

**PERBANDINGAN METODE NAÏVE BAYES DAN KNN DALAM  
PENERAPAN DATA MINING UNTUK KLASIFIKASI MENU  
POTENSIAL ( STUDI KASUS : ATEKU KOPI MEDAN )**

**SKRIPSI**

**NIKEN ALFRIDO DONATUS SIBORO  
171401115**



**PROGRAM STUDI S1 ILMU KOMPUTER  
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI  
UNIVERSITAS SUMATERA UTARA  
MEDAN  
2024**

**PERBANDINGAN METODE NAÏVE BAYES DAN KNN DALAM  
PENERAPAN DATA MINING UNTUK KLASIFIKASI MENU  
POTENSIAL ( STUDI KASUS : ATEKU KOPI MEDAN )**

**SKRIPSI**

Diajukan untuk melengkapi dan memenuhi syarat memperoleh ijazah Sarjana  
Ilmu Komputer

**NIKEN ALFRIDO DONATUS SIBORO**

**171401115**



**PROGRAM STUDI S1 ILMU KOMPUTER  
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI  
UNIVERSITAS SUMATERA UTARA  
MEDAN  
2024**

**PERSETUJUAN**

Judul : PERBANDINGAN METODE NAÏVE BAYES  
DAN KNN DALAM PENERAPAN DATA  
MINING UNTUK KLASIFIKASI MENU  
POTENSIAL ( STUDI KASUS : ATEKU  
KOPI MEDAN )

Kategori : SKRIPSI

Nama : NIKEN ALFRIDO DONATUS SIBORO

Nomor Induk Mahasiswa : 171401115

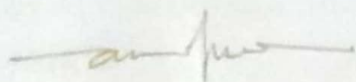
Program Studi : SARJANA (S-1) ILMU KOMPUTER

Fakultas : FAKULTAS ILMU KOMPUTER DAN  
TEKNOLOGI INFORMASI UNIVERSITAS  
SUMATERA UTARA

Telah diuji dan dinyatakan lulus di Medan, 9 Juli 2024

Komisi Pembimbing :

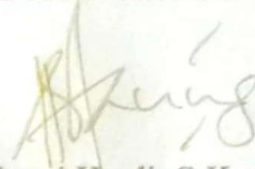
Dosen Pembimbing 2



Dr. Amalia S.T., M.T.

NIP. 197812212014042001

Dosen Pembimbing 1



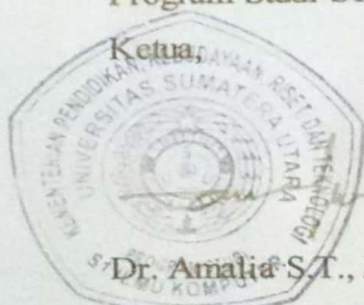
Sri Melvani Hardi, S.Kom., M.Kom

NIP. 198805012015042006

Diketahui/Disetujui oleh

Program Studi S1 Ilmu Komputer

Ketua,



Dr. Amalia S.T., M.T.

NIP. 197812212014042001

**PERNYATAAN**

PERBANDINGAN METODE NAÏVE BAYES DAN KNN DALAM  
PENERAPAN DATA MINING UNTUK KLASIFIKASI MENU POTENSIAL (   
STUDI KASUS : ATEKU KOPI MEDAN )

**SKRIPSI**

Saya mengakui bahwa Skripsi ini adalah hasil karya saya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing telah disebutkan sumbernya.

Medan, 9 Juli 2024



Niken Alfrido Donatus Siboro

171401115

## PENGHARGAAN

Segala puji syukur penulis panjatkan kepada Tuhan Yesus Kristus atas hikmat dan berkat-Nya sehingga penulis dapat menyelesaikan penulisan skripsi ini, yang merupakan bagian dari persyaratan untuk menyelesaikan pendidikan Program Studi Sarjana Ilmu Komputer di Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara, serta meraih gelar sarjana.

Tidak lupa penulis juga ingin mengucapkan penghargaan kepada semua pihak yang telah memberikan dukungan dalam penyusunan skripsi ini. Pihak-pihak tersebut antara lain:

1. Bapak Prof. Dr. Muryanto Amin, S.Sos., M.Si. Rektor Universitas Sumatera Utara.
2. Ibu Dr. Maya Silvi Lydia B.Sc. M.Sc., Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara.
3. Ibu Sri Melvani Hardi S.Kom., M.Kom. selaku Dosen Pembimbing I yang selalu sabar dan ikhlas memberikan bimbingannya kepada penulis selama proses pengerjaan skripsi.
4. Ibu Dr. Amalia ST., M.T. selaku Ketua Program Studi S-1 Ilmu Komputer Universitas Sumatera Utara, selaku Pembimbing Akademik yang sudah bersedia membimbing penulis selama menempuh perkuliahan dan Dosen Pembimbing II yang bersedia memberikan bimbingannya kepada penulis selama proses pengerjaan skripsi.
5. Tenaga pendidik beserta para staf pegawai Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara.
6. Keluarga serta orang tua peneliti, teman dan sahabat yang selalu memberi doa, dukungan, serta motivasi kepada penulis dalam menyelesaikan skripsi ini.
7. Semua Pihak yang telah membantu penulis dalam penyelesaian skripsi ini yang tidak bisa disebutkan satu per satu.

Penulis sadar bahwa skripsi ini tidak luput akan kesalahan serta kekurangan. Meskipun begitu, penulis berharap agar skripsi ini tetap bermanfaat

setelah dibaca dan mungkin dapat digunakan sebagai referensi bagi penelitian dimasa mendatang.

Medan, 9 Juli 2024

Penulis,

A handwritten signature in black ink, appearing to read 'Niken Alfrido'.

Niken Alfrido Donatus Siboro

171401115

**PERBANDINGAN METODE NAÏVE BAYES DAN KNN DALAM  
PENERAPAN DATA MINING UNTUK KLASIFIKASI MENU  
POTENSIAL ( STUDI KASUS : ATEKU KOPI MEDAN )**

**ABSTRAK**

Perkembangan usaha kuliner, khususnya di sektor kafe, menghadapi persaingan yang ketat sehingga memerlukan strategi pemasaran yang efektif. Sistem Promosi yang dilakukan tiap pelaku usaha untuk meningkatkan penjualan terkadang masih kurang efisien. Penelitian ini bertujuan untuk membangun sistem berbasis web yang menggunakan metode Naïve Bayes dan K-Nearest Neighbor (KNN) dalam klasifikasi menu potensial di Ateku Kopi Medan. Metodologi yang digunakan meliputi studi pustaka, pengumpulan data penjualan dari Januari 2024 hingga April 2024, implementasi sistem menggunakan PHP, serta pengujian sistem dengan evaluasi menggunakan confusion matrix. Data penjualan sebanyak 428 transaksi digunakan sebagai dataset yang dibagi menjadi data training (80%) dan data testing (20%). Sistem ini mempertimbangkan kriteria seperti harga, jumlah terjual dan ada atau tidaknya diskon. Hasil penelitian menunjukkan bahwa sistem yang dikembangkan dapat memberikan hasil yang akurat dan efektif dalam mencari menu potensial yang ada, mengurangi risiko kesalahan dan meningkatkan efisiensi dan dapat mengidentifikasi menu dengan potensi tinggi, sedang, atau rendah, yang berguna untuk meningkatkan penjualan dan efisiensi stok bahan makanan dengan akurasi yang didapat sebesar 90.58% untuk metode Naïve Bayes dan 88.23% untuk K-Nearest Neighbor (KNN).

**Kata kunci:** Naïve Bayes, K-Nearest Neighbour, Klasifikasi, *Data Mining*

**COMPARISON OF NAÏVE BAYES AND KNN METHODS IN DATA MINING  
APPLICATION FOR POTENTIAL MENU CLASSIFICATION (CASE  
STUDY: ATEKU KOPI MEDAN)**

**ABSTRACT**

The growth of the culinary business, particularly in the café sector, faces intense competition, necessitating effective marketing strategies. The promotional systems employed by business operators to increase sales are sometimes inefficient. This research aims to develop a web-based system using Naïve Bayes and K-Nearest Neighbor (KNN) methods to classify potential menus at Ateku Kopi Medan. The methodology includes literature review, sales data collection from January 2024 to April 2024, system implementation using PHP, and system testing with evaluation using a confusion matrix. A dataset of 428 sales transactions is used, divided into training data (80%) and testing data (20%). The system considers criteria such as price, number sold, and whether a discount is present. The results of the study show that the developed system can provide accurate and effective results in finding potential menus, reducing the risk of errors and increasing efficiency and can identify menus with high, medium, or low potential, which is useful for increasing sales and efficiency of food stock with an accuracy of 90.58% for the Naïve Bayes method and 88.23% for K-Nearest Neighbor (KNN).

**Keywords:** Naïve Bayes, K-Nearest Neighbor, Classification, Data Mining



## DAFTAR ISI

PERSETUJUAN .....	ii
PERNYATAAN .....	iii
PENGHARGAAN .....	iv
ABSTRAK .....	vi
ABSTRACT .....	vii
DAFTAR ISI .....	viii
BAB 1 PENDAHULUAN .....	1
1.1.    Latar Belakang.....	1
1.2.    Rumusan Masalah.....	3
1.3.    Batasan Masalah .....	3
1.4.    Tujuan Penelitian.....	3
1.5.    Manfaat Penelitian.....	4
1.6.    Metodologi Penelitian.....	4
1.7.    Sistematika Penulisan .....	5
BAB 2 LANDASAN TEORI.....	6
2.1. <i>Data Mining</i> .....	6
2.2.    K-Nearest Neighbor.....	7
2.3.    Naïve Bayes .....	9
2.4.    Pelanggan.....	10
2.5.    Menu.....	11
2.6.    Tinjauan Penelitian .....	12
BAB 3 ANALISIS DAN PERANCANGAN .....	14
3.1.    Analisis Sistem .....	14
3.1.1 <i>Analisis Masalah</i> .....	14
3.2.    Jenis dan Sumber Data .....	15
3.3.    Dataset .....	15

3.4.	<i>Pre-processing</i> Dataset.....	15
3.5.	Klasifikasi dan Training Model .....	15
3.6.	Hasil Training .....	16
3.7.	Metode Penelitian .....	16
3.8.	Evaluasi .....	18
3.9.	Pemodelan Sistem.....	20
3.8.1	<i>Use Case Diagram</i> .....	20
3.8.2	<i>Class Diagram</i> .....	21
3.8.3	<i>Activity Diagram</i> .....	22
3.10.	Perancangan Basis Data.....	24
3.10.1	<i>Rancangan Relasi Antar Tabel</i> .....	25
3.11.	Perancangan Antarmuka ( <i>Interface</i> ).....	28
BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM.....		33
4.1.	Implementasi Kebutuhan Sistem.....	33
4.1.1	<i>Perangkat Keras (Hardware)</i> .....	33
4.3.2	<i>Perangkat Lunak (Software)</i> .....	33
4.2.	Implementasi Data .....	33
4.3.	Implementasi Sistem.....	34
4.3.1	<i>Halaman Utama</i> .....	34
4.3.2	<i>Halaman Atribut</i> .....	35
4.3.3	<i>Halaman Nilai Atribut</i> .....	35
4.3.4	<i>Halaman Dataset</i> .....	36
4.3.5	<i>Implementasi Tambah Dataset</i> .....	36
4.3.6	<i>Halaman Update Dataset</i> .....	37
4.3.7	<i>Halaman Perhitungan Data</i> .....	38
4.4.	Perhitungan Metode dalam Sistem .....	39
4.4.1	<i>Perhitungan dengan Naïve Bayes</i> .....	41

4.4.2	<i>Perhitungan dengan KNN</i> .....	44
4.5	Pengujian Black Box .....	51
4.6	Hasil Evaluasi .....	52
4.6.1	<i>Akurasi Naïve Bayes</i> .....	52
4.6.2	<i>Akurasi K-Nearest Neighbor</i> .....	54
BAB 5 KESIMPULAN DAN SARAN .....		55
5.4	Kesimpulan .....	55
5.5	Saran .....	56
Daftar Pustaka .....		57

## Daftar Tabel

Tabel 3. 1 Tabel <i>Confusion Matrix</i> .....	19
Tabel 3. 2 Struktur Tabel Admin .....	25
Tabel 3. 3 Struktur Tabel Atribut Naïve Bayes .....	26
Tabel 3. 4 Struktur Tabel Atribut KNN .....	26
Tabel 3. 5 Struktur Tabel Dataset Naïve Bayes .....	27
Tabel 3. 6 Struktur Tabel Dataset KNN.....	27
Tabel 3. 7 Struktur Tabel Nilai Atribut Naïve Bayes .....	28
Tabel 3. 8 Struktur Tabel Nilai Atribut KNN .....	28
Tabel 4. 1 Tabel Data Transaksi Penjualan.....	40
Tabel 4. 2 Tabel Atribut Perhitungan Kedua Metode .....	41
Tabel 4. 3 Tabel Data Setelah Nilai Atribut Diskon Dirubah.....	45
Tabel 4. 4 Tabel Data Setelah Normalisasi.....	47
Tabel 4. 5 Tabel Pengurutan Jarak Terdekat.....	50
Tabel 4. 6 Tabel 5 Data dengan Jarak Terdekat.....	51
Tabel 4. 7 Tabel Pengujian <i>Black Box</i> .....	51
Tabel 4. 8 Tabel Pengujian <i>Confusion Matrix</i> Naïve Bayes.....	53
Tabel 4. 9 Tabel Pengujian <i>Confusion Matrix</i> K-Nearest Neighbor.....	54

## Daftar Gambar

Gambar 2. 1 Tahapan <i>Knowledge Discovery in Database</i> .....	6
Gambar 3. 1 Diagram <i>Ishikawa</i> .....	14
Gambar 3. 2 Alur Penelitian Naïve Bayes .....	16
Gambar 3. 3 Alur Penelitian K-Neares Neighbor .....	18
Gambar 3. 4 <i>Use Case Diagram</i> .....	21
Gambar 3. 5 <i>Class Diagram</i> .....	21
Gambar 3. 6 <i>Activity Diagram</i> Atribut.....	22
Gambar 3. 7 <i>Activity Diagram</i> Nilai Atribut.....	23
Gambar 3. 8 <i>Activity Diagram</i> Dataset.....	23
Gambar 3. 9 <i>Activity Diagram</i> Perhitungan.....	24
Gambar 3. 10 Rancangan Relasi Antar Tabel .....	25
Gambar 3. 11 Desain Data Atribut.....	29
Gambar 3. 12 Desain Nilai Atribut .....	30
Gambar 3. 13 Desain Halaman Dataset .....	31
Gambar 3. 14 Desain Halaman Perhitungan.....	32
Gambar 4. 1 Halaman Utama.....	34
Gambar 4. 2 Halaman Atribut.....	35
Gambar 4. 3 Halaman Nilai Atribut .....	35
Gambar 4. 4 Halaman Dataset .....	36
Gambar 4. 5 Halaman Tambah Dataset .....	37
Gambar 4. 6 Halaman <i>Update</i> Dataset.....	37
Gambar 4. 7 Halaman Input Form Perhitungan.....	38
Gambar 4. 8 Halaman Hasil Perhitungan Data <i>Testing</i> .....	39

## **BAB 1**

### **PENDAHULUAN**

#### **1.1. Latar Belakang**

Perkembangan dunia usaha di bidang kuliner berkembang sangat pesat di era *modern* ini. Strategi pemasaran yang tepat dalam persaingan bisnis kuliner dituntut bagi para pelaku bisnis untuk bisa melakukan inovasi pada perkembangan dunia kuliner secara konsisten agar dapat bertahan di bisnis kuliner. Berhasil atau tidak bisnis kuliner dilihat dari kegiatan dalam pemasaran untuk meningkatkan pendapatan, keuntungan usaha serta memperoleh pangsa pasar konsumen (Putri, 2022). Salah satu bidang kuliner yang kemunculannya sangat *massive* adalah usaha di bidang kafe yang menyasar segmen anak muda.

