

## Penerapan Naive Bayes untuk Memilih Produk Berdasarkan Jenis Kulit di Toko Kosmetik

Gabriela Honestya<sup>✉</sup>, Sarjon Defit, Gunadi W Nurcahyo

Fakultas Ilmu Komputer, Universitas Putra Indonesia "YPTK" Padang, Padang, 25145, Indonesia

[gabrielahonestya12@gmail.com](mailto:gabrielahonestya12@gmail.com)

### Abstract

*Choosing skincare and cosmetic products that suit your facial skin type is an important factor in maintaining healthy and beautiful skin. There are still many consumers who have difficulty choosing the right product, which causes several problems with facial skin. Mistakes in choosing products can cause more serious skin problems. Therefore, a recommendation system is needed that can help consumers choose the right product. This research aims to provide a percentage of accuracy and suitability for selecting skincare and cosmetic products using the Naive Bayes method, which can help consumers choose products that suit their facial skin type. The case study was conducted at Inti Makeup Store Padang, where data was collected from purchase history and direct customer surveys. The Naive Bayes method was chosen because of its ability to classify data based on probability, allowing the system to provide accurate recommendations based on the consumer's facial skin type. The data used in this research amounted to 1074 records from 87 correspondent customers of Inti Makeup Store which included information about the suitability of using products they had previously used. The research results show that the recommendation system developed is able to provide recommendations for skincare and cosmetic products with a suitability accuracy rate of 93.18%, thereby increasing consumer satisfaction in choosing the right product. The application of the Naive Bayes method in selecting skincare and cosmetic products is not only beneficial for consumers, but can also provide added value for stores in understanding customer needs, optimizing stock and increasing product sales. It is hoped that this research can become a basis for further development in the field of beauty product recommendation systems.*

**Keywords:** Information Technology, Product Selection, Naive Bayes, Facial Skin Type, Inti Makeup Store

### Abstrak

Pemilihan produk skincare dan kosmetik yang sesuai dengan jenis kulit wajah merupakan salah satu faktor penting dalam menjaga kesehatan dan kecantikan kulit. Masih banyak ditemukan konsumen yang mengalami kesulitan dalam menentukan produk yang tepat sehingga menyebabkan beberapa permasalahan terhadap kulit wajah. Kesalahan dalam memilih produk dapat menyebabkan masalah kulit yang lebih serius. Sistem rekomendasi diperlukan untuk dapat membantu konsumen dalam memilih produk yang tepat. Penelitian ini bertujuan untuk memberikan persentase ketepatan dan kecocokan pemilihan produk skincare dan kosmetik dengan menggunakan metode Naive Bayes, yang dapat membantu konsumen dalam memilih produk yang sesuai dengan jenis kulit wajah mereka. Studi kasus dilakukan di Inti Makeup Store Padang, di mana data dikumpulkan dari riwayat pembelian dan survei langsung kepada pelanggan. Metode Naive Bayes dipilih karena kemampuannya dalam mengklasifikasikan data berdasarkan probabilitas, sehingga memungkinkan sistem untuk memberikan rekomendasi yang akurat berdasarkan jenis kulit wajah konsumen. Data yang digunakan dalam penelitian ini berjumlah 1074 records dari 87 koresponden pelanggan Inti Makeup Store yang mencakup informasi tentang kecocokan pemakaian produk yang pernah mereka gunakan sebelumnya. Hasil penelitian menunjukkan bahwa sistem rekomendasi yang dikembangkan mampu memberikan rekomendasi produk skincare dan kosmetik dengan tingkat akurasi kecocokan sebesar 93.18%, sehingga dapat meningkatkan kepuasan konsumen dalam memilih produk yang tepat. Penerapan metode Naive Bayes dalam pemilihan produk skincare dan kosmetik tidak hanya bermanfaat bagi konsumen, tetapi juga dapat memberikan nilai tambah bagi toko dalam memahami kebutuhan pelanggan, mengoptimalkan stok dan meningkatkan penjualan produk. Penelitian ini diharapkan dapat menjadi landasan bagi pengembangan lebih lanjut dalam bidang sistem rekomendasi produk kecantikan.

**Kata kunci:** Teknologi Informasi, Pemilihan Produk, Naive Bayes, Jenis Kulit Wajah, Inti Makeup Store.

*KomtekInfo is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.*



### 1. Pendahuluan

Pembelajaran mesin (*machine learning*) dianggap sebagai cabang kecerdasan buatan, yang dianggap sebagai metode analisis data yang mengotomatiskan proses pembuatan model. Identifikasi pola yang mudah di antara data dan kemampuan untuk meningkatkan pembelajaran mesin dari waktu ke waktu adalah dua keunggulan utama pembelajaran mesin dibandingkan

teknik klasifikasi data tradisional [1]. Penggunaan teknologi canggih, seperti pembelajaran mesin (*machine learning*) dan kecerdasan buatan (*artificial intelligence*), dapat membantu meningkatkan keakuratan diagnosis dan prognosis penyakit. [2]. Data Mining merupakan proses menganalisa data dari perspektif yang berbeda dan menyimpulkannya menjadi informasi-informasi penting yang digunakan untuk keperluan kerja, prediksi dan meningkatkan keuntungan [3]. Metode klasifikasi

ini menggunakan beberapa kelas, dengan mempertimbangkan akurasi berdasarkan penelitian sebelumnya, mempunyai potensi yang baik, mudah dan cepat dalam penerapannya, serta banyak digunakan untuk klasifikasi [4].

