. DOI: <https://doi.org/10.33096/ijodas.v1i2.13>.
- [7] Rizal, M. (2019). *Klasifikasi Mutu Biji Kopi Menggunakan Metode K-Nearest Neighbor Berdasarkan Warna dan Tekstur* (Doctoral dissertation, University of Technology Yogyakarta).
- [8] Thalib, M. A. (2022). Pelatihan Teknik Pengumpulan Data dalam Metode Kualitatif untuk Riset Akuntansi Budaya. *Seandanan: Jurnal Pengabdian Pada Masyarakat*, 2(1), 44-50. <https://doi.org/10.23960/seandanan.v2i1.29>.
- [9] Hansen, S. (2020). Investigasi teknik wawancara dalam penelitian kualitatif manajemen konstruksi. *Jurnal Teknik Sipil*, 27(3), 283.
- [10] Najjichah, H., Syukur, A., & Subagyo, H. (2019). Pengaruh text preprocessing dan kombinasinya pada peringkas dokumen otomatis teks berbahasa Indonesia. *Jurnal Cyberku*, 15(1), 1-11.
- [11] Simanjuntak, H. E., & Windarto, W. (2020). Analisa Data Mining Menggunakan Frequent Pattern Growth pada Data Transaksi Penjualan PT Mora Telematika Indonesia untuk Rekomendasi Strategi Pemasaran Produk Internet. *Jurnal Media Informatika Budidarma*, 4(4), 914-923. DOI: <http://dx.doi.org/10.30865/mib.v4i4.2300>.



- [12] Damuri, A., Riyanto, U., Rusdianto, H., & Aminudin, M. (2021). Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako. *JURIKOM (Jurnal Riset Komputer)*, 8(6), 219-225. DOI: <http://dx.doi.org/10.30865/jurikom.v8i6.3655>.
- [13] Christianto, M., Andjarwirawan, J., & Tjondrowiguno, A. N. (2020). Aplikasi analisa sentimen pada komentar berbahasa Indonesia dalam objek video di website YouTube menggunakan metode Naïve Bayes classifier. *Jurnal Infra*, 8(1), 255-259.
- [14] Ginting, V. S., Kusriani, K., & Taufiq, E. (2020). Implementasi Algoritma C4. 5 untuk Memprediksi Keterlambatan Pembayaran Sumbangan Pembangunan Pendidikan Sekolah Menggunakan Python. *Inspiration: Jurnal Teknologi Informasi dan Komunikasi*, 10(1), 36-44. DOI: <http://dx.doi.org/10.35585/inspir.v10i1.2535>.
- [15] Suhet, W. K., & Nugroho, A. (2023). Analisis Sentimen Penilaian Masyarakat Indonesia Terhadap Konferensi G20 di Bali dengan Menggunakan Metode Naive Bayes. *Jurnal Indonesia: Manajemen Informatika dan Komunikasi*, 4(3), 904-912. DOI: <https://doi.org/10.35870/jimik.v4i3.290>.