Ateku Kopi merupakan salah satu pelaku usaha kuliner di bidang kafe yang sudah berjalan di kota Medan sejak tahun 2019. Dalam perkembangan usahanya Ateku Kopi mengalami berbagai tantangan usaha khususnya di bidang kafe yang memiliki banyak pesaing sehingga harus tetap *exist* dan berjalan agar mendapatkan peningkatan laba usaha.

Berdasarkan penelitian independen yang dilakukan Toffin, jumlah kedai kopi / kafe di Indonesia hingga Agustus 2019 mencapai lebih dari 2.950 gerai, meningkat hampir tiga kali lipat dibandingkan pada 2016, yang hanya 1.000 gerai. Hasil laporan didapatkan market value yang dihasilkan mencapai Rp.4,8 triliun. Dari data tersebut menjadikan *coffee shop* akhir-akhir ini membuat kalangan muda secara *massive* menjadi penikmat kopi dadakan.

Segmen pelanggan pada kafe Ateku Kopi menyasar segmen anak muda dan keluarga, sehingga menu yang ditawarkan tersedia berbagai macam jenis dari jenis minuman kopi dan minuman lainnya, camilan, makanan ringan hingga berat. Dalam hal ketersediaan menu Ateku Kopi ingin menganalisa ketersediaan menu selama ini apakah efektif sebagai salah satu *support* penjualan yang disukai konsumen atau tidak. Oleh karena itu perlu dilakukan langkah analisa produk menu yang mampu melakukan perhitungan berdasarkan data penjualan selama ini, apakah penjualan yang dilakukan efektif tidak. Hal ini bertujuan agar dari data yang didapat nantinya dapat dijadikan strategi penjualan kedepannya, sehingga ketersediaan menu pada Ateku

Kopi Medan benar-benar dibutuhkan dan disukai oleh konsumen sehingga dapat meningkatkan laba penjualan usaha.

Proses analisa penjualan tentu harus dilakukan berbagai langkah cara baik secara teknis maupun non teknis untuk meningkatkan penjualan ke konsumen. Salah satu langkah yang dapat dilakukan adalah dengan melakukan strategi analisa minat konsumen dalam melakukan pembelian varian menu makanan. Dengan mengetahui menu makanan mana yang diminati konsumen tentu akan membuat usaha Ateku Kopi Medan bisa meningkatkan kualitas layanan terhadap menu tersebut.

Metode yang dapat diterapkan untuk melakukan analisa minat konsumen terhadap menu di Ateku Kopi Medan adalah dengan menerapkan metode klasifikasi Naïve Bayes. Naïve bayes merupakan salah satu metode klasifikasi yang sangat sederhana. Teorema bayes terbukti memiliki akurasi dengan kecepatan tinggi saat diaplikasikan ke dalam basis data dengan jumlah data yang besar. Teorema bayes mengamsusikan ada atau tidak adanya fitur dari kelas tidak saling berhubungan. Oleh karena itu peneliti menggunakan teorema ini karena metode ini sangat cocok digunakan dalam klasifikasi dengan membutuhkan data *training* yang kecil untuk menentukan jumlah parameter yang diperlukan dalam proses pengklasifikasian (Putro, 2020).

Selain metode Naïve Bayes akan diterapkan juga metode K-Nearest Neighbor (KNN) untuk melihat perbandingan akurasi data, seberapa tinggi akurasi antara kedua metode sehingga dapat ditetapkan untuk analisa penjualan menu produk nantinya. Algoritma K-Nearest Neighbor (KNN) merupakan algoritma *machine learning* sederhana dan mudah diterapkan yang dapat digunakan untuk pemecahan masalah regresi dan klasifikasi.

Berdasarkan paparan di atas, maka penulis ingin melakukan penelitian di Ateku Kopi Medan dengan melakukan pengambilan data penjualan untuk dilakukan klasifikasi dengan Naïve Bayes dan algoritma pembanding K-Nearest Neighbor (KNN). Diharapkan dari hasil penelitian tersebut dapat bermanfaat sebagai masukan di perusahaan sehingga kedepan dapat dijadikan analisa penjualan perusahaan untuk meningkatkan omset penjualan perusahaan.

### 1.2. Rumusan Masalah

Melakukan proses peninjauan penjualan produk yang ada membutuhkan waktu dan belum adanya acuan untuk dilakukannya dasar untuk sistem promosi untuk meningkatkan penjualan kedepannya serta mengurangi resiko kekurangan ataupun kelebihan stok bahan makanan penjualan. Rumusan masalah ini mengarah pada pencarian dan untuk pengklasifikasian untuk mencari produk potensial yang ada.

### 1.3. Batasan Masalah

Agar penelitian yang dilakukan tidak terlalu melebar, maka dilakukan pembatasan masalah yang akan diteliti. Adapun batasan masalah dalam penelitian ini adalah:

1. Sistem yang dibangun menampilkan hasil analisa sesuai dengan metode masing-masing yaitu metode Naïve Bayes dan K-Nearest Neighbor (KNN).
2. Objek penelitian dilakukan di Ateku Kopi Medan.
3. Data yang diambil adalah data penjualan Ateku Kopi Medan periode bulan Januari 2024 – April 2024.
4. Data penjualan yang diambil sebagai atribut dalam penelitian ini berupa total penjualan, diskon dan harga dari penjualan tiap produk.
5. Implementasi program dilakukan dengan pemrograman PHP dan DBMS.
6. Hasil dari penerapan program merupakan analisa klasifikasi berdasarkan perhitungan Naive Bayes dengan algoritma pembanding K-Nearest Neighbor (KNN).

### 1.4. Tujuan Penelitian

1. Membangun sistem berbasis *website* untuk dapat menentukan menu potensial berdasarkan penjualan produk yang ada sebelumnya.
2. Mengimplementasikan algoritma Naïve Bayes dan K-Nearest Neighbor dalam sistem untuk klasifikasi produk potensial.
3. Melakukan analisa algoritma Naïve Bayes dengan algoritma perbandingan K-Nearest Neighbor (KNN) untuk klasifikasi menu produk potensial di Ateku Kopi Medan.



### **1.5. Manfaat Penelitian**

Penelitian ini juga diharapkan memiliki manfaat bagi beberapa pihak, antara lain:

1. Membantu manajemen perusahaan khususnya di Ateku Kopi Medan untuk melakukan analisa penjualan usaha.
2. Membantu manajemen perusahaan dalam melakukan strategi promosi kedepan berdasarkan penjualan produk yang sudah dilakukan.

### **1.6. Metodologi Penelitian**

Adapun metodologi yang akan digunakan dalam penelitian ini adalah sebagai berikut:

#### **1. Studi Pustaka**

Tahapan pertama dalam penelitian ini, akan dilakukan studi pustaka yang bertujuan untuk mendapatkan pemahaman yang lebih mendalam terkait metode algoritma Naïve Bayes dan K-Nearest Neighbor (KNN).

#### **2. Pengumpulan Data**

Pengumpulan data dilakukan setelah tahap studi pustaka, pada tahap ini dilakukan pengumpulan data transaksi di Ateku Kopi Medan.

#### **3. Implementasi Sistem**

Pada tahapan ini dilakukan penerapan kedalam bentuk program untuk masing-masing algoritma. Pembuatan program dilakukan dengan bahasa pemrograman PHP dengan database Mysql dengan *tools* Xampp dan Visual Studio Code.

#### **4. Pengujian Sistem**

Untuk pengujian sistem, perlu dilakukan evaluasi agar sistem dapat berjalan sesuai dengan yang diharapkan.

## 5. Dokumentasi

Pada tahap ini dilakukan dokumentasi dan penulisan laporan dari hasil tahapan sebelumnya untuk menunjukkan proses dan hasil penelitian.

### 1.7. Sistematika Penulisan

Terdiri dari beberapa bagian utama yang dijelaskan, sebagai berikut:

#### **BAB 1 PENDAHULUAN**

Membahas latar belakang penelitian dan juga terdapat rumusan masalah yang akan diselesaikan, batasan masalah yang sudah ditentukan, tujuan penelitian, metode penelitian, dan juga sistematika penulisan.

#### **BAB 2 LANDASAN TEORI**

Penjelasan tentang definisi dan teori-teori yang bersangkutan dengan penelitian ini, yaitu *data mining*, metode Naïve Bayes dan K-Nearest Neighbor (KNN).

#### **BAB 3 ANALISIS DAN PERANCANGAN**

Membahas tentang analisis dan perancangan sistem terhadap masalah yang diteliti dan perancangan sistem yang dibangun untuk solusi terhadap masalah tersebut.

#### **BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM**

Membahas tentang implementasi sistem yang dibuat berdasarkan perancangan yang ada. Kemudian dilanjutkan dengan pengujian untuk evaluasi terhadap sistem yang dibuat.

#### **BAB 5 KESIMPULAN DAN SARAN**

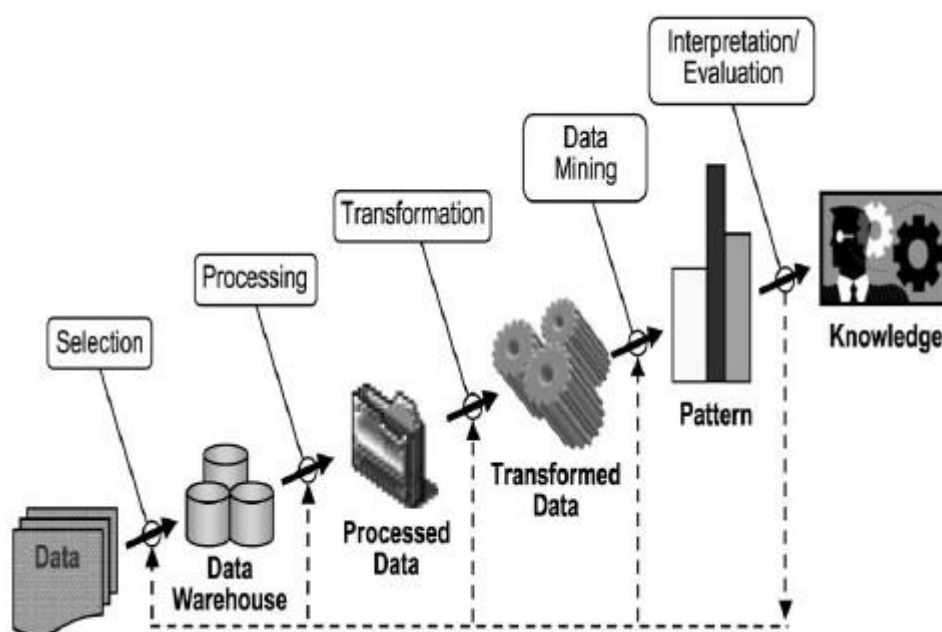
Penulis akan memberikan kesimpulan dari keseluruhan penelitian yang sudah dilakukan dan saran dari hasil penelitian yang bermanfaat untuk pengembangan penelitian berikutnya.

## BAB 2

### LANDASAN TEORI

#### 2.1. Data Mining

*Data mining* adalah salah satu teknik dalam pengolahan data yang mengidentifikasi hubungan tersembunyi dari data yang tidak diketahui oleh pengguna dan menyajikannya dalam bentuk yang mudah dipahami. Dengan demikian, hubungan data tersebut dapat digunakan sebagai dasar untuk pengambilan keputusan. *Data mining* terutama digunakan untuk menemukan pengetahuan dalam basis data yang besar, sehingga sering disebut *Knowledge Discovery in Databases* (KDD) (Putro et al., 2020), atau dikenal juga sebagai penemuan pengetahuan.



**Gambar 2. 1** Tahapan *Knowledge Discovery in Database*

##### 1. Seleksi Data

Pemilihan (seleksi) data yang berasal dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil

seleksi yang digunakan untuk proses *data mining* akan disimpan dalam suatu berkas, terpisah dari basis data operasional yang ada.

## 2. *Pre-processing* / Pembersihan

Sebelum proses *data mining* akan dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD. Proses *cleaning* mencakup antara lain membuang duplikasi data yang ada, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan pada data.

## 3. Transformasi

*Coding* adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

## 4. *Data Mining*

*Data mining* adalah proses mencari pola-pola atau informasi menarik dalam data yang diteliti dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

## 5. Evaluasi / Interpretasi

Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang lebih mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut interpretasi. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

### 2.2. K-Nearest Neighbor

K-Nearest Neighbor (KNN) adalah metode yang menggunakan algoritma supervised untuk mengklasifikasikan *instance* baru berdasarkan mayoritas kategori dari K tetangga terdekat. Tujuan dari algoritma KNN adalah untuk mengklasifikasikan objek

baru berdasarkan atributnya dan data pelatihan. Hasil klasifikasi dari sampel uji baru ditentukan oleh mayoritas kategori dari K tetangga terdekat tersebut.