Pada penelitian Ariyani tentang prediksi dan penentuan faktor anteseden pengembangan desa wisata, penilaian kinerja dengan menggunakan nilai presisi diketahui paling baik terdapat pada model Naïve Bayes, memiliki nilai sebesar 0,933, dan pohon keputusan memiliki nilai sebesar 0,878 [5]. Penelitian Dadang Priyanto dkk menunjukkan perbandingan akurasi prediksi sebesar 94,5% sedangkan KNN menunjukkan akurasi prediksi sebesar 92,5% terhadap data penyakit kecanduan narkoba [6]. Penelitian oleh Nazar dkk dalam mengklasifikasikan air bersih mencapai hasil akhir pada pengujian menggunakan RapidMiner menghasilkan akurasi sebesar 97,35% [7]. Hasil penelitian oleh Okikiola menunjukkan bahwa model mencapai F-measure sebesar 87% dan menunjukkan kinerja yang lebih baik dalam uji AUC dibandingkan dengan metode lain, menyarankan efektivitasnya dalam prediksi diabetes [8]. Penggunaan algoritma Naïve Bayes Classifier dalam klasifikasi kelayakan penerima bantuan pangan non tunai dan program keluarga harapan di kota Sukabumi terbukti efektif. Penelitian yang telah dilakukan oleh Okikiola dkk ini mampu mengklasifikasikan data dengan tingkat akurasi tinggi sebesar 98,77% [9]. Algoritma pembelajaran mesin seperti pohon keputusan dapat digunakan untuk menentukan faktor anteseden yang menghasilkan pola keberhasilan pengembangan desa wisata [5].

Naive Bayes menggunakan cabang matematika yang dikenal dengan teori probabilitas untuk mencari peluang terbesar dari kemungkinan klasifikasi dengan cara melihat frekuensi tiap klasifikasi pada data training [10]. Sebagai seorang praktisi Teknologi Informasi yang juga aktif di dunia Makeup, penulis berinisiatif mempelajari dan menerapkan Algoritma Naïve Bayes untuk membantu para pengguna aktif skincare dan kosmetik di Inti Makeup Store Padang dalam memilih dan menentukan produk yang cocok dan sesuai jenis kulit wajahnya. Penelitian ini diharapkan dapat memecahkan masalah pemilihan produk skincare dan kosmetik bagi konsumen, dan dapat membantu toko dalam memahami kebutuhan pelanggan, mengoptimalkan stok dan meningkatkan penjualan produk. Penelitian ini diharapkan dapat menjadi landasan bagi pengembangan lebih lanjut dalam bidang sistem rekomendasi produk kecantikan.

## 2. Metodologi Penelitian

Pada penelitian ini ada beberapa tahapan yang diperlukan untuk mendapatkan hasil yang terbaik, yang mana tahapan tersebut meliputi pengumpulan data dan pelabelan data, dilanjutkan dengan *preprocessing*. Data yang telah diproses sebelumnya akan diklasifikasikan menggunakan algoritma Naïve Bayes [11]. Tahapan

yang akan dilakukan dapat dilihat pada Gambar 1. di bawah ini



Gambar 1. Kerangka Penelitian

Berdasarkan kerangka kerja penelitian pada Gambar 1, terdapat 6 tahapan proses yang diperlukan. Tahapan-tahapan ini saling berhubungan dan harus diperhatikan dengan seksama untuk menjaga kualitas penelitian. Proses tahapan kerangka penelitian ini dapat dijelaskan sebagai berikut:

### 1. Mengidentifikasi Masalah

Salah satu masalah utama dalam memilih produk skincare dan makeup yang tepat adalah bagaimana cara mengelompokkan produk yang cepat dan tepat sesuai dengan jenis kulit wajah. Selanjutnya masih ditemukan kesulitan dalam merekomendasikan atau penentuan produk yang tepat oleh pelanggan Inti Makeup Store. Ditemukan banyak kasus kulit yang rusak akibat ketidak-sesuaian produk dengan jenis kulit wajah pelanggan

### 2. Menganalisa Masalah

Pada tahap ini dilanjutkan dengan pemahaman dan merencanakan langkah berikutnya yang sesuai dengan masalah-masalah yang telah ditemukan. Pemahaman ini kemudian digunakan untuk merancang solusi yang tepat dan merencanakan strategi yang efektif dalam mengatasi masalah-masalah tersebut. Teknik analisa yang dapat dilakukan dengan beberapa tahap berikut:

- a. Tahap *identify* yaitu mengidentifikasi permasalahan yang terjadi.
- b. Tahap *understand* yaitu memahami lebih lanjut tentang permasalahan yang ada dengan cara melakukan pengumpulan data yang dibutuhkan.
- c. Tahapan *analyze* yaitu mencari kelemahan-kelemahan sistem yang ada dan mengumpulkan informasi tentang kebutuhan-kebutuhan lebih lanjut yang diperlukan.

