

## Klasifikasi Data Mining Dalam Menentukan Produk Vapor-Juice Terlaris Menggunakan Naïve Bayes

Akbar Aldino Wijaya<sup>1</sup>, Kurniabudi<sup>2</sup>, Hendri<sup>3</sup>

<sup>1</sup>Fakultas Ilmu Komputer, Teknik Informatika, Universitas Dinamika Bangsa, Jambi, Indonesia

Email: <sup>1</sup>akbaraldino32@gmail.com, <sup>2</sup>kbudiz@stikom-db.ac.id, <sup>3</sup>hendri\_boy94@yahoo.com

Email Penulis Korespondensi: [akbaraldino32@gmail.com](mailto:akbaraldino32@gmail.com)

Artikel Info :

Artikel History :

Submitted : 17-09-2024

Accepted : 29-049-2024

Published : 30-09-2024

**Kata Kunci:**

Data Mining,  
Klasifikasi, Kelarisan,  
Naïve Bayes.

**Abstrak**– Vape Shop merupakan salah satu toko Vapor (rokok elektrik) yang berada di kota Jambi yang menjual berbagai perlengkapan Vapor. Banyaknya produk Vapor-Juice yang ada, membuat Vape Shop mengalami kesulitan untuk menentukan Vapor-Juice apa saja yang terlaris dalam 8 bulan terakhir. Tujuan dari studi ini merupakan buat mengklasifikasikan serta mengupas data-data produk Vapor-Juice, yang diharapkan memberi manfaat untuk mengetahui tingkat kelarisan pada produk-produk yang ada di Vape Shop. Studi ini mengenakan prosedur Naïve Bayes. Hasil dari penelitian ini didapatkan bahwa klasifikasi Naïve Bayes menggunakan Tools WEKA didapatkan akurasi teratas yang diperoleh dengan menggunakan Test Option, Use Training Set dan Percentage Split 60%, dengan persentase akurasi yang sama yaitu 80% untuk Correctly Classified Instances, dan 20% untuk Incorrectly Classified Instances.

**Abstract**– Vape Shop is one of the Vapor (e-cigarette) shops located in Jambi city that sells various Vapor equipment. The large number of Vapor-Juice products available, makes it difficult for the Vape Shop to determine which Vapor-Juice are the best-selling in the last 8 months. The purpose of this study is to classify and analyze data on Vapor-Juice products, which is expected to be useful in determining the level of sales of products in the Vape Shop. This study used the Naïve Bayes procedure. The results of this study showed that the Naïve Bayes classification using the WEKA Tools obtained the highest accuracy obtained using the Test Option, Use Training Set and Percentage Split 60%, with the same percentage of accuracy, namely 80% for Correctly Classified Instances, and 20% for Incorrectly Classified Instances.

**Keywords:**

Data Mining,  
Classification, Selling,  
Naïve Bayes

### 1. PENDAHULUAN

Dalam periode globalisasi, kelanjutan kecanggihan teknologi data yang kian kencang ialah penilaian yang sanggup digunakan buat menjangkau kemudahan. Perkembangan teknologi informasi memberikan pengaruh sangat cepat pada pertumbuhan jumlah data yang disimpan dalam basis data, dibutuhkan sebuah metode yang dapat merubah data tersebut menjadi sebuah informasi yang berharga dan bermanfaat dalam mendukung pengambilan sebuah keputusan, terlebih pada aspek bidang usaha yang semua itu menuntut mengenakan sistem data.

Toko Vape Shop adalah salah satu toko Vapor (rokok elektrik) yang berada di kota Jambi yang beralamat di Jalan H. Agus Salim, dengan menjual berbagai perlengkapan Vapor, mulai dari Device Vapor, Vapor-Juice, maupun perlengkapan Vapor lainnya. Diketahui bahwa peningkatan pada penjualan Vapor-Juice ini meningkat sangat pesat. Dengan banyaknya produk Vapor-Juice yang ada, toko Vape Shop mengalami kesulitan untuk menentukan Vapor-Juice apa saja yang terlaris berdasarkan data 8 (delapan) bulan terakhir, sehingga didapatkan sebuah permasalahan yang ada bahwa toko Vape Shop kesulitan dalam menentukan Vapor-Juice apa saja yang paling diminati pembeli.

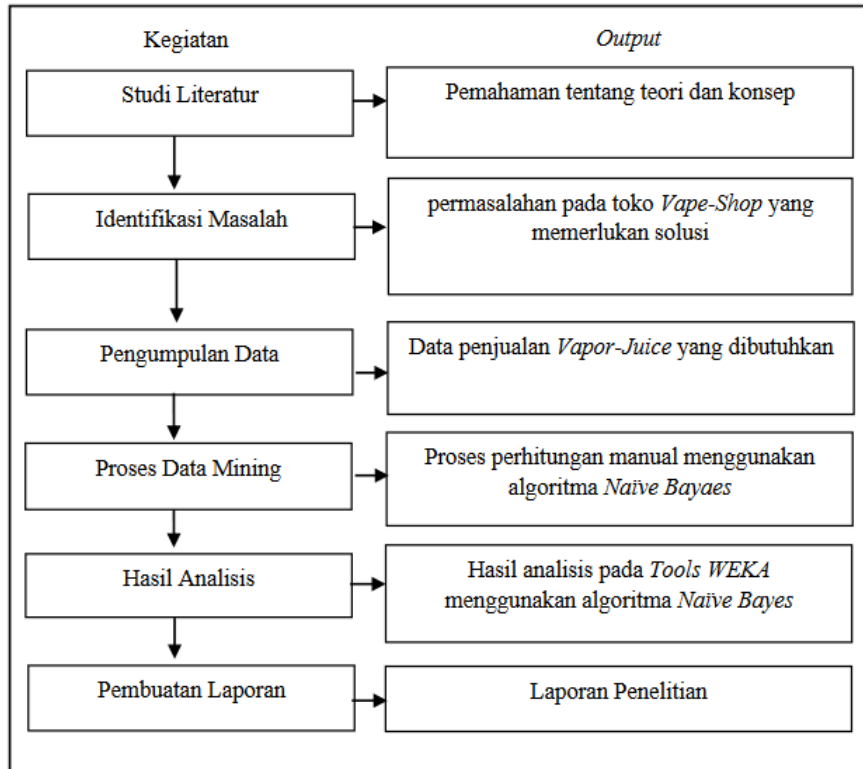
Berdasarkan penelitian sebelumnya yang dilakukan oleh Ismasari Nawangsih, Asti Setyaningsih [1] memaparkan hasil Dengan menggunakan metode algoritma Naïve Bayes mendapatkan tingkat akurasi klasifikasi menghasilkan nilai Accuracy sebesar 97,50%, nilai Precision 100%, dan nilai Recall 93,48%. Ikhsan Romli et al [2] hasil penelitian menunjukkan bahwa tingkat akurasi klasifikasi dengan Metode Naïve Bayes menghasilkan nilai accuracy sebesar 92,19%, nilai Precision 98,39% dan nilai Recall 87,14%. Herry Derajad Wijaya, Saruni Dwiasnati [3] hasil penelitian yang didapat nilai accuracy menunjukkan dinilai 88.00%. Novi Pransiska, et al [4] Pada hasil simulasi Naïve Bayes, informasi yang didapatkan adalah class recall kipas angin 60,57%, air conditioner 2,99%, mesin cuci 6,10%, setrika 13,34%, blender 3,52%, TV 8,39%, rice cooker 3,57% dan DVD player 4,58%. Muhammad Makmun Effendi, et al [5] Dalam pengujian akurasi data menggunakan rapidminer, akurasi yang dihasilkan dari data mining dalam prediksi stok tshirt pada distro crunchz co ini adalah 55%.

Penelitian ini bertujuan membantu toko Vape Shop untuk mengetahui produk Vapor-Juice apa saja yang terlaris yang ada pada toko Vape Shop.

## 2. METODOLOGI PENELITIAN

### 2.1 Kerangka Kerja Penelitian

Didalam melakukan sebuah penelitian, terdapat tahapan- tahapan kegiatan penelitian yang harus dilakukan supaya tujuan studi sanggup terlaksana dengan positif. Tahapan – tahapan kegiatan tersebut disajikan pada gambar 1 berikut ini:



Gambar 1 Kerangka Kerja Penelitian

Bersumber pada kerangka kerja penelitian yang pernah ditafsirkan diatas, sehingga bisa dijabarkan ulasan masing-masing langkah dalam studi yakni sebagai berikut:

1. Studi Literatur  
Pada langkah ini penulis mengerjakan pencarian pada landasan-landasan skema yang dihasilkan dari bermacam asal usul semacam buku, jurnal keilmuan serta rujukan yang lain, guna mempelajari baik tentang rancangan serta skema maka ada cerminan yang baik serta relevan.
2. Identifikasi Masalah  
Pada langkah ini penulis mengenali tentang apa saja yang sebagai perkara yang ada pada toko Vape Shop. Hasil dari identifikasi masalah ini, ditemukannya sebuah permasalahan yang ada pada toko Vape Shop yaitu, belum memanfaatkannya data-data transaksi penjualan *Vapor-Juice* yang ada di dalam *database*.
3. Pengumpulan Data  
Mengumpulkan data-data terkait dengan kebutuhan pada penelitian ini yaitu data penjualan *Vapor-Juice* yang ada pada toko Vape Shop.
4. Proses Data Mining  
Pada tingkatan ini bermuatan ulasan tentang tingkatan-tahapan pada jalan teknik data mining.
5. Hasil Analisis  
Hasil dari kajian hendak ditaksir ketepatan data model yang telah ditemukan dengan kalkulasi *Naïve Bayes* selanjutnya dicoba memakai model pengecekan *Use Training Set, 5&10 K-Fold Cross Validation*, serta *Percentage Split*.
6. Pembuatan Laporan

sesudah seluruh tahapan studi dijalani, sehingga hendak dibuat penjelasan selaku pengumpulan studi supaya sanggup digunakan pada periode yang hendak terlihat bagus oleh peneliti sendiri ataupun peneliti yang lain.

## 2.2 Metode Pengumpulan Data

Dalam studi ini penulis memakai 2 cara pengumpulan data, selanjutnya ini ialah sistem pengumpulan data yang penulis terapkan.:

### 1. Pengamatan (Observasi)

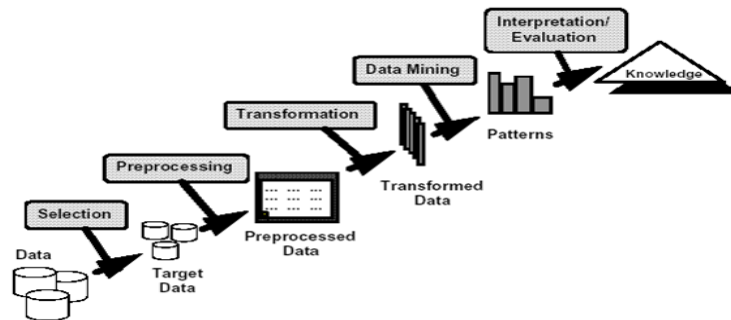
Teknik ini digeluti dengan observasi secara langsung pada subjek yang hendak diteliti. Dalam perihal ini penulis mengobservasi bagaimana pengurusan data yang ada pada toko Vape Shop, tentang data-data penjualan *Vapor-Juice*, maka penulis mendapati data-data *Vapor-Juice* yang pernah terjual pada toko Vape Shop.

### 2. Wawancara

Pada teknik ini penulis melaksanakan tanya jawab pendek secara langsung dengan pemilik toko Vape Shop guna memperoleh data serta data akurat tentang data penjualan *Vapor-Juice* yang penulis butuhkan.

## 2.3 Alur Proses Data Mining

Teknik Data Mining dijalani dengan sebagian tingkatan. Adapula tahapan dalam teknik Data Mining bisa dilihat pada gambar 2.



Gambar 2 Alur Proses Data Mining [6]

Berikut ini penjabaran tentang tahapan-tahapan pada alur proses data mining:

### a. Pengumpulan Data

Pengumpulan data yakni langkah dimana penulis melaksanakan riset serta tanya jawab guna mendapati data data-data yang dipunyai oleh toko Vape Shop. Dari sekian banyak atribut data *Vapor-Juice*, atribut data yang dipakai ialah *Variant*, *Brand*, isi, kadar nikotin, harga, jumlah terjual. Data yang akan digunakan nantinya ada 150 data, dari penjualan *Vapor-Juice* tahun 2021 pada bulan januari hingga agustus.

### b. Preprocessing

*Preprocessing* yakni salah satu tahapan yang penting pada teknik Data Mining, karna tidak seluruh data ataupun atribut data dalam data dipakai dalam teknik data mining. teknik ini dijalani supaya data yang hendak digunakan sesuai keinginan.

### c. Transformasi Data

Transformasi data merupakan proses mentransformasi data yang digunakan untuk proses Mining. Pada tahap ini dilakukan pembersihan atau pemisahan data dari *Noise* data dan data yang tidak konsisten seperti data *Vapor-Juice* yang memiliki atribut tidak lengkap dan juga menentukan kumpulan data dalam suatu kisaran pada standar rentang tertentu.

### d. Pemodelan (Modelling)

Pada tahap ini dilakukan pemodelan terhadap data yang sudah dilakukan *preprocessing*. Pada tahap ini akan diuji menggunakan model tes *Use Training Set*, *5&10 K-Fold Cross Validation*, dan *Percentage Split 10% & 60%*.

### e. Evaluasi (Evaluation)

Pada tahapan ini, dilakukan proses metode klasifikasi dengan mengukur performa terhadap algoritma *Naïve Bayes*.

## 2.2 Naïve Bayes

*Naïve Bayes* termasuk ke dalam pembelajaran *Supervised*, sehingga pada tahapan pembelajaran dibutuhkan data awal berupa data pelatihan untuk dapat mengambil keputusan. [7] *Naïve Bayes* yakni semacam

pengklasifikasian probabilistik simpel yang mengestimasi sekumpulan peluang dengan menambah saluran serta campuran angka dari *dataset* yang diberikan. Adapun formula dari *Naïve Bayes*:

$$P(H | X) = \frac{P(X | H)P(H)}{P(X)} \quad \dots(\text{Persamaan 2.1})$$

Keterangan :

- X : Data dengan class yang belum diketahui
- H : Hipotesis data X merupakan suatu class spesifik
- P(H | X) : Probabilitas hipotesis H berdasarkan kondisi X (posteriori prob.)
- P(H) : Probabilitas hipotesis H (prior prob.)
- P(X | H) : Probabilitas X berdasarkan kondisi tersebut
- P(X) : Probabilitas dari X

## 2.3 Klasifikasi

Klasifikasi ialah sebuah prosedur mendeteksi kerumunan pola maupun guna yang menggambarkan dan memisahkan kasta data yang satu dengan yang yang lain buat melaporkan entitas itu masuk pada golongan terpilih yang telah ditetapkan [8].