Algoritma K-Nearest Neighbor (KNN) adalah metode untuk mengklasifikasikan objek berdasarkan data pelatihan yang memiliki jarak terdekat dengan objek tersebut. KNN adalah algoritma pembelajaran *supervised* di mana hasil klasifikasi *instance* baru ditentukan oleh mayoritas kategori dari K tetangga terdekat. Kelas yang paling sering muncul di antara tetangga-tetangga ini akan menjadi kelas hasil klasifikasi. Kedekatan diukur menggunakan metrik jarak, seperti jarak *euclidean*, yang dapat dihitung menggunakan persamaan berikut:

$$D_{xy} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Keterangan :

$D$  : jarak kedekatan

$x$  : data training

$y$  : data testing

$i$  : Atribut individu antara 1 sampai dengan  $n$

$n$  : jumlah atribut individu antara 1 sampai dengan ke  $n$

Berikut adalah Langkah untuk melakukan perhitungan KNN :

1. Tentukan nilai dari K (jumlah tetangga terdekat).
2. Hitung kuadrat jarak *euclidean* dari setiap objek terhadap data sampel yang diberikan menggunakan persamaan 1.
3. Urutkan objek-objek tersebut ke dalam kelompok yang memiliki jarak *euclidean* terkecil.
4. Kumpulkan kategori  $Y$  (klasifikasi tetangga terdekat).
5. Prediksi nilai *query instance* berdasarkan kategori tetangga terdekat yang paling sering muncul

### 2.3. Naïve Bayes

Naïve Bayes adalah metode klasifikasi yang sederhana dimana menggunakan teorema Bayes dengan asumsi bahwa semua fitur yang ada tidak saling berhubungan. Algoritma Naïve Bayes ini memanfaatkan probabilitas keseluruhan, yaitu probabilitas dokumen terhadap kategori (prior). Teks kemudian akan dikategorikan berdasarkan probabilitas maksimum (posterior). Metode ini mengasumsikan bahwa keberadaan atau ketiadaan suatu fitur dalam sebuah kelas tidak bergantung pada keberadaan atau ketiadaan dari fitur lainnya (Kurniawan et al., 2017). Klasifikasi yang mampu memprediksi probabilitas setiap kelas adalah Naïve Bayes. Naïve Bayes merupakan klasifikasi yang cukup sederhana dan merupakan algoritma klasifikasi yang efisien secara komputasi dan memiliki akurasi yang baik untuk dataset yang besar.

Naïve Bayes memiliki kelemahan karena asumsi bahwa atribut atau fitur bersifat independen yang sering kali tidak akurat, yang mengakibatkan estimasi probabilitas yang kurang optimal. Dengan menggunakan metode pembobotan terhadap tiap atribut bisa menjadi salah satu opsi untuk mengatasi kelemahan ini yang nantinya diharapkan dapat meningkatkan akurasi dari Naïve Bayes. Kegunaan dari Naïve Bayes:

1. Mengklasifikasikan dokumen teks, seperti artikel berita atau makalah akademis.
2. Sebagai metode pembelajaran mesin yang didasarkan pada probabilitas.
3. Untuk deteksi diagnosis otomatis.
4. Mendeteksi spam atau menyaring email spam.

Rumus yang digunakan dalam naïve bayes adalah :

$$P(H | X) = \frac{P(X | H).P(H)}{P(X)}$$

dimana:

X : Data dengan kelas yang tidak diketahui

H : Hipotesis bahwa data tersebut termasuk dalam kelas tertentu

$P(H)$  : Probabilitas awal dari hipotesis H (prior probability)

$P(X)$  : Probabilitas X

$P(H/X)$  : Probabilitas hipotesis H berdasarkan kondisi X (probabilitas posterior)

$P(X/H)$  : Probabilitas X diberikan hipotesis H

## 2.4. Pelanggan

Pelanggan adalah individu yang rutin membeli dan menggunakan produk, baik berupa barang maupun jasa. Mereka adalah konsumen yang berinteraksi, baik secara langsung maupun tidak langsung, dengan perusahaan bisnis yang ada.

Menurut Haryono Budi (2016:24), pelanggan adalah individu atau organisasi yang telah melakukan pembelian secara efektif. Pelanggan adalah setiap orang yang menuntut penyediaan layanan dari perusahaan dengan standar kualitas tertentu, yang pada gilirannya mempengaruhi kinerja perusahaan tersebut. Dengan kata lain, pelanggan adalah orang-orang atau pembeli yang tidak bergantung pada produk, melainkan produk yang bergantung pada mereka. Oleh karena itu, karena pelanggan adalah pembeli atau pengguna suatu produk, mereka harus diberikan kepuasan.

Jenis-Jenis pelanggan secara garis besar dapat dibagi menjadi tiga jenis pelanggan, yaitu:

### 1. Pelanggan internal

Pelanggan internal (*internal customer*) merupakan individu atau pengguna produk yang berada di dalam perusahaan dimana individu ini memiliki pengaruh terhadap maju mundurnya sebuah perusahaan.

### 2. Pelanggan perantara

Pelanggan perantara (*intermediate customer*) adalah individu atau entitas yang bertindak sebagai perantara produk, bukan sebagai konsumen akhir. Contoh pelanggan perantara mencakup distributor seperti agen koran yang menjual koran dan toko buku yang memasarkan buku.

### 3. Pelanggan eksternal

Pelanggan eksternal (*external customer*) adalah individu atau kelompok yang menggunakan produk (barang atau jasa) yang dihasilkan oleh sebuah perusahaan.

Pelanggan eksternal ini berfungsi sebagai pelanggan nyata atau konsumen akhir. (Dewa & Setyohadi, 2017).

## **2.5. Menu**

Dalam bisnis restoran, menu merupakan alat komunikasi antara restoran dan pelanggannya. Ini tercermin dari tingginya tingkat penjualan yang dicapai oleh restoran. Manajemen dari pemasaran, yang merupakan seni dan ilmu dalam memilih target dari pasar serta untuk meraih, mempertahankan, dan mengembangkan pelanggan dengan cara menciptakan dan mengkomunikasikan nilai-nilai unggul bagi pelanggan, membantu manajemen dalam merencanakan dan mempromosikan produk atau menu yang ditawarkan. Bauran pemasaran adalah satu perangkat dimana terdiri daripada produk, harga, promosi, serta distribusi, yang semuanya dapat menentukan tingkat keberhasilan dari pemasaran dengan tujuan mendapatkan respon yang diinginkan dari pasar sasaran yang ada. Perusahaan juga dapat mempengaruhi konsumen dengan produk yang mereka tawarkan, sehingga nantinya produk tersebut harus dibuat dengan baik untuk menarik minat konsumen agar membeli atau mengonsumsi produk tersebut (Suartana et al., 2020).

Menu adalah daftar atau rangkaian pilihan makanan dan minuman yang ditawarkan oleh sebuah restoran, kafe, atau tempat makan lainnya kepada pelanggan. Menu biasanya mencakup berbagai jenis hidangan, seperti hidangan pembuka, hidangan utama, hidangan penutup, dan minuman. Menu dapat disajikan dalam bentuk fisik seperti buku menu, papan tulis, atau dalam bentuk digital melalui aplikasi atau website. Menu berfungsi sebagai panduan bagi pelanggan untuk mengetahui pilihan makanan dan minuman yang tersedia. Menu memberikan informasi mengenai nama hidangan, bahan-bahan utama, cara penyajian, dan harga .

Menu memainkan peran penting dalam membentuk pengalaman pelanggan. Menu yang dirancang dengan baik dapat meningkatkan kepuasan pelanggan dengan memberikan pilihan yang beragam dan memenuhi berbagai selera. Sebaliknya, menu yang membingungkan atau kurang informatif dapat menurunkan kualitas pengalaman makan dan mempengaruhi reputasi restoran .



## 2.6. Tinjauan Penelitian

Dalam tahapan ini dicantumkan jurnal penelitian terdahulu tentang metode Naïve Bayes dan metode K-Nearest Neighbor, berikut adalah jurnal penelitian terdahulu yang menjadi acuan penulis :

1. Penelitian yang dilakukan oleh Izzan Arimi mengenai metode KNN untuk memprediksi penjualan produk UMKM pengolahan ikan menunjukkan bahwa metode KNN dengan peramalan time series menghasilkan uji performa dengan nilai RMSE lebih rendah dibandingkan metode lainnya. Rekomendasi yang diusulkan menyebutkan bahwa investasi untuk membeli laptop sebagai alat bantu, dengan nilai *Net Benefit Cost Ratio* sebesar 2,82, adalah layak. Penelitian tersebut juga menggunakan 7 model peramalan, termasuk metode KNN dengan time series, yang menghasilkan model peramalan data *time series* berupa prediksi penjualan selama 12 bulan dan uji performa setiap model peramalan.
2. Penelitian lainnya dilakukan oleh Vania Ariyani Prilia Putri (2022) yang meneliti tentang metode Naïve bayes dan KNN untuk memprediksi harga rumah. Dalam penelitian ini, prediksi harga rumah menggunakan *machine learning* memanfaatkan algoritma Naïve Bayes sebagai representasi metode pembelajaran *eager learning*, dan K-Nearest Neighbor sebagai representasi metode pembelajaran *lazy learning*. Hasil penelitian menunjukkan bahwa model pembelajaran *lazy learning* unggul dalam nilai *accuracy score* dan kecepatan waktu dalam proses pelatihan data dibandingkan dengan model pembelajaran *eager learning*. Meskipun demikian, kedua algoritma yang digunakan dalam penelitian ini belum mampu memprediksi harga rumah dengan akurasi tinggi, karena nilai *mean absolute error percentage* (MAPE) yang diperoleh masuk dalam kategori "cukup" dan bukan "sangat baik".

Berdasarkan tujuan penelitian “Perbandingan Kinerja Algoritme Naïve Bayes dan K-Nearest Neighbor (KNN) Untuk Prediksi Harga Rumah”, diperoleh kesimpulan bahwa model pembelajaran *lazy learning* memiliki kinerja yang lebih unggul dalam nilai akurasi skor serta kecepatan untuk *training* data dengan nilai *accuracy score* dan waktu yang dibutuhkan adalah 0,5714 dan 0,0839 detik menggunakan algoritma K-Nearest Neighbor. Sedangkan *accuracy score* tertinggi yang didapatkan model pembelajaran *eager learning* adalah 0,4 dengan

waktu *training* data tercepat selama 0,1615 detik dengan menggunakan algoritma Naïve Bayes.

3. Penelitian lainnya dilakukan oleh Saifur Rohman Cholil (2020) yang melakukan penelitian tentang Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa. Penelitian tersebut mengungkap bahwa Algoritma KNN dapat menyediakan data yang akurat dan informasi yang diperlukan untuk menyeleksi calon penerima beasiswa. Hasil penelitian menunjukkan bahwa dari 89 data yang diklasifikasikan, 30 orang berhasil terseleksi. Pengujian sistem menggunakan metode *confusion matrix* menunjukkan akurasi sebesar 90,5%. Ini mengindikasikan bahwa Algoritma KNN dapat digunakan untuk mengklasifikasikan seleksi penerimaan beasiswa dengan baik.
4. Penelitian lain oleh Syarli tentang metode Naive Bayes untuk memprediksi kelulusan mahasiswa baru di perguruan tinggi menjelaskan bahwa Naive Bayes dapat melakukan klasifikasi menggunakan metode probabilitas dan statistik, yaitu memprediksi peluang masa depan berdasarkan data masa lalu. Persentase akurasi menunjukkan efektivitas dataset Penerimaan Mahasiswa Baru ketika diterapkan pada metode Naive Bayes *classification*. Implementasi Naive Bayes menggunakan aplikasi WEKA dapat menelusuri karakteristik atribut dari dataset dengan output berupa pilihan kelulusan. Pengelompokan pilihan kelulusan dilakukan berdasarkan atribut terpilih yaitu program studi, pilihan pertama, pilihan kedua, dan nilai rata-rata.

## BAB 3

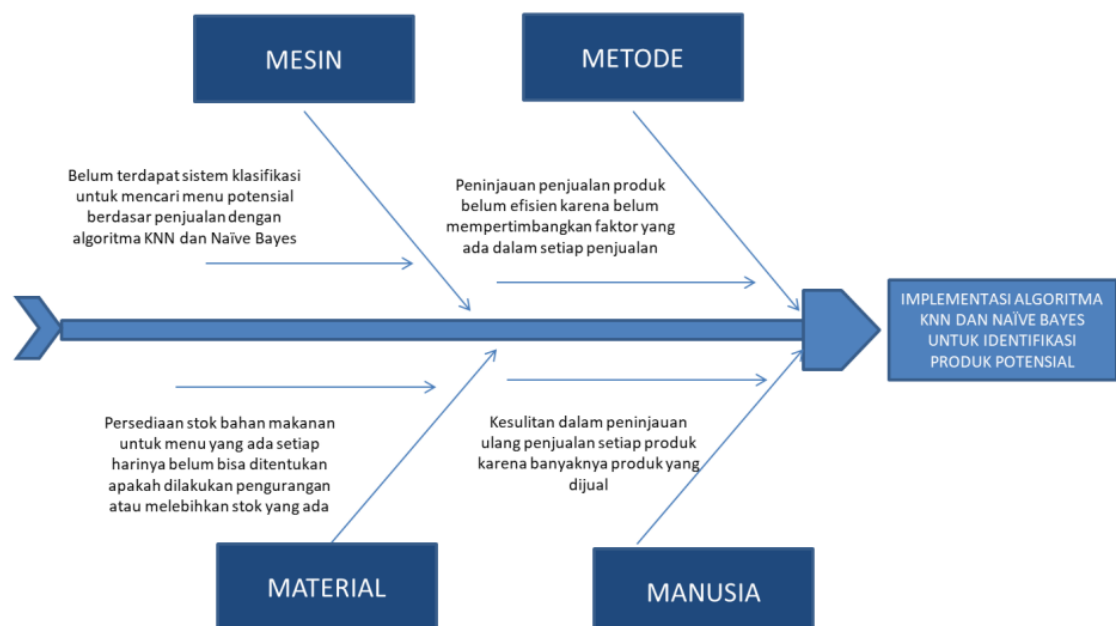
### ANALISIS DAN PERANCANGAN

#### 3.1. Analisis Sistem

Analisis sistem adalah proses mengevaluasi kebutuhan yang diperlukan dalam sebuah sistem untuk mencapai tujuan keberhasilan pembangunan sistem yang dilakukan.

##### 3.1.1 Analisis Masalah

Analisis masalah bertujuan untuk mengidentifikasi setiap faktor-faktor penyebab terjadinya masalah dan menguraikan faktor yang ditemukan untuk menemukan solusi terbaik. Masalah yang ditemukan di penelitian ini yaitu belum ada sebuah sistem yang dapat melakukan klasifikasi menu potensial untuk meningkatkan penjualan kedepannya. Masalah tersebut kemudian dinyatakan dalam diagram *ishikawa* untuk menguraikan sebab-akibat dari setiap faktor yang terlibat dari sebuah sistem.



**Gambar 3. 1** Diagram *Ishikawa*

### 3.2. Jenis dan Sumber Data

Jenis data yang digunakan dalam penelitian ini merupakan data primer. Data Primer yaitu data yang cara memperolehnya langsung ke sumber objek penelitian. Dalam hal ini nantinya penelitian dilakukan secara langsung di Ateku Kopi Medan dengan mengambil data transaksi penjualan selama beberapa periode untuk kebutuhan penelitian.

### 3.3. Dataset

Dataset dalam penelitian ini yang diperoleh dari Ateku Kopi Medan berupa data transaksi penjualan perbulannya dengan periode mulai dari Januari 2024 sampai pada April 2024. Dataset yang digunakan untuk penelitian ini berjumlah 428 dataset dengan atribut seperti harga, diskon dan jumlah transaksi dan dengan pelabelan rendah, sedang, dan tinggi.