### 3. Melakukan Studi Literatur dan Lapangan

Studi literatur dilakukan dengan cara mempelajari dan memahami teori yang digunakan dengan cara mengumpulkan jurnal, buku dan browsing internet untuk mendapatkan informasi yang berhubungan dengan masalah yang dibahas dalam penelitian ini. Informasi yang diperoleh dari sumber-sumber tersebut kemudian dianalisis kembali untuk memastikan bahwa semua aspek yang berkaitan

dengan masalah penelitian telah teridentifikasi dan dipahami dengan baik. Selanjutnya studi lapangan dilakukan dengan mencari dan menganalisis ulasan produk yang sesuai dengan rekomendasi untuk jenis kulit wajah tertentu untuk mendapatkan wawasan yang dapat mendukung teori yang telah dipelajari.

#### 4. Mengumpulkan Data

Tahap ini merupakan cara mengumpulkan data yang dilakukan dengan cara observasi dan wawancara kepada Pemilik, Karyawan dan Pelanggan Inti Makeup Store untuk mengumpulkan data produk dan dilanjutkan ke Klink Kecantikan Hers Padang untuk mengumpulkan data kandungan bahan yang direkomendasikan sesuai dengan jenis kulit wajah. Setelah selesai mengklasifikasikan produk sesuai jenis rekomendasi kulit wajah, penulis membuat dan mengumpulkan data *survey* terhadap pelanggan toko Inti Makeup Store.

#### 5. Mengolah Data

Mengolah data merupakan tahap lanjutan dari proses pengumpulan data, di mana data yang telah terkumpul dimasukkan ke dalam *formula Excel* yang disesuaikan dengan metode Naive Bayes. Sebelum analisis dan perhitungan dilakukan, data harus melalui proses seleksi yang cermat untuk memastikan keakuratan yang akan dianalisis. Jika ditemukan kesalahan atau inkonsistensi, perbaikan dilakukan terlebih dahulu untuk menjaga hasil analisis yang akan dihasilkan dari metode Naive Bayes.

#### 6. Implementasi

Setelah melakukan pengolahan data selanjutnya dilakukan implementasi metode Naive Bayes terhadap data tersebut. Implementasi ini melibatkan penerapan langkah-langkah spesifik dari metode Naive Bayes, yang dirancang untuk mengklasifikasikan data berdasarkan probabilitas dan pola yang telah diidentifikasi sebelumnya. Langkah-langkah dari metode Naive Bayes adalah sebagai berikut :

- a. Menentukan Data Set
- b. Memasukkan Data Uji
- c. Menghitung Probabilitas Prior

Probabilitas prior adalah ukuran peluang awal dari suatu kejadian atau kelas sebelum mempertimbangkan bukti atau informasi tambahan. Dalam klasifikasi, probabilitas prior mencerminkan frekuensi relatif dari suatu kelas dalam data set secara keseluruhan. Probabilitas prior dapat dilihat dari Persamaan 1.

$$P(C_i) = \frac{\sum C_i}{n} \quad (1)$$

Keterangan lanjutan:

$P(C_i)$  : Probabilitas label pada,  
 $\sum C_i$  : Jumlah data dengan label pada kelas,  
 $n$  : Jumlah total pada data latih

Probabilitas prior  $P(C_i)$  menggambarkan peluang awal dari suatu kelas label  $C_i$  dalam keseluruhan data set sebelum mempertimbangkan fitur-fitur lainnya. Nilai  $P(C_i)$  dihitung dengan membagi jumlah data yang memiliki label  $C_i$  ( $\sum C_i$ ) dengan jumlah total data dalam data set latih ( $n$ ). Probabilitas prior ini digunakan sebagai dasar untuk menghitung probabilitas posterior dalam metode Naive Bayes, yang menggabungkan informasi ini dengan data fitur untuk membuat prediksi.

#### d. Menghitung Probabilitas Posterior

Probabilitas posterior adalah peluang suatu kejadian atau kelas setelah mempertimbangkan bukti atau informasi tambahan. Dalam metode Naive Bayes, probabilitas posterior dihitung dengan menggabungkan probabilitas prior dengan data fitur yang diamati, memberikan penilaian yang lebih akurat berdasarkan bukti yang ada. Probabilitas ini digunakan untuk membuat keputusan atau prediksi tentang kelas yang paling mungkin dari data yang dianalisis. Probabilitas prior dapat dilihat dari Persamaan 2.

$$P(X_i|C_i) = \frac{\sum X_i|C_i}{\sum X_i} \quad (2)$$

Keterangan lanjutan:

$P(X_i|C_i)$  : Probabilitas fitur dengan label pada kelas  
 $\sum X_i|C_i$  : Jumlah data fitur dengan label pada kelas  
 $\sum X_i$  : Jumlah data dengan label pada kelas

Persamaan probabilitas posterior menghitung peluang suatu kelas  $C_i$  berdasarkan fitur yang diamati  $X_i$ . Probabilitas posterior  $P(C_i|X_i)$  dihitung dengan mengalikan probabilitas fitur  $P(X_i|C_i)$  dengan probabilitas prior kelas  $P(C_i)$  dan kemudian dibagi dengan probabilitas total fitur. Dalam perhitungan ini,  $P(X_i|C_i)$  adalah frekuensi fitur dalam kelas  $C_i$ , sementara  $\sum X_i|C_i$  dan  $\sum X_i$  adalah jumlah data fitur dan jumlah data dengan label kelas, masing-masing.

