

# Klasifikasi Omset ATK menggunakan Algoritma Naïve Bayes

<sup>1\*</sup>Mar'atuttahirah, <sup>2</sup>Miftahulhairah, <sup>3</sup>Syafira Fatwa, <sup>4</sup>Arsan Kumala Jaya.

<sup>1,2,3</sup>Institut Teknologi Bacharuddin Jusuf Habibie

<sup>4</sup> Institut Bisnis dan Keuangan Nitro

Email: <sup>1</sup>maratuttahirah.ir@ith.ac.id, <sup>2</sup>miftahulhairah@ith.ac.id, <sup>3</sup>firafatwa@gmail.com,  
<sup>4</sup>arsanptik05@gmail.com

## Abstrak

Salah satu faktor yang mempengaruhi dinamika bisnis modern adalah kemajuan teknologi. Klasifikasi omset adalah pendekatan yang efektif untuk bisnis dapat menyesuaikan strategi pemasaran yang akan berdampak pada pendapatan bisnis. Keadaan stok suatu barang sangat mempengaruhi omset penjualan. Tujuan penelitian ini adalah klasifikasi omset untuk barang yang terjual di toko apakah banyak terjual atau kurang terjual untuk memaksimalkan stok barang per waktu. Metode yang diterapkan pada penelitian ini adalah algoritma *Naive Bayes*. Alur penelitian adalah dimulai dari data penelitian yang dilakukan pengelompokan atribut, *preprocessing*, transformasi data, *modeling* dengan algoritma *Naive Bayes* dengan bahasa pemrograman python dan evaluasi. Dari hasil penelitian mendapatkan tingkat akurasi 90% dalam klasifikasi omset ATK untuk menjadi solusi bagi pimpinan toko lebih mudah mengambil strategi bisnis dengan cepat dan optimal yang tentunya akan berpengaruh terhadap omset.

Kata Kunci: *Atribut, Klasifikasi, Modeling, Naïve Bayes, Omset,*

## Abstract

*One of the factors that influences the dynamics of modern business is technological progress. Profit classification is an effective approach for businesses to adapt marketing strategies and the stock condition of an item greatly affect to income. The purpose of this study is to classify profit for item sold in the shop, whether they are sold a lot or not enough to maximize the stock of item per time. The method used in this study is the Naive Bayes algorithm. The plot of this study is starts from data study carried out by grouping attributes, preprocessing, data transformation, modeling with the Naive Bayes algorithm in the Python programming language and evaluation. From the research results obtained an accuracy rate of 90% in the classification of omset to become a solution for shop leaders for easily adopt business strategies quickly and optimally which certainly will have an impact on omset.*

*Keywords: Attributes, Classification, Modeling, Naïve Bayes, Profit*

## 1. Pendahuluan

Peran usaha kecil semakin meningkat di perekonomian berbagai Negara. Usaha kecil berperan dalam penyerapan tenaga kerja, pemberian pelayanan ekonomi kepada masyarakat dan dalam proses pemerataan dan peningkatan pendapatan masyarakat. Oleh karena itu, usaha kecil merupakan sesuatu hal penting yang dapat mengatasi pengangguran dan kemiskinan (Yemelyanov et al., 2020). Berdasarkan Badan Pusat

Statistik (BPS) tahun 2019, pelaku usaha di Indonesia 99,99% tergolong sebagai pelaku Usaha Mikro Kecil (UMK) yang terdiri dari usaha mikro sebanyak 57,53 juta unit (98,79%) dan usaha kecil 629,4 ribu unit (1,11%). Jumlah UMK pun mengalami pertumbuhan yang begitu pesat, dari yang semula pada tahun 2017 sebesar 56,20 juta unit dan pada tahun 2018 menjadi 57,53 juta unit. Gambaran tersebut menunjukkan bahwa sesungguhnya perekonomian Indonesia secara riil digerakkan oleh para pelaku usaha mikro dan kecil. BPS juga melansir bahwa pelaku usaha mikro, kecil dan menengah memberikan andil besar dalam perekonomian nasional dan daerah, kontribusinya secara total dalam PDB sebesar 59,08%, dan mampu menyerap tenaga kerja sebesar 97,16% dari total tenaga kerja nasional. Data tersebut mengisyaratkan bahwa UMK dalam perekonomian nasional telah menunjukkan peran sebagai berikut: (1) Pemain utama dalam kegiatan ekonomi di berbagai sektor usaha, (2) Penyedia lapangan kerja terbesar, (3) Pelaku penting dalam pemberdayaan dan pengembangan ekonomi masyarakat, (4) Pencipta pasar baru dan sumber inovasi (Sudrartono, 2019).

