

Penerapan Algoritma Naïve Bayes untuk Rekomendasi Pakaian Wanita

Ratih Yulia Hayuningtyas

STMIK Nusa Mandiri/Teknik Informatika
Email: ratih.ryl@nusamandiri.ac.id

Abstrak

Banyak kegiatan jual beli yang dilakukan secara *online* yang tentunya sangat memudahkan bagi para pembeli. Salah satunya yaitu pembelian pakaian wanita, dalam memilih sebuah pakaian terkadang sebagai wanita merasa sangat membingungkan, dikarenakan banyak pilihan yang menarik. Selain itu ada beberapa faktor yang menjadi pertimbangan seperti tipe tubuh, bahan, model, warna dan harga. Dari faktor tersebut dibutuhkan sistem untuk rekomendasi pada saat pembelian pakaian wanita terutama yang dilakukan secara *online*. Algoritma Naïve Bayes merