

Model Klasifikasi Produk Paling Diminati Pada Penjualan Sandal Wanita Menggunakan Metode *Naïve Bayes* (Studi Kasus pada UMKM Ann-D'Mello Sandals Krian Sidoarjo)

Tiffany Maulida Innayah¹, I Kadek Dwi Nuryana²

^{1,2}Jurusan Teknik Informatika/Program Studi S1 Sistem Informasi, Universitas Negeri Surabaya)

¹tiffany.20020@mhs.unesa.ac.id

²dwinuryana@unesa.ac.id

Abstrak— UMKM Ann-d'Mellow Sandals di Krian, Sidoarjo, menghadapi tantangan dalam menentukan model sandal yang paling diminati untuk menghindari kelebihan stok penjualan. Dalam dunia fashion yang selalu berubah, manajemen persediaan menjadi kompleks, dengan beberapa produk laris terjual sementara yang lain tetap di rak. Untuk mengatasi ketidakseimbangan ini, teknik data mining digunakan untuk menganalisis pola pembelian pelanggan dan mengidentifikasi tren pasar sehingga membantu UMKM dalam merancang strategi pemasaran dan penjualan yang lebih efektif. Penelitian ini bertujuan menerapkan metode *Naïve Bayes* untuk mengklasifikasikan produk sandal wanita yang paling diminati di UMKM Ann-D'Mellow Sandals dan mengevaluasi akurasi model tersebut. Berdasarkan hasil *cross-validation* dengan fold 3, 5, dan 10, model menunjukkan tingkat akurasi rata-rata sebesar 0.62 pada data pelatihan dan 0.61 pada data pengujian. Model ini juga mencapai akurasi keseluruhan sebesar 96.15%, yang menunjukkan bahwa metode *Naïve Bayes* cukup efektif dalam klasifikasi produk meskipun ada variasi akurasi antar fold. Hasil analisis menunjukkan bahwa jenis sandal 'Hills A' adalah yang paling laku dengan penjualan sebanyak 72 kali orderan. Temuan ini sangat berharga untuk strategi pemasaran dan manajemen stok, memungkinkan UMKM untuk mengalokasikan sumber daya lebih efisien, meningkatkan produksi dan persediaan produk 'Hills A', serta merencanakan kampanye pemasaran yang lebih terarah. Pengetahuan ini juga mendukung pengembangan produk yang lebih sesuai dengan preferensi konsumen, yang pada akhirnya meningkatkan kepuasan pelanggan dan mendukung pertumbuhan bisnis yang berkelanjutan.

Kata Kunci— UMKM, *Naive Bayes*, manajemen persediaan, preferensi pelanggan, strategi pemasaran

I. PENDAHULUAN

Indonesia menempati peringkat kesepuluh di dunia dalam hal daya beli masyarakat, yang tercermin dalam perubahan signifikan di industri fashion (Kemlu, 2018). Kebiasaan konsumtif masyarakat Indonesia tidak hanya dipengaruhi oleh keadaan ekonomi kelompok tertentu, tetapi juga berdampak besar pada pertumbuhan ekonomi nasional secara keseluruhan. Masyarakat Indonesia memiliki berbagai pilihan konsumsi sesuai dengan keinginan dan kapasitas finansial mereka [1]. Perkembangan dunia fashion sangat pesat, mendorong munculnya tren-tren baru dan model-model yang terus

berkembang. Tren sandal dan sepatu yang populer saat ini sebenarnya merupakan evolusi dari berbagai bentuk alas kaki yang telah ada di masa lampau. Transformasi signifikan dalam hal material dan desain terjadi seiring berjalannya waktu sebagai respons terhadap kebutuhan manusia, menjadikan alas kaki sebagai elemen penting dalam kehidupan sehari-hari [2].

Setiap individu memiliki standar dan preferensi masing-masing dalam memilih sandal, berdasarkan jenis, model, atau bahkan merek tertentu. Sandal tidak hanya dipilih berdasarkan aspek fashion, tetapi juga karena manfaat lain yang dimilikinya. Sandal menjadi bagian integral dalam industri mode, dengan beragam varian yang hadir untuk memenuhi berbagai kebutuhan. Pemilihan sandal harus disesuaikan dengan situasi tertentu, termasuk pertimbangan usia, ukuran, jenis kelamin, bahan, harga, dan kenyamanan [3]. Indonesia menghadapi krisis ekonomi pada tahun 1998 yang menyebabkan kemerosotan ekonomi nasional. Sejumlah perusahaan besar di berbagai sektor mengalami stagnasi bahkan berhenti beroperasi. Namun, Usaha Mikro, Kecil, dan Menengah (UMKM) berhasil bertahan dan berperan sebagai penyelamat ekonomi di tengah krisis moneter yang melanda berbagai sektor pada waktu tersebut. Aktivitas UMKM mampu tumbuh dan konsisten dalam perekonomian nasional, menciptakan lapangan pekerjaan yang produktif [4].

Tren fashion yang terus berkembang menjadi pendorong utama dalam mendukung pertumbuhan ekonomi melalui sektor UMKM. UMKM menjadi usaha yang potensial dan mendukung diversifikasi ekonomi, memberikan kontribusi signifikan terhadap perekonomian nasional. Perkembangan industri fashion juga dipengaruhi oleh faktor sosial, karena esensi kemajuan industri ini mengikuti arus perubahan sosial di masyarakat [5]. Faktor sosial memiliki dampak signifikan pada perkembangan industri fashion. Loyalitas terhadap merek, misalnya, tidak hanya terkait dengan aspek ekonomi tetapi juga dipengaruhi oleh jaringan sosial. Setiap jaringan sosial, sekecil apapun, memainkan peran penting, baik secara langsung maupun tidak langsung, dalam perkembangan industri fashion [6]. Adaptasi terhadap perkembangan teknologi menjadi kunci dalam pertumbuhan UMKM fashion di Indonesia. Layanan e-commerce menjadi platform penting bagi pelaku bisnis UMKM fashion, memungkinkan mereka mencapai pasar yang lebih luas dan meningkatkan efisiensi teknis bisnis mereka. Perkembangan UMKM fashion di

Indonesia mencerminkan kemajuan dalam desain dan kreativitas serta adaptasi cerdas terhadap dinamika pasar global dan kemajuan teknologi (Muhammad Deovan,2022).

Di Krian, Sidoarjo, industri UMKM seperti Ann-d'Mello Sandals memainkan peran penting dalam perekonomian lokal dengan menawarkan beragam model dan jenis sandal yang dapat disesuaikan dengan keinginan pelanggan. Penggunaan data penjualan dan teknik pengolahan data, seperti data mining, membantu UMKM menganalisis pola pembelian pelanggan dan mengoptimalkan stok produk. Dengan demikian, UMKM dapat mengambil keputusan pemasaran yang lebih cerdas, menghindari kelebihan stok, dan meningkatkan efisiensi bisnis mereka [7].

Perkembangan data mining saat ini menjadi kunci sukses dalam meningkatkan efisiensi UMKM fashion. Dengan kemampuannya untuk menghitung dan menggabungkan data secara kompleks, data mining tidak hanya membantu mengelola persediaan dengan lebih efisien tetapi juga memberikan landasan yang kokoh untuk pengambilan keputusan bisnis. Implementasi strategis dalam memanfaatkan data dan data mining menjadi langkah penting dalam menghadapi dinamika pasar fashion yang cepat berubah (Indriyani,2023).

UMKM Ann-d'Mello Sandals di Krian, Sidoarjo, menghadapi tantangan dalam menentukan model sandal yang paling diminati oleh pelanggan guna menghindari kelebihan stok. Solusi untuk masalah ini melibatkan strategi pengelolaan persediaan yang cerdas dan berbasis data. Ann-d'Mello Sandals dapat memanfaatkan data penjualan historis dan preferensi pelanggan untuk mengidentifikasi pola pembelian yang konsisten. Dengan menggunakan metode Naive Bayes, UMKM dapat memprediksi model sandal yang kemungkinan besar diminati pelanggan berdasarkan faktor-faktor seperti warna, desain, atau tipe sandal.

Implementasi data mining Naive Bayes memungkinkan Ann-d'Mello Sandals untuk mengidentifikasi korelasi antara karakteristik tertentu dari model sandal dengan preferensi pelanggan. Dengan mengevaluasi data seperti umpan balik pelanggan atau data survei, UMKM dapat memahami faktor-faktor yang paling mempengaruhi keputusan pembelian pelanggan. Hal ini membantu UMKM menyesuaikan produksi mereka agar lebih sesuai dengan kebutuhan pasar. Melalui penerapan metode ini, UMKM Ann-d'Mello Sandals dapat secara proaktif mengelola persediaan mereka. Dengan memprediksi model yang akan diminati oleh pelanggan, UMKM dapat menghindari kelebihan stok dan mengoptimalkan produksi. Dengan demikian, UMKM dapat meningkatkan efisiensi bisnis mereka dan meminimalkan risiko kerugian akibat penumpukan barang yang tidak terjual.

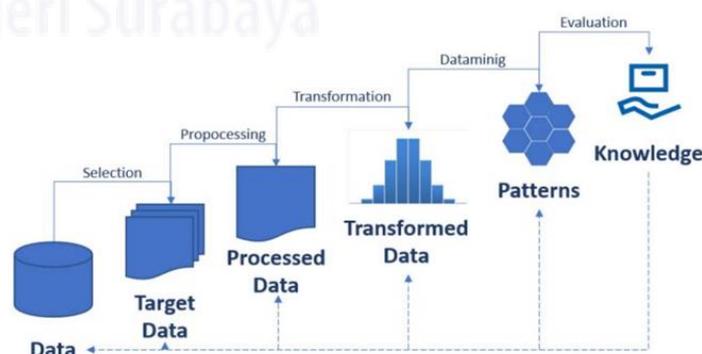
Penelitian sebelumnya oleh Edy Widodo (2019) menggunakan metode Naive Bayes untuk klasifikasi penjualan alat tulis kantor menunjukkan hasil akurasi tertinggi sebesar 99.26% dan akurasi terendah sebesar 94.39%. Hasil evaluasi kurva ROC/AUC menunjukkan kategori klasifikasi sangat baik, dengan skor tertinggi sebesar 0.997. Penelitian ini

menunjukkan potensi metode Naive Bayes dalam membantu UMKM mengoptimalkan strategi pemasaran [8].

Metode Naive Bayes dapat digunakan untuk melakukan klasifikasi real-time saat pelanggan berbelanja dengan menggabungkan data historis dan informasi aktual dari setiap transaksi, memungkinkan UMKM Ann-d'Mello Sandals memberikan rekomendasi model sandal yang sesuai dengan preferensi pelanggan secara cepat. Hal ini meningkatkan pengalaman berbelanja pelanggan, mengoptimalkan penjualan, dan mengurangi kelebihan stok. Dengan menerapkan metode ini, UMKM Ann-d'Mello Sandals dapat meningkatkan presisi dalam merespons kebutuhan pasar, mempercepat pengambilan keputusan, dan secara efektif mengelola inventaris. Berdasarkan latar belakang tersebut, penulis tertarik untuk mengkaji lebih lanjut masalah klasifikasi produk sandal yang paling diminati untuk meningkatkan penjualan dan mengurangi penimbunan bahan di gudang, yang akan dibahas dalam penelitian berjudul "Model Klasifikasi Produk Paling Diminati Pada Penjualan Sandal Wanita Menggunakan Metode Naive Bayes (Studi Kasus: UMKM Ann-d'Mello Sandals Krian Sidoarjo)".

II. METODE PENELITIAN

Metode penelitian adalah serangkaian langkah yang dijalankan dalam penelitian ini untuk mencapai hasil yang diinginkan sesuai dengan tujuan penelitian, dengan memperhatikan batasan yang telah ditetapkan. Dalam penelitian ini, digunakan pendekatan KDD (Knowledge Discovery in Database) untuk menganalisis data secara mendalam. KDD merupakan proses yang terdiri dari beberapa tahapan, termasuk pembersihan data, transformasi data, penambangan data, dan evaluasi pola atau pengetahuan yang ditemukan. Dengan menerapkan KDD, penelitian ini bertujuan untuk mengungkapkan pola-pola tersembunyi, tren, dan hubungan dalam dataset yang ada, sehingga dapat memberikan wawasan yang berharga dan mendukung pengambilan keputusan yang lebih baik. Pendekatan ini memungkinkan analisis data yang lebih komprehensif dan sistematis, memastikan bahwa hasil yang diperoleh dapat diandalkan dan relevan dengan tujuan penelitian.



Gbr. 1 KDD (Knowledge Discovery in Database)

A. Data Selection

Data Selection adalah fase penting dalam penelitian yang melibatkan identifikasi dan pengumpulan data yang relevan untuk mencapai tujuan penelitian. Proses ini dimulai dengan penentuan tujuan penelitian yang spesifik, diikuti oleh identifikasi sumber data yang dapat mendukung analisis yang diperlukan. Sumber data dapat berupa data penjualan, data pelanggan, atau data inventaris yang diperoleh dari internal organisasi atau sumber eksternal yang relevan. Setelah sumber data diidentifikasi, data dikumpulkan menggunakan metode yang sesuai seperti pengambilan data langsung dari sistem informasi internal, wawancara dengan pemangku kepentingan, atau survei. Tahap ini memastikan bahwa dataset yang dikumpulkan representatif, akurat, dan relevan untuk analisis lebih lanjut.

Dalam penelitian ini, data dikumpulkan langsung dari UMKM Ann-d'Mellow Sandals, mencakup data historis penjualan dari Januari 2021 hingga Desember 2023, yang meliputi informasi rinci tentang jumlah penjualan setiap produk sandal wanita, preferensi pelanggan, dan tren penjualan tahunan. Pengumpulan data historis ini penting untuk memahami dinamika penjualan dan mengidentifikasi pola yang dapat digunakan untuk membuat prediksi yang akurat di masa mendatang, dengan periode 2021 hingga 2023 dipilih karena mencakup rentang waktu yang cukup untuk mengamati perubahan signifikan dalam preferensi pelanggan dan tren pasar.

Bulan	Tahun	Minggu	Jenis Sandal	Permintaan_kodi	Produksi_kodi	Harga	Kategori		
Januari	2021	1	wedges	10	5	Mahal	Laku		
			hills A	15	10	Standar	Laku		
			flat	10	12	Standar	Kurang Laku		
			Hills A	15	12	Standar	Laku		
		2	flat	10	12	Mahal	Kurang Laku		
			wedges	8	10	Standar	Kurang Laku		
			hills B	13	10	Standar	Laku		
			flat	10	13	Mahal	Kurang Laku		
		3	wedges	10	9	Standar	Laku		
			Hills A	17	13	Standar	Laku		
			4	flat	12	15	Mahal	Kurang Laku	
				Hills A	18	16	Mahal	Laku	
		Hills B		15	10	Standar	Laku		
		flat							
		Februari	2021	1	Hills A	15	12	Standar	Laku
					wedges			Standar	Kurang Laku
flat	8				10	Mahal	Kurang Laku		
Hills A	17				13	Standar	Laku		
3	hills B			9	12	Standar	Kurang Laku		
	wedges			12	15	Standar	Kurang Laku		

Gbr. 2 Data Awal

B. Preprocessing

Tahap preprocessing melibatkan beberapa langkah krusial, seperti pengumpulan dan pembersihan data, di mana data penjualan historis diperiksa untuk menghilangkan data yang hilang atau tidak konsisten. Pada tahap preprocessing, terdapat 3 langkah yang akan dilakukan:

1. Menghapus kolom tahun, bulan, dan minggu

Pada tahap preprocessing ini langkah penting adalah menghapus fitur yang tidak berguna, seperti bulan, tahun, dan minggu. Penghapusan fitur-fitur ini dilakukan setelah evaluasi menyeluruh untuk memastikan bahwa

informasi temporal tersebut tidak memiliki korelasi signifikan dengan variabel target atau tidak menyumbang pada peningkatan performa model. Menghapus fitur-fitur temporal dapat menyederhanakan model, meningkatkan efisiensi komputasi, dan fokus pada fitur-fitur yang lebih relevan dan informatif, sehingga model yang dihasilkan menjadi lebih akurat dan generalis.

TABEL I
DATA SETELAH PREPROCESSING

Jenis Sandal	Permintaan_kodi	Produksi_kodi	Harga	Kategori
wedges	10	5	Mahal	Laku
hills A	15	10	Mahal	Laku
Flat	10	12	Standar	Kurang Laku
Hills A	15	12	Mahal	Laku
Flat	10	12	Standar	Kurang Laku
wedges	8	10	Standar	Kurang Laku
hills B	13	10	Standar	Laku
Flat	10	13	Mahal	Kurang Laku
wedges	10	9	Standar	Laku
Hills A	17	13	Mahal	Laku
Flat	12	15	Mahal	Kurang Laku
Hills A	18	16	Mahal	Laku
Hills B	15	10	Standar	Laku
Hills A	15	12	Mahal	Laku
Flat	8	10	Standar	Kurang Laku
Hills A	17	13	Mahal	Laku

2. Pengecekan missing value

Pada sub bab ini dijelaskan tahapan penanganan missing value untuk memastikan analisis data yang akurat dan efisien. Langkah pertama adalah mengidentifikasi missing value dalam dataset menggunakan teknik eksplorasi data seperti visualisasi atau ringkasan statistik. Selanjutnya, dilakukan evaluasi untuk menentukan metode penanganan yang sesuai, seperti menghapus baris atau kolom dengan missing value yang banyak, atau menggantinya dengan nilai substitusi seperti mean, median, atau modus. Penanganan missing value ini penting untuk menjaga kualitas dan akurasi model prediktif, sehingga langkah-langkah ini bertujuan untuk meminimalkan dampak negatif dari missing value pada hasil analisis.

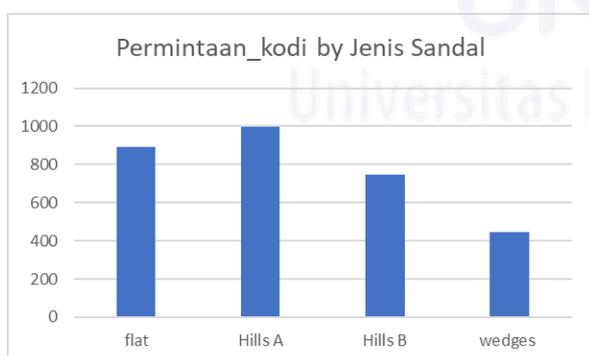
```
Jumlah nilai yang hilang per kolom:
Jenis Sandal      0
Permintaan_kodi   94
Produksi_kodi     37
Kategori          26
dtype: int64
```

Gbr. 3 Pengecekan Missing Value

Hasil kesimpulan dari analisis missing value menunjukkan jumlah nilai yang hilang per kolom dalam dataset sebagai berikut: kolom 'Jenis Sandal' tidak memiliki nilai yang hilang (0), kolom 'Permintaan_kodi' memiliki 94 nilai yang hilang, kolom 'Produksi_kodi' memiliki 37 nilai yang hilang, dan kolom 'Kategori' memiliki 26 nilai yang hilang. Dari data ini, terlihat bahwa kolom 'Permintaan_kodi' memiliki jumlah missing value yang paling signifikan, yang dapat mempengaruhi analisis dan model prediktif jika tidak ditangani dengan tepat. Sebaliknya, kolom 'Jenis Sandal' tidak memerlukan penanganan khusus untuk missing value. Penting untuk mengatasi missing value ini dengan metode yang sesuai agar data yang digunakan dalam analisis lebih lengkap dan representatif, sehingga menghasilkan model yang lebih akurat.