Pengambilan keputusan adalah salah satu proses yang paling penting dalam bisnis (Yaman & Bilgic, 2019). Proses bisnis yang aktif dapat mencegah masalah dengan mengambil langkah untuk secara proaktif mengidentifikasi, mencegah, dan mengurangi risiko serta kendala yang mungkin timbul selama pelaksanaan proses bisnis. Dengan mengimplementasikan proses bisnis yang aktif, organisasi dapat meningkatkan keefektifan operasional mereka, mengurangi potensi kerugian, dan menghindari masalah yang dapat muncul di masa depan. Hal ini juga dapat meningkatkan reputasi bisnis dan kepercayaan pelanggan serta pemangku kepentingan lainnya (Zhang & Liu, 2020).

Di era digital ini, laju perubahan sangatlah pesat, teknologi yang berkembang telah membuka pintu untuk inovasi yang cepat. Masyarakat dan bisnis harus bersiap-siap untuk beradaptasi dengan perubahan tersebut agar tetap relevan dan kompetitif dalam era digital yang terus berubah (Strutyńska et al., 2019).

Dalam beberapa tahun terakhir, persaingan antara bisnis telah mengalami peningkatan yang signifikan dalam upaya mereka untuk mempertahankan posisi mereka di pasar. Fenomena ini terjadi karena beberapa faktor yang mempengaruhi dinamika bisnis modern. Salah satu faktor yang berperan penting adalah kemajuan teknologi (Pramono et al., 2019). Penerapan teknologi dalam sebuah perusahaan sangat penting untuk efisiensi pelaksanaan aktivitas manajemen dalam meningkatkan kinerja bisnis dan meningkatkan daya saing dalam bisnis global (Yoon, 2020). Dalam era modern, inovasi menjadi hal penting dimana setiap usaha terlibat dalam kompetisi untuk menjadi lebih baik dari yang lainnya. Bisnis saat ini bergantung pada inovasi yang memiliki kemampuan untuk menarik pelanggan dengan produk-produknya (Kilari et al., 2022).

Segmentasi pasar adalah pendekatan yang banyak digunakan dalam menyelidiki heterogenitas perilaku pembelian dan profitabilitas pelanggan. Salah satu tujuan dari segmentasi ini adalah untuk menentukan setiap kategori dan meningkatkan keuntungan bisnis. Salah satu hal penting adalah bahwa model segmentasi pasar dalam industri sebagian besar adalah metode deskriptif, kurangnya wawasan pasar yang memadai, dan seringkali gagal mengidentifikasi segmen yang meskipun segmen cukup kecil (Yoseph et al., 2020). Segmentasi pelanggan adalah pendekatan yang populer dan efektif dalam memahami dan membagi pelanggan. Metode ini memungkinkan bisnis untuk mengidentifikasi kelompok pelanggan yang memiliki karakteristik dan perilaku yang sama, memungkinkan strategi pemasaran dan layanan yang lebih terfokus (Shihab et al., 2019). Berbagai penelitian tentang klasifikasi omset dengan metode yang berbeda, klasifikasi omset memperkirakan pelanggan terbanyak yang membeli barang dengan

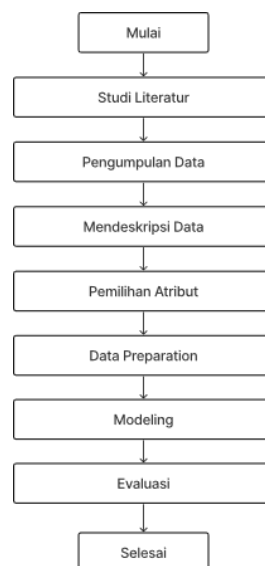
metode *K-Means* dan *LRFM - Average Item* (Pramono et al., 2019), klasifikasi omset dengan *Cluster Analysis* (Muchardie et al., 2019), klasifikasi omset dengan segmentasi pelanggan untuk identifikasi target konsumen dengan *K-Means* (Nandapala & Jayasena, 2020).

Pemilihan model algoritma berdasarkan penelitian yang mengimplementasikan algoritma *Naïve Bayes*, segmentasi untuk deteksi leukemia (Das & Dutta, 2020), segmentasi gelombang pasang air laut (Sianipar et al., 2021), pengambilan keputusan dalam pembelian mobil (Harahap et al., 2018), prediksi kelulusan (Peling et al., 2017), dan segmentasi penyakit kulit (Balaji et al., 2020). Algoritma *Naïve Bayes* mengasumsikan atribut objek secara independent, berperan sangat penting karena metode di dalam algoritma ini sangat baik dalam pengklasifikasian, algoritma *Naïve Bayes* juga hanya membutuhkan jumlah data *training* yang lebih sedikit untuk menyelesaikan masalah klasifikasi dan masalah *biner multiclass* (Novitalia et al., 2021).