#### 7. Menghitung probabilitas Posterior terhadap tiap class data

Menghitung probabilitas posterior terhadap tiap kelas data melibatkan penilaian seberapa besar kemungkinan data termasuk dalam kelas tertentu setelah mempertimbangkan fitur yang ada. Proses ini dilakukan dengan mengalikan probabilitas fitur yang diberikan kelas dengan probabilitas prior kelas tersebut, kemudian dibagi dengan probabilitas total fitur yang diamati. Hasil dari perhitungan ini adalah nilai probabilitas posterior untuk setiap kelas, yang digunakan untuk menentukan kelas yang paling mungkin menjadi hasil prediksi. Probabilitas posterior pada tiap kelas dapat dilihat dari Persamaan 3

$$P(C_i|X) = P(C_i) \prod_{i=1}^n P(X_i|C_i) \quad (3)$$

Keterangan lanjutan:

$P(C_i|X)$  : Probabilitas pada kelas dalam data,

$P(C_i)$  : Probabilitas pada label kelas,

$P(X_i|C_i)$  : Probabilitas pada fitur dengan label kelas

Persamaan probabilitas posterior  $P(C_i|X)$  mengukur peluang bahwa data  $X$  termasuk dalam kelas  $C_i$  setelah mempertimbangkan fitur yang diamati. Ini dihitung dengan mengalikan probabilitas fitur  $P(X_i|C_i)$  dengan probabilitas prior kelas  $P(C_i)$ , dan kemudian dinormalisasi dengan membagi hasilnya dengan probabilitas total fitur. Dengan demikian,  $P(C_i|X)$  memberikan nilai yang menunjukkan seberapa besar kemungkinan data  $X$  berada di kelas  $C_i$  berdasarkan informasi fitur yang tersedia.

### 8. Pengujian Hasil

Pada pengujian ini data skincare dan kosmetik yang diolah berupa data hasil perhitungan manual yang akan dibandingkan dengan *Software Rapid Miner*. Hasil dari pengujian sistem ini diharapkan dapat membantu pelanggan memilih produk yang tepat sesuai dengan jenis kulit wajahnya. Dari hasil penelitian ini, maka penerapan dengan metode algoritma Naive Bayes juga dapat membantu pihak toko untuk merekomendasikan produk terbaik sesuai jenis kulit wajah pelanggan, dan membantu pengelolaan prioritas stok produk.

### 3. Hasil dan Pembahasan

Tahap selanjutnya setelah mendapatkan data produk adalah mencari kandungan bahan yang dapat mengarahkan rekomendasi produk terhadap jenis kulit wajah yang tepat. Data acuan didapat sesuai dengan pendapat pakar kulit dr. Bella Miralda pada Klinik Kecantikan Hers Padang. Berikut unsur data produk yang teridentifikasi yang telah dikelompokkan berdasarkan spesifikasi produknya dapat dilihat pada Tabel 1.

Tabel 1. Data produk dan spesifikasi bahan

No	Nama Produk	Nama Bahan	Rekomendasi Kulit
1	Loreal Infalible Matte Cover Foundation	Dimethicone, qua,Cyclohexasiloxane, Isododecane, Aluminum, Tocopherol, Carica	Normal, Berminyak
2	Lucas Papaw Ointment	Papaya, Petrolatum, Gum Balsam Peru	Normal, Kering
3	Emina Moist in A Bottle	Hyaluronic Acid ,Glycerin ,Sodium Hyaluronate ,Niacinamide (Vitamin B3)	Normal, Kering
...	...	...	...

20	Pond's Miracle Cream	Age Night	Niacinamide (Vitamin B3),Retinol, Glycerin, Lactic Acid, Sodium PCA, Tocopheryl Acetate (Vitamin E), Allantoin	Normal, Kering
----	----------------------	-----------	--	----------------

Hasil klasifikasi yang didapat merupakan data penjualan yang diminati dari penjualan produk Inti Makeup Store. Produk yang diambil adalah produk yang memiliki spesifikasi khusus yang memiliki rekomendasi jenis kulit. Selanjutnya dari daftar produk yang telah dikumpulkan dikelompokkan sesuai dengan arahan yang diberikan oleh pakar kulit dan kecantikan.

### 1. Mengidentifikasi Data Set

Data set pada penelitian ini memiliki jumlah sebanyak 1074 data sampel, dimana data ini nantinya akan dijadikan sebagai data training. Untuk mempermudah dalam melakukan analisa data, maka dilakukan transformasi data. Berikut adalah transformasi data set untuk delapan kelas terdapat pada Tabel 2 dan Tabel 3.