## 2.4 WEKA

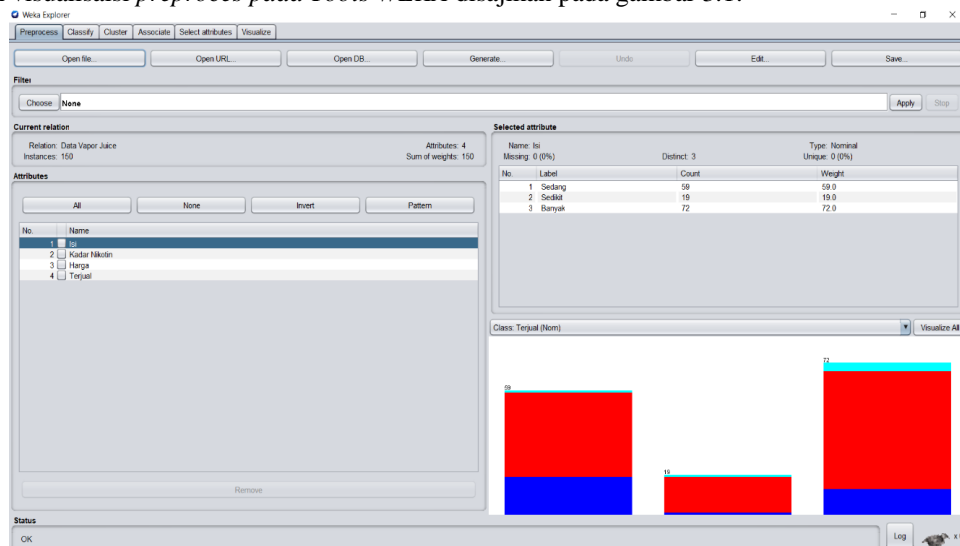
*WEKA* merupakan sebuah *system* data mining yang dikembangkan oleh Universitas Waikato di Selandia Baru yang mengimplementasikan algoritma data mining. *WEKA* menyediakan implementasi dari pembelajaran algoritma yang dapat dengan mudah untuk diterapkan pada dataset. Pengamalan itu pula meliputi bermacam perlengkapan guna mengalihkan *dataset*, *Pre-Process dataset*, memberikan garis pembelajaran, serta menyelidiki pengelompokan yang dibentuk serta kinerjanya tanpa perlu mencatatkan petunjuk program.

## 3. HASIL DAN PEMBAHASAN

Klasifikasi ini menggunakan *Tools WEKA*, dimana dataset berjumlah 150 data dengan 4 atribut yaitu: isi, kadar nikotin, harga, dan jumlah terjual.

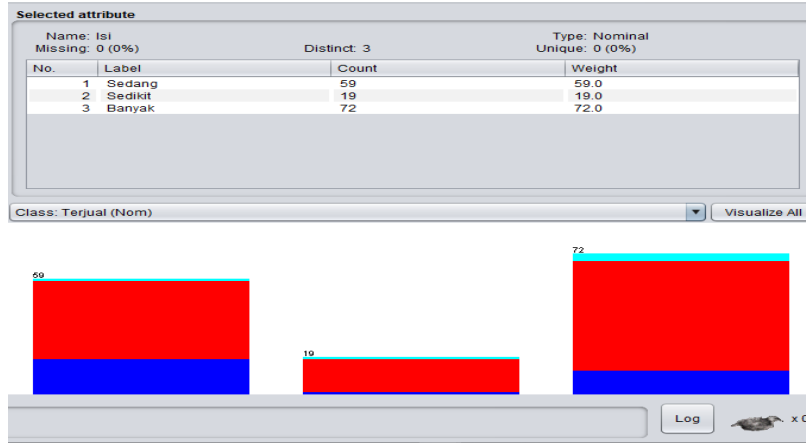
### 3.1 Implementasi Pada *Tools WEKA*

1. Pada penelitian ini pengelompokkan berdasarkan atribut terpilih yaitu isi, kadar nikotin, harga, terjual. Adapun visualisaisi *preproces pada Tools WEKA* disajikan pada gambar 3.1.



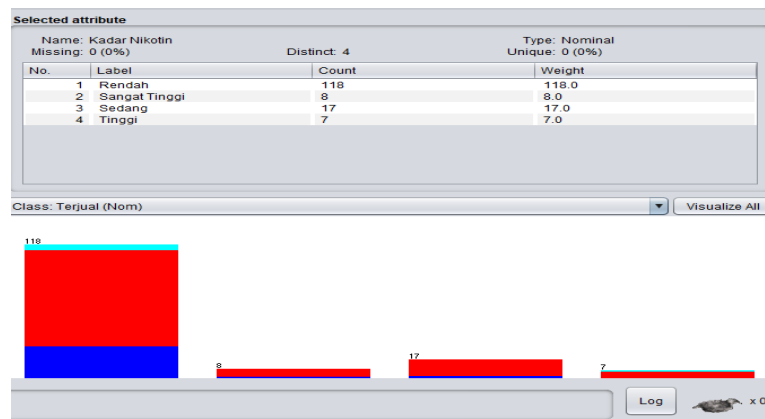
Gambar 3.1 Visualisasi *Preproces Tools WEKA*

2. Gambar 3.2 menyajikan visualisasi dari atribut isi. Diketahui bahwa dari 150 data *Vapor Juice*, terdiri dari 59 atribut isi yang berlabel sedang, 19 atribut isi yang berlabel sedikit, dan 72 atribut isi yang berlabel banyak, maka dapat disimpulkan bahwa jumlah atribut isi *Vapor Juice* yang berlabel banyak, lebih tinggi dari pada atribut isi yang berlabel sedang dan sedikit.



