

# Perbandingan Tingkat Optimalisasi Metode K-Nearest Neighbor Dan Naive Bayes Dalam Klasifikasi Kelayakan Alat Laboratorium Kimia

Sri Mulya<sup>1\*</sup>, Gunadi Widi Nurcahyo<sup>2</sup>, Billy Hendrik<sup>3</sup>  
<sup>1,2,3</sup>Fakultas Ilmu Komputer, Universitas Putra Indonesia "YPTK", Padang,  
Indonesia  
E-mail: srimulya@adm.unand.ac.id<sup>1\*</sup>, gunadiwidi@yahoo.co.id<sup>2</sup>  
billy\_hendrik@upiypk.ac.id<sup>3</sup>

## Abstract

Classification of the appropriateness of equipment in the laboratory is needed by university management to determine future laboratory development steps. The suitability of laboratory equipment can be influenced by various factors, so it is necessary to know which variables are crucial in influencing the condition of the laboratory equipment's suitability. Data mining techniques can be used to explore new knowledge so that it can produce appropriate laboratory equipment. Some algorithms that can be used are K-Nearest Neighbor and Naive Bayes. The aim of this research is to compare the level of optimization of two methods in classifying the suitability of Chemistry laboratory equipment at FMIPA Unand using the K-Nearest Neighbor and Naive Bayes methods. The attributes used are year of procurement, level of use, level of damage, length of use of the tool, and condition of tool accessories. The data used is Materials Chemistry laboratory equipment, FMIPA, Andalas University from 2010-2023 with a total of 105 data. The research results show that the accuracy level of the Naive Bayes Method is better than the K-Nearest Neighbor Method. This is proven by the results of the Rapidminer test, which obtained the highest accuracy of 94.74% at a total testing data of 30% of the total data, while for the K-Nearest Neighbor method, the highest accuracy was obtained at 79.03% at a total testing data of 50% of the total data. It is hoped that the results of the tool classification can serve as guidance and evaluation to support the development of the FMIPA Chemistry laboratory at Andalas University.

**Keywords:** classification, feasibility, laboratory tools, K-Nearest-Neighbor Method, Naive Bayes Method

## Abstrak

Klasifikasi kelayakan alat yang ada di laboratorium dibutuhkan oleh manajemen perguruan tinggi untuk menentukan langkah-langkah pengembangan laboratorium kedepannya. Kelayakan alat laboratorium dapat dipengaruhi oleh berbagai faktor sehingga perlu diketahui variabel mana yang sangat menentukan dalam mempengaruhi kondisi kelayakan alat laboratorium. Teknik data mining dapat digunakan untuk menggali suatu pengetahuan baru sehingga dapat menghasilkan kelayakan alat laboratorium. Beberapa Algoritma yang dapat digunakan adalah K-Nearest Neighbor dan Naive Bayes. Tujuan penelitian ini adalah untuk membandingkan tingkat optimalisasi dari dua metode dalam mengklasifikasi kelayakan alat laboratorium Kimia FMIPA Unand menggunakan Metode K-Nearest Neighbor dan Naive Bayes. Atribut yang digunakan yaitu tahun pengadaan, tingkat pemakaian, tingkat kerusakan, lama pakai alat, dan kondisi aksesoris alat. Data yang digunakan adalah alat laboratorium Kimia Material FMIPA Universitas Andalas dari tahun 2010-2023 dengan jumlah data 105. Hasil penelitian menunjukkan tingkat akurasi Metode Naive Bayes lebih baik dari pada Metode K-Nearest Neighbor. Hal ini dibuktikan dari hasil pengujian Rapidminer didapatkan akurasi tertinggi 94.74% pada jumlah data testing 30% dari total

keseluruhan data sedangkan untuk Metode K-Nearest Neighbors didapatkan akurasi tertinggi 79.03% pada jumlah data testing 50% dari total keseluruhan data. Hasil klasifikasi alat diharapkan dapat menjadi pedoman dan evaluasi untuk mendukung pengembangan laboratorium Kimia FMIPA Universitas Andalas.

**Kata Kunci:** klasifikasi, kelayakan, alat laboratorium, Metode K-Nearest-Neighbor, Metode Naïve Bayes

## 1. Pendahuluan

Data Mining adalah rangkaian atau aktivitas yang digunakan untuk mencari hubungan yang bermakna dengan menggunakan pola dan kecenderungan dalam sejumlah besar data yang direkam dengan metode. Teknik dan metode di dalam Data Mining sangat beragam [1]. Data Mining dikenal dengan nama Knowledge Discovery in Database (KDD) dalam jurnal ilmiah sehingga hasil Data Mining dapat digunakan untuk memperbaiki pengambilan keputusan di masa yang akan datang [2]. Proses yang biasanya dilakukan oleh Data Mining adalah deskripsi, prediksi, evaluasi, klasifikasi, pengelompokan, dan asosiasi [3]. Dalam Data Mining prediksi dan klasifikasi banyak digunakan untuk menganalisis data yang dapat menggambarkan kategori data atau memprediksi data di masa mendatang [4].