Tabel 2. Kriteria Kelas Data Set

No	Kriteria Unsur	Kode
1.	Nama Produk	A
2.	Spesifikasi Produk	B
3.	Spesifikasi Bahan	C
4.	Harga	D
5.	Jenis Kelamin	E
6.	Usia	F
7.	Jenis Kulit	G
8.	Kecocokan	H

Tabel 3. Transformasi data setKelas Nama Produk

No	Kriteria Unsur	Kode
1.	Loreal Infalible Matte ...	A1
2.	Makeover Matte ...	A2
3.	Npure Centella ...	A3
4.	Make Up For Ever...	A4
5.	LT Pro Perfect ...	A5
...	...	...
20.	Hada Labo Perfect...	A20

Hasil data yang telah diolah kemudian dibuatkan ke dalam bentuk kode agar lebih mudah diklasifikasikan. Terdapat 8 kelas data set yang dibutuhkan, dan diberikan kode dari A sampai H. Kelas data set ini akan diberikan kode numerik sesuai dengan banyak jenis data yang ada.

### 2. Data Set dan Data Uji

Data Set adalah kumpulan data yang digunakan untuk melatih dan menguji model dalam proses pembelajaran mesin. Data uji adalah bagian dari data set yang sengaja dipisahkan dan digunakan untuk mengevaluasi kinerja model setelah pelatihan selesai, memastikan bahwa model dapat membuat prediksi yang akurat pada data baru yang belum pernah dilihat

sebelumnya. Data set yang digunakan pada penelitian ini berjumlah 1074 baris data dan 45 baris data uji. Berikut ini adalah beberapa tabel data set dan perhitungan manual pertama dari data uji pada tabel 5:

Tabel 4. Transformasi Data Set Kelas Nama Produk

A	B	C	D	E	F	G	H
A1	B2	C1	D1	E2	E2	G2	H3
A2	B2	C1	D2	E2	E2	G2	H1
A3	B2	C2	D1	E2	E2	G2	H1
A4	B2	C1	D3	E2	E2	G2	H1
A5	B2	C1	D1	E2	E2	G2	H2
A6	B2	C1	D3	E2	E2	G2	H2
...	...	...	...	...	...	...	...
<b>A12</b>	<b>B1</b>	<b>C7</b>	<b>D1</b>	<b>E1</b>	<b>E2</b>	<b>G2</b>	<b>H1</b>

Tabel 5. Data Uji

A	B	C	D	E	F	G	H
A13	B1	C4	D2	E2	F2	G3	???

Langkah selanjutnya setelah mendapatkan data set dan data uji adalah menghitung nilai setiap data train terhadap data uji. Metode yang dipilih pada penelitian ini adalah Naive Bayes. Langkah-langkah perhitungan yang digunakan untuk mendapatkan klasifikasi dengan metode Naive Bayes adalah sebagai berikut:

## 2. Perhitungam Naive Bayes

Mengitung nilai probabilitas prior untuk setiap kelas/ unsur, di mana kelas H1 (Cocok) adalah sebanyak 700 data, kelas H2 (Kurang Cocok) Sebanyak 233 data, dan H3 (Tidak Cocok) sebanyak 141 data. Maka perhitungan probabilitas prior-nya adalah sebagai berikut :

$$P(\text{Cocok} [H1]) = 700 / 1074 = 0.651769088$$

$$P(\text{Kurang Cocok} [H2]) = 233 / 1074 = 0.216945996$$

$$P(\text{Tidak Cocok} [H3]) = 141 / 1074 = 0.131284916$$

Langkah pertama dalam menghitung nilai probabilitas untuk setiap kriteria kelas penilaian indeks kecocokan adalah mengidentifikasi dan menghitung frekuensi kemunculan setiap kriteria dalam kelas yang relevan. Setelah frekuensi diketahui, nilai probabilitas untuk setiap kriteria dihitung dengan membagi jumlah kemunculan kriteria dalam kelas tertentu dengan total jumlah data dalam kelas tersebut. Nilai probabilitas ini kemudian digunakan untuk menentukan seberapa besar kecocokan antara data baru dan setiap kelas yang ada, yang akhirnya membantu dalam pengambilan keputusan atau prediksi.

Menghitung probabilitas kriteria kelas Nama Produk

$$P(A13) | [H1] = 38 / 700 = 0.054285714$$

$$P(A13) | [H2] = 8 / 233 = 0.034334764$$

$$P(A13) | [H3] = 6 / 141 = 0.042553191$$

Menghitung probabilitas kriteria kelas spesifikasi produk pada metode Naive Bayes melibatkan penilaian seberapa besar kemungkinan suatu produk memiliki

kriteria tertentu berdasarkan data yang ada. Probabilitas ini dihitung dengan membagi jumlah produk dalam kelas spesifikasi tertentu yang memenuhi kriteria tersebut dengan total jumlah produk dalam kelas tersebut. Hasilnya digunakan untuk menentukan kelas spesifikasi yang paling sesuai untuk produk baru berdasarkan fitur-fitur yang diamati.