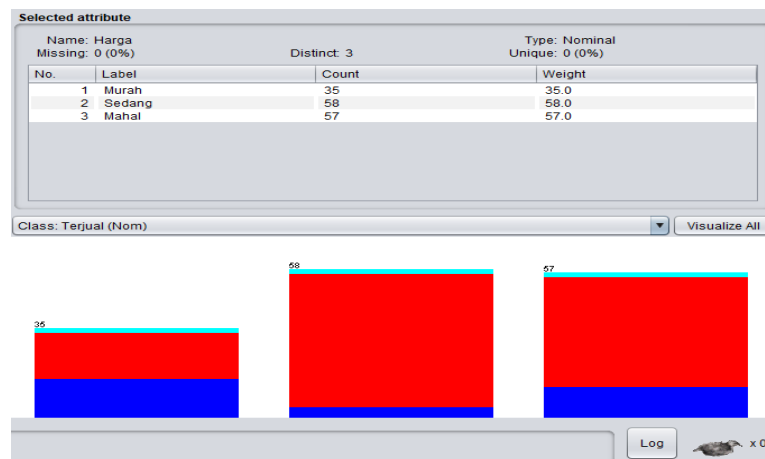
Gambar 3.2 Visualisasi Atribut Isi

3. Gambar 3.3 menyajikan visualisasi dari atribut kadar nikotin. Diketahui bahwa dari 150 data *Vapor Juice*, terdiri dari 118 atribut kadar nikotin yang berlabel rendah, 8 atribut kadar nikotin yang berlabel sangat tinggi, 17 atribut kadar nikotin yang berlabel sedang, dan 7 atribut kadar nikotin yang berlabel tinggi maka dapat disimpulkan bahwa jumlah atribut kadar nikotin *Vapor Juice* yang berlabel rendah lebih tinggi dari pada atribut kadar nikotin yang berlabel sangat tinggi, sedang, dan tinggi.



Gambar 3.3 Visualisasi Atribut Kadar Nikotin

4. Gambar 3.4 menyajikan visualisasi dari atribut harga. Diketahui bahwa dari 150 data *Vapor Juice*, terdiri dari 35 atribut harga yang berlabel murah, 58 atribut harga yang berlabel sedang, dan 57 atribut harga yang berlabel mahal, maka dapat disimpulkan bahwa jumlah atribut harga *Vapor Juice* yang berlabel sedang, lebih tinggi dari pada atribut isi yang berlabel harga mahal dan murah.



Gambar 3.4 Visualisasi Atribut Harga

3.2 Hasil Output Klasifikasi Naïve Bayes Dengan Menggunakan Tools WEKA

Klasifikasi menggunakan algoritma Naïve Bayes dilakukan dengan 5 Test Option yaitu, menggunakan Use Training Set, 5-Cross Validation, 10-Cross Validation, Percentage Split 10%, dan Percentage Split 60%.

1. Hasil Output Klasifikasi Naïve Bayes Dengan Menggunakan Tools WEKA (Use Training Set).

```

Classifier output
Time taken to test model on training data: 0.01 seconds

=== Summary ===
Correctly Classified Instances      120      80 %
Incorrectly Classified Instances    30       20 %
Kappa statistic                    0.3881
Mean absolute error                 0.2235
Root mean squared error             0.3293
Relative absolute error             95.1539 %
Root relative squared error        91.5467 %
Total Number of Instances          150

=== Detailed Accuracy By Class ===
              TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
              0.452  0.067  0.636     0.452  0.528     0.440   0.769    0.454    Sangat Laris
              0.938  0.595  0.828     0.938  0.880     0.419   0.712    0.859    LARIS
              0.000  0.000  ?         0.000  ?         ?       0.611    0.070    Kurang Laris
Weighted Avg.  0.800  0.462  ?         0.800  ?         ?       0.720    0.744

=== Confusion Matrix ===
  a  b  c  <-- classified as
14 17  0  a = Sangat Laris
 7 106 0  b = LARIS
 1  5  0  c = Kurang Laris
    
```

Gambar 3.5 Output Klasifikasi Naïve Bayes (Use Training Set) Menggunakan WEKA

Gambar 3.5 merupakan hasil Output klasifikasi Naïve Bayes pada Tools WEKA dengan menggunakan Test Option (Use Training Set).

Table 3.1 Hasil Klasifikasi (Use Training Set)

Test Option	Akurasi	Jumlah Data	Persentase (%)
Use Training Set	Correctly Classified Instances	120	80
	Incorrectly Classified Instances	30	20

Tabel 3.1 menunjukkan hasil klasifikasi dengan menggunakan Test Option Use Training Set, 120 data dengan prediksi benar terklasifikasi, dengan akurasi sebesar 80%, dan 30 data dengan prediksi salah tidak terklasifikasi dengan persentase 20%, dengan waktu klasifikasi selama 0,01 detik.

Tabel 3.2 Performa Klasifikasi (Use Training Set)

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.452	0.067	0.636	0.452	0.528	0.440	0.769	0.454	Sangat Laris
0.938	0.595	0.828	0.938	0.880	0.419	0.712	0.859	Laris
0.000	0.000	?	0.000	?	?	0.611	0.070	Kurang Laris
0.800	0.462	?	0.800	?	?	0.720	0.744	Weighted Avg

Tabel 3.2 merupakan performa klasifikasi pada Tools WEKA dengan menggunakan Test Option (Use Training Set), yang menunjukkan hasil detail performa setiap kelas.