3. Eksplorasi data

Pada tahap eksplorasi data untuk klasifikasi penjualan sandal wanita yang paling diminati, dilakukan langkah-langkah untuk memahami pola dan tren dalam data penjualan, dimulai dengan pengumpulan data penjualan yang relevan dan dilanjutkan dengan preprocessing data untuk memastikan format yang konsisten dan mengatasi missing value. Setelah data siap, visualisasi seperti histogram digunakan untuk menggambarkan distribusi penjualan per jenis sandal, membantu mengidentifikasi jenis sandal yang paling diminati oleh konsumen. Misalnya, sandal jenis hills A mungkin menunjukkan permintaan tertinggi. Informasi ini memungkinkan analisis lebih lanjut untuk memahami faktor-faktor yang mempengaruhi popularitas sandal tertentu, seperti desain, harga, atau musim, dan sangat berguna untuk menginformasikan strategi penjualan dan produksi agar stok sandal yang paling diminati tersedia untuk memenuhi permintaan pasar.



Gbr. 4 Histogram Permintaan Jenis Sandal

Hasil analisis dari histogram menunjukkan jumlah permintaan untuk empat jenis sandal: flat, hills A, hills B, dan wedges. Sandal flat memiliki permintaan sekitar 900 kodi, sementara hills A memimpin dengan 1100 kodi,

menunjukkan tingginya minat konsumen. Hills B memiliki permintaan sekitar 800 kodi, dan wedges memiliki permintaan terendah sekitar 600 kodi. Informasi ini memberikan wawasan berharga mengenai preferensi konsumen, memungkinkan perusahaan menyesuaikan strategi produksi dan pemasaran untuk memenuhi permintaan pasar.

C. Transformasi Data

Sub bab ini menjelaskan transformasi teks menjadi vektor menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency). TF-IDF mengubah teks mentah menjadi representasi numerik yang digunakan dalam algoritma pembelajaran mesin. Prosesnya melibatkan penghitungan frekuensi kemunculan kata dalam dokumen (term frequency) dan menyesuainya dengan seberapa sering kata tersebut muncul di seluruh dokumen dalam dataset (inverse document frequency). Hasilnya adalah bobot untuk setiap kata, dengan kata-kata umum memiliki bobot lebih rendah dan kata-kata jarang namun relevan memiliki bobot lebih tinggi. Transformasi ini penting untuk menangkap makna dan informasi dari teks secara efektif, memungkinkan model lebih akurat dalam tugas seperti klasifikasi teks atau analisis sentimen.

TABEL II
 DATA HASIL TRANSFORMASI TF-IDF

flat	hills	wedges
0	0	1
0	1	0
1	0	0
0	1	0
1	0	0
0	0	1
0	1	0
1	0	0
0	1	0
0	1	0
1	0	0
0	1	0
0	1	0
0	1	0
1	0	0
0	1	0
0	1	0
0	0	1
0	1	0

D. Data Mining

Sub bab ini menjelaskan implementasi model Naive Bayes untuk mengklasifikasikan sandal wanita yang paling diminati oleh pelanggan UMKM Ann-D'Mello Sandals. Model ini, yang menggunakan data penjualan dan preferensi pelanggan dari tahun 2020 hingga 2023, menganalisis atribut seperti jenis sandal, harga, dan fitur lainnya yang mempengaruhi kepopuleran produk. Setelah data disiapkan, model Naive Bayes dilatih untuk menghitung probabilitas masing-masing kelas atau label dari data yang diberikan. Ini memungkinkan klasifikasi produk berdasarkan desain, ukuran, dan harga untuk memprediksi produk yang paling diminati pelanggan. Hasilnya membantu UMKM merencanakan strategi pemasaran yang lebih tepat dan efektif serta meningkatkan pengelolaan stok barang sesuai permintaan pasar. Pendekatan ini memberikan rekomendasi tentang produk yang perlu dipromosikan atau diperluas produksinya, mendukung pengambilan keputusan yang lebih cerdas dalam strategi bisnis UMKM.

1. Pembangunan Model

Pada sub bab ini dijelaskan proses tahapan modelling menggunakan algoritma Naive Bayes. Mulai dengan membagi dataset menjadi data pelatihan dan data pengujian, umumnya dengan perbandingan 80:20. Kemudian, algoritma Naive Bayes digunakan pada data pelatihan untuk membangun model prediktif. Ini melibatkan menghitung probabilitas bersyarat untuk setiap fitur terhadap kelas target, dengan asumsi fitur independen. Setelah model terbentuk, evaluasi dilakukan pada data pengujian dengan metrik seperti akurasi, presisi, recall, dan F1-score untuk mengukur kinerja model terhadap data baru.

2. Perhitungan Probabilitas Posterior dan Prediksi Kelas

Langkah berikutnya adalah mengestimasi probabilitas prior untuk setiap kelas dari data pelatihan. Setelah itu, dihitunglah probabilitas kondisional dari setiap fitur terhadap kelas yang bersangkutan. Dengan menerapkan teorema Bayes, kita dapat menghitung probabilitas posterior untuk setiap kelas berdasarkan fitur-fitur yang diamati. Langkah ini merupakan inti dari algoritma Naive Bayes dan menjadi dasar dari prediksi yang dilakukan oleh model.