Menghitung probabilitas kriteria kelas Spesifikasi Produk.

$$P(B1) | [H1] = 252 / 700 = 0.36$$

$$P(B1) | [H2] = 144 / 233 = 0.618025751$$

$$P(B1) | [H3] = 55 / 141 = 0.390070922$$

Menghitung probabilitas kriteria kelas Spesifikasi Bahan Kandungan Produk pada metode Naive Bayes melibatkan penilaian peluang bahwa suatu produk mengandung bahan tertentu berdasarkan data pelatihan yang ada. Probabilitas ini dihitung dengan membagi jumlah produk dalam kelas tertentu yang mengandung bahan tersebut dengan total produk dalam kelas itu. Hasil probabilitas ini kemudian digunakan untuk memperkirakan kelas bahan kandungan yang paling mungkin bagi probabilitas baru yang sedang dianalisis.

Menghitung probabilitas kriteria kelas Spesifikasi Bahan

$$P(C4) | [H1] = 99 / 700 = 0.141428571$$

$$P(C4) | [H2] = 95 / 233 = 0.407725322$$

$$P(C4) | [H3] = 11 / 141 = 0.078014184$$

Menghitung probabilitas kriteria kelas Harga pada metode Naive Bayes melibatkan evaluasi seberapa besar kemungkinan suatu produk termasuk dalam rentang harga tertentu berdasarkan data yang ada. Probabilitas ini dihitung dengan membagi jumlah produk dalam kelas harga tertentu dengan total jumlah produk dalam dataset. Hasilnya digunakan untuk memperkirakan rentang harga yang paling mungkin untuk produk baru, berdasarkan fitur atau karakteristik yang telah diamati.

Menghitung probabilitas kriteria kelas Harga

$$P(D2) | [H1] = 197 / 700 = 0.281428571$$

$$P(D2) | [H2] = 44 / 233 = 0.188841202$$

$$P(D2) | [H3] = 47 / 141 = 0.333333333$$

Menghitung probabilitas kriteria kelas Jenis Kelamin pada metode Naive Bayes melibatkan penilaian kemungkinan bahwa data atau individu tertentu termasuk dalam kategori jenis kelamin tertentu (pria atau wanita) berdasarkan informasi yang tersedia. Probabilitas ini dihitung dengan membagi jumlah individu dalam kategori jenis kelamin tertentu dengan total jumlah individu dalam dataset. Hasil probabilitas ini digunakan untuk memprediksi jenis kelamin yang paling mungkin bagi data baru yang dianalisis, berdasarkan fitur-fitur yang ada.

Menghitung probabilitas kriteria kelas Jenis Kelamin

$$P(E2) | [H1] = 693 / 700 = 0.99$$

$$P(E2) | [H2] = 225 / 233 = 0.965665236$$

$$P(E2) | [H3] = 141 / 141 = 1$$

Menghitung probabilitas kriteria kelas Usia pada metode Naive Bayes melibatkan penilaian seberapa besar kemungkinan data atau individu termasuk dalam kelompok usia tertentu berdasarkan remaja, dewasa, dan paruh baya. Probabilitas ini dihitung dengan membagi jumlah individu dalam kelompok usia tersebut dengan total jumlah individu dalam dataset. Hasil ini digunakan untuk memprediksi kelompok usia yang paling mungkin sesuai untuk data baru, berdasarkan fitur-fitur yang telah diamati.

Menghitung probabilitas kriteria kelas Usia

$$P(F2) | [H1] = 489 / 700 = 0.698571429$$

$$P(F2) | [H2] = 169 / 233 = 0.725321888$$

$$P(F2) | [H3] = 110 / 141 = 0.780141844$$

Menghitung probabilitas kriteria kelas Jenis Kulit pada metode Naive Bayes melibatkan penilaian kemungkinan bahwa data atau individu tertentu termasuk dalam kategori jenis kulit tertentu (normal, kering, dan berminyak) berdasarkan data pelatihan. Probabilitas ini dihitung dengan membagi jumlah individu dalam kategori jenis kulit tertentu dengan total jumlah individu dalam dataset. Hasil probabilitas ini digunakan untuk memprediksi jenis kulit yang paling mungkin sesuai untuk data baru berdasarkan fitur-fitur yang telah dianalisis.

Menghitung probabilitas kriteria kelas Jenis Kulit

$$P(G3) | [H1] = 249 / 700 = 0.355714286$$

$$P(G3) | [H2] = 60 / 233 = 0.25751073$$

$$P(G3) | [H3] = 71 / 141 = 0.503546099$$

Setelah menghitung probabilitas prior untuk semua kelas dalam data uji, langkah berikutnya adalah menghitung probabilitas posterior untuk setiap kelas berdasarkan fitur-fitur yang ada dalam data tersebut. Perhitungan probabilitas posterior ini melibatkan penggabungan probabilitas prior dengan probabilitas fitur-fitur yang diamati, menggunakan rumus Naive Bayes, untuk setiap kelas yang sedang dianalisis. Hasil dari perhitungan probabilitas posterior ini akan menunjukkan kelas yang paling mungkin sesuai dengan data uji, memungkinkan untuk membuat prediksi yang lebih akurat:

$$P(X | Cocok [H1]) = 0.054285714 * 0.36 * 0.141428571 * 0.281428571 * 0.99 * 0.698571429 * 0.355714286 * 700 = 0.133948779$$

$$P(X | Kurang Cocok [H2]) = 0.034334764 * 0.618025751 * 0.407725322 * 0.188841202 *$$

$$0.965665236 * 0.725321888 * 0.25751073 * 233 = 0.068661567$$

$$P(X | Tidak Cocok [H3]) = 0.042553191 * 0.390070922 * 0.078014184 * 0.333333333 * 1 * 0.780141844 * 0.503546099 * 141 = 0.02390892$$