### 3.4. *Pre-processing* Dataset

Pada tahap *pre-processing*, dataset yang ada dilakukan pembersihan data. Lalu dilakukan pengintegrasian data dimana data transaksi penjualan tiap bulannya digabungkan. Lalu dilakukan pelabelan terhadap keseluruhan data yang ada dimana dibuat label tinggi untuk penjualan diatas 129, dikarenakan rata-rata penjualan dari keseluruhan menu yaitu berjumlah 129, dan label sedang antara 64-129 yang merupakan rentang setengah dari nilai label tinggi serta label rendah yang jumlah penjualannya dibawah 64.

### 3.5. Klasifikasi dan Training Model

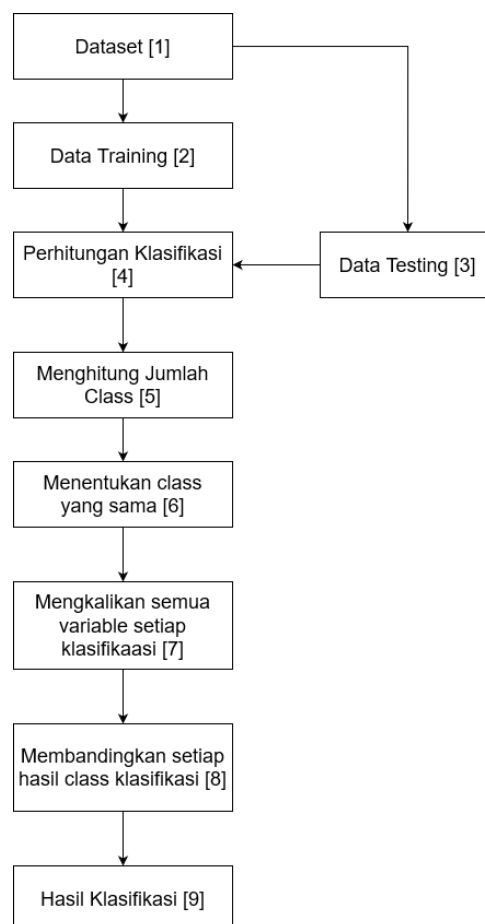
Dataset yang ada akan dilatih dan diuji menggunakan algoritma klasifikasi Naïve Bayes dan KNN. Pelatihan dan pengujian dalam penelitian ini menggunakan rasio data 80:20 dimana jumlah daripada data *training* sebesar 80% dan jumlah dari data *testing* sebesar 20% dari keseluruhan data. Berdasarkan pembagian tersebut maka dari jumlah data sebanyak 428 data, maka akan digunakan sebanyak 343 sebagai data *training* dan 85 data *testing*.

### 3.6. Hasil Training

Untuk mengetahui kesimpulan performa model selama proses pelatihan maka dilakukan evaluasi dengan menggunakan *confusion matrix* untuk memperoleh nilai akurasi dari model tersebut. Proses ini diperlukan agar mengetahui apakah model yang digunakan relevan dan mampu melakukan tugasnya dengan baik untuk mencapai tujuan yang diharapkan.

### 3.7. Metode Penelitian

Pada tahapan ini dilakukan penggunaan metode untuk proses penelitian, seperti yang telah dijelaskan bahwa dalam penelitian ini menggunakan KNN dengan perbandingan metode Naive Bayes dan berikut yaitu alur proses yang dilakukan dengan Naive Bayes dan KNN.

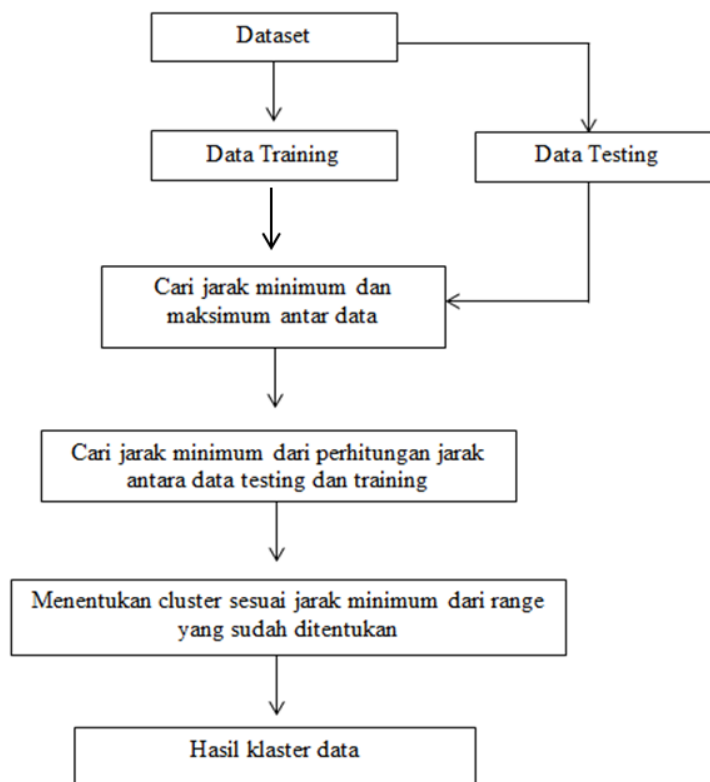


**Gambar 3. 2** Alur Penelitian Naïve Bayes

Berdasarkan proses diatas tahapan – tahapan dapat dijelaskanantara lain:

1. Dataset pada penelitian ini berisi data transaksi penjualan di ateku kopi medan.
2. Data *training* pada penelitian ini yaitu data yang menjadi acuan untuk melakukan perhitungan metode naïve bayes.
3. Data *testing* yaitu data yang diujikan yang nantinya data yang sudah diuji akan dilakukan perhitungan dengan data training.
4. Perhitungan klasifikasi menggunakan metode Naïve Bayes dengan cara menghitung data *training* dan data *testing* yang akan digunakan dalam penelitian ini.
5. Melakukan perhitungan data *class* pada dataset.
6. Menyaring jumlah *class* yang sama untuk diketahui hasil dari nilai angkanya.
7. Menghitung dengan cara mengalikan seluruh variabel dalam setiap klasifikasi yang sudah terbaca dalam sistem.
8. Membandingkan jumlah *class* dan dicari hasil nilai tertinggi dari seluruh *class* yang ada.
9. Nilai tertinggi dari seluruh *class* akan menjadi hasil akhir dari perhitungan Naïve Bayes.

Dan berikut adalah proses penelitian menggunakan metode K-Nearest Neighbor:



**Gambar 3. 3** Alur Penelitian K-Nearest Neighbor

Proses metode K-Nearest Neighbor ini menggunakan dataset dari transaksi penjualan produk. Langkah awal dalam proses perhitungan adalah menggunakan dataset *training* dan *testing* sebagai data untuk dilakukan perhitungan setelah itu sistem akan menentukan data *training* dan data *testing*. Setelah menentukan data *training* dan data *testing*, langkah selanjutnya adalah mencari jarak minimum dan maksimum dari perhitungan jarak pada data *training*. Setelah itu, dilakukan pencarian jarak minimum dari perhitungan tersebut. Setelah mendapatkan jarak antar data, langkah berikutnya adalah menentukan kelas sesuai dengan jarak minimum dalam rentang yang telah ditentukan. Hasil akan di tampilkan setelah semua perhitungan selesai dan menghasilkan klasifikasi data.

### 3.8. Evaluasi

Setelah semua penelitian selesai dilakukan tahapan selanjutnya akan dilakukan

proses evaluasi. Dalam melakukan evaluasi dilakukan dengan menerapkan metode *confusion matrix*. Berikut adalah rumus pengujian dengan metode *Confusion Matrix*:

**Tabel 3. 1** Tabel *Confusion Matrix*

	Benar Negatif	Benar Positif
Prediksi Negatif	(TN)	(FP)
Prediksi Positif	(FN)	(TP)

$$Akurasi = \frac{TN + TP}{TN + TP + FN + FP} \times 100 \%$$



Dimana:

**TP** adalah True Positif, merupakan jumlah dari data positif yang terklasifikasikan dengan benar oleh sistem.

**TN** adalah True Negatif, merupakan jumlah daripada data negatif yang terklasifikasikan dengan benar oleh sistem.

**FN** adalah False Negatif, yaitu jumlah dari data negatif namun terklasifikasikan salah oleh sistem yang ada.

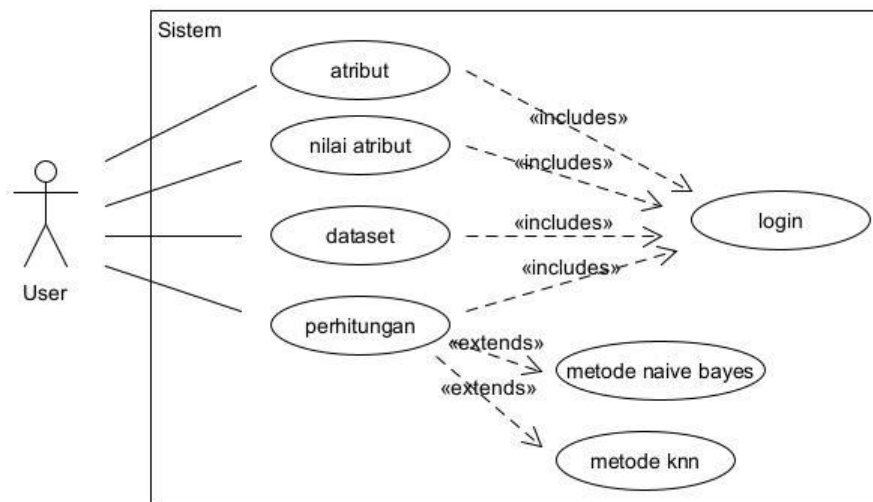
**FP** adalah False Positif, yaitu jumlah daripada data positif namun terklasifikasikan salah oleh sistem.

### 3.9. Pemodelan Sistem

Pemodelan sistem merupakan pengetahuan hubungan dari komponen dalam sebuah sistem dan bagaimana interaksi dari antar komponen tersebut.

#### 3.8.1 Use Case Diagram

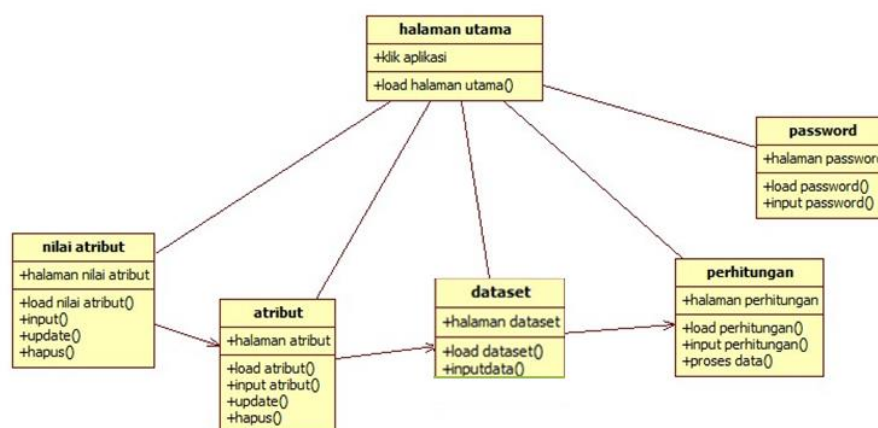
*Use Case Diagram* adalah jenis diagram yang termasuk dalam *Unified Modeling Language* (UML) yaitu bahasa pemodelan sistem yang menggambarkan hubungan antara aktor (pengguna) dengan sistem yang dirancang yang umumnya digambarkan sebagai objek di luar sistem yang memiliki interaksi dengan sistem tersebut, sedangkan usecase menjelaskan tindakan actor yang dapat dilakukan pada sistem.



**Gambar 3.4** Use Case Diagram

### 3.8.2 Class Diagram

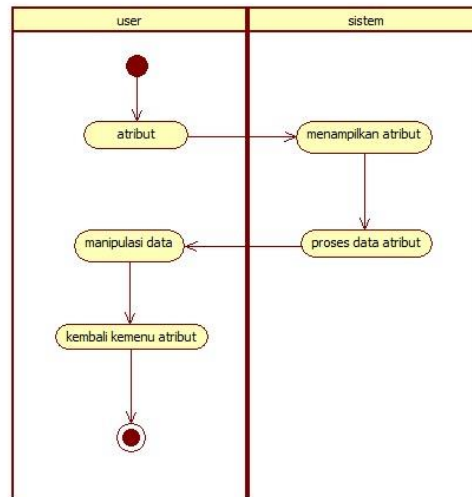
Diagram Kelas digunakan untuk menggambarkan struktur objek dalam sebuah sistem, mencakup objek-objek yang ada dalam sistem tersebut serta hubungan antara objek-objek tersebut.



**Gambar 3.5** Class Diagram

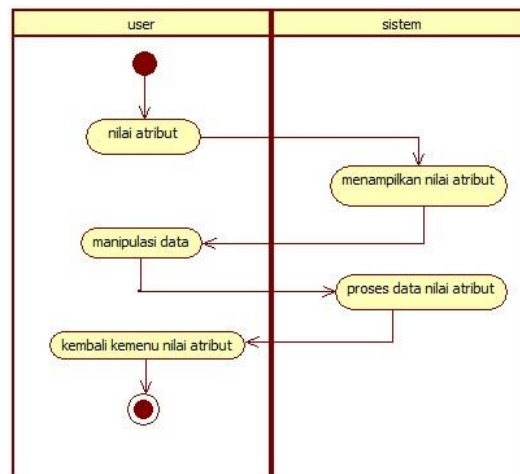
### 3.8.3 Activity Diagram

Diagram aktivitas adalah diagram penggambaran dari proses jalannya sistem dan urutan aktivitas-aktivitas yang mendukung penggambaran tindakan sistem, baik yang bersifat kondisional maupun paralel.



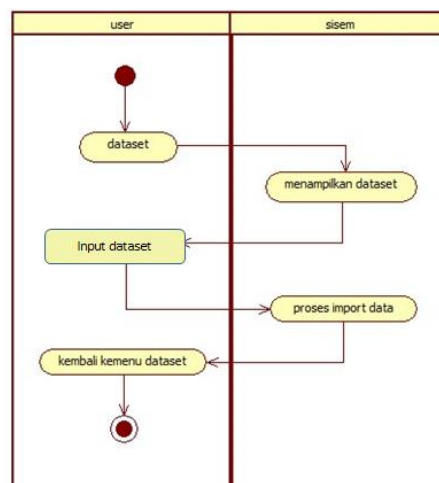
**Gambar 3. 6** Activity Diagram Atribut

Pada gambar diatas jalannya aktivitas sistem dan *user* ketika dijalankan. Aktivitas diagram tersebut berfungsi untuk menggambarkan menu atribut pada aplikasi yang akan dibuat.