Langkah 1 : Menghitung probabilitas prior untuk setiap kategori

Jumlah sampel:

$$\text{Total sampel} = 5$$

$$\text{Jumlah sampel "Laku"} = 3$$

$$\text{Jumlah sampel "Kurang Laku"} = 2$$

Probabilitas prior:

$$P(\text{Laku}) = \frac{3}{5} = 0.6$$

$$P(\text{Kurang Laku}) = \frac{2}{5} = 0.4$$

Langkah 2 : Menghitung probabilitas atribut untuk setiap kategori

Probabilitas untuk Kategori "Laku":

1. Jenis Sandal :

$$P(\text{Jenis Sandal} = \text{hills A} \mid \text{Laku}) = \frac{2}{3} = 0.667$$

2. Permintaan_kodi :

$$P(\text{Permintaan_kodi} = 15 \mid \text{Laku}) = \frac{2}{3} = 0.667$$

3. Produksi_kodi :

$$P(\text{Produksi_kodi} = 12 \mid \text{Laku}) = \frac{1}{3} = 0.333$$

4. Harga :

$$P(\text{Harga} = \text{Mahal} \mid \text{Laku}) = \frac{2}{3} = 0.667$$

Probabilitas untuk Kategori " Kurang Laku":

1. Jenis Sandal :

$$P(\text{Jenis Sandal} = \text{hills A} \mid \text{Kurang Laku}) = \frac{0}{2} = 0$$

2. Permintaan_kodi :

$$P(\text{Permintaan_kodi} = 15 \mid \text{Kurang Laku}) = \frac{0}{2} = 0$$

3. Produksi_kodi :

$$P(\text{Produksi_kodi} = 12 \mid \text{Kurang Laku}) = \frac{2}{2} = 1$$

4. Harga :

$$P(\text{Harga} = \text{Mahal} \mid \text{Kurang Laku}) = \frac{0}{2} = 0$$

Langkah 3 : Menghitung probabilitas posterior

Misalkan kita ingin memprediksi kategori untuk sampel baru dengan atribut berikut:

Jenis Sandal	: hills A
Permintaan_kodi	: 15
Produksi_kodi	: 12
Harga	: Mahal

Probabilitas Posterior untuk Kategori "Laku":

$$P(\text{Laku} \mid \text{hills A, 15, 12, Mahal}) = \frac{P(\text{hills A, 15, 12, Mahal} \mid \text{Laku}) \cdot P(\text{Laku})}{P(\text{hills A, 15, 12, Mahal})}$$

$$P(\text{hills A, 15, 12, Mahal} \mid \text{Laku}) = P(\text{hills A} \mid \text{Laku}) \cdot P(15 \mid \text{Laku}) \cdot P(12 \mid \text{Laku}) \cdot P(\text{Mahal} \mid \text{Laku})$$

$$= 0.667 \times 0.667 \times 0.333 \times 0.667 = 0.099$$

$$P(\text{Laku} \mid \text{hills A, 15, 12, Mahal}) = \frac{0.099 \times 0.6}{0.0594} = 1$$

Probabilitas Posterior untuk Kategori "Kurang Laku":

$$P(\text{Kurang Laku} \mid \text{hills A, 15, 12, Mahal}) = \frac{P(\text{hills A, 15, 12, Mahal} \mid \text{Kurang Laku}) \cdot P(\text{Kurang Laku})}{P(\text{hills A, 15, 12, Mahal})}$$

$$P(\text{hills A, 15, 12, Mahal} \mid \text{Kurang Laku}) = P(\text{hills A} \mid \text{Kurang Laku}) \cdot P(15 \mid \text{Kurang Laku}) \cdot P(12 \mid \text{Kurang Laku}) \cdot P(\text{Mahal} \mid \text{Kurang Laku})$$

$$= 0 \times 0 \times 1 \times 0 = 0$$

$$P(\text{Laku} \mid \text{hills A, 15, 12, Mahal}) = \frac{0 \times 0.4}{0.0594} = 0$$

Hitung P(X)

$$P(X) = P(\text{hills A, 15, 12, Mahal}) = P(X \mid \text{Laku}) \cdot P(\text{Laku}) + P(X \mid \text{Kurang Laku}) \cdot P(\text{Kurang Laku})$$

$$P(X) = (0.099 \times 0.6) + (0 \times 0.4) = 0.0594$$

Hasil dari Probabilitas posterior untuk kategori "Laku" adalah 1, yang berarti ada keyakinan 100% bahwa sampel dengan atribut tersebut adalah dalam kategori "Laku". Sebaliknya, probabilitas posterior untuk kategori "Kurang Laku" adalah 0, yang berarti ada keyakinan 0% bahwa sampel tersebut dalam kategori "Kurang Laku".

3. Optimasi Model

Pada langkah ini, model terbaik dipilih menggunakan GridSearchCV, sebuah teknik untuk mencari parameter optimal dalam model machine learning. Proses ini melibatkan penentuan nilai-nilai potensial untuk beberapa parameter model, di mana GridSearchCV secara sistematis mengevaluasi semua kombinasi parameter untuk menemukan yang menghasilkan kinerja terbaik.

```
Best parameters found by grid search: {'clf__alpha': 0.1, 'vect__max_features': 500, 'vect__ngram_range': (1, 1)}
```

Kombinasi parameter ini memberikan hasil terbaik dalam akurasi model pada data validasi silang selama proses GridSearchCV. Dengan parameter ini, model diharapkan dapat mengklasifikasikan data baru dengan lebih akurat dan efisien. Temuan ini menunjukkan bahwa dengan pengaturan smoothing yang sesuai, jumlah fitur yang terbatas, dan menggunakan unigram, model Naive Bayes dapat memberikan kinerja optimal untuk dataset ini.

E. Evaluation

Pada tahap evaluasi ini, model atau hasil dari tahapan sebelumnya seperti pemilihan data, preprocessing, transformasi, dan pemodelan dievaluasi secara kritis untuk mengukur kualitas dan efektivitasnya. Evaluasi dilakukan dengan menguji model terhadap dataset terpisah atau menggunakan teknik validasi silang untuk mencegah overfitting dan memastikan generalisasi yang baik. Tujuannya adalah untuk menilai sejauh mana model memenuhi tujuan penelitian awal serta untuk mengevaluasi performa objektif model dengan metrik-metrik seperti akurasi, presisi, recall, dan F1-score. Evaluasi kinerja model Naive Bayes ini memberikan wawasan penting untuk memahami kemampuan model dalam memprediksi preferensi pelanggan, membantu UMKM Ann-D'Mello Sandals merancang strategi pemasaran yang lebih efektif, dan mengoptimalkan manajemen stok produk.

F. Knowledge

Dengan menerapkan model Naive Bayes pada data UMKM Ann-D'Mello Sandals, tahap ini akan menghasilkan pemahaman mendalam tentang preferensi pelanggan terhadap produk sandal wanita. Model ini mengidentifikasi pola-pola yang menunjukkan produk sandal yang paling diminati, berdasarkan desain, harga, ukuran, dan atribut lainnya. Analisis ini memberikan wawasan tentang faktor-faktor yang mempengaruhi keputusan pembelian pelanggan dalam kategori produk ini.