Dari hasil perhitungan di atas diperoleh nilai probabilitas prediksi indeks kecocokan produk dengan hasil 0.133948779 yang diklasifikasikan pada kategori cocok, untuk klasifikasi kurang cocok diperoleh nilai 0.068661567 dan untuk indeks tidak cocok nilainya 0.02390892.. Perhitungan akurasi dan persentase model Naive Bayes diuji menggunakan aplikasi RapidMiner. Hasilnya ditemukan bahwa akurasi penerapan model Naive Bayes terhadap data testing sebesar 93.18%.

	true Cocok	true Kurang Cocok	true Tidak Cocok	class precision
pred. Cocok	19	1	0	95.00%
pred. Kurang Cocok	2	20	0	90.91%
pred. Tidak Cocok	0	0	2	100.00%
class recall	90.48%	95.24%	100.00%	

Gambar 2. Hasil Akurasi RapidMiner

Evaluasi akurasi dan performa model Naive Bayes dilakukan menggunakan perangkat lunak RapidMiner, yang merupakan alat canggih untuk analisis data dan pemodelan statistik. Dalam evaluasi ini, model Naive Bayes diterapkan pada data setuju untuk menilai sejauh mana model dapat mengklasifikasikan data yang belum digunakan selama pelatihan. RapidMiner menghitung akurasi model dengan membandingkan hasil prediksi model terhadap label yang benar dalam data setuju. Hasil analisis menunjukkan bahwa model Naive Bayes mencapai akurasi sebesar 93.18%, menunjukkan proporsi kasus di mana model berhasil mengklasifikasikan data dengan benar. Tingkat akurasi tersebut mengindikasikan bahwa model Naive Bayes memiliki kinerja yang sangat memuaskan dalam konteks data uji, menggarisbawahi efektivitasnya dalam menghasilkan prediksi yang akurat. Dengan akurasi yang tinggi, model ini menunjukkan potensi yang signifikan untuk aplikasi praktis dan pengambilan keputusan berbasis data yang lebih andal.

#### 4. Kesimpulan

Penerapan metode Naive Bayes dalam pemilihan produk berdasarkan jenis kulit di toko kosmetik menunjukkan hasil yang signifikan dan jelas dalam meningkatkan akurasi rekomendasi produk. Metode ini memungkinkan pengklasifikasian produk yang lebih tepat sesuai dengan karakteristik jenis kulit konsumen, Naive Bayes mampu memprediksi produk yang paling sesuai dengan menggunakan data historis dan preferensi pelanggan, sehingga dapat mengurangi risiko kesalahan dalam pemilihan produk yang dapat menyebabkan masalah kulit.

Penelitian ini menggunakan hasil dari 87 responden yang telah mengisi survey di Inti Makeup Store. Data produk mengacu dari data penjualan 3 bulan terakhir yang merupakan produk terspesifikasi yang diminati konsumen. Hasil dari penelitian ini dapat memprediksi dan merekomendasikan kepada pelanggan yang tidak tahu atau kurang mengenali produk. Selain itu, toko juga mendapatkan manfaat dengan peningkatan penjualan dan pengelolaan stok yang lebih baik dengan memahami kebutuhan pelanggan. Penelitian selanjutnya diharapkan dapat lebih mendalami jenis dan permasalahan kulit lainnya, atau penambahan metode yang dapat dikolaborasi dengan metode Naive Bayes untuk mendapatkan hasil akurasi yang lebih baik.