Knowledge Discovery In Databases (KDD) didefinisikan ekstraksi informasi potensial implisit dan tidak dikenal dari sekumpulan data. Proses Knowledge Discovery melibatkan hasil dari proses Data Mining (proses mengekstrak kecenderungan pola suatu data), kemudian mengubah hasilnya secara akurat menjadi informasi yang mudah dipahami [5]. Knowledge Discovery in Database (KDD) adalah proses secara keseluruhan dalam menentukan pengetahuan yang berguna dari suatu kumpulan data. Adapun tahapan-tahapan dalam KDD antara lain selection (Menyeleksi data yang relevan), preprocessing (menghilangkan noise dan inkonsisten data, menggabungkan data yang bersumber dari banyak sumber), transformation (mentransformasi data kedalam bentuk yang sesuai untuk proses Data Mining), Data Mining (memilih Metode Data Mining yang sesuai dengan pattern data; ekstraksi pola dari data), interpretation/evaluation (menginterpretasi pola menjadi pengetahuan dengan menghilangkan pola yang redundant dan tidak relevan [6]. Data Mining sebagai suatu rangkaian proses dapat dibagi menjadi beberapa tahap proses. Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantara knowledge base [2][7]. Tahap-tahap yang ada dalam Data Mining adalah sebagai berikut:

- a. Seleksi data (data selection)  
Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database.
- b. Preprocessing  
Merupakan proses yang dilakukan dengan cara memeriksa kembali data antara lain untuk pembersihan data-data redundansi, outlier, dan bernilai null (data yang kosong). Hal ini dilakukan agar input data yang akan diolah merupakan data yang telah “bersih” sehingga hasil dari perhitungan Metode Data Mining pun akan memberikan hasil yang sesuai.
- c. Transformasi data (data transformation)  
Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam Data Mining.
- d. Proses mining  
Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.
- e. Evaluasi pola (pattern evaluation)  
Untuk mengidentifikasi pola-pola menarik ke dalam Knowledge Based yang ditemukan.

- f. Presentasi pengetahuan (knowledge presentation)  
Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

Laboratorium Kimia Material merupakan laboratorium penelitian yang berfungsi untuk menyelenggarakan penelitian dan pengabdian kepada masyarakat dalam pengembangan material dan bahan-bahan anorganik. Laboratorium Kimia Material digunakan oleh dosen dan mahasiswa setiap harinya, sehingga dapat menimbulkan tingkat kerusakan alat yang ada di laboratorium semakin tinggi. Kerusakan yang terjadi pada alat-alat laboratorium akan dapat mengganggu proses penelitian dan pengabdian kepada masyarakat yang dilakukan di Laboratorium Kimia Material.

Klasifikasi adalah salah satu ide penambangan data yang banyak digunakan pada tahap awal pengambilan keputusan [8]. Klasifikasi dalam Data Mining sering digunakan untuk menemukan pola yang berkaitan dengan deskripsi dan perbedaan kelas mencoba menebak kelas dari objek yang ID kelasnya tidak diketahui [9]. Ketepatan dalam memprediksi dan mengklasifikasikan kelayakan alat laboratorium membutuhkan data yang akurat sehingga dibutuhkan sebuah Metode terutama dalam klasifikasi yang mudah diinterpretasikan dan diimplementasikan dalam Sistem Pendukung Keputusan [6]. Dalam Data Mining ada beberapa teknik pengolahan data yaitu: deskripsi, estimasi, prediksi, klasifikasi, klusterisasi, dan asosiasi [10]. Beberapa Metode-Metode klasifikasi yang termasuk dalam kategori mudah diinterpretasikan yaitu, C4.5, K-means, Priori, Support Vector Machines, K-Nearest Neighbor dan Naive Bayes [11][12][13].

Penelitian sebelumnya tentang akurasi tingkat kelayakan pakai terhadap peralatan perangkat keras menggunakan metoda rough Set menghasilkan 10 rule keputusan [5][9]. Penelitian serupa juga sudah dilakukan yaitu penerapan Data Mining Rough Set dalam menentukan tingkat kerusakan alat di laboratorium yang mengatakan bahwa metoda ini sangat cocok untuk penelitian ini [14]. Metode klasifikasi lainnya yang sudah dipakai untuk penelitian yang serupa adalah identifikasi tingkat kerusakan peralatan labor teknik komputer jaringan menggunakan metode decision tree dengan hasil nilai variabel kondisi memiliki nilai Gain tertinggi yaitu 0.4734353 [15].

Beberapa penelitian terdahulu sudah melakukan perbandingan metoda tapi belum ada mengenai klasifikasi kelayakan alat laboratorium. Tejawati et al telah melakukan penelitian tentang klasifikasi konten berita di mana dari hasil penelitian tersebut didapatkan KNN memperoleh nilai akurasi 86% dan 51% untuk Metode Naive Bayes [16]. Penelitian lain yaitu Rinanda et al melakukan perbandingan metoda KNN dan Naive Bayes untuk klasifikasi resiko diabetes pada ibu hamil memperoleh nilai akurasi 75,78% pada Metode Naive Bayes dan 74,48% pada Metode KNN [4]. Hasil penelitian lain menyatakan bahwa tingkat akurasi Metode decision tree C4.5 yaitu sebesar 98,00% untuk kasus penentuan kelayakan pemberian kredit, nilai tersebut lebih besar dibandingkan dengan akurasi Metode K-Nearest Neighbor (K-NN) yang hanya memperoleh tingkat akurasi sebesar 93.33% dan akurasi Metode Naive Bayes sebesar 86,67% [12].