Selain itu, pengetahuan yang dihasilkan juga mencakup pemahaman tentang produk sandal wanita yang kurang diminati atau tidak populer di pasar. Dengan mengidentifikasi pola-pola ini, UMKM Ann-D'Mello Sandals dapat mengambil langkah-langkah strategis untuk meningkatkan atau mengubah produk yang kurang diminati agar lebih sesuai dengan preferensi pelanggan, atau mengurangi produksi produk dengan tingkat permintaan rendah. Pengetahuan ini mendukung pengambilan keputusan yang lebih cerdas dalam mengelola portofolio produk dan meningkatkan strategi pemasaran untuk pertumbuhan bisnis yang berkelanjutan..

III. HASIL DAN PEMBAHASAN

A. Hasil Penerapan Naive Bayes

Pada sub bab ini dibahas hasil penerapan algoritma Naive Bayes untuk klasifikasi produk sandal wanita yang paling diminati. Hasil menunjukkan bahwa model Naive Bayes mampu mengklasifikasikan produk dengan akurasi memuaskan, mengidentifikasi produk yang paling diminati berdasarkan data historis permintaan dan produksi. Evaluasi metrik seperti akurasi, presisi, recall, dan F1-score menunjukkan kinerja model yang baik dalam memprediksi kategori produk dengan permintaan tinggi. Hasil ini memberikan wawasan berharga bagi perusahaan untuk strategi pemasaran dan pengelolaan stok barang, sehingga lebih efektif memenuhi permintaan konsumen dan mengoptimalkan penjualan. Model ini juga membantu memahami preferensi pelanggan, yang bermanfaat untuk pengembangan produk dan promosi di masa depan.

a. Evaluasi Model

Pada sub bab ini, dilakukan evaluasi kinerja model menggunakan metrik F1-score, akurasi, presisi, dan recall. F1-score mengukur keseimbangan antara presisi dan recall, berguna ketika ada ketidakseimbangan kelas pada data, dan menggabungkan keduanya menjadi satu nilai untuk menilai kemampuan model dalam mengidentifikasi kelas positif. Akurasi mengukur seberapa baik model mengklasifikasikan keseluruhan data dengan benar, presisi mengukur proporsi prediksi positif yang benar dari total prediksi positif, dan recall mengukur proporsi prediksi positif yang benar dari total kasus positif sebenarnya. Menggunakan ketiga metrik ini bersama-sama memberikan pemahaman komprehensif tentang kinerja model dalam

mengklasifikasikan data. Berikut hasil akurasi yang diperoleh.

```

Classification Report:
              precision    recall  f1-score   support

   Kurang Laku      0.97      0.97      0.97        32
     Laku          0.95      0.95      0.95        20

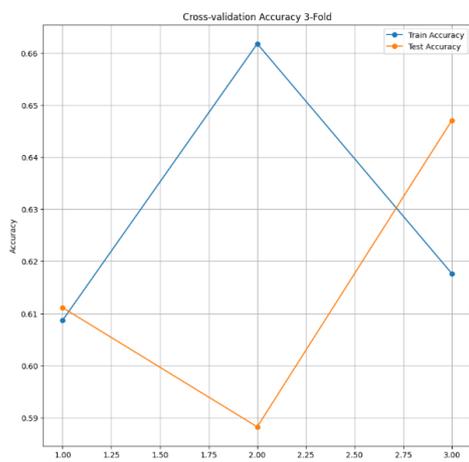
 accuracy                   0.96        52
 macro avg              0.96      0.96      0.96        52
 weighted avg          0.96      0.96      0.96        52
    
```

b. Evaluasi Model Menggunakan Crossvalidation

Pada sub bab ini, dilakukan evaluasi kinerja model menggunakan metode cross-validation untuk mengukur akurasi. Cross-validation adalah teknik umum dalam pembelajaran mesin yang mengevaluasi performa model secara objektif dan mencegah overfitting atau underfitting. Dalam metode ini, data dibagi menjadi beberapa subset (fold) yang tidak tumpang tindih. Model dilatih menggunakan k-1 subset dan diuji pada subset yang tersisa. Proses ini diulang beberapa kali dengan pembagian berbeda, dan rata-rata akurasi dari semua iterasi dihitung untuk mendapatkan akurasi akhir.

Hasil akurasi dari cross-validation memberikan gambaran kinerja model yang lebih stabil dibandingkan menggunakan satu set data pengujian, karena setiap data diuji pada model yang dilatih dengan data berbeda-beda. Ini memastikan bahwa evaluasi kinerja model tidak dipengaruhi oleh pembagian data tertentu dan memberikan gambaran konsisten tentang seberapa baik model dapat memprediksi data baru.

- **Fold 3**
 Pada sub bab ini, dilakukan evaluasi kinerja model Naive Bayes menggunakan metode cross-validation dengan 3 fold. Data dibagi menjadi tiga subset yang sama besar; model dilatih pada dua subset dan diuji pada subset yang tersisa, diulang tiga kali sehingga setiap subset digunakan sebagai data uji sekali. Hasil akurasi dari setiap iterasi memberikan perkiraan kinerja model yang lebih andal dibandingkan penggunaan satu pembagian data train-test. Metode ini memastikan penilaian model yang objektif dan memberikan perkiraan akurasi yang konsisten dari model Naive Bayes yang disesuaikan dengan parameter terbaik dari Grid Search.



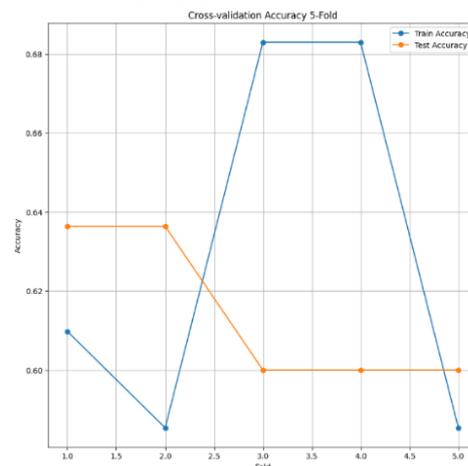
Gbr. 5 Hasil Grafik Cross-Validation Fold 3

TABEL III
 HASIL AKURASI CROSS-VALIDATION FOLD 3

Fold Ke	Train	Test
1	0.60	0.61
2	0.66	0.58
3	0.61	0.64
Mean	0.62	0.61

Hasil cross-validation menunjukkan bahwa model Naive Bayes menghasilkan akurasi yang bervariasi pada setiap fold data latih, berkisar antara 0.608 hingga 0.662, dengan rata-rata akurasi 0.629. Untuk data uji, akurasi bervariasi antara 0.611 hingga 0.647 per fold, dengan rata-rata akurasi 0.615. Variabilitas ini mencerminkan konsistensi model dalam menangani data yang berbeda, meskipun ada sedikit fluktuasi performa pada setiap iterasi cross-validation.