Tabel 3.3 Confusion Matrix (Use Training Set)

A	B	C	Classified as
14	17	0	A=Sangat Laris
7	106	0	B=Laris
1	5	0	C=Kurang Laris

Tabel 3.3 merupakan hasil *Output Confusion Matrix* pada *Tools WEKA* dengan menggunakan *Test Option (Use Training Set)*.

- Hasil *Output* Klasifikasi Naïve Bayes Dengan Menggunakan *Tools WEKA (5-Cross Validation)*.

```

Classifier output
Time taken to build model: 0 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      113          75.3333 %
Incorrectly Classified Instances    37           24.6667 %
Kappa statistic                    0.2361
Mean absolute error                 0.2399
Root mean squared error             0.355
Relative absolute error             91.0955 %
Root relative squared error         98.6136 %
Total Number of Instances          150

=== Detailed Accuracy By Class ===
          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          0.323   0.059   0.588     0.323   0.417     0.337   0.667    0.384    Sangat Laris
          0.912   0.730   0.792     0.912   0.848     0.231   0.586    0.798    LARIS
          0.000   0.021   0.000     0.000   0.000    -0.029  0.306    0.038    Kurang Laris
Weighted Avg.   0.753   0.563   0.718     0.753   0.725     0.242   0.592    0.682

=== Confusion Matrix ===
  a  b  c  <-- classified as
10  21  0  |  a = Sangat Laris
 7 103  3  |  b = LARIS

```

Gambar 3.6 *Output* Klasifikasi Naïve Bayes (*5-Cross Validation*) Menggunakan *WEKA*

Tabel 3.4 Hasil Klasifikasi (*5-Cross Validation*)

Test Option	Akurasi	Jumlah Data	Persentasi (%)
5-Cross Validation	Correctly Classified Instances	113	75.3333
	Incorrectly Classified Instances	37	24.6667

Tabel 3.4 menunjukkan hasil klasifikasi dengan menggunakan *Test Option (5-Cross Validation)*, yang menunjukkan hasil 113 data dengan prediksi benar terklasifikasi, dengan akurasi sebesar 75.3333%, dan 37 data dengan prediksi salah tidak terklasifikasi dengan persentase 24.6667%, dengan waktu klasifikasi selama 0 detik.

Tabel 3.5 Performa Klasifikasi (*5-Cross Validation*)

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.323	0.059	0.588	0.323	0.417	0.337	0.667	0.384	Sangat Laris
0.912	0.730	0.792	0.912	0.848	0.231	0.586	0.798	Laris
0.000	0.021	0.000	0.000	0.000	-0.029	0.306	0.038	Kurang Laris
0.753	0.563	0.718	0.753	0.725	0.242	0.592	0.682	Weighted Avg

Tabel 3.5 merupakan performa klasifikasi pada *Tools WEKA* dengan menggunakan *Test Option (5-Cross Validation)*, yang menunjukkan hasil detail performa setiap kelas.

Tabel 3.6 *Confusion Matrix*

A	B	C	Classified as
10	21	0	A=Sangat Laris
7	103	3	B=Laris
0	6	0	C=Kurang Laris

Tabel 3.6 merupakan *Confusion Matrix* pada *Tools WEKA* dengan menggunakan *Test Option (5-Cross Validation)*.

- Hasil *Output* klasifikasi Naïve Bayes Dengan Menggunakan *Tools WEKA (10-Cross Validation)*.

```

Classifier output
Time taken to build model: 0 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      117          78 %
Incorrectly Classified Instances    33           22 %
Kappa statistic                    0.355
Mean absolute error                 0.2409
Root mean squared error             0.3632
Relative absolute error             91.5622 %
Root relative squared error         98.1067 %
Total Number of Instances          150

=== Detailed Accuracy By Class ===
                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.452  0.067  0.636  0.452  0.528  0.440  0.655  0.375  Sangat Laris
0.912  0.595  0.824  0.912  0.866  0.367  0.591  0.803  LARIS
0.000  0.021  0.000  0.000  0.000  -0.029  0.299  0.036  Kurang Laris
Weighted Avg.  0.780  0.463  0.752  0.780  0.761  0.366  0.593  0.684

=== Confusion Matrix ===
 a  b  c  <-- Classified as
14 17  0  a = Sangat Laris
 7 103 3  b = LARIS
 1  5  0  c = Kurang Laris
    
```

Gambar 3.7 Output Klasifikasi Naïve Bayes (10-Cross Validation) Menggunakan WEKA

Tabel 3.7 Hasil Klasifikasi (10-Cross Validation)

Test Option	Akurasi	Jumlah Data	Persentasi (%)
10-Cross Validation	Correctly Classified Instances	117	78
	Incorrectly Classified Instances	33	22

Tabel 3.7 menunjukkan hasil klasifikasi dengan menggunakan *Test Option (10-Cross Validation)*, yang menunjukkan hasil 117 data dengan prediksi benar terklasifikasi, dengan akurasi sebesar 78%, dan 33 data dengan prediksi salah tidak terklasifikasi dengan persentase 22%, dengan waktu klasifikasi selama 0 detik.