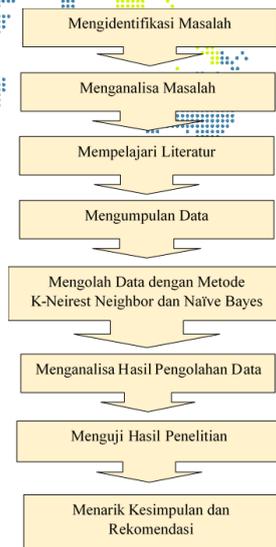
Tujuan dari penelitian ini adalah untuk membandingkan tingkat optimalisasi Metode K-Nearest Neighbor dan Naive Bayes dalam klasifikasi kelayakan alat laboratorium, untuk mengetahui perbandingan tingkat akurasi klasifikasi dan mengetahui variabel yang menentukan dalam klasifikasi tersebut, sehingga hasil dari penelitian ini dapat memberikan informasi dan dapat digunakan sebagai pendukung keputusan.

## 2. Metodologi Penelitian

### 2.1. Kerangka Kerja Penelitian

Kerangka kerja penelitian merupakan langkah-langkah yang dibuat dalam kerangka kerja penelitian disusun secara sistematis guna menyelesaikan permasalahan yang ada. Bab ini akan diuraikan metodologi penelitian dan kerangka kerja yang akan digunakan

dalam penyelesaian masalah. Berikut ini kerangka kerja penelitian dapat digambarkan pada Gambar 1.



**Gambar 1.** Kerangka Kerja Penelitian

## 2.2. Tahapan Penelitian dengan Metode K-Nearest Neighbor

Sumber data penelitian kelayakan alat laboratorium ini diperoleh dari Departemen Kimia Fakultas MIPA Universitas Andalas sebanyak 105 data. Metode K-NN merupakan algoritma untuk mencari jarak terdekat antara data training dengan data testing. Menghitung jarak terdekat atau jauhnya tetangga dapat menggunakan rumus euclidean distance [17]. Proses identifikasi menggunakan metode K-NN akan dilakukan dengan beberapa tahapan [18]. Tahapan tersebut dapat dilihat pada Gambar 2. yaitu:



**Gambar 2.** Tahapan Metode K-Nearest Neighbord

### a) Menentukan nilai k

Dalam metode K-NN, nilai k ditentukan berdasarkan kebutuhan yang akan menjadi acuan untuk mendapatkan hasil dalam klasifikasi data penelitian.

### b) Menghitung Jarak Query Instance

Metode K-NN merupakan metode untuk mencari jarak terdekat antara data training dengan data testing. Untuk menghitung jarak terdekat dapat menggunakan rumus persamaan:

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

**c) Mengurutkan Hasil Query Instance**

Setelah Query Instance didapatkan dari masing-masing data, langkah selanjutnya adalah mengurutkan data-data tersebut ke dalam kelompok yang mempunyai jarak euclid terkecil hingga terbesar.

**d) Mengumpulkan Kategori Y**

Dengan menggunakan kategori Nearest Neighbor yang paling mayoritas maka dapat diidentifikasi nilai query instance yang telah dihitung. Pada urutan jarak yang terdekat dari satu sampai tiga (nilai k=3).

**2.3. Tahapan Penelitian dengan Metode Naïve Bayes**

Metode Naïve Bayes merupakan salah satu metode klasifikasi perhitungan probabilitas dan statistik. Berikut langkah-langkah penyelesaian yang dilakukan dalam menentukan klasifikasi kelayakan alat laboratorium menggunakan Metode Naïve Bayes adalah:



**Gambar 3.** Alur Proses Metode Naïve Bayes

**a) Baca Data Training**

Data yang digunakan dalam tahap ini sama dengan sebelumnya yang digunakan pada proses Metode K-Nearest Neighbord yaitu data alat laboratorium kimia

**b) Menghitung Probabilitas Kelas/Label**

Pada tahap awal ini nilai probabilitas dilakukan untuk menghitung jumlah probabilitas keseluruhan. Adapun perhitungan dilakukan dengan menggunakan rumus:

$$P(H | X) = \frac{P(X|H)P(H)}{P(X)} \tag{2}$$

**c) Menghitung Probabilitas Atribut Tiap Kelas**

Menghitung probabilitas pada setiap kelas terhadap atribut yang digunakan dalam penelitian ini. Masing-masing atribut akan dilakukan perhitungan terhadap jumlah label klasifikasi yang ditetapkan.

**d) Hasil Probabilitas Akhir**

Hasil akhir perhitungan Naive Bayes digunakan untuk menentukan kelas atau label yang menjadi keputusan. Kelas atau label ditentukan berdasarkan nilai probabilitas yang paling tinggi.

**3. Hasil dan Pembahasan**

Data merupakan informasi yang diperoleh melalui pengamatan atau pencarian dari beberapa sumber. Sumber data penelitian kelayakan alat laboratorium ini diperoleh dari Departemen Kimia Fakultas MIPA Universitas Andalas sebanyak 105 data. Data ini dibuat menjadi dua data yaitu data training dan data testing. Dalam penelitian ini, data training yang digunakan sebanyak 100 data dan 5 data sebagai data testing. Data tersebut merupakan data mentah yang belum dilakukan preprocessing dengan atribut nama alat,

tahun pengadaan, tingkat pemakaian, tingkat kerusakan, lama pakai alat, kondisi aksesoris alat.

### 3.1. Analisa Data dengan Metode K-Nearest Neighbor

#### 3.1.1. Transformasi Data

Sebelum data dengan langkah-langkah Metode K-Nearest Neighbor akan dilakukan proses transformasi data. Agar data dapat diolah menggunakan Metode K-NN maka data yang berjenis data huruf harus diinisialisasikan menjadi bentuk angka. Cara mentransformasikan data yaitu dengan cara membuat range untuk data yang berbentuk huruf. Hasil transformasi data training dapat dilihat pada Tabel 1.