- **Fold 5**
 Pada sub bab ini, dilakukan validasi silang (cross validation) menggunakan metode Naive Bayes dengan pembagian data menjadi 5 lipatan (fold-5). Proses ini melibatkan pembagian data pelatihan menjadi lima bagian yang sama besar, di mana pada setiap iterasi, empat bagian digunakan untuk melatih model dan satu bagian digunakan untuk menguji model. Tujuan dari validasi silang ini adalah untuk mengevaluasi kinerja model secara lebih akurat dengan mengurangi bias yang mungkin terjadi jika hanya menggunakan satu set pelatihan dan pengujian.



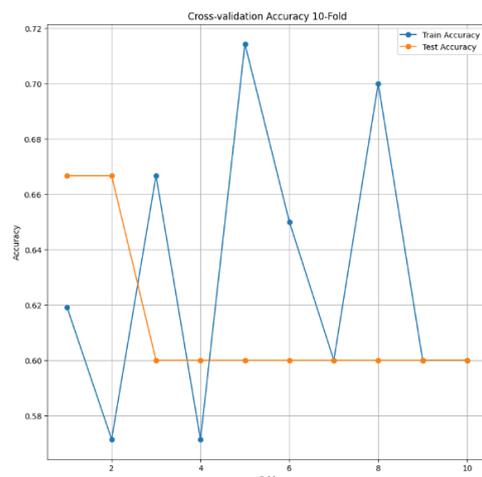
Gbr. 6 Hasil Grafik Cross-Validation Fold 5

TABEL IV
 HASIL AKURASI CROSS-VALIDATION FOLD 5

Fold Ke	Train	Test
1	0.60	0.63
2	0.58	0.63
3	0.68	0.6
4	0.68	0.6
5	0.58	0.6
Mean	0.62	0.61

Hasil dari evaluasi cross-validation menunjukkan variasi performa model Naive Bayes pada data latih, dengan akurasi yang berkisar antara 0.585 hingga 0.683 per fold. Rata-rata akurasi dari cross-validation train adalah sekitar 0.629. Pada data uji, model menunjukkan akurasi antara 0.6 hingga 0.636 per fold, dengan rata-rata akurasi sekitar 0.615. Hasil ini mengindikasikan bahwa model memiliki konsistensi yang relatif baik dalam menangani data uji yang belum pernah dilihat sebelumnya, meskipun ada sedikit fluktuasi dalam performa antar iterasi cross-validation.

- **Fold 10**
 Pada sub bab ini, dilakukan evaluasi kinerja model menggunakan metode cross-validation dengan 10 fold, yang diterapkan pada model klasifikasi Naive Bayes. Cross-validation membagi data latih menjadi 10 subset yang sama besar, di mana setiap subset digunakan secara bergantian sebagai data validasi sementara yang tersisa digunakan sebagai data latih. Proses ini diulang untuk setiap subset, sehingga setiap observasi dalam data latih diuji dengan akurasi yang dapat diandalkan. Hasil akhir dari proses ini memberikan estimasi yang lebih konsisten tentang kinerja model dibandingkan dengan sekadar membagi data menjadi satu set latih dan satu set uji saja.



Gbr. 7 Hasil Grafik Cross-Validation Fold 10

TABEL V
 HASIL AKURASI CROSS-VALIDATION FOLD 10

Fold Ke	Train	Test
1	0.61	0.66
2	0.57	0.66
3	0.66	0.6
4	0.57	0.6
5	0.71	0.6
6	0.65	0.6
7	0.6	0.6
8	0.7	0.6
9	0.6	0.6
10	0.6	0.6
Mean	0.62	0.61

Hasil dari cross-validation dengan 10 fold menggunakan metode Naive Bayes menunjukkan bahwa model memiliki tingkat akurasi yang bervariasi antara data pelatihan (train) dan data pengujian (test). Pada data pelatihan, akurasi berkisar antara 0.57 hingga 0.71 untuk setiap iterasi fold, dengan rata-rata akurasi sebesar 0.629. Sedangkan pada data pengujian, akurasi juga bervariasi antara 0.6 hingga 0.67 untuk setiap iterasi fold, dengan rata-rata akurasi sebesar 0.613.

Meskipun terdapat variasi dalam akurasi antar fold, rata-rata akurasi yang diperoleh menunjukkan bahwa model memiliki kinerja yang stabil dan mampu melakukan klasifikasi dengan baik

c. Hasil Klasifikasi

Pada sub bab ini, dilakukan pembagian data Pada sub bab ini, dilakukan analisis untuk mengidentifikasi produk sandal wanita yang paling diminati berdasarkan data penjualan. Dengan menggunakan algoritma seperti Naive Bayes, dilakukan prediksi terhadap produk dalam dataset untuk menentukan produk mana yang paling sering dipilih oleh konsumen. Hasil prediksi ini dievaluasi untuk mengetahui produk dengan permintaan

tertinggi. Informasi ini sangat berharga bagi perusahaan dalam merencanakan strategi pemasaran, manajemen stok, dan pengembangan produk, sehingga mereka dapat fokus pada produk yang paling diminati untuk meningkatkan keuntungan dan kepuasan pelanggan. Sub bab ini memberikan wawasan penting tentang preferensi konsumen yang mendukung pengambilan keputusan bisnis yang lebih efektif. Berikut hasil klasifikasi dan prediksi yang didapatkan.