Tabel 3.8 Performa Klasifikasi (10-Cross Validation)

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.452	0.067	0.636	0.452	0.528	0.440	0.655	0.375	Sangat Laris
0.912	0.595	0.824	0.912	0.866	0.367	0.591	0.803	Laris
0.000	0.021	0.000	0.000	0.000	-0.029	0.299	0.036	Kurang Laris
0.780	0.463	0.752	0.780	0.761	0.366	0.593	0.684	Weighted Avg

Tabel 3.8 merupakan Performa klasifikasi pada *Tools WEKA* dengan menggunakan *Test Option (10-Cross Validation)*, yang menunjukkan hasil detail performa setiap kelas.

Tabel 3.9 Confusion Matrix (10-Cross Validation)

A	B	C	Classified as
14	17	0	A=Sangat Laris
7	103	3	B=Laris
1	5	0	C=Kurang Laris

Tabel 3.9 merupakan *Confusion Matrix* pada *Tools WEKA* dengan menggunakan *Test Option (10-Cross Validation)*.

- Hasil *Output* klasifikasi dengan Naïve Bayes menggunakan *Tools WEKA (Percentage Split 10%)*.



```

Classifier output
Time taken to test model on test split: 0.01 seconds

=== Summary ===
Correctly Classified Instances      79      58.5185 %
Incorrectly Classified Instances    56      41.4815 %
Kappa statistic                    0.1038
Mean absolute error                0.2973
Root mean squared error            0.3964
Relative absolute error            94.064 %
Root relative squared error        107.0947 %
Total Number of Instances         135

=== Detailed Accuracy By Class ===
          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          0.519   0.324   0.286     0.519   0.368     0.162   0.587    0.412    Sangat Laris
          0.631   0.531   0.793     0.631   0.703     0.087   0.535    0.769    LARIS
          0.000   0.031   0.000     0.000   0.000    -0.034  0.452    0.040    Kurang Laris
Weighted Avg.   0.585   0.471   0.662     0.585   0.610     0.097   0.543    0.670

=== Confusion Matrix ===
 a  b  c  <-- classified as
14 13  0 | a = Sangat Laris
34 65  4 | b = LARIS
 1  4  0 | c = Kurang Laris
    
```

Gambar 3.8 Output Klasifikasi Naïve Bayes (Percentage Split 10%) Menggunakan WEKA

Tabel 3.10 Hasil Klasifikasi (Percentage Split 10%)

Test Option	Akurasi	Jumlah Data	Persentasi (%)
Percentage Split 10%	Correctly Classified Instances	79	58.5185
	Incorrectly Classified Instances	56	41.4815

Tabel 3.10 menunjukkan hasil klasifikasi dengan menggunakan *Test Option (Percentage Split 10%)*, yang menunjukkan hasil 79 data dengan prediksi benar terklasifikasi, dengan akurasi sebesar 58.5185%, dan 56 data dengan prediksi salah tidak terklasifikasi dengan persentase 41.4815%, dengan waktu klasifikasi selama 0 detik.

Tabel 3.11 Performa Klasifikasi (Percentage Split 10%)

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.519	0.324	0.286	0.519	0.368	0.162	0.587	0.412	Sangat Laris
0.631	0.531	0.793	0.631	0.703	0.087	0.535	0.769	Laris
0.000	0.031	0.000	0.000	0.000	-0.034	0.452	0.040	Kurang Laris
0.585	0.471	0.662	0.585	0.610	0.097	0.543	0.670	Weighted Avg

Tabel 3.11 merupakan performa klasifikasi pada *Tools WEKA* dengan menggunakan *Test Option (Percentage Split 10%)* yang menunjukkan hasil detail performa setiap kelas.

Tabel 3.12 Confusion Matrix (Percentage Split 10%)

A	B	C	Classified as
14	13	0	A=Sangat Laris
34	65	4	B=Laris
1	4	0	C=Kurang Laris

Tabel 3.12 merupakan *Confusion Matrix* pada *Tools WEKA* dengan menggunakan *Test Option (Percentage Split 10%)*.

- Hasil *Output* klasifikasi Naïve Bayes Dengan Menggunakan *Tools WEKA (Percentage Split 60%)*.

```

Classifier output
Time taken to test model on test split: 0 seconds

=== Summary ===
Correctly Classified Instances      48      80 %
Incorrectly Classified Instances    12      20 %
Kappa statistic                    0.322
Mean absolute error                 0.2361
Root mean squared error             0.3286
Relative absolute error             90.5741 %
Root relative squared error         96.9551 %
Total Number of Instances          60

=== Detailed Accuracy By Class ===
              TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.400  0.100  0.444  0.400  0.421  0.313  0.692  0.395  Sangat Laris
0.917  0.583  0.863  0.917  0.889  0.373  0.689  0.882  LARIS
0.000  0.000  ?      0.000  ?      ?      0.457  0.067  Kurang Laris
Weighted Avg.  0.800  0.483  ?      0.800  ?      ?      0.682  0.774

=== Confusion Matrix ===
 a  b  c  <-- classified as
 4  6  0  | a = Sangat Laris
 4  44 0  | b = LARIS
 1  1  0  | c = Kurang Laris
    
```

Gambar 3.9 Output Klasifikasi Naïve Bayes (Percentage Split 60%) Menggunakan WEKA

Tabel 3.13 Hasil Klasifikasi (Percentage Split 60%)

Test Option	Akurasi	Jumlah Data	Persentasi (%)
Percentage Split 60%	Correctly Classified Instances	48	80
	Incorrectly Classified Instances	12	20

Tabel 3.13 menunjukkan hasil klasifikasi dengan menggunakan *Test Option (Percentage Split 60%)*, yang menunjukkan hasil 48 data dengan prediksi benar terklasifikasi, dengan akurasi sebesar 80%, dan 12 data dengan prediksi salah tidak terklasifikasi dengan persentase 20%, dengan waktu klasifikasi selama 0 detik.