Ketersediaan produk dalam pemenuhan konsumen merupakan permasalahan yang harus di atasi. Keadaan stok suatu barang sangat mempengaruhi omset penjualan. Jumlah permintaan barang yang meningkat, maka akan memperoleh omset yang besar. Penelitian ini berfokus pada sebuah sistem klasifikasi omset, dalam hal ini segmentasi frekuensi pendapatan untuk masing-masing barang yang terjual di toko untuk memaksimalkan stok barang per waktu dengan menggunakan Algoritma *Naïve Bayes*. Sistem yang diusulkan diharapkan mampu menjadi solusi bagi pimpinan toko untuk lebih mudah mengambil strategi bisnis dengan cepat dan optimal yang tentunya akan berpengaruh terhadap omset.

## 2. Metode

Tahapan penelitian dalam penelitian ini diperlihatkan pada Gambar 1.



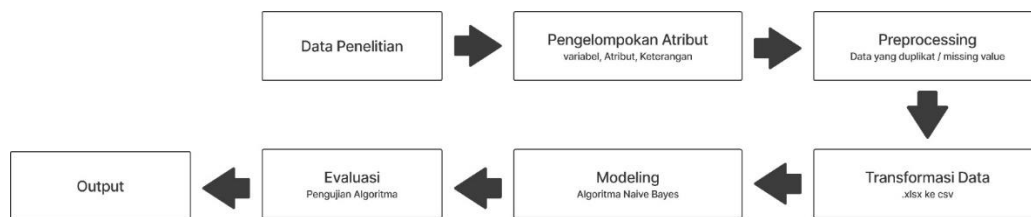
Gambar 1. Tahapan Penelitian

Tahapan awal dalam penelitian ini study literatur klasifikasi omset dengan algoritma *Naïve Bayes*. Data yang dikumpulkan adalah data transaksi pelanggan ATK, selanjutnya pemilihan atribut dari data transaksi pelanggan. Atribut yang dimaksud dalam penelitian ini adalah kriteria-kriteria yang ada dalam data transaksi pelanggan. Untuk tahap data preparation disini mencakup proses pembersihan dan transformasi data untuk

kemudian dijadikan input. Selanjutnya adalah tahap *modeling*, pada tahap ini akan dilakukan pemilihan dan penerapan algoritma dengan parameternya yang akan disesuaikan untuk mendapatkan hasil yang optimal. Dan tahap terakhir adalah tahap evaluasi yang dilakukan setelah didapatkan output dari sistem untuk mengetahui bagaimana keberhasilan sistem klasifikasi omset penelitian ini.

## 2.1. Alur Perancangan Sistem

Alur perancangan system algoritma *Naïve Bayes* untuk Segmentasi Pendapatan Penjualan ATK diperlihatkan pada Gambar 2.



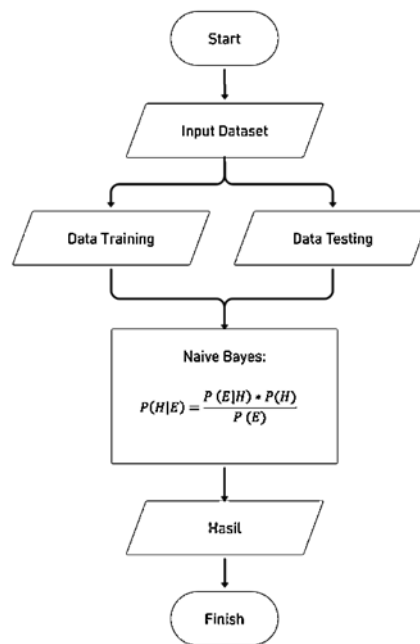
Gambar 2. Alur Perancangan Sistem

Alur perancangan system diuraikan sebagai berikut:

- Step 1 : Data penelitian yang akan menjadi input pada sistem. Data penelitian ini di ambil dari data penjualan ATK.
- Step 2 : Pengelompokan atribut dikategorikan berdasarkan nama barang, unit, satuan, harga, dan total.
- Step 3 : *Preprocessing*, dalam hal ini dilakukan dengan melakukan pengecekan pada data yang duplikat atau missing value pada data tersebut.
- Step 4 : Transformasi data, disini dilakukan proses pengubahan format file untuk kebutuhan dalam input dataset
- Step 5 : *Modeling*, pada tahap awal disini dilakukan pambagian data training dan data testing, selanjutnya implementasi algoritma *Naïve Bayes* untuk data input penjualan toko ATK.
- Step 6 : Evaluasi, pada tahap ini dilakukan pengujian algoritma yang diterapkan untuk klasifikasi omset atk menggunakan algoritma *Naïve Bayes*