**Tabel 1.** Tabel *Data Training* Transformasi

No.	Nama Alat	Tahun Pengadaan	Tingkat pemakaian (%)	Tingkat Kerusakan	Lama Pakai Alat	Kondisi Aksesoris alat	Klasifikasi
1.	Hotplate Stirer (1)	2012	60	2	7	2	Perbaikan
2.	pH Meter (2)	2017	70	2	4	2	Perbaikan
3.	Oven Vacum Selecta Vaciotem-T	2010	10	2	8	2	Perbaikan
...	....	...	..	...	..	...	...
10.	Neraca Analitik	2010	50	3	13	1	Perawatan
11.	Water Bath Shaker	2010	30	3	13	1	Perawatan
...	....	...	..	...	..	...	...
98.	Heating Mantle	2010	80	3	13	1	ganti
99.	Bom calorimeter	2010	60	1	5	2	ganti
100	Hot plate	2010	100	3	13	1	ganti

Selanjutnya data testing yang digunakan pada penelitian ini sebanyak 5 data seperti yang terlihat pada Tabel 2.

**Tabel 2.** Data *Testing*

No.	Nama Alat	Tahun Pengadaan	Tingkat pemakaian (%)	Tingkat Kerusakan	Lama Pakai Alat	Kondisi Aksesoris alat	Klasifikasi
1	Spektrofotometer Shimadzu	2021	60	3	3	1	Perawatan
2	Centrifuge	2010	70	1	8	2	Ganti
3	Spektrofotometer Genesys	2010	70	1	10	2	Ganti
4	Ultrasonic Cleaner	2010	20	3	13	1	Perawatan
5	Autoklave	2010	100	3	13	1	Ganti

#### 3.1.2. Menentukan Nilai k

Dalam metode K-NN, nilai k ditentukan berdasarkan kebutuhan yang akan menjadi acuan untuk mendapatkan hasil dalam mengidentifikasi seleksi penerimaan. Dalam penelitian ini nilai k yang digunakan adalah k=3.

#### 3.1.3. Menghitung Jarak Euclid (Query Instance)

Metode K-NN merupakan metode untuk mencari jarak terdekat antara data training dengan data testing. Untuk menghitung jarak terdekat dapat menggunakan rumus persamaan (2.1). Query instance data training no=1 untuk data testing 1 adalah:

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

$$\begin{aligned}
 &= \sqrt{(2010 - 2010)^2 + (100 - 20)^2 + (3 - 3)^2 + (13 - 13)^2 + (1 - 1)^2} \\
 &= \sqrt{(-0)^2 + (80)^2 + (0)^2 + (0)^2 + (0)^2} \\
 &= \sqrt{6400} = 80
 \end{aligned}$$

Demikian seterusnya perhitungan dilakukan sampai data *training* ke – 100  
*Query instance* data *training* no=1 untuk data *testing* 2 adalah:

$$\begin{aligned}
 &= \sqrt{(2010 - 2021)^2 + (100 - 60)^2 + (3 - 3)^2 + (13 - 3)^2 + (1 - 1)^2} \\
 &= \sqrt{(-11)^2 + (40)^2 + (0)^2 + (10)^2 + (0)^2} \\
 &= \sqrt{1821} = 42,6731
 \end{aligned}$$

Demikian seterusnya perhitungan dilakukan sampai data *training* ke – 100 dan data *testing* yang ke-5 yang dapat dilihat pada Tabel 3.

**Tabel 3.** Hasil *Query Instance*

No.	Nama Alat	<i>Query Instance 1</i>	<i>Query Instance 2</i>	<i>Query Instance 3</i>	<i>Query Instance 4</i>	<i>Query Instance 5</i>
1	Economy incubator	80.0000	30.2324	42.6732	30.4959	0.0000
2	Hot Plate	30.0000	20.3470	17.9165	20.7364	50.0000
3	Orbital Shaker	60.2495	10.1980	23.4734	10.0000	20.7364
4	Oven	30.0000	20.3470	17.9165	20.7364	50.0000
..	...	...	...	...	...	...
97	Pompa Vakum Eyela A-1000S	2.8284	50.0999	41.7732	50.1797	80.0500
98	Pompa Vakum Eyela A-1000S	2.8284	50.0999	41.7732	50.1797	80.0500
99	Pompa Vakum V-100 Buchi	51.7784	11.6190	10.1980	10.5357	32.8786
100	Pompa Vakum V-100 Buchi	52.3641	13.7840	10.0499	12.7279	33.7935

### 3.1.4. Mengurutkan Hasil *Query Instance*

Setelah *Query Instance* didapatkan dari masing-masing data, langkah selanjutnya adalah mengurutkan data-data tersebut ke dalam kelompok yang mempunyai jarak euclid terkecil hingga terbesar. Setelah Hasil *Query Instance* diurutkan, maka diambil nilai jarak yang terkecil dengan ketentuan  $k = 3$ .

**Tabel 4.** Hasil Data *Testing* dengan Metode K-Nearest Neighbor

No.	Nama Alat	Tahun Pengadaan	Tingkat pemakaian (%)	Tingkat Kerusakan	Lama Pakai Alat	Kondisi Aksesoris alat	Klasifikasi
1	Spektrofotometer Shimadzu	2021	60	Baik	3	Lengkap	Perawatan
2	Centrifuge	2010	70	Rusak	8	Tidak Lengkap	Perawatan
3	Spektrofotometer Genesys	2010	70	Rusak	10	Tidak Lengkap	Perawatan
4	Ultrasonic Cleaner	2010	20	Baik	13	Lengkap	Perawatan
5	Autoklave	2010	100	Baik	13	Lengkap	Perawatan

Pada Tabel 4. dapat dilihat hasil identifikasi dari metode K-NN terhadap data *testing* 1-5 yang menghasilkan identifikasi bahwa Alat Ultrasonic Cleaner, Spektrofotometer Genesys, Spektrofotometer Shimadzu, Centrifuge, dan Autoklave dapat diklasifikasikan pada tingkat kelayakan “perawatan”.