TABEL VI
 PERBANDINGAN HASIL AKURASI DENGAN 3 SKENARIO

Jenis Sandal	Permintaan_kodi	Produksi_kodi	Kategori	Prediksi
wedges	10	5	Laku	Laku
hills A	15	10	Laku	Laku
flat	10	12	Kurang Laku	Kurang Laku
Hills A	15	12	Laku	Laku
Flat	10	12	Kurang Laku	Kurang Laku
wedges	8	10	Kurang Laku	Laku
hills B	13	10	Laku	Laku
flat	10	13	Kurang Laku	Kurang Laku
wedges	10	9	Laku	Laku
Hills A	17	13	Laku	Laku
Flat	12	15	Kurang Laku	Kurang Laku
Hills A	18	16	Laku	Laku
Hills B	15	10	Laku	Laku
Hills A	15	12	Laku	Laku
flat	8	10	Kurang Laku	Kurang Laku
hilss A	17	13	Laku	Laku
hills B	9	12	Kurang Laku	Kurang Laku
wedges	12	15	Kurang Laku	Kurang Laku
hills A	15	13	Laku	Laku
hills B	9	12	Kurang Laku	Kurang Laku
hills A	15	10	Laku	Laku
flat	15	10	Laku	Laku
wedges	10	8	Laku	Laku
hills B	10	8	Laku	Laku
flat	13	15	Kurang Laku	Kurang Laku
hills A	15	12	Laku	Laku

Analisis menunjukkan bahwa sandal 'Hills A' adalah yang paling laku dengan 72 kali order. Ini mengindikasikan permintaan tertinggi untuk 'Hills A' di

pasar sandal wanita. Informasi ini penting untuk strategi pemasaran dan manajemen stok, memungkinkan perusahaan untuk mengalokasikan sumber daya lebih efisien, meningkatkan produksi dan persediaan produk yang diminati, serta merencanakan kampanye pemasaran yang lebih terarah. Pengetahuan ini juga dapat digunakan untuk pengembangan atau penyesuaian produk agar lebih sesuai dengan kebutuhan pelanggan, meningkatkan kepuasan pelanggan, dan mendukung pertumbuhan bisnis yang berkelanjutan. Informasi ini memberikan wawasan tentang tren pasar dan dasar untuk pengambilan keputusan yang lebih baik dalam manajemen bisnis.

IV. PENUTUP

A. Kesimpulan

Metode Naïve Bayes terbukti cukup efektif dalam mengklasifikasikan produk sandal wanita yang paling diminati, dengan rata-rata akurasi sebesar 0.62 pada data pelatihan dan 0.61 pada data pengujian, serta akurasi keseluruhan sebesar 96.15%. Hal ini menunjukkan bahwa model ini dapat memberikan hasil yang cukup akurat meskipun terdapat variasi dalam akurasi antar fold. Selain itu, jenis sandal 'hills A' tercatat sebagai produk paling laku dengan 72 kali orderan, menunjukkan permintaan tertinggi di pasar sandal wanita. Temuan ini memberikan informasi berharga bagi perusahaan untuk strategi pemasaran dan manajemen stok yang lebih efisien, memungkinkan alokasi sumber daya yang lebih baik, peningkatan produksi dan persediaan produk 'hills A', serta perencanaan kampanye pemasaran yang lebih terarah. Pengetahuan ini juga dapat digunakan untuk pengembangan atau penyesuaian produk, yang pada akhirnya meningkatkan kepuasan pelanggan dan mendukung pertumbuhan bisnis yang berkelanjutan.

B. Saran

Berdasarkan kesimpulan yang telah diuraikan, berikut beberapa saran yang dapat diterapkan:

1. Optimalisasi Model Naïve Bayes, Meskipun metode Naïve Bayes menunjukkan hasil yang cukup akurat dalam mengklasifikasikan produk sandal wanita yang paling diminati, terdapat variasi dalam akurasi antara data pelatihan dan pengujian. Oleh karena itu, disarankan untuk mengeksplorasi dan menguji berbagai teknik pra-pemrosesan data tambahan atau fitur engineering yang dapat meningkatkan konsistensi dan akurasi model. Selain itu, pertimbangkan untuk mengkombinasikan metode Naïve Bayes dengan algoritma lain melalui teknik ensemble untuk meningkatkan performa model secara keseluruhan.
2. Peneliti selanjutnya dapat melakukan analisis lebih mendalam mengenai faktor-faktor yang mempengaruhi permintaan produk. Misalnya, melakukan analisis sentimen dari ulasan pelanggan, mempelajari pengaruh harga, musim, tren mode, dan

promosi terhadap penjualan produk. Dengan memahami faktor-faktor ini, perusahaan dapat mengembangkan strategi pemasaran dan penjualan yang lebih efektif serta dapat merespons perubahan preferensi konsumen dengan lebih cepat dan tepat.

REFERENSI

- [1] M. Fathrezza Imani, M. K. (2022). Pengaruh Kreatif, Inovasi dan Strategi Pemasaran Terhadap Peningkatan Daya Beli Konsumen. *Jurnal Ilmu Manajemen, Ekonomi dan Kewirausahaan*, 2, 118-139.
- [2] Rianty Buana Pertiwi, Y. R. (2020). Pengaruh Desain Produk, Kualitas Produk Dan Harga Kompetitif Terhadap Keputusan Pembelian Sandal Wanita Merek Cardinal Di Wilayah Kota Bandung. *Jurnal Sain Manajemen*, 2.
- [3] Bintang Sahala Marpaung, A. M. (2023). Digital Marketing dan Inovasi Produksi UMKM Ssaeka Angelina (Sepatu Sandal Dampingan UMKN Kota Bogor) dalam Meningkatkan Penjualan. *Jurnal Pengabdian Masyarakat*, 3, 110-121.
- [4] Amin Dwi Ananda, D. S. (2017). Pengembangan Usaha Mikro Kecil dan Menengah (UMKM) Berbasis Industri Kreatif di Kota Malang. *Jurnal Ilmu Ekonomi*, 1.
- [5] Wardhana, B. (2022). Perancangan Pola Baru Pemasaran Usaha Fashion di Indonesia Melalui Fenomena Citayam Fashion Week. *Jurnal Ekonomi dan Bisnis*, 7.
- [6] Muhammad Deovan Reondy Putra, H. S. (2022). Counterfeit Culture dalam Perkembangan UMKM: Suatu Kajian Kekayaan Intelektual. *KRTHA BHAYANGKARA*, 16, 297-314.
- [7] Indriyani, A. B. (2023). Implementasi Data Mining Untuk Mengklasifikasikan Data Penjualan. *Jurnal Manajemen dan Bisnis Ekonomi*, 1, 207-220.
- [8] Edy Widodo, P. (2019). Klasifikasi Data Penjualan Alat Tulis Kantor (ATK) Terlaris Untuk Optimasi Strategi Pemasaran di Toko Citramedia Menggunakan Metode Naive Bayes. *Jurnal Teknologi Pelita Bangsa*, 10, 138-152.