## 2.2. Model Penelitian

Pada penelitian ini metode yang digunakan yaitu algoritma *Naïve Bayes* untuk segmentasi frekuensi pendapatan masing-masing barang yang terjual di toko untuk memaksimalkan stok barang per waktu menjadi solusi bagi pimpinan toko untuk lebih mudah mengambil strategi bisnis dengan cepat dan optimal yang tentunya akan berpengaruh terhadap pendapatan. *Flowchart* perhitungan secara menyeluruh algoritma *Naïve Bayes* diperlihatkan pada Gambar 3.



Gambar 2. Flowchart Algoritma Naive Bayes

Prediksi *Naive Bayes* didasarkan pada teorema *Bayes* dengan formula umum sebagai berikut (Romadhon & Kurniawan, 2021).

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)} \quad (1)$$

Keterangan:

$P(H/E)$  : Probabilitas akhir bersyarat (*conditional probability*) suatu hipotesis  $H$  terjadi jika diberikan bukti (*evidence*)  $E$  terjadi.

$P(E/H)$  : Probabilitas sebuah bukti  $E$  terjadi akan mempengaruhi hipotesis  $H$ .

$P(H)$  : Probabilitas Awal (priori) hipotesis  $H$  terjadi tanpa memandang bukti apapun.

$P(E)$  : Probabilitas Awal (priori) bukti  $E$  terjadi tanpa memandang hipotesis/bukti yang lain.

Ide dasar dari aturan *Bayes* adalah bahwa hasil dari hipotesis atau peristiwa ( $H$ ) dapat diperkirakan berdasarkan pada beberapa bukti ( $E$ ) yang diamati. Ada beberapa hal penting dari aturan *Bayes* tersebut, yaitu :

1. Sebuah probabilitas awal/priori  $H$  atau  $P(H)$  adalah probabilitas dari suatu hipotesis sebelum bukti diamati.
2. Sebuah probabilitas akhir  $H$  atau  $P(H/E)$  adalah probabilitas dari suatu hipotesis setelah bukti diamati.

Kaitan antara *Naive Bayes* dengan klasifikasi, korelasi hipotesis, dan bukti dengan klasifikasi adalah bahwa hipotesis dalam teorema *Bayes* merupakan label kelas menjadi target pemetaan dalam klasifikasi, sedangkan bukti merupakan fitur – fitur yang menjadi masukan dalam model klasifikasi. Jika  $X$  adalah vektor masukan yang berisi fitur dan  $Y$  adalah label kelas, *Naive Bayes* dituliskan dengan  $P(Y/X)$ . Notasi tersebut berarti probabilitas label kelas  $Y$  didapatkan setelah fitur-fitur diamati. Notasi ini disebut juga

probabilitas akhir (*posterior probability*) untuk  $Y$ , sedangkan  $P(Y)$  disebut dengan probabilitas awal (*prior probability*)  $Y$  (Peling et al., 2017).

### 3. Hasil dan Pembahasan

Hasil dari perancangan system menggunakan *Naïve Bayes* untuk segmentasi pendapatan penjualan ATK. Dengan pengelompokan atribut dan modeling dengan algoritma *Naïve Bayes*.

#### 3.1. Data Penelitian

Data yang digunakan dalam penelitian ini adalah data penjualan toko INTEC sebanyak 345 data.

#### 3.2. Pengelompokan Atribut

Pengelompokan atribut pada penelitian ini dikategorikan berdasarkan nama barang, unit, satuan, harga, dan total. Pengelompokan atribut diperlihatkan pada Tabel 1.