### 3.2. Analisa Data dengan Metode Naive Bayes

#### 3.2.1. Menghitung Probabilitas Kelas/Label

Berdasarkan dataset yang ada pada Tabel 1, maka pada tahap ini dilakukan proses menghitung jumlah probabilitas dengan menggunakan Rumus Naive Bayes.

$$\text{Jumlah data label "perawatan"} = 74/100 = 0,74$$

$$\text{Jumlah data label "perbaikan"} = 8/100 = 0,08$$

$$\text{Jumlah data label "ganti"} = 18/100 = 0,18$$

#### 3.2.2. Menghitung Probabilitas Atribut Tiap Kelas

Pada tahap berikutnya dilakukan proses perhitungan probabilitas atribut untuk setiap kelas yang digunakan dalam penelitian ini. Adapun pada masing-masing atribut kelas tersebut dilakukan proses perhitungan terhadap jumlah label "perawatan" jumlah label "perbaikan" dan jumlah label "ganti". Demikian seterusnya perhitungan dilakukan sampai semua atribut dan kelas didapatkan nilai probabilitasnya.

Perhitungan nilai probabilitas Atribut "Tahun Pengadaan" dapat dilakukan dengan cara sebagai berikut:

Atribut Kelas "2010"

$$\text{a. } P(2010 | \text{perawatan}) = 13/74 = 0,175$$

$$\text{b. } P(2010 | \text{perbaikan}) = 2/8 = 0,25$$

$$\text{c. } P(2010 | \text{ganti}) = 17/18 = 0,944$$

#### 3.2.3. Hasil Probabilitas Akhir

Berikut ini adalah hasil perhitungan dari data testing dengan menggunakan metode Naive Bayes dapat dilihat pada Tabel 5.

**Tabel 5.** Hasil Data Testing dengan Metode Naive Bayes

No.	Nama Alat	Tahun Pengadaan	Tingkat pemakaian (%)	Tingkat Kerusakan	Lama Pakai Alat	Kondisi Aksesoris alat	Nilai Probabilitas	Hasil Prediksi
1	Spektrofotometer Shimadzu	2021	60	Baik	3	Lengkap	0.00037983	Perawatan
2	Centrifuge	2010	70	Rusak	8	Tidak Lengkap	0.00010364	Ganti
3	Spektrofotometer Genesys	2010	70	Rusak	10	Tidak Lengkap	0.00010364	Ganti
4	Ultrasonic Cleaner	2010	20	Baik	13	Lengkap	0.00401205	Perawatan
5	Autoklave	2010	100	Baik	13	Lengkap	0.00888736	Ganti

Pada Tabel 5 dapat dilihat hasil identifikasi dari metode Naive Bayes terhadap data testing 1-5 yang menghasilkan identifikasi bahwa Alat Ultrasonic Cleaner dan Spektrofotometer Shimadzu dapat diklasifikasikan pada tingkat kelayakan "Perawatan" sedangkan alat Spektrofotometer Genesys, Centrifuge, dan Autoklave diklasifikasikan pada tingkat kelayakan "ganti".

### 3.3. Hasil Perbandingan Metode K-Nearest Neighbor dan Naive Bayes

Berdasarkan hasil pengolahan data pada Tabel 4 untuk Metode K-Nearest Neighbor dan Tabel 5 untuk Metode Naive Bayes maka didapatkan hasil perbandingan pengolahan data dengan Metode K-Nearest Neighbor dan Metode Naive Bayes dapat dilihat pada Tabel 6 yaitu:

**Tabel 6.** Hasil Perbandingan Metoda K-Neirest Neighbor dan Metode Naive Bayes

No.	Nama Alat	Tahun Pengadaan	Tingkat pemakaian (%)	Tingkat Kerusakan	Lama Pakai Alat	Kondisi Aksesoris alat	Fakta	Hasil Metode K-NN	Hasil Metode Naive Bayes
1	Spektrofotometer Shimadzu	2021	60	Baik	3	Lengkap	Perawatan	Perawatan	Perawatan

No.	Nama Alat	Tahun Pengadaan	Tingkat pemakaian (%)	Tingkat Kerusakan	Lama Pakai Alat	Kondisi Aksesoris alat	Fakta	Hasil Metode K-NN	Hasil Metode Naïve Bayes
2	Centrifuge	2010	70	Rusak	8	Tidak Lengkap	Ganti	Perawatan	Ganti
3	Spektrofotometer Genesys	2010	70	Rusak	10	Tidak Lengkap	Ganti	Perawatan	Ganti
4	Ultrasonic Cleaner	2010	20	Baik	13	Lengkap	Perawatan	Perawatan	Perawatan
5	Autoklave	2010	100	Baik	13	Lengkap	Ganti	Perawatan	Ganti