**Gambar 3. 7** Activity Diagram Nilai Atribut

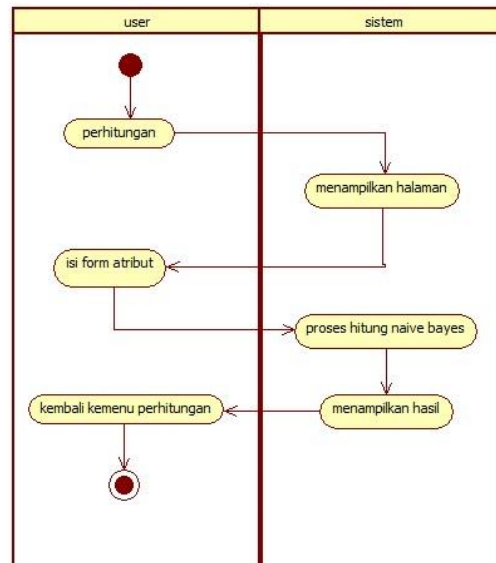
Aktivitas diagram diatas menggambarkan jalannya sistem dan *user* didalam menu nilai atribut. Dalam proses tersebut berfungsi untuk menginputkan data nilai atribut dari data atribut yang telah di inputkan kedalam aplikasi.



**Gambar 3. 8** Activity Diagram Dataset

Pada *activity* diagram diatas menggambarkan fungsi jalannya aktivitas pada menu dataset, menu dataset berfungsi untuk menginputkan data transaksi dirawat dan

nantinya akan disimpan dalam *database* dan ditampilkan pada menu tersebut.



**Gambar 3. 9** Activity Diagram Perhitungan

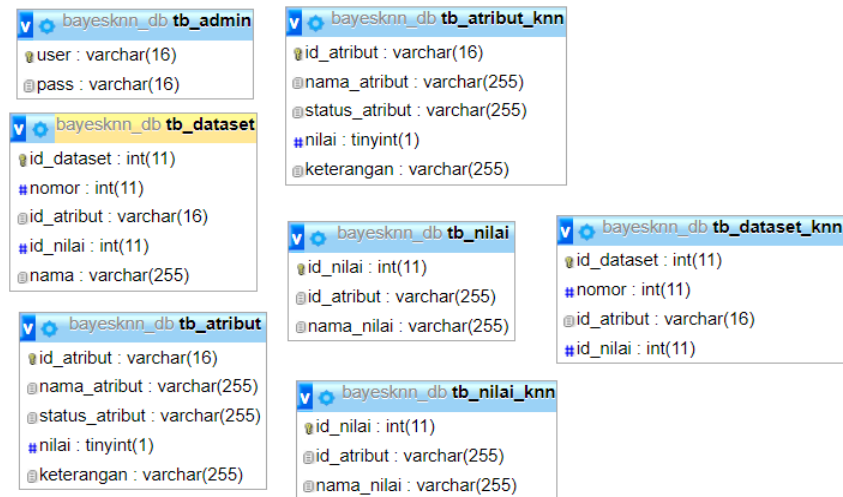
Pada activity diagram diatas menggambarkan fungsi jalannya aktivitas pada menu perhitungan Naïve bayes dan K-Nearest Neighbor. Menu perhitungan berfungsi untuk melakukan data uji perhitungan dimana *user* nantinya menginputkan form yang telah dibuat dalam system, Setelah itu sistem menampilkan hasil perhitungan sesuai dengan jalannya metode Naïve Bayes dan K-nearest Neighbor.

### 3.10. Perancangan Basis Data

Perancangan basis data adalah langkah untuk menetapkan bagaimana data disimpan dan diakses dalam sistem. Dalam penelitian ini, perancangan basis data dilakukan dengan menggunakan database MySQL yang terdiri dari tujuh tabel yang mencakup tabel admin, tabel atribut Naïve Bayes, tabel atribut KNN, tabel dataset Naïve Bayes, tabel dataset KNN, tabel nilai atribut Naïve Bayes dan tabel nilai atribut KNN.

### 3.10.1 Rancangan Relasi Antar Tabel

Dalam rancangan relasi antar tabel akan ditampilkan koneksi antara satu tabel dengan tabel lain yang terlibat dalam hubungan yang saling berinteraksi.



**Gambar 3. 10** Rancangan Relasi Antar Tabel

#### 1. Rancangan Struktur Tabel Admin

Tabel admin berperan sebagai tempat penyimpanan data login, struktur tabel admin akan disajikan dalam Tabel 3.2.

Primary Key : user

**Tabel 3. 2** Struktur Tabel Admin

<i>Field</i>	<i>Type</i>	<i>Length</i>
user	varchar	16
pass	varchar	16

#### 2. Rancangan Struktur Tabel Atribut Naïve Bayes

Tabel atribut Naïve Bayes berperan sebagai tempat penyimpanan data atribut Naïve Bayes, struktur tabel atribut Naïve Bayes akan disajikan dalam Tabel 3.3.

**Tabel 3. 3** Struktur Tabel Atribut Naïve Bayes

<i>Field</i>	<i>Type</i>	<i>Length</i>
id_atribut	varchar	16
nama_atribut	varchar	255
status_atribut	varchar	255
nilai	tinyint	1
keterangan	varchar	255

Primary Key : id\_atribut

### 3. Rancangan Struktur Tabel Atribut KNN

Tabel atribut KNN berperan sebagai tempat penyimpanan data atribut KNN, struktur tabel atribut KNN akan disajikan dalam Tabel 3.4.

**Tabel 3. 4** Struktur Tabel Atribut KNN

<i>Field</i>	<i>Type</i>	<i>Length</i>
id_atribut	varchar	16
nama_atribut	varchar	255
status_atribut	varchar	255
nilai	tinyint	1
keterangan	varchar	255

Primary Key : id\_atribut

### 4. Rancangan Struktur Tabel Dataset Naïve Bayes

Tabel dataset Naïve Bayes berperan sebagai tempat penyimpanan dataset Naïve Bayes, struktur tabel atribut Naïve Bayes akan disajikan dalam Tabel 3.5.

**Tabel 3. 5** Struktur Tabel Dataset Naïve Bayes

<i>Field</i>	<i>Type</i>	<i>Length</i>
id_dataset	int	11
nomor	int	11
id_atribut	varchar	16
id_nilai	int	11

Primary Key : id\_dataset

5. Rancangan Struktur Tabel Dataset KNN

Tabel dataset KNN berperan sebagai tempat penyimpanan data atribut KNN, struktur tabel dataset KNN akan disajikan dalam Tabel 3.6.

**Tabel 3. 6** Struktur Tabel Dataset KNN

<i>Field</i>	<i>Type</i>	<i>Length</i>
id_dataset	int	11
nomor	int	11
id_atribut	varchar	16
id_nilai	int	11

Primary Key : id\_dataset

6. Rancangan Struktur Tabel Nilai Atribut Naïve Bayes

Tabel nilai atribut Naïve Bayes berperan sebagai tempat penyimpanan data nilai atribut Naïve Bayes, struktur tabel nilai atribut Naïve Bayes akan disajikan dalam Tabel 3.7.



**Tabel 3. 7** Struktur Tabel Nilai Atribut Naïve Bayes

<i>Field</i>	<i>Type</i>	<i>Length</i>
id_nilai	int	11
id_atribut	varchar	255
nama_nilai	varchar	255

Primary Key : id\_nilai

#### 7. Rancangan Struktur Tabel Nilai Atribut KNN

Tabel nilai atribut KNN berperan sebagai tempat penyimpanan data nilai atribut KNN, struktur tabel nilai atribut KNN akan disajikan dalam Tabel 3.8.

**Tabel 3. 8** Struktur Tabel Nilai Atribut KNN

<i>Field</i>	<i>Type</i>	<i>Length</i>
id_nilai	int	11
id_atribut	varchar	255
nama_nilai	varchar	255

Primary Key: id\_nilai

### 3.11. Perancangan Antarmuka (*Interface*)

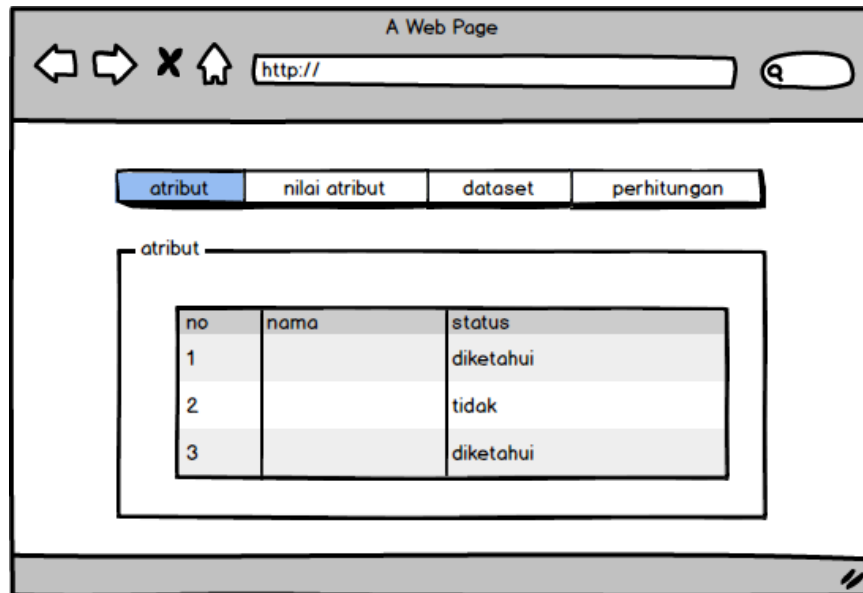
Perancangan sistem adalah fase dalam pengembangan aplikasi yang bertujuan untuk menyusun deskripsi mendetail tentang sistem yang akan dibuat. Tujuan utama dari perancangan sistem adalah untuk menciptakan kerangka aplikasi yang terstruktur dan teratur, memastikan bahwa operasional aplikasi dapat berjalan dengan optimal dan efisien.

Pada tahapan ini dilakukan desain tampilan sistem yang akan dibuat, desain tampilan berfungsi untuk menggambarkan tampilan sistem yang akan dibuat nantinya sebagai gambaran fitur apa saja yang didapati didalam system.

Berikut adalah desain tampilan yang sudah dibuat :

### 1. Halaman atribut

Pada halaman ini nantinya ditampilkan data berupa table data atribut kriteria yang telah ditentukan.



**Gambar 3. 11** Desain Data Atribut

Dalam desain diatas juga nantinya dapat dilakukan manipulasi data berupa input data edit dan hapus data atribut penjualan.

### 2. Desain halaman nilai atribut

Pada desain halaman nilai atribut dilakukan desain dengan menampilkan data nilai atribut Naïve Bayes dan K-Nearest Neighbor, berikut adalah desain yang sudah dibuat.

A Web Page

← → × ↗ http:// 🔍

atribut nilai atribut dataset perhitungan

nilai atribut

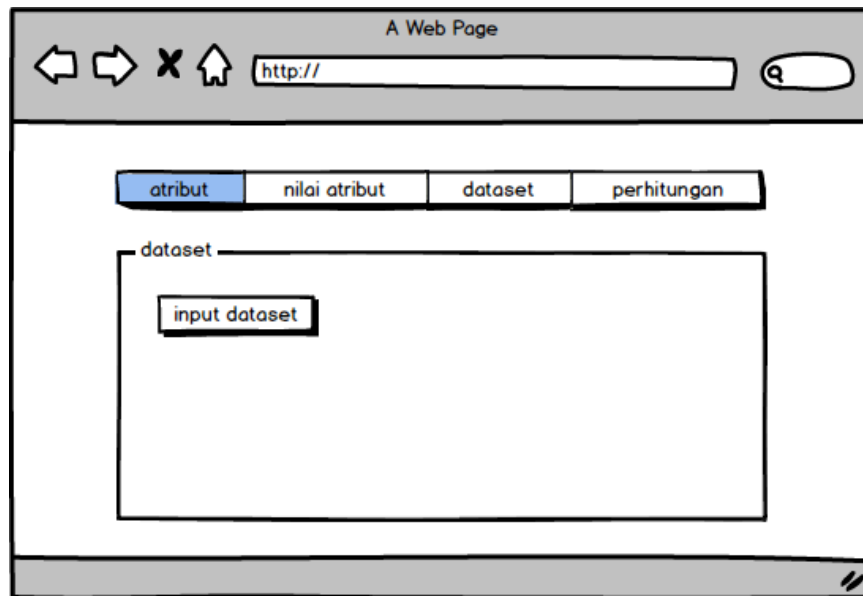
no	nama atribut	nama nilai atribut
1		
2		
3		

**Gambar 3. 12** Desain Nilai Atribut

Desain nilai atribut nantinya berfungsi untuk menginputkan nilai-nilai atribut yang akan dibutuhkan untuk proses perhitungan menggunakan Naïve Bayes dan K-Nearest Neighbor pada dataset.

### 3. Desain halaman dataset

Pada desain ini nantinya difungsikan sebagai halaman untuk menampilkan dataset, berikut adalah desain yang telah dibuat.



**Gambar 3. 13** Desain Halaman Dataset

#### 4. Desain Halaman Perhitungan

Pada desain halaman ini nantinya ditampilkan halaman proses perhitungan Naïve Bayes dan K-Nearest Neighbor, proses perhitungan dijalankan sesuai dengan metode yang telah ditentukan.

Penelitian Naive Bayes

A Web Page

http://

atribut nilai atribut dataset perhitungan

perhitungan

variabel 1

Vaeiabel 2

Vaeiabel 3

hitung

**Gambar 3. 14** Desain Halaman Perhitungan

Setelah proses perhitungan menggunakan metode Naïve Bayes atau K-Narest Neighbor selesai dilakukan maka akan tampil hasil klasifikasi berupa status hasil perhitungan apakah menu produk tersebut masuk kedalam kategori tinggi sedang atau rendah hasil klasifikasinya dimana untuk metode KNN menggunakan nilai  $k = 9$ .