Tabel 1. Pengelompokan Atribut

Varibel	Atribut	Keterangan
X1	Nama Barang	Foto Copy
		Jilid
		Kertas F4
		Pulpen
		Quantum
		Pulpen
		Snowman
		Double Tip
		Tip Ex
		Spidol
		Whiteboard
		Stapler HD-10
		Etona No. 10
		Lem Fox
		Pulpen
		Quantum
		38
		Lem Gelu
		Panjang
		Gunting Kecil
X2	Unit	Tinta Epson 664
		Baterai AA
		Panasonic
		Amplop
		Baterai AA
		Panasonic
		Gunting
		Binder Klip
		Tinta Stempel
		1-20
21-50		
50-100		
100-1000		

<b>Varibel</b>	<b>Atribut</b>	<b>Keterangan</b>
X3	Satuan	Lbr Rim Buah Dos Pak
		Rangkap
X4	Harga	<1.000 (Rendah)
		1.100-35.000 (Sedang)
		>36.000 (Tinggi)
X5	Total	<10.000 (Sangat Rendah)
		10.100-50.000 (Rendah)
		50.100-100.00 (Tinggi)
		10.100-500.000 (Sangat Tinggi)

### 3.3. Preprocessing

Pada tahap *preprocessing* ini, dilakukan memastikan kualitas data yang telah dipilih dengan membersihkan data yang duplikat dan *missing value*.

### 3.4. Transformasi Data

Pada tahap ini, proses pengubahan format file untuk kebutuhann input dataset, yaitu (.xlsx) ke file csv.

### 3.5. Modeling dengan Algoritma Naïve Bayes

Hasil pengumpulan data penjualan, digunakan sebagai data *training* dan data *testing* untuk penentuan sangan banyak atau kurang banyak terjual untuk satu jenis barang di toko. Sebagian data *training* ditunjukkan pada Tabel 2. Selanjutnya untuk data *testing* ditunjukkan pada Tabel 3.

Tabel 2. Data *Traning*

<b>No.</b>	<b>Nama Barang</b>	<b>Unit</b>	<b>Satuan</b>	<b>Harga</b>	<b>Kelas</b>
1.	Foto copy	100	Lbr	Rp 350	Banyak Terjual
2.	Kertas F4	1	Rim	Rp 70.000	Banyak Terjual
3.	Notulen Lokmin Bulanan	53	Lbr	Rp 350	Kurang Terjual
4	Foto Copy	30	Lbr	Rp 350	Kurang Terjual
5	Jilid Tebal	1	Buah	Rp 10.000	Kurang Terjual
6	Kertas F4	2	Rim	Rp 70.000	Banyak Terjual
7	Foto Copy	350	Lbr	Rp 350	Banyak Terjual
8	Pulpen Quantum	2	Dos	Rp 2.000	Kurang Terjual

No.	Nama Barang	Unit	Satuan	Harga	Kelas
9	Double Tip	1	Buah	Rp 10.000	Kurang Terjual
10.	Tip Ex	1	Buah	Rp 8.000	Kurang Terjual
11.	Stapler HD-10	1	Buah	Rp 15.000	Kurang Terjual
12.	Etona No. 10	1	Buah	Rp 3.000	Kurang Terjual
13.	Lem Fox	1	Buah	Rp 15.000	Kurang Terjual
14.	Tissu Montis	1	Buah	Rp 15.000	Kurang Terjual
15.	Foto Copy	30	Lbr	Rp 350	Kurang Terjual

Tabel 3. Data *Testing*

No.	Nama Barang	Unit	Satuan	Harga	Total	Kelas
1.	Foto copy KIA	570	Lbr	Rp 350	Rp 199.500	?
2.	Pulpen Nevada	17	Buah	Rp 1.000	Rp 17.000	?
3.	Kertas A4	2	Rim	Rp 60.000	Rp 120.000	?
4	Map L	7	Lbr	Rp 2.000	Rp 14.000	?
5	Buku Nota 3 Ply	5	Buah	Rp 5.000	Rp 25.000	?

### 3.6. Menghitung Jumlah *Class/Label*

Pada tahap ini, menghitung jumlah *class/label* untuk data banyak terjual dan kurang terjual. Metode pengumpulan data yang digunakan dalam penelitian ini yaitu metode dokumentasi, dengan data penjualan Toko INTEC periode Januari 2023-Mei 2023.

$$P(H = \text{Banyak Terjual}) = \frac{207 \text{ (Jml data "BT" pada data training)}}{340 \text{ (Jml keseluruhan data training)}} \quad (2)$$

$$P(H = \text{Kurang Terjual}) = \frac{133 \text{ (Jml data "KT" pada data training)}}{340 \text{ (Jml keseluruhan data training)}} \quad (3)$$

### 3.8. Membandingkan Probabilitas

Membandingkan hasil probabilitas tiap kelas dengan mengalikan semua hasil perhitungan variabel banyak terjual dan kurang terjual.