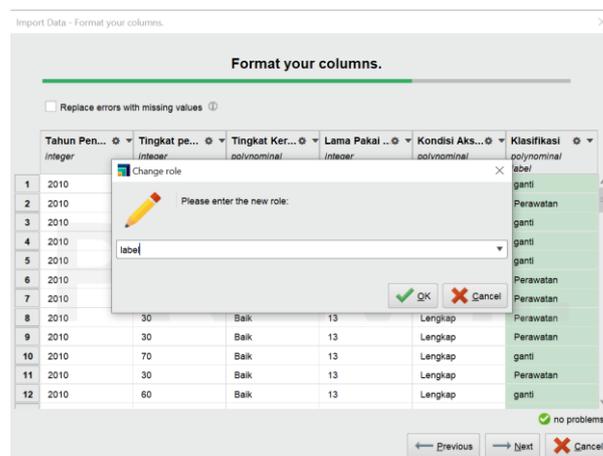
### 3.4. Implementasi Metode K-Nearest Neighbord dan Naïve Bayes pada Rapid Miner Studi 10.3

Pada tahap ini akan dilakukan proses pengujian dengan menggunakan tool Rapid Miner 10.3 serta langkah-langkah dalam menjalankan tool sampai didapatkan hasil akurasi dari nilai klasifikasi kelayakan alat Laboratorium Kimia Material Fakultas MIPA Universitas Andalas. Pada tahapan ini akan dijelaskan cara memasukkan data baru yang akan dieksekusi lebih lanjut. Dalam hal ini data yang akan dieksekusi berupa data excel. Tahap untuk memasukkan data dapat dilihat pada Gambar 4.



Gambar 4. Tampilan awal Import Data

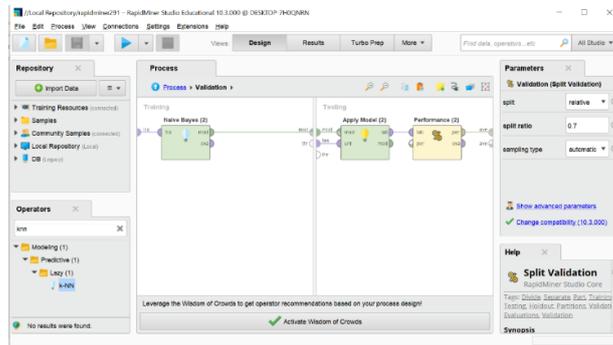
Pada tahapan ini cara memasukkan data dengan dapat dilakukan dengan mengklik pada bagian kiri yaitu import data lalu akan muncul pilihan import data dari tempat kita mengambil data yang akan dimasukkan. Setelah data dimasukkan maka akan muncul tampilan seperti pada Gambar 5 yaitu :



Gambar 5. Tampilan import data pada aplikasi Rapid Miner

Gambar 5 menunjukkan isi data yang kita pilih pada tahap sebelumnya dan akan diproses. Pada halaman ini kita juga bisa memilih sheet dan blok data yang akan digunakan. Setelah

ditentukan sheet dan blok yang digunakan lalu klik next, kemudian pada tahapan berikutnya kita dapat melakukan pemilihan tipe pada data yang dijadikan sebagai label target yaitu klasifikasi. Pada tahapan ini kita juga melakukan perubahan role pada atribut klasifikasi menjadi label. Selanjutnya tahap pemrosesan aplikasi digunakan untuk menjelaskan tahap-tahap penggunaan data yang sudah di-import menggunakan metode K-Nearest Neighbor dan Naive Bayes dalam RapidMiner 10.3.



Gambar 6. Tampilan Proses pada aplikasi *Rapid Miner*

Setelah data dimasukkan, kita perlu membagi data menjadi data training dan data testing dengan mencari split validation pada bagian operators. Setting parameters split validation disebelah kanan sebesar 0,7 dan hubungkan juga port output pada data. Split validation 0,7 adalah jumlah data training dan sisanya 0,3 dijadikan sebagai data testing. Output dari proses yang dilakukan adalah klasifikasi. Pada tahap ini akan menampilkan hasil akhir serta langkah terakhir dalam penggunaan tool RapidMiner. Berikut adalah output dari tool RapidMiner yang dapat dilihat pada Gambar 7.

...	Klasifikasi	prediction(Klasifi...	confid...	confi...	confi...	Tahun P...	Tingka...	Tingkat Ke...	Lama P...	Kondisi Aks...
1	Perawatan	Perawatan	0	1	0	2010	20	Baik	13	Lengkap
2	ganti	ganti	0.737	0.263	0	2010	70	Rusak	10	Tidak Lengkap
3	Perawatan	Perawatan	0	0.753	0.247	2021	60	Baik	3	Lengkap
4	ganti	Perawatan	0.481	0.519	0	2010	70	Rusak	8	Tidak Lengkap
5	ganti	ganti	0.856	0.144	0	2010	100	Baik	13	Lengkap
6	Perawatan	Perawatan	0	1	0	2022	0	Baik	0	Lengkap
7	Perawatan	Perawatan	0	1	0	2021	50	Baik	3	Lengkap
8	ganti	Perawatan	0	0.736	0.264	2018	70	Rusak	3	Tidak Lengkap
9	Perbaikan	ganti	0.517	0.483	0	2010	50	Rusak Ringan	10	Tidak Lengkap
10	Perawatan	Perawatan	0	1	0	2021	0	Baik	0	Lengkap
11	Perawatan	Perawatan	0	1	0	2020	30	Baik	3	Lengkap
12	Perawatan	Perawatan	0	1	0	2019	70	Baik	3	Lengkap
13	Perawatan	Perawatan	0	1	0	2021	70	Baik	2	Lengkap

Gambar 7. Hasil Klasifikasi Kelayakan Alat Laboratorium Kimia Material

Dari hasil perhitungan klasifikasi kelayakan alat Laboratorium Kimia Material menggunakan Rapidminer 10.30, akurasi Metode K-NN dalam pengklasifikasian terpapar pada Gambar 8 dan hasil klasifikasi dengan Metode Naive Bayes terlihat pada Gambar 9.