Tabel 3.14 Performa Klasifikasi (Percentage Split 60%)

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.400	0.100	0.444	0.400	0.421	0.313	0.692	0.395	Sangat Laris
0.917	0.583	0.863	0.917	0.889	0.373	0.689	0.882	Laris
0.000	0.000	?	0.000	?	?	0.457	0.067	Kurang Laris
0.800	0.483	?	0.800	?	?	0.682	0.774	Weighted Avg

Tabel 3.14 merupakan performa klasifikasi pada *Tools WEKA* dengan menggunakan *Test Option (Percentage Split 60%)*, yang menunjukkan hasil detail performa setiap kelas.

Tabel 3.15 Confusion Matrix (Percentage Split 60%)

A	B	C	Classified as
4	6	0	A=Sangat Laris
4	44	0	B=Laris
1	1	0	C=Kurang Laris

Tabel 3.15 merupakan *Confusion Matrix* pada *Tools WEKA* dengan menggunakan *Test Option (Percentage Split 60%)*.

#### 4. KESIMPULAN

Berdasarkan hasil yang didapat setelah melakukan penelitian ini, maka dapat diambil kesimpulan bahwa berdasarkan proses klasifikasi *Naïve Bayes* menggunakan *Tools WEKA* terhadap dataset penjualan *Vapor-Juice* pada toko Vape Shop, didapatkan akurasi teratas dari beberapa *Test Option* yang digunakan, diperoleh hasil *Test Option* tertinggi yaitu, *Use Training Set* dan *Percentage Split 60%*.

## REFERENCES

- [1] I. Nawangsih dan Setyaningsih Asti, "PENERAPAN ALGORITMA NAIVE BAYES UNTUK MENENTUKAN KLASIFIKASI PRODUK TERLARIS PADA PENJUALAN PULSA," *Teknol. Pelita Bangsa*, vol. 10, hal. 195–207, 2020.
- [2] I. Romli, E. Pusnawati, dan U. P. Bangsa, "Penentuan Tingkat Penjualan Mobil Di Indonesia," *Pros. Semin. Nas. Teknol. dan Sains*, vol. x, no. x, 2019.
- [3] H. D. Wijaya dan S. Dwiasnati, "Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat," *J. Inform.*, vol. 7, no. 1, hal. 1–7, 2020, doi: 10.31311/ji.v7i1.6203.
- [4] N. Pransiska, A. H. Mirza, dan Andri., "PENERAPAN DATA MINING PREDIKSI PENJUALAN BARANG ELEKTRONIK TERLARIS MENGGUNAKAN ALGORITMA NAÏVE BAYES ( Study Kasus : Planet Cash And Credit Cabang Muara Enim )," *Bina Darma Conf. Comput. Sci.*, hal. 2157–2169, 2016, [Daring]. Tersedia pada: [https://scholar.google.com/scholar?hl=en&as\\_sdt=0%2C5&q=PENERAPAN+DATA+MINING+PREDIKSI+PENJUALAN+BARANG+ELEKTRONIK+TERLARIS+MENGGUNAKAN+ALGORITMA+NAÏVE+BAYES+%28+Study+Kasus+%3A+Planet+Cash+And+Credit+Cabang+Muara+Enim+%29+Bina+Darma+Conference+on+Compute](https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=PENERAPAN+DATA+MINING+PREDIKSI+PENJUALAN+BARANG+ELEKTRONIK+TERLARIS+MENGGUNAKAN+ALGORITMA+NAÏVE+BAYES+%28+Study+Kasus+%3A+Planet+Cash+And+Credit+Cabang+Muara+Enim+%29+Bina+Darma+Conference+on+Compute).
- [5] M. M. Effendi, W. Hadikristanto, dan R. Adelia, "PENERAPAN DATA MINING DALAM MEMPREDIKSI PEMBELIAN T-SHIRT MENGGUNAKAN ALGORITMA C4.," *Teknol. Pelita Bangsa*, no. 5, hal. 613–616, 2018, doi: 10.7868/s0869565218050249.
- [6] Nurdin dan D. Astika, "Penerapan Data Mining Untuk Menganalisis Penjualan Barang Dengan Pada Supermarket Sejahtera Lhokseumawe," vol. 6, no. 1, hal. 134–155, 2015, doi: 10.29103/TECHSI.V7I1.184.
- [7] A. Saleh, "Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," *Creat. Inf. Technol. J.*, vol. 2, no. 3, hal. 207–217, 2015.
- [8] N. Adhika, "Penerapan Algoritma Klasifikasi Data Mining Dalam," *KNiST*, vol. XIV, no. 2, hal. 120–129, 2012.