$$\text{Banyak terjual} = P(X|H = \text{BT}) \times P(H = \text{BT}) \quad (4)$$

$$\begin{aligned} &= P(X1 = \text{Kertas A4} | H = \text{BT}) \\ &= P(X2 = 2 | H = \text{BT}) \\ &= P(X3 = \text{Rim} | H = \text{BT}) \\ &= P(X4 = 60.000 | H = \text{BT}) \\ &= P(X5 = 120.000 | H = \text{BT}) \\ &= P(H = \text{Banyak Terjual}) \\ P &= \frac{2}{207} \times \frac{19}{207} \times \frac{22}{207} \times \frac{1}{207} \times \frac{2}{207} \times \frac{207}{340} \\ &= 0.0000000296948 \end{aligned}$$



$$\begin{aligned}
 \text{Kurang terjual} &= P(X|H = KT) \times P(H = KT) \\
 &= P(X1 = \text{Kertas A4} | H = KT) \\
 &= P(X2 = 2 | H = KT) \\
 &= P(X3 = \text{Rim} | H = KT) \\
 &= P(X4 = 60.000 | H = KT) \\
 &= P(X5 = 120.000 | H = KT) \\
 &= P(H = \text{Kurang Terjual}) \\
 P &= \frac{1}{133} \times \frac{19}{133} \times \frac{3}{133} \times \frac{1}{133} \times \frac{1}{133} \times \frac{133}{340} \\
 &= 0.00000000065843
 \end{aligned}
 \tag{5}$$

Probabilitas dapat ditentukan:

$$\begin{aligned}
 \text{Probabilitas Banyak Terjual} &= \frac{\text{Banyak Terjual}}{\text{Kurang Terjual}} + \text{Banyak Terjual} \\
 &= \frac{0.00000000296948}{0.00000000065843} + 0.00000000296948 \\
 &= 4.509936198
 \end{aligned}
 \tag{6}$$

$$\begin{aligned}
 \text{Probabilitas Kurang Terjual} &= \frac{\text{Kurang Terjual}}{\text{Banyak Terjual}} + \text{Kurang Terjual} \\
 &= \frac{0.00000000065843}{0.00000000296948} + 0.00000000065843 \\
 &= 0.221732628
 \end{aligned}
 \tag{7}$$

Perbandingan hasil perkalian dan probabilitas aktif dan tidak aktif, hasil yang diperoleh dari perhitungan manual untuk salah satu data *testing* menunjukkan bahwa nilai probabilitas tertinggi terdapat pada kelas (P=Banyak Terjual), sehingga dapat disimpulkan bahwa hasil untuk salah satu data *testing* untuk nama barang Kertas A4, unit 2, satuan rim, harga 60.000, dan total 120.000 merupakan jenis barang yang banyak terjual.

Selanjutnya pada pemrograman Python, data dari Tabel 2 yang diinput dengan menggunakan format file *Comma Separated Values* (CSV) selanjutnya di ubah dengan menggunakan model data *integer*. Data *training* pada kolom kelas terdapat 2 data yaitu 0 dan 1. Keterangan untuk nilai 1 adalah nilai kelas data yang banyak terjual, sedangkan untuk nilai 0 adalah nilai kelas data yang kurang terjual. Tampilan data *training* ditunjukkan pada Tabel 4.

Tabel 4. Data *Training* dalam Program Python

<i>Data Training</i>						
	<i>Nama Barang</i>	<i>Unit</i>	<i>Satuan</i>	<i>Harga</i>	<i>Total</i>	<i>Kelas</i>
0.	Foto copy	100	Lbr	Rp 350	Rp 199.500	1
1.	Kertas A4	1	Rim	Rp 70000	Rp 17.000	1
2.	Notulen lokmin bulanna	53	Lbr	Rp 350	Rp 120.000	0

<i>Data Training</i>						
	<i>Nama Barang</i>	<i>Unit</i>	<i>Satuan</i>	<i>Harga</i>	<i>Total</i>	<i>Kelas</i>
3.	Foto Copy	30	Lbr	Rp 350	Rp 14.000	0
4.	Jilid Tebal	1	Buah	Rp 10000	Rp 25.000	0
...	...	...	...	...	...	...
338	Jilid	1	Buah	Rp 10000	Rp 10000	1
339	Foto Copy	180	Lbr	350	63000	1

Matriks data  $X$  training merupakan data latih pada variable  $X$  yang di ambil dari 272 baris data untuk dilatih pada ke-5 kolom selain kolom target (kelas). Matriks data  $X$  testing merupakan data uji pada variable  $X$  yang diambil dari 68 baris data untuk di uji pada ke-5 kolom selain kolom target (kelas). Sedangkan matriks data  $Y$  adalah matriks untuk kolom target (kelas). Jumlah matriks data hasil training ditunjukkan pada Tabel 5.