Table View Plot View

accuracy: 79.03%

	true ganti	true Perawatan	true Perbaikan
pred. ganti	5	2	0
pred. Perawatan	6	43	5
pred. Perbaikan	0	0	1
class recall	45.45%	95.56%	16.67%

Gambar 8. Hasil Akurasi Data Klasifikasi Kelayakan Alat Labortatorium dengan Metode *K-Neirest Neighbor*

● Table View ○ Plot View

accuracy: 94.74%

	true ganti	true Perawatan	true Perbaikan
pred. ganti	6	1	0
pred. Perawatan	0	26	0
pred. Perbaikan	1	0	4
class recall	85.71%	96.30%	100.00%

**Gambar 9.** Hasil Akurasi Data Klasifikasi Kelayakan Alat Labortatorium dengan Metode *Naïve Bayes*

Berdasarkan Gambar 8. dan Gambar 9. maka dapat diketahui perhitungan hasil akurasi klasifikasi dari metode K-Nearest Neighbord dan Naïve Bayes dalam klasifikasi kelayakan alat Laboratorium Kimia Material dengan menggunakan software Rapidminer 10.3. Menghitung akurasi atau tingkat keberhasilan pada sistem ini maka digunakan rumus di bawah ini:

Nilai Akurasi Metode K-Neirest Neighbort dengan jumlah *data testing* 50% dari total keseluruhan data adalah:

$$Accuracy = \frac{\sum \text{Hasil pengujian bernilai benar (kecocokan)}}{\sum \text{Banyak data testing}} \times 100\%$$

$$Accuracy = \frac{49}{62} \times 100\% = 79,03\%$$

Nilai Akurasi Metode Naïve bayes dengan jumlah *data testing* 30% dari total keseluruhan data adalah:

$$Accuracy = \frac{\sum \text{Hasil pengujian bernilai benar (kecocokan)}}{\sum \text{Banyak data testing}} \times 100\%$$

$$Accuracy = \frac{36}{38} \times 100\% = 94,74\%$$

Hasil klasifikasi kelayakan laboratorium Kimia Material untuk 105 dataset dari data alat Laboratorium Kimia Material Fakultas MIPA Universitas Andalas dari tahun 2010 sampai 2023 dengan menggunakan Metode K-Neirest Neighbor mencapai tingkat akurasi tertinggi 79.03% pada jumlah data testing 50% dari total keseluruhan data dengan k=3. Kemudian dengan data yang sama hasil klasifikasi menggunakan Metode Naïve Bayes mencapai tingkat akurasi terbesar 94.74% pada jumlah data testing 30% dari total keseluruhan data. Berdasarkan hasil akurasi tersebut, maka Metode Naïve Bayes dapat digunakan sebagai penentu keputusan atau klasifikasi yang baik dalam hal klasifikasi pada kelayakan alat Laboratorium Kimia Material.

#### 4. Kesimpulan

Berdasarkan penelitian dan pembahasan yang telah dilakukan didapatkan bahwa hasil implementasi Metode Naive Bayes pada tool RapidMiner untuk menguji klasifikasi kelayakan alat laboratorium Kimia FMIPA Universitas Andalas dengan akurasi tertinggi 94.74% pada jumlah data testing 30% dari total keseluruhan data sedangkan dengan metode K-Nearest Neighbor (K-NN) dengan akurasi tertinggi 79.03% pada jumlah data testing 50% dari total keseluruhan data. Perbandingan akurasi metode K-Nearest Neighbor (K-NN) dan Naive Bayes menggunakan rasio data 50%:50%, 60%:40%, 70%:30%, 80%:20% dan 90%:10% sehingga dihasilkan akurasi terbaik untuk klasifikasi kelayakan alat laboratorium Kimia Material FMIPA Universitas Andalas. Pengujian

menggunakan tool RapidMiner dihasilkan akurasi terbaik adalah metode Naive Bayes. Hal ini dibuktikan dari nilai akurasi yang lebih tinggi dibandingkan K-NN. Berdasarkan nilai akurasi yang diperoleh maka dapat disimpulkan bahwa tingkat optimalisasi metode Naive Bayes lebih tinggi dibandingkan Metode K-Nearest Neighbor untuk penelitian penentuan klasifikasi kelayak alat laboratorium kimia.