## **BAB 4**

### **IMPLEMENTASI DAN PENGUJIAN SISTEM**

#### **4.1. Implementasi Kebutuhan Sistem**

Dalam pelaksanaan penelitian ini, beberapa komponen *hardware* dan *software* menjadi elemen penting dalam implementasi sistem. Berikut ini adalah perangkat keras dan perangkat lunak yang digunakan dalam penelitian ini:

##### *4.1.1 Perangkat Keras (Hardware)*

Berikut adalah spesifikasi dari perangkat keras dalam tahap sistem implementasi sistem dalam penelitian ini:

1. Processor : Intel Core i5-8265U CPU 1.60GHz 1.80 GHz
2. Sistem Operasi : Windows 10 Home 64-bit
3. SSD : 512 GB
4. RAM : 12 GB

##### *4.3.2 Perangkat Lunak (Software)*

Perangkat lunak yang digunakan untuk pembangunan sistem adalah seperti berikut:

1. Windows 10 Home 64-bit
2. PHP 5.6.5
3. XAMPP Control Panel v3.2.1
4. Visual Studio Code
5. Microsoft Office 2010

#### **4.2. Implementasi Data**

Dalam Dataset yang digunakan untuk penelitian ini merupakan data transaksi penjualan perbulan di Ateku Kopi Medan periode Januari 2024 – April 2024. Dari

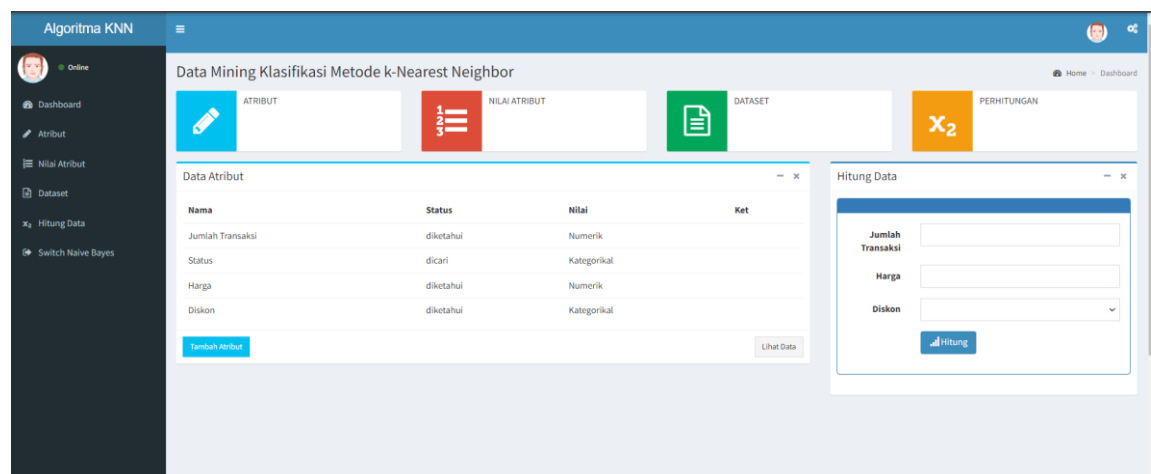
data yang berhasil dikumpulkan, terdapat 428 dataset dengan harga, diskon dan jumlah transaksi sebagai atribut dengan label yang dibagi menjadi rendah, sedang, dan tinggi.

### 4.3. Implementasi Sistem

Pada tahapan ini merupakan tahapan implementasi kedalam bentuk system menggunakan PHP dengan metode Naïve Bayes dan K-Nearest Neighbor. Untuk tampilan masing-masing metode fiturnya sama baik dari segi atribut, parameter dan datasetnya. Berikut ini merupakan bentuk program dan juga keterangan gambar dari aplikasi yang telah dibuat dengan algoritma Naïve Bayes dan K-Nearest Neighbor.

#### 4.3.1 Halaman Utama

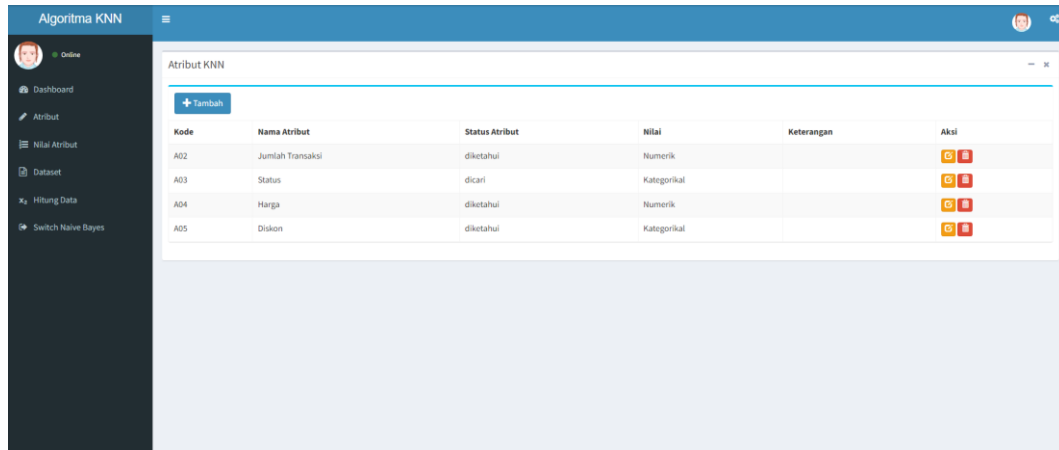
Halaman utama merupakan sebagai halaman awal ketika pengguna berhasil melakukan login, halaman utama berisi beberapa menu antara metode Naïve Bayes dan K-Nearest Neighbor. Berikut merupakan tampilan halaman utama.











**Gambar 4. 1** Halaman Utama

#### 4.3.2 Halaman Atribut

Menu atribut menampilkan atribut yang digunakan didalam sistem yang dibangun, berikut adalah tampilan menu atribut yang dibuat:

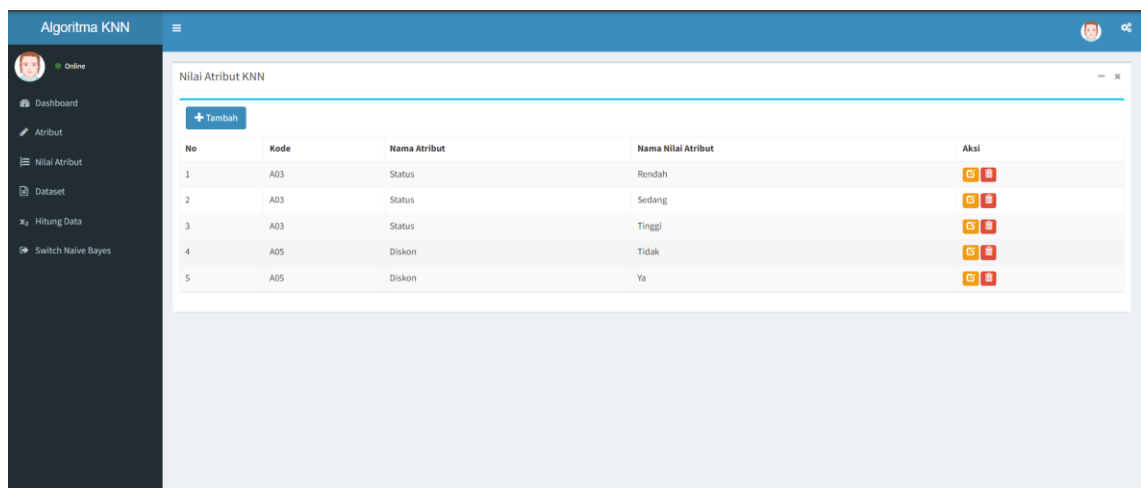












Kode	Nama Atribut	Status Atribut	Nilai	Keterangan	Aksi
A02	Jumlah Transaksi	diketahui	Numerik		 
A03	Status	dicari	Kategorikal		 
A04	Harga	diketahui	Numerik		 
A05	Diskon	diketahui	Kategorikal		 

**Gambar 4. 2** Halaman Atribut

#### 4.3.3 Halaman Nilai Atribut

Di halaman ini berguna untuk input data nilai atribut yang telah ditentukan dalam penelitian, berikut adalah merupakan tampilan nilai atribut yang telah dibuat:



No	Kode	Nama Atribut	Nama Nilai Atribut	Aksi
1	A03	Status	Rendah	 
2	A03	Status	Sedang	 
3	A03	Status	Tinggi	 
4	A05	Diskon	Tidak	 
5	A05	Diskon	Ya	 

**Gambar 4. 3** Halaman Nilai Atribut



Pada halaman diatas ditampilkan informasi atribut nilai yang sudah di inputkan, dapat juga dilakukan pengubahan data berupa tambah data, update data serta hapus data nilai atribut. Halaman tambah nilai atribut terdapat pada metode Naïve Bayes dan K-Nearest Neighbor.

#### 4.3.4 Halaman Dataset

Pada halaman ini berfungsi untuk menampilkan dataset yang sudah di inputkan kedalam aplikasi, Berikut adalah tampilan dataset yang sudah dibuat.

Nomor	Jumlah Transaksi	Harga	Diskon	Status	Aksi
1	54	10000	Tidak	Rendah	
2	42	15000	Ya	Rendah	
3	33	12000	Tidak	Rendah	
4	36	17000	Tidak	Rendah	
5	24	15000	Tidak	Rendah	
6	101	13000	Tidak	Sedang	
7	42	12000	Tidak	Rendah	
8	198	12000	Ya	Tinggi	
9	149	15000	Ya	Sedang	
10	60	12000	Tidak	Rendah	
11	24	14000	Tidak	Rendah	
12	24	11000	Tidak	Rendah	
13	2	15000	Tidak	Rendah	

**Gambar 4.4** Halaman Dataset

Pada halaman ini ditampilkan atribut yang sudah diisi data, terdapat juga fitur merubah data seperti penambahan data, hapus dan *update* data.

#### 4.3.5 Implementasi Tambah Dataset

Pada halaman ini berfungsi untuk menambah dataset dari data pokok yang didapatkan dari objek penelitian, Data set berisi isian form nama menu, harga dan labelnya, berikut adalah hasil halaman yang sudah dibuat.

The screenshot shows the 'Tambah Dataset KNN' form. The left sidebar contains navigation links: Dashboard, Atribut, Nilai Atribut, Dataset, Hitung Data, and Switch Naive Bayes. The main form area has the following fields:

- Nomor \***: Text input with value 429.
- Jumlah Transaksi \***: Text input.
- Harga \***: Text input.
- Diskon \***: Dropdown menu.
- Status \***: Dropdown menu.

At the bottom of the form are two buttons: 'Simpan' (Save) and 'Kembali' (Back).

**Gambar 4. 5** Halaman Tambah Dataset

Pada halaman diatas terdiri dari beberapa form inputan sesuai dengan atribut yang telah ditentukan form tersebut berupa form menu dan jumlah transaksi, Data yang sudah diisi akan masuk kedalam database aplikasi untuk kemudian nantinya di olah menggunakan metode Naïve Bayes maupun K-Nearest Neighbor.

#### 4.3.6 Halaman Update Dataset

Pada halaman ini berfungsi untuk mengupdate dataset jika diperlukan untuk melakukan perubahan data.

The screenshot shows the 'Update Dataset' form. The left sidebar is the same as in the previous image. The main form area has the following fields:

- Nomor \***: Text input with value 1.
- Jumlah Transaksi \***: Text input with value 54.
- Harga \***: Text input with value 18000.
- Diskon \***: Dropdown menu with value 'Tidak'.
- Status \***: Dropdown menu with value 'Rendah'.

At the bottom of the form are two buttons: 'Simpan' (Save) and 'Kembali' (Back).

**Gambar 4. 6** Halaman *Update* Dataset

Pada halaman diatas bisa dilakukan update data jika di inginkan dalam penginputan dataset, data yang ditampilkan adalah data yang dipilih untuk proses perubahan data, perubahan data ini juga bisa dilakukan antara metode Naïve Bayes dan KNN.

#### 4.3.7 Halaman Perhitungan Data

Pada halaman ini berfungsi untuk menguji data dengan melakukan data *testing* didalam aplikasi yang sudah dibuat baik untuk melakukan testing dari metode Naïve Bayes maupun KNN, berikut adalah halaman yang telah dibuat.

**Gambar 4. 7** Halaman Input Form Perhitungan

Dari halaman diatas di masukan nama menu kemudian harga, diskon, jumlah transaksi setiap item menunya. Setelah itu diklik tombol hitung maka akan melakukan proses perhitungan data seperti tampilan dibawah ini.

9 Nearest	
Nomor	Jarak
166	0.009
378	0.051
60	0.062
57	0.063
162	0.066
165	0.073
375	0.076
174	0.08
272	0.08

Total	
Status	Total
Tinggi	5
Sedang	4

Berdasarkan perhitungan, dengan Menu: **KOPI SUSU BROWN**, Jumlah Transaksi: **189**, Harga: **16000**, Diskon: **Tidak**, maka hasilnya: **Tinggi**.  
 Penjualan dari produk sudah baik dan dapat dipertahankan.

**Gambar 4. 8** Halaman Hasil Perhitungan Data *Testing*

Setelah dilakukan proses perhitungan data maka akan menampilkan hasil klasifikasi dengan metode KNN didalam aplikasi sehingga diketahui hasil dari data testing apakah menu tersebut masuk kedalam kategori penjualan tinggi, sedang atau rendah dan akan diberikan pesan dari hasil klasifikasi untuk membantu apakah produk dari hasil penjualan bulan ini sudah baik dan dapat dipertahankan atau dapat meninjau produk dan mungkin bisa melakukan promosi seperti melakukan pemotongan harga untuk menarik atensi pelanggan untuk penjualan kedepan atau bahkan jika penjualan selalu tergolong rendah walaupun sudah melakukan pendiskonan ataupun peninjauan ulang harga, akan disarankan untuk memikirkan kembali apakah menu ini layak dipertahankan atau melakukan penghapusan dari daftar menu di penjualan kedepan. Hal ini juga berlaku jika ingin melakukan pengujian di metode Naïve Bayes.

#### 4.4. Perhitungan Metode dalam Sistem

Pada tahapan ini dilakukan contoh perhitungan dengan metode KNN dan Naïve Bayes sebagai perbandingan, perhitungan dilakukan dengan mengambil beberapa sampel dataset dimana digunakan 15 data untuk contoh perhitungan manual

menggunakan kedua metode. Berikut adalah contoh dataset dan label yang akan digunakan dalam sampel perhitungan antara kedua metode.

**Tabel 4. 1** Tabel Data Transaksi Penjualan

No	Menu	Harga	Diskon	Jumlah Transaksi	Label
1	NASI GORENG ATEKU	18000	Ya	421	Tinggi
2	KENTANG GORENG	12000	Ya	198	Tinggi
3	NASI GORENG TERI CABE HIJAU	16000	Ya	143	Tinggi
4	PAKET HOTMAN	29000	Tidak	30	Rendah
5	INDOMIE GORENG	12000	Ya	88	Sedang
6	PAKET HOTMAN	29000	Tidak	35	Rendah
7	NASI GORENG ANDALIMAN	18000	Tidak	79	Sedang
8	Carmilla	20000	Tidak	152	Tinggi
9	JUS JERUK	11000	Ya	196	Tinggi
10	THAI TEA SUPER CREAM	13000	Ya	126	Sedang
11	CHOCO LATTE	15000	Ya	77	Rendah
12	BURGER ROYAL	12000	Tidak	52	Rendah
13	NASI AYAM PENYET CABE HIJAU	15000	Ya	269	Tinggi
14	Iced Strawberry Tea	12000	Tidak	61	Rendah
15	NUGGET GORENG	11000	Tidak	79	Sedang

Dari dataset diatas ada beberapa variabel yang akan di gunakan untuk perhitungan dengan metode KNN dan Naïve Bayes. Berikut adalah variabel yang akan di gunakan.