Tabel 5. Jumlah Matriks Hasil Data *Training*

No.	<i>Nama Barang</i>
1.	Matriks Data x Training = (272,5)
2.	Matrix Data x Testing = (68,5)
3.	Matriks Data y Training = (272,)
4.	Matriks Data y Testing = (68,)

Hasil *testing* pada data *testing* yang telah diinput ke dalam system menunjukkan nilai kelas 0 (Kurang Terjual), dan 1 (Banyak Terjual). Misalnya untuk data *testing* pertama dan ketiga mendapatkan nilai kelas 1 artinya barang tersebut banyak terjual, dan untuk data *testing* kedua, keempat dan kelima mendapatkan nilai kelas 0 yang artinya barang tersebut kurang terjual. Hasil pengujian ditunjukkan pada Tabel 6.

Tabel 6. Hasil *Testing*

No.	<i>Nama Barang</i>	<i>Unit</i>	<i>Satuan</i>	<i>Harga</i>	<i>Total</i>	<i>Kelas</i>
1.	Foto copy KIA	570	Lbr	Rp 350	Rp 199.500	1
2.	Pulpen Nevada	17	Buah	Rp 1.000	Rp 17.000	0
3.	Kertas A4	2	Rim	Rp 60.000	Rp 120.000	1
4	Map L	7	Lbr	Rp 2.000	Rp 14.000	0
5	Buku Nota 3 Ply	5	Buah	Rp 5.000	Rp 25.000	0

Hasil pengujian dari algoritma *Naïve Bayes* dengan tingkat akurasi 90% ditunjukkan pada tabel 7.

Tabel 7. Hasil Pengujian Algoritma Naive Bayes

Test Report	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Accuracy</i>
0	0.76	1.00	0.86	0.90
1	1.00	0.85	0.92	

## 4. Kesimpulan

Klasifikasi omset dilakukan untuk mengetahui berapa banyak pendapatan tambahan yang dihasilkan oleh suatu usaha. Persaingan usaha yang ketat mengharuskan perusahaan untuk selalu mampu memahami sebuah pasar yang terus berkembang. Peningkatan usaha menjadi prioritas pemerintah dalam upaya meningkatkan perekonomian Nasional. Ketersediaan produk dalam pemenuhan konsumen merupakan permasalahan yang harus di atasi. Keadaan stok suatu barang sangat mempengaruhi pendapatan penjualan. Jumlah permintaan barang yang meningkat, maka akan memperoleh pendapatan yang besar.

Penelitian ini berfokus pada sebuah system untuk klasifikasi omset, dalam hal ini klasifikasi omset ATK untuk masing-masing barang yang terjual di toko apakah banyak terjual atau kurang terjual untuk memaksimalkan stok barang per waktu dengan menggunakan Algoritma Naive Bayes. Jumlah data yang digunakan dalam penelitian ini sebanyak 345 data penjualan ATK. Dari data penelitian dilakukan pengelompokan atribut, *preprocessing*, transformasi data dan *modeling* dengan algoritma *Naive Bayes* dengan bahasa pemrograman python. Pengujian algoritma *Naive Bayes* mendapatkan hasil akurasi 90% untuk klasifikasi omset ATK.

## Daftar Pustaka

- Balaji, V. R., Suganthi, S. T., Rajadevi, R., Krishna Kumar, V., Saravana Balaji, B., & Pandiyan, S. (2020). Skin disease detection and segmentation using dynamic graph cut algorithm and classification through Naive Bayes classifier. *Measurement*, *163*, 107922. <https://doi.org/10.1016/j.measurement.2020.107922>
- Das, B. K., & Dutta, H. S. (2020). GFNB: Gini index–based Fuzzy Naive Bayes and blast cell segmentation for leukemia detection using multi-cell blood smear images. *Medical & Biological Engineering & Computing*, *58*(11), 2789–2803. <https://doi.org/10.1007/s11517-020-02249-y>
- Harahap, F., Harahap, A. Y. N., Ekadiansyah, E., Sari, R. N., Adawiyah, R., & Harahap, C. B. (2018). Implementation of Naïve Bayes Classification Method for Predicting Purchase. *2018 6th International Conference on Cyber and IT Service Management (CITSM)*, 1–5. <https://doi.org/10.1109/CITSM.2018.8674324>
- Kilari, H., Edara, S., Yarra, G. R. S., & Gadhiraju, D. V. (2022). Customer Segmentation using K-Means Clustering. *International Journal of Engineering Research*, *11*(03).
- Muchardie, B. G., Gunawan, A., & Aditya, B. (2019). E-Commerce Market Segmentation Based On The Antecedents Of Customer Satisfaction and Customer Retention. *2019 International Conference on Information Management and Technology (ICIMTech)*, 103–108. <https://doi.org/10.1109/ICIMTech.2019.8843792>
- Nandapala, E. Y. L., & Jayasena, K. P. N. (2020). The practical approach in Customers segmentation by using the K-Means Algorithm. *2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS)*, 344–349. <https://doi.org/10.1109/ICIIS51140.2020.9342639>
- Novitalia, N., Mawasgenti, P. D., Apriani, T., S., A. P., & Saifudin, A. (2021). Penggunaan Metode Naive Bayes Classifier untuk Mengevaluasi Kinerja Akademik Mahasiswa di Perguruan Tinggi. *Jurnal Teknologi Sistem Informasi dan Aplikasi*, *4*(2), 65. <https://doi.org/10.32493/jtsi.v4i2.7752>