## Daftar Pustaka

- [1] S. Umam and F. Wahyu Christanto, "Algoritma C4.5 Pada Sistem Analisis Data Untuk Klasifikasi Nasabah Sebagai Dasar Promosi Penjualan Produk Asuransi," *J. Tek. Inform. dan Sist. Inf.*, vol. 10, no. 1, pp. 875–884, 2023, [Online]. Available: <http://jurnal.mdp.ac.id>
- [2] A. Purwanto and H. W. Nugroho, "Analisa Perbandingan Kinerja Algoritma C4.5 Dan Algoritma K-Nearest Neighbors Untuk Klasifikasi Penerima Beasiswa," *J. Teknoinfo*, vol. 17, no. 1, p. 236, 2023, doi: 10.33365/jti.v17i1.2370.
- [3] M. M. Baharuddin, H. Azis, and T. Hasanuddin, "Analisis Performa Metode K-Nearest Neighbor Untuk Identifikasi Jenis Kaca," *Ilk. J. Ilm.*, vol. 11, no. 3, pp. 269–274, 2019, doi: 10.33096/ilkom.v11i3.489.269-274.
- [4] P. D. Rinanda, B. Delvika, S. Nurhidayarnis, N. Abror, and A. Hidayat, "Perbandingan Klasifikasi Antara Naive Bayes dan K-Nearest Neighbor Terhadap Resiko Diabetes pada Ibu Hamil," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 2, pp. 68–75, 2022, doi: 10.57152/malcom.v2i2.432.
- [5] Nurhidayat, S. Defit, and Sumijan, "Data Mining dalam Akurasi Tingkat Kelayakan Pakai terhadap Peralatan Perangkat Keras," *J. Inf. dan Teknol.*, vol. 2, pp. 83–88, 2020, doi: 10.37034/jidt.v2i3.67.
- [6] H. Juliansa, S. Defit, and S. Sumijan, "Identifikasi Tingkat Kerusakan Peralatan Laboratorium Komputer Menggunakan Metode Rough Set," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 2, no. 1, pp. 410–415, 2018, doi: 10.29207/resti.v2i1.274.
- [7] S. Kasus *et al.*, "Penerapan Data Mining Metode K-Nearest Neighbor Untuk Memprediksi Kelulusan Siswa Sekolah Menengah Pertama," *SMARTICS J.*, vol. 9, no. 1, pp. 14–19, 2023, [Online]. Available: <https://doi.org/10.21067/smartics.v9i1.8088>
- [8] T. Hardiani, "Comparison of Naive Bayes Method, K-NN (K-Nearest Neighbor) and Decision Tree for Predicting the Graduation of 'Aisyiyah University Students of Yogyakarta," *Int. J. Heal. Sci. Technol.*, vol. 2, no. 1, pp. 75–85, 2021, doi: 10.31101/ijhst.v2i1.1829.
- [9] Munarsih and B. A. Ningsi, "Performance Comparison of Data Mining Classification Algorithms on Student Academic Achievement Prediction," *Indones. J. Artif. Intell. Data Min.*, vol. 6, no. 1, pp. 29–39, 2023.
- [10] Y. I. Kurniawan, "Perbandingan Algoritma Naive Bayes dan C.45 dalam Klasifikasi Data Mining," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, pp. 455–464, 2018, doi: 10.25126/jtiik.201854803.
- [11] E. I. Siburian, G. Ginting, and R. D. Sianturi, "Penerapan Algoritma C4.5 Untuk Prediksi Kelayakan Pemakaian Alat Laboratorium Komputer," *Maj. Ilm. INTI*, vol. 13, no. September, pp. 256–262, 2018.
- [12] T. T. Muryono, A. Taufik, and I. Irwansyah, "Perbandingan Algoritma K-Nearest Neighbor, Decision Tree, Dan Naive Bayes Untuk Menentukan Kelayakan Pemberian Kredit," *Infotech J. Technol. Inf.*, vol. 7, no. 1, pp. 35–40, 2021, doi: 10.37365/jti.v7i1.104.
- [13] E. Elisa, "Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti," *J. Online Inform.*, vol. 2, no. 1, p. 36, 2017, doi: 10.15575/join.v2i1.71.

- [14] R. I. Salam and S. Defit, "Penentuan Tingkat Kerusakan Peralatan Labor Komputer Menggunakan Data Mining Rough Set," *J. Sistim Inf. dan Teknol.*, vol. 1, pp. 36–41, 2019, doi: 10.37034/jsisfotek.v1i4.7.
- [15] D. P. Sukma, S. Defit, and G. W. Nurcahyo, "Identifikasi Tingkat Kerusakan Peralatan Labor Teknik Komputer Jaringan Menggunakan Metode Decision Tree," *J. Sistim Inf. dan Teknol.*, vol. 3, pp. 275–280, 2021, doi: 10.37034/jsisfotek.v3i4.78.
- [16] A. Tejawati, A. Septiarini, R. Rismawati, and N. Puspitasari, "Perbandingan Metode K-Nearest Neighbor dan Naive Bayes untuk Klasifikasi Konten Berita," *J. Tek. Inform.*, vol. 4, no. 2, pp. 401–412, 2023.
- [17] I. dan A. Mutiara, "Penerapan K-Optimal Pada Algoritma Knn Untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam Berdasarkan Ip Sampai Dengan Semester 4," *Klik - Kumpul. J. Ilmu Komput.*, vol. 2, no. 2, pp. 159–173, 2015.
- [18] S. Wiyono, D. S. Wibowo, M. F. Hidayatullah, and D. Dairoh, "Comparative Study of KNN, SVM and Decision Tree Algorithm for Student's Performance Prediction," *Int. J. Comput. Sci. Appl. Math.*, vol. 6, no. 2, p. 50, 2020, doi: 10.12962/j24775401.v6i2.4360.
- [19] A. Damayunita, R. S. Fuadi, and C. Juliane, "Comparative Analysis of Naive Bayes, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) Algorithms for Classification of Heart Disease Patients," *J. Online Inform.*, vol. 7, no. 2, pp. 219–225, 2022, doi: 10.15575/join.v7i2.919.
- [20] M. E. Febrian, F. X. Ferdinan, G. P. Sendani, K. M. Suryanigrum, and R. Yunanda, "Diabetes prediction using supervised machine learning," *Procedia Comput. Sci.*, vol. 216, no. 2022, pp. 21–30, 2022, doi: 10.1016/j.procs.2022.12.107.