#### Daftar Rujukan

- [1] Lemons, K. (2020). A Comparison Between Naive Bayes And Random Forest To Predict Breast Cancer. *International Journal Of Undergraduate Research And Creative Activities*, 12(1), 1. <https://doi.org/10.7710/2168-0620.0287>
- [2] Yulhandri, Malabay., & Kartini. (N.D.). Correlated Naive Bayes Algorithm To Determine Healing Rate Of Hepatitis Patients. In *International Journal Of Science*. <http://ijstm.inarah.co.id>
- [3] Zidane, P. A. M., Zer, R. W. P. P., & Gunawan, I. (2022). *Jurnal Media Informatika [Jumin] Penerapan Data Mining Naive Bayes Dalam Klasifikasi Kepuasan Mahasiswa Berlangganan Wifi Indihome*.
- [4] Jefriyanto, J., Ainun, N., Arif, M., & Ardha, A. (N.D.). Application Of Naive Bayes Classification To Analyze Performance Using Stopwords. *Journal Of Information System*, 1(2), 49–53. <http://gemapublisher.com/index.php/jiste>
- [5] Ariyani, N., Fauzi, A., & Umar, F. (2023). Predicting And Determining Antecedent Factors Of Tourist Village Development Using Naive Bayes And Tree Algorithm. *International Journal Of Applied Sciences In Tourism And Events*, 7(1), 1–15. <https://doi.org/10.31940/ijaste.v7i1.1-15>
- [6] Priyanto, D., Iman, A. R., & Jollyta, D. (2023). Naive Bayes and K-Nearest Neighbor Algorithm Approach in Data Mining Classification of Drugs Addictive Diseases. *ILKOM Jurnal Ilmiah*, 15(2), 262–270. <https://doi.org/10.33096/ilkom.v15i2.1544.262-270>
- [7] Nazar Yuniar, M. (2023). Klasifikasi Kualitas Air Bersih Menggunakan Metode Naive Baiyes. *Jurnal Sains Dan Teknologi*, 5(1), 243–246. <https://doi.org/10.55338/saintek.v5i1.1383>
- [8] Nazmi, A. M., & Pambudi, A. (2024). *Journal Of Artificial Intelligence And Engineering Applications Classification Analysis Of The Eligibility Of Recipients Of Non-Cash Food Assistance And Family Hope Programs In The City Of Sukabumi Using The Naive Bayes Classifier Algorithm* (Vol. 3, Issue 2). <https://ioinformatic.org/>
- [9] Okikiola, F. M., Adewale, O. S., & Obe, O. O. (2023). A Diabetes Prediction Classifier Model Using Naive Bayes Algorithm. *Fudma Journal Of Sciences*, 7(1), 253–260. <https://doi.org/10.33003/fjs-2023-0701-1301>
- [10] Halila Nasution, R., & Humaira, S. (2024). Implementasi Algoritma Naive Bayes Untuk Memprediksi Kelayakan Pemberian Kredit Sepeda Motor. In *Journal Of Science And Social Research* (Issue 2). <http://jurnal.goretanpena.com/index.php/jssr>
- [11] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, & Fitri Nurapriani. (2023). Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naive Bayes dan KNN. *Jurnal KomtekInfo*, 1–7. <https://doi.org/10.35134/komtekinfo.v10i1.330>
- [12] Muharrom, M. (2023). *Bulletin of Information Technology (BIT) Analisis Komparasi Algoritma Data Mining Naive Bayes, K-Nearest Neighbors dan Regresi Linier Dalam Prediksi Harga Emas*. 4(4), 430–438. <https://doi.org/10.47065/bit.v3i1>
- [13] Nurdina, A., & Puspita, A. B. I. (2023). Naive Bayes and KNN for Airline Passenger Satisfaction Classification: Comparative Analysis. *Journal of Information System Exploration and Research*, 1(2). <https://doi.org/10.52465/joiser.v1i2.167>
- [14] Nurjanah, I., Karaman, J., Widaningrum, I., & Mustikasari, D. (2023). Penggunaan Algoritma Naive Bayes Untuk Menentukan Pemberian Kredit Pada Koperasi Desa. In *Journal of Computer Science and Information Technology E-ISSN* (Vol. 3, Issue 2).
- [15] Rastogi, R., & Bansal, M. (2023). Diabetes prediction model using data mining techniques. *Measurement: Sensors*, 25. <https://doi.org/10.1016/j.measen.2022.100605>
- [16] Wijaya, D., & Abdillah, L. A. (2023). Sentiment Analysis of Omicron COVID-19 Variant using Naive Bayes Classifier and RapidMiner. In *JOURNAL OF DATA SCIENCE* / (Vol. 2023).
- [17] Amijoyo, T., & Mumtaza, A. (2023). Implementation Naive Bayes Method To Prediction Of Student Achievement at Bimbel Sonai West Jakarta. *Jurnal Inovatif: Inovasi Teknologi Informasi Dan Informatika*, 6(2), 139–144. <https://doi.org/10.32832/inovatif>
- [18] Esa Unggul, U., & Indonesia, J. (n.d.). Correlated Naive Bayes Algorithm To Determine Healing Rate Of Hepatitis Patients. In *International Journal of Science*. <http://ijstm.inarah.co.id>
- [19] Faroek, D., Yusuf, M., & Syatauw, G. (2023). Sentiment Analysis of the Popularity of Parties Supporting the 2024 Presidential Candidates on Twitter Using the Naive Bayes Classifier Algorithm. *Antivirus: Jurnal Ilmiah Teknik Informatika*, 17(2), 216–227. <https://doi.org/10.35457/antivirus.v17i2.3261>
- [20] Fauzi Fayyad, M., Takratama Savra, D., Kurniawan, V., & Hilmi Estanto, B. (2023). *Sentiment Analysis of Towards Electric Cars using Naive Bayes Classifier and Support Vector Machine Algorithm*. 1(1), 1–9. <https://journal.irpi.or.id/index.php/predatecs/article/view/814>
- [21] Febriyanti, L., & Zakaria, H. (n.d.). *Implementasi Data Mining Untuk Memprediksi Produktivitas Pada Tanaman Kacang Tanah Menggunakan Metode Naive Bayes (Studi Kasus: Perkebunan Kacang Tanah Di Kota Bogor)*. <https://journal.mediapublikasi.id/index.php/logic>
- [22] Kaur, J., Bhambri, P., & Sharma, K. (n.d.). *Wheat Production Analysis based on Naive Bayes Classifier*.

□