**Tabel 4. 2** Tabel Atribut Perhitungan Kedua Metode

No.	Variabel	Keterangan / <i>value</i>
1	Harga	Angka
2	Diskon	Ya / Tidak
3	Jumlah Transaksi	Angka
4	Label	Tinggi Sedang Rendah

Dari dataset dan variabel diatas akan dilakukan perhitungan manual dengan sebuah studi kasus yang akan dilakukan. Studi kasus dilakukan untuk mencari apakah menu masuk kedalam kategori menu yang penjualannya ke dalam label Tinggi/Sedang/Rendah dari dataset yang telah disajikan. Berikut adalah studi kasus atau data testing yang dilakukan jika menggunakan metode Naïve Bayes.

Menu = NASI GORENG TERI CABAI HIJAU  
 Harga = 16000  
 Diskon = Ya  
 Jumlah Transaksi = 154

#### 4.4.1 Perhitungan dengan Naïve Bayes

Dari data diatas, akan dilakukan perhitungan dengan Naïve Bayes dengan rumus sebagai berikut:

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

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n-1}}$$

Dimana:

$f(x)$  = probabilitas

$x$  = nilai dari variabel

$\mu$  = nilai dari rata-rata

$\sigma$  = nilai untuk standar deviasi

Hitung probabilitas dari masing-masing kelas dan atribut:

I. Hitung jumlah kelas pada data *training*

$$P(\text{Label} = \text{Tinggi}) = 6/15 = 0.4$$

$$P(\text{Label} = \text{Sedang}) = 4/15 = 0.26$$

$$P(\text{Label} = \text{Rendah}) = 5/15 = 0.33$$

II. Hitung jumlah kasus yang sama dengan kasus yang sama

$$P(\text{Diskon} = \text{Ya} \mid \text{Label} = \text{Tinggi}) = 5/6 = 0.83$$

$$P(\text{Diskon} = \text{Ya} \mid \text{Label} = \text{Sedang}) = 2/4 = 0.5$$

$$P(\text{Diskon} = \text{Ya} \mid \text{Label} = \text{Rendah}) = 1/5 = 0.2$$

$$P(\text{Diskon} = \text{Tidak} \mid \text{Label} = \text{Tinggi}) = 1/6 = 0.16$$

$$P(\text{Diskon} = \text{Tidak} \mid \text{Label} = \text{Sedang}) = 2/4 = 0.5$$

$$P(\text{Diskon} = \text{Tidak} \mid \text{Label} = \text{Rendah}) = 4/5 = 0.8$$

III. Atribut jumlah

$$\mu_{\text{Label}=\text{Tinggi}} = (421 + 198 + 143 + 152 + 196 + 269) / 6 = 229.83$$

$$\mu_{\text{Label}=\text{Sedang}} = (88 + 79 + 126 + 79) / 4 = 93$$

$$\mu_{\text{Label}=\text{Rendah}} = (30 + 35 + 77 + 52 + 61) / 5 = 51$$

$$\sigma_{\text{Label}=\text{Tinggi}} =$$

$$\sqrt{\frac{\sum (421-229.83)^2 + (198-229.83)^2 + (143-229.83)^2 + (152-229.83)^2 + (196-229.83)^2 + (269-229.83)^2}{6-1}} =$$

$$103.76$$

$$\sigma_{\text{Label}=\text{Sedang}} = \sqrt{\frac{\sum (88-93)^2 + (79-93)^2 + (126-93)^2 + (79-93)^2}{4-1}} = 22.41$$

$$\sigma_{Label=Rendah} = \sqrt{\frac{\Sigma(30-51)^2 + (35-51)^2 + (77-51)^2 + (52-93)^2 + (61-51)^2}{5-1}} = 19.20$$

$$P(\text{Jumlah} = 154 \mid \text{Label} = \text{Tinggi}) = \frac{1}{\sqrt{2\pi} \cdot 103.76} \exp \frac{-(154-229.83)^2}{2(103.76)^2} = 0.002$$

$$P(\text{Jumlah} = 154 \mid \text{Label} = \text{Sedang}) = \frac{1}{\sqrt{2\pi} \cdot 22.41} \exp \frac{-(154-93)^2}{2(22.41)^2} = 0.144$$

$$P(\text{Jumlah} = 154 \mid \text{Label} = \text{Rendah}) = \frac{1}{\sqrt{2\pi} \cdot 19.20} \exp \frac{-(154-51)^2}{2(19.20)^2} = 0.652$$

#### IV. Atribut harga

$$\mu_{Label=Tinggi} = (18000 + 12000 + 16000 + 20000 + 11000 + 15000) / 6 = 15333.33$$

$$\mu_{Label=Sedang} = (12000 + 18000 + 13000 + 11000) / 4 = 13500$$

$$\mu_{Label=Rendah} = (29000 + 29000 + 15000 + 12000 + 12000) / 5 = 19400$$

$$\begin{aligned} \sigma_{Label=Tinggi} &= \sqrt{\frac{\Sigma (18000-15333.33)^2 + (12000-15333.33)^2 + (16000-15333.33)^2 \\ &\quad + (20000-15333.33)^2 + (11000-15333.33)^2 + (15000-15333.33)^2}{6-1}} \\ &= 3444.80 \end{aligned}$$

$$\sigma_{Label=Sedang} = \sqrt{\frac{\Sigma (12000-13500)^2 + (18000-13500)^2 \\ + (13000-13500)^2 + (11000-13500)^2}{4-1}} = 3109.12$$

$$\sigma_{Label=Rendah} = \sqrt{\frac{\Sigma (29000-19400)^2 + (29000-19400)^2 + (15000-19400)^2 \\ + (12000-19400)^2 + (12000-19400)^2}{5-1}} = 7211.10$$

$$P(\text{Harga}=16000 \mid \text{Label}= \text{Tinggi}) = \frac{1}{\sqrt{2\pi} \cdot 103.76} \exp \frac{-(16000-15333.33)^2}{2(103.76)^2} = 0.09387$$



$$P(\text{Harga}=16000 \mid \text{Label}=\text{Sedang}) = \frac{1}{\sqrt{2\pi \cdot 22.41}} \exp \frac{-(16000-13500)^2}{2(22.41)^2} = 0.0001493$$

$$P(\text{Harga} = 16000 \mid \text{Label} = \text{Rendah}) = \frac{1}{\sqrt{2\pi \cdot 19.20}} \exp \frac{-(16000-19400)^2}{2(19.20)^2} = 0.000243$$

V. Lakukan perkalian berdasarkan data *testing* yang akan dicari kelasnya:

$$P(X \mid \text{Label} = \text{Tinggi}) = 0.4 \times 0.83 \times 0.002 \times 0.09387 = 0.00006232968$$

$$P(X \mid \text{Label} = \text{Sedang}) = 0.26 \times 0.5 \times 0.144 \times 0.0001493 = 0.000002794896$$

$$P(X \mid \text{Label} = \text{Rendah}) = 0.33 \times 0.2 \times 0.652 \times 0.000243 = 0.000010456776$$

Berdasarkan probabilitas terbesar, maka hasil klasifikasi dari data testing adalah Tinggi.

#### 4.4.2 Perhitungan dengan KNN

Pada perhitungan di KNN, digunakan normalisasi berupa *z-score* dan melakukan perubahan nilai di atribut diskon dimana ya menjadi 1 dan tidak menjadi 2 agar nantinya dapat dilakukan perhitungan dengan metode KNN. Untuk menghitung manual klasifikasi dengan KNN menggunakan normalisasi *z-score*, kita perlu melakukan beberapa langkah:

1. Ubah nilai diskon (Ya = 1, Tidak = 2).
2. Normalisasi dengan *z-score* untuk setiap atribut (harga, diskon, jumlah transaksi).
3. Hitung jarak *euclidean* dari data yang akan diklasifikasikan ke seluruh data dalam dataset.
4. Pilih k data terdekat (k = 5).
5. Klasifikasikan berdasarkan mayoritas label dari k data terdekat.

Langkah 1: Ubah Nilai Diskon:

**Tabel 4. 3** Tabel Data Setelah Nilai Atribut Diskon Diubah

No	Menu	Harga	Diskon	Jumlah Transaksi	Label
1	NASI GORENG ATEKU	18000	1	421	Tinggi
2	KENTANG GORENG	12000	1	198	Tinggi
3	NASI GORENG TERI CABE HIJAU	16000	1	143	Tinggi
4	PAKET HOTMAN	29000	2	30	Rendah
5	INDOMIE GORENG	12000	1	88	Sedang
6	PAKET HOTMAN	29000	2	35	Rendah
7	NASI GORENG ANDALIMAN	18000	2	79	Sedang
8	Carmilla	20000	2	152	Tinggi
9	JUS JERUK	11000	1	196	Tinggi
10	THAI TEA SUPER CREAM	13000	1	126	Sedang
11	CHOCO LATTE	15000	1	77	Rendah
12	BURGER ROYAL	12000	2	52	Rendah
13	NASI AYAM PENYET CABE HIJAU	15000	1	269	Tinggi
14	Iced Strawberry Tea	12000	2	61	Rendah
15	NUGGET GORENG	11000	2	79	Sedang

Langkah 2: Normalisasi dengan *Z-Score*.

Z-score dihitung dengan rumus:

$$z = \frac{(x - \mu)}{\sigma}$$

dimana  $\mu$  adalah rata-rata dan  $\sigma$  adalah standar deviasi dari data.

Hitung rata-rata dan standar deviasi:

**Harga:**

$$\text{Rata-rata harga}(\mu) = \frac{18000+12000+16000+29000+12000+29000+18000+20000+11000+13000+15000+12000+15000+12000+11000}{15} = 15866.67$$

$$\text{Standar Deviasi harga}(\sigma) = \sqrt{\frac{\sum(x_i - \mu)^2}{n}}$$

$$= \sqrt{\frac{(18000-15866.67)^2 + (12000-15866.67)^2 + (16000-15866.67)^2 + (29000-15866.67)^2 + (12000-15866.67)^2 + (29000-15866.67)^2 + \sum (18000-15866.67)^2 + (20000-15866.67)^2 + (11000-15866.67)^2 + (13000-15866.67)^2 + (15000-15866.67)^2 + (12000-15866.67)^2 + (15000-15866.67)^2 + (12000-15866.67)^2 + (11000-15866.67)^2}{15}} = 5485.57$$

**Diskon:**

$$\text{Rata-rata diskon}(\mu) = \frac{1+1+1+2+1+2+2+2+1+1+1+2+1+2+2}{15} = 1.4$$

$$\text{Standar Deviasi diskon}(\sigma) = \sqrt{\frac{\sum (x_i - \mu)^2}{n}}$$

$$= \sqrt{\frac{(1-1.4)^2 + (1-1.4)^2 + (1-1.4)^2 + (2-1.4)^2 + (1-1.4)^2 + (2-1.4)^2 + \sum (2-1.4)^2 + (2-1.4)^2 + (1-1.4)^2 + (1-1.4)^2 + (1-1.4)^2 + (2-1.4)^2 + (1-1.4)^2 + (2-1.4)^2 + (2-1.4)^2}{15}} = 0.507$$

**Jumlah Transaksi :**

$$\text{Rata-rata transaksi}(\mu) = \frac{421+198+143+30+88+35+79+152+196+126+77+52+269+61+79}{15} = 133.07$$

$$\text{Standar Deviasi Transaksi}(\sigma) = \sqrt{\frac{\sum (x_i - \mu)^2}{n}}$$

$$= \sqrt{\frac{(421-133.07)^2 + (198-133.07)^2 + (143-133.07)^2 + (30-133.07)^2 + (88-133.07)^2 + (35-133.07)^2 + \sum (79-133.07)^2 + (152-133.07)^2 + (196-133.07)^2 + (126-133.07)^2 + (77-133.07)^2 + (52-133.07)^2 + (269-133.07)^2 + (61-133.07)^2 + (79-133.07)^2}{15}} = 110.06$$

Hitung *Z-Score* untuk semua data:

**Tabel 4. 4** Tabel Data Setelah Normalisasi

No	Menu	Harga	Diskon	Jumlah Transaksi	Label
1	NASI GORENG ATEKU	0.338	-0.789	2.61	Tinggi
2	KENTANG GORENG	-0.893	-0.789	0.59	Tinggi
3	NASI GORENG TERI CABE HIJAU	0.024	-0.789	0.09	Tinggi
4	PAKET HOTMAN	2.398	1.184	-0.94	Rendah
5	INDOMIE GORENG	-0.893	-0.789	-0.41	Sedang
6	PAKET HOTMAN	2.398	1.184	-0.89	Rendah
7	NASI GORENG ANDALIMAN	0.338	1.184	-0.49	Sedang
8	Carmilla	0.750	1.184	0.17	Tinggi
9	JUS JERUK	-1.073	-0.789	0.57	Tinggi
10	THAI TEA SUPER CREAM	-0.714	-0.789	-0.06	Sedang
11	CHOCO LATTE	-0.159	-0.789	-0.51	Rendah
12	BURGER ROYAL	-0.893	1.184	-0.74	Rendah
13	NASI AYAM PENYET CABE HIJAU	-0.159	-0.789	1.23	Tinggi
14	Iced Strawberry Tea	-0.893	1.184	-0.65	Rendah
15	NUGGET GORENG	-1.073	1.184	-0.49	Sedang

Hitung nilai *z-score* untuk menu nasi goreng teri cabe hijau(data *testing*) :

$$Z_{harga} = \frac{16000 - 15866.67}{5485.57} = 0.024$$

$$Z_{diskon} = \frac{1 - 1.4}{0.507} = -0.789$$

$$Z_{transaksi} = \frac{154 - 133.07}{110.06} = 0.190$$

Langkah 3: Hitung jarak *euclidean*

Distance=

$$\sqrt{(harga_1 - harga_2)^2 + (diskon_1 - diskon_2)^2 + (transaksi_1 - transaksi_2)^2}$$

1. Nasi Goreng Ateku :

$$\sqrt{(0.338 - 0.024)^2 + (-0.789 - (-0.789))^2 + (2.61 - 0.190)^2} = 2.44$$

2. Kentang Goreng :

$$\sqrt{(-0.893 - 0.024)^2 + (-0.789 - (-0.789))^2 + (0.59 - 0.190)^2} \\ = 1.00$$

3. Nasi Goreng Teri Cabe Hijau :

$$\sqrt{(0.024 - 0.024)^2 + (-0.789 - (-0.789))^2 + (0.09 - 0.190)^2} = 0.1$$

4. Paket Hotman :

$$\sqrt{(2.398 - 0.024)^2 + (1.184 - (-0.789))^2 + (-0.94 - 0.190)^2} = 3.29$$

5. Indomie Goreng :

$$\sqrt{(-0.893 - 0.024)^2 + (-0.789 - (-0.789))^2 + (-0.41 - 0.190)^2} \\ = 1.10$$