- Peling, I. B. A., Arnawan, I. N., Arthawan, I. P. A., & Janardana, I. G. N. (2017). Implementation of Data Mining To Predict Period of Students Study Using Naive Bayes Algorithm. *International Journal of Engineering and Emerging Technology*, 2(1), 53. <https://doi.org/10.24843/IJEET.2017.v02.i01.p11>
- Pramono, P. P., Surjandari, I., & Laoh, E. (2019). Estimating Customer Segmentation based on Customer Lifetime Value Using Two-Stage Clustering Method. *2019 16th International Conference on Service Systems and Service Management (ICSSSM)*, 1–5. <https://doi.org/10.1109/ICSSSM.2019.8887704>
- Romadhon, M. R., & Kurniawan, F. (2021). A Comparison of Naive Bayes Methods, Logistic Regression and KNN for Predicting Healing of Covid-19 Patients in Indonesia. *2021 3rd East Indonesia Conference on Computer and Information Technology (EIconCIT)*, 41–44. <https://doi.org/10.1109/EIconCIT50028.2021.9431845>
- Shihab, S. H., Afroge, S., & Mishu, S. Z. (2019). RFM Based Market Segmentation Approach Using Advanced K-means and Agglomerative Clustering: A Comparative Study. *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 1–4. <https://doi.org/10.1109/ECACE.2019.8679376>
- Sianipar, J. P., Saputra, R. E., & Setianingsih, C. (2021). *Waves With Multi-Sensor System Based on Web Application Using Naive Bayes Algorithm*.
- Strutynska, I., Kozbur, G., Dmytrotsa, L., Sorokivska, O., & Melnyk, L. (2019). Influence of Digital Technology on Roadmap Development for Digital Business Transformation. *2019 9th International Conference on Advanced Computer Information Technologies (ACIT)*, 333–337. <https://doi.org/10.1109/ACITT.2019.8780056>
- Sudrartono, T. (2019). *Pengaruh Segmentasi Pasar Terhadap Tingkat Penjualan Produk Fashion Usaha Mikro Kecil*.
- Yaman, T. T., & Bilgic, E. (2019). Is Business Analytics Education Sufficient in Business Schools? The Case in Turkish Business Schools. *2019 3rd International Conference on Data Science and Business Analytics (ICDSBA)*, 152–156. <https://doi.org/10.1109/ICDSBA48748.2019.00040>
- Yemelyanov, O., Petrushka, T., Lesyk, L., Symak, A., & Vovk, O. (2020). Modelling and Information Support for the Development of Government Programs to Increase the Accessibility of Small Business Lending. *2020 IEEE 15th International Conference on Computer Sciences and Information Technologies (CSIT)*, 229–232. <https://doi.org/10.1109/CSIT49958.2020.9322040>
- Yoon, C. Y. (2020). Analyzing Enterprise Smart Business Capability in a Smart Technology Environment. *2020 3rd International Conference on Information and Computer Technologies (ICICT)*, 224–229. <https://doi.org/10.1109/ICICT50521.2020.00097>
- Yoseph, F., Ahamed Hassain Malim, N. H., Heikkilä, M., Brezilianu, A., Geman, O., & Paskhal Rostam, N. A. (2020). The impact of big data market segmentation using data mining and clustering techniques. *Journal of Intelligent & Fuzzy Systems*, 38(5), 6159–6173. <https://doi.org/10.3233/JIFS-179698>
- Zhang, J., & Liu, G. (2020). Prediction of Incompliance With Business Goals With Business-Related Data and Context Data. *IEEE Access*, 8, 187008–187020. <https://doi.org/10.1109/ACCESS.2020.3030809>