6. Paket Hotman :

$$\sqrt{(2.398 - 0.024)^2 + (1.184 - (-0.789))^2 + (-0.89 - 0.190)^2} = 3.27$$

7. Nasi Goreng Andaliman :

$$\sqrt{(0.338 - 0.024)^2 + (1.184 - (-0.789))^2 + (-0.49 - 0.190)^2} = 2.11$$

8. Carmilla :

$$\sqrt{(0.75 - 0.024)^2 + (1.184 - (-0.789))^2 + (0.17 - 0.190)^2} = 2.10$$

9. Jus Jeruk :

$$\sqrt{(-1.073 - 0.024)^2 + (-0.789 - (-0.789))^2 + (0.57 - 0.190)^2} \\ = 1.16$$

10. Thai Tea Super Cream :

$$\sqrt{(-0.714 - 0.024)^2 + (-0.789 - (-0.789))^2 + (-0.06 - 0.190)^2} \\ = 0.78$$

11. Choco Latte :

$$\sqrt{(-0.159 - 0.024)^2 + (-0.789 - (-0.789))^2 + (-0.51 - 0.190)^2} \\ = 0.72$$

12. Burger Royal :

$$\sqrt{(-0.893 - 0.024)^2 + (1.184 - (-0.789))^2 + (-0.74 - 0.190)^2}$$

$$= 2.37$$

13. Nasi Ayam Penyet Cabe Hijau :

$$\sqrt{(-0.159 - 0.024)^2 + (-0.789 - (-0.789))^2 + (1.23 - 0.190)^2}$$

$$= 1.06$$

14. Iced Strawberry Tea:

$$\sqrt{(-0.893 - 0.024)^2 + (1.184 - (-0.789))^2 + (-0.65 - 0.190)^2}$$

$$= 2.33$$

15. Carmilla :

$$\sqrt{(-1.073 - 0.024)^2 + (1.184 - (-0.789))^2 + (-0.49 - 0.190)^2}$$

$$= 2.36$$

Langkah 4: Pilih data terdekat dengan  $k = 5$ :

**Tabel 4. 5** Tabel Pengurutan Jarak Terdekat

No	Menu	Jarak	Ranking	Label
1	NASI GORENG ATEKU	2.44	13	Tinggi
2	KENTANG GORENG	1.00	4	Tinggi
3	NASI GORENG TERI CABE HIJAU	0.1	1	Tinggi
4	PAKET HOTMAN	3.29	15	Rendah
5	INDOMIE GORENG	1.10	6	Sedang
6	PAKET HOTMAN	3.27	14	Rendah
7	NASI GORENG ANDALIMAN	2.11	9	Sedang
8	Carmilla	2.10	8	Tinggi
9	JUS JERUK	1.16	7	Tinggi
10	THAI TEA SUPER CREAM	0.78	3	Sedang
11	CHOCO LATTE	0.72	2	Rendah
12	BURGER ROYAL	2.37	12	Rendah
13	NASI AYAM PENYET CABE HIJAU	1.06	5	Tinggi
14	Iced Strawberry Tea	2.33	10	Rendah
15	NUGGET GORENG	2.36	11	Sedang

Langkah 5: Klasifikasikan berdasarkan mayoritas label dari k data terdekat.

Berdasarkan pengurutan jarak pada tabel diatas dengan nilai  $K = 5$  maka hasil dari klasifikasi ialah sebagai berikut:

**Tabel 4. 6** Tabel 5 Data dengan Jarak Terdekat

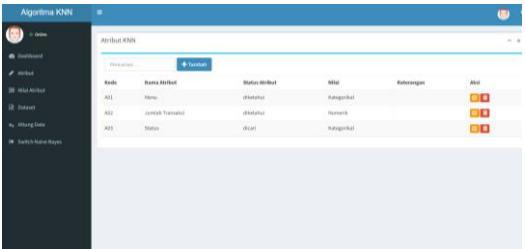
No	Menu	Jarak	Ranking	Label
1	NASI GORENG TERI CABE HIJAU	0.1	1	Tinggi
2	CHOCO LATTE	0.72	2	Rendah
3	THAI TEA SUPER CREAM	0.78	3	Sedang
4	KENTANG GORENG	1.00	4	Tinggi
5	NASI AYAM PENYET CABE HIJAU	1.06	5	Tinggi

Dari tabel diatas berdasarkan jarak terdekat, terdapat 3 dengan label tinggi, 1 Rendah, dan 1 sedang. Jadi hasil klasifikasi untuk data testing adalah tinggi dengan  $k = 5$ .

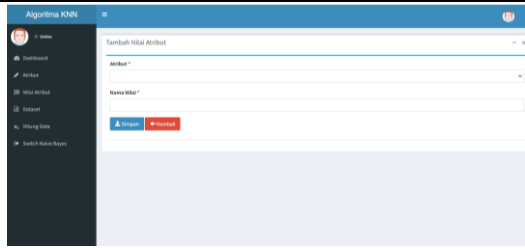
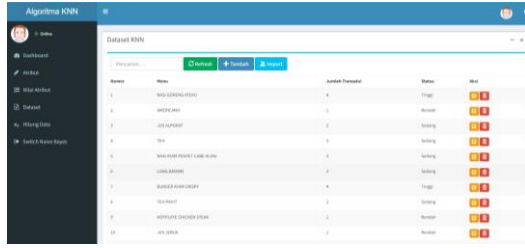
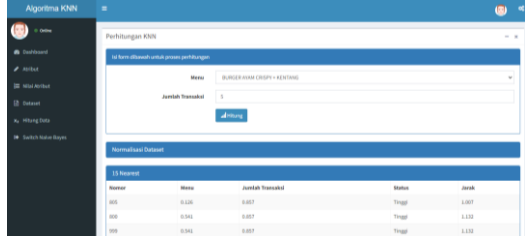
#### 4.5 Pengujian Black Box

Dalam pengembangan sistem ini, penulis menggunakan pengujian *Black Box*. Pengujian *Black Box* adalah metode yang diterapkan pada antarmuka perangkat lunak untuk memastikan bahwa fungsinya beroperasi dengan benar. Pengujian ini bertujuan untuk memastikan bahwa masukan yang diterima sesuai dan keluaran yang dihasilkan benar serta tepat. Dibawah ini merupakan hasil dari pengujian *Black Box*:

**Tabel 4. 7** Tabel Pengujian *Black Box*

Lay out	Input	Proses	Output	Hasil
Halaman atribut	Klik tombol atribut	<a href="http://localhost/nbaye_skn/index.php?m=atribut">http://localhost/nbaye_skn/index.php?m=atribut</a>	Halaman atribut 	Clear
Tombol nilai atribut	Klik nilai atribut	<a href="http://localhost/nbaye_skn/index.php?m=">http://localhost/nbaye_skn/index.php?m=</a>	Halaman nilai atribut	Clear



Lay out	Input	Proses	Output	Hasil
		nilai		
Tombol dataset	Klik menu dataset	<a href="http://localhost/nbaye sknn/index.php?m=dataset">http://localhost/nbaye sknn/index.php?m=dataset</a>		Clear
Tombol Perhitungan	Klik tombol perhitungan	<a href="http://localhost/nbaye sknn/index.php?m=hitung">http://localhost/nbaye sknn/index.php?m=hitung</a>		Clear

## 4.6 Hasil Evaluasi

### 4.6.1 Akurasi Naïve Bayes

Dari hasil total keseluruhan perhitungan dataset yang dimiliki yaitu sebanyak 428 data akan dilakukan pengujian data *testing* sebanyak 85 data untuk metode Naïve Bayes. Dan berikut adalah tabel pengujian *confusion matrix* yang telah dilakukan.

**Tabel 4. 8** Tabel Pengujian *Confusion Matrix* Naïve Bayes

Kelas	Prediksi Benar (Rendah)	Prediksi Benar (Sedang)	Prediksi Benar (Tinggi)
<b>Rendah</b>	42 TP / True Rendah	2 FP / False Sedang	0 FP / False Tinggi
<b>Sedang</b>	4 FN / False Rendah	10 TN / True Sedang	1 FN / False Tinggi
<b>Tinggi</b>	0 FN / False Rendah	1 FN / False Sedang	25 TN / True Tinggi

Dari hasil pengujian confusion matrix diatas kemudian dilakukan perhitungan akurasi dengan rumus sebagai berikut.

$$\text{Nilai Akurasi} = \frac{TN + TP}{TN + TP + FN + FP} \times 100 \%$$

Keterangan :

TP atau True Positive, yaitu jumlah dari data positif yang terklasifikasikan dengan benar oleh sistem yang ada.

TN atau True Negative, yaitu jumlah dari data negatif yang terklasifikasikan dengan benar oleh sistem yang ada.

FN atau False Negative adalah jumlah data negatif yang salah diklasifikasikan sebagai positif oleh sistem.

FP atau False Positive adalah jumlah data positif yang salah diklasifikasikan sebagai negatif oleh sistem.

$$\text{Akurasi} = \frac{42 + 10 + 25}{42 + 2 + 0 + 4 + 10 + 1 + 0 + 1 + 25} \times 100 \%$$

$$\text{Akurasi} = \frac{78}{85} \times 100\% = 90.58 \%$$

Dari hasil perhitungan *confusion matrix* diatas didapatkan nilai akurasi sebanyak 90.58% untuk metode Naïve Bayes dengan jumlah data testing 85 sampel data.

#### 4.6.2 Akurasi K-Nearest Neighbor

Perhitungan akurasi metode K-Nearest Neighbor juga dilakukan dengan menggunakan 85 data *testing* dengan 428 data training. Pengujian dilakukan dengan menggunakan tabel *confusion matrix*, berikut adalah hasil pengujian yang dilakukan.

**Tabel 4. 9** Tabel Pengujian *Confusion Matrix* K-Nearest Neighbor

Kelas	Prediksi Benar (Rendah)	Prediksi Benar (Sedang)	Prediksi Benar (Tinggi)
<b>Rendah</b>	35 TP / True Rendah	0 FP / False Sedang	1 FP / False Tinggi
<b>Sedang</b>	6 FN / False Rendah	17 TN / True Sedang	1 FN / False Tinggi
<b>Tinggi</b>	1 FN / False Rendah	1 FN / False Sedang	23 TN / True Tinggi

$$\text{Akurasi} = \frac{35 + 17 + 23}{35 + 0 + 1 + 6 + 17 + 1 + 1 + 1 + 23} \times 100 \%$$

$$\text{Akurasi} = \frac{75}{85} \times 100\% = 88.23 \%$$

Dari hasil perhitungan *confusion matrix* diatas didapatkan nilai akurasi untuk metode K-Nearest Neighbor sebanyak 88.23% dengan jumlah data testing 85 sampel data.

## **BAB 5**

### **KESIMPULAN DAN SARAN**

#### **5.4 Kesimpulan**

Dari penelitian yang telah dilakukan memperoleh beberapa kesimpulan seperti:

1. Telah berhasil dibangun sistem berbasis web untuk klasifikasi menu berdasarkan penjualan secara cepat dan mudah dengan implementasi algoritma Naïve Bayes dan KNN.
2. Jumlah dataset berperan penting dalam menentukan kualitas representasi data dan hasil yang ada.
3. Hasil penelitian yang dilakukan dengan penerapan metode Naïve Bayes dan KNN didapatkan perhitungan yang sesuai dengan kriteria atribut yang telah ditentukan. Dimana dari atribut tersebut dilakukan perhitungan dengan Naïve Bayes dan KNN didapatkan hasil yang dapat mengklasifikasikan data menu berdasarkan kategori tinggi, sedang maupun rendah.
4. Berdasarkan hasil pengujian sistem untuk evaluasi model dengan dataset yang ada, pengujian akurasi untuk masing-masing metode mendapatkan angka 90.58% untuk metode Naïve Bayes dan 88.23% untuk metode K-Nearest Neighbor dengan persentase 80% data latih dan 20% untuk data pengujian.

### **5.5 Saran**

Dalam implementasi dilapangan tentunya masih banyak hal yang harus dikembangkan untuk kemajuan penelitian, dalam penelitian ini tentunya masih banyak kekurangan yang ada dan penulis menyarankan untuk kedepannya agar pengoptimalan algoritma Naïve Bayes dan K-Nearest Neighbor dapat lebih ditingkatkan lagi, dengan menguji menggunakan dataset yang lebih banyak dan bervariasi dari dataset yang digunakan penulis, sehingga didapatkan hasil implementasi yang maksimal dan tingkat akurasinya lebih tinggi, mengimplementasikan metode evaluasi lain untuk pengujian sistem dan melakukan perhitungan atribut lain ke dalam sistem klasifikasi seperti jenis menu ketika sudah memiliki banyak dataset yang dibutuhkan.

### Daftar Pustaka

- Ajat Sudrajat. (2022). Penerapan Metode Naïve Bayes Untuk Menentukan Penilaian Kinerja Karyawan PT.Sinergi Guna Solusindo. *Jurnal Ilmu Komputer dan Science Volume 99, No. 99*.
- Aryani, E., Zanaria, Y., & Kurniawan, A. (2022). ANALISIS PERKEMBANGAN COFFEE SHOP SEBAGAI SALAH SATU PERANAN UMKM DI KOTA METRO. *Jurnal Akuntansi AKTIVA*, 3(2), 139-145.
- Dewa, B. P., & Setyohadi, B. (2017). Analisis Dampak Faktor Customer Relationship Management dalam Melihat Tingkat Kepuasan Dan Loyalitas Pada Pelanggan Marketplace Di Indonesia. *Telematika: Jurnal Informatika dan Teknologi Informasi*, 14(1), 33-38.
- Farrel Putro, Muchammad et al. PENERAPAN DATA MINING DENGAN NAÏVE BAYES UNTUK KLASIFIKASI SISWA SEKOLAH MENENGAH ATAS DALAM PENENTUAN PERGURUAN TINGGI. *Akrab Juara : Jurnal Ilmu-ilmu Sosial*, [S.1.], v.6, n. 2, may 2021
- Jeanne Fransisca Bokko; Abedneigo.C. Rambulangi; Mey.E. Limbongan. Strategi Promosi Rumah Makan Soponyono Dalam Mempertahankan Pelanggan Di Kecamatan Kesu' Kabupaten Toraja Utara. *semnas 2023*, 2, 28-38.
- Kurniawan, B., Fauzi, M. A., & Widodo, A. W. (2017). Klasifikasi Berita Twitter Menggunakan Metode Improved Naive Bayes. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 1(10), 1193–1200. Diambil dari <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/361>
- Mardi, Y. (2017). Data mining: Klasifikasi menggunakan algoritma c4. 5. *Jurnal Edik Informatika Penelitian Bidang Komputer Sains dan Pendidikan Informatika*, 2(2), 213-219.

- Putro, H. F., Vlandari, R. T., & Saptomo, W. L. Y. (2020). Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan. *Jurnal Teknologi Informasi dan Komunikasi (TIKomSiN)*, 8(2).
- Senika, A., Rasiban., & Iskandar, D. (2022). Implementasi Metode Naïve Bayes Dalam Penilaian Kinerja Sales Marketing Pada PT. Pachira Distrinusa. *JURNAL MEDIA INFORMATIKA BUDIDARMA Volume 6, Nomor 1, Januari 2022, Page 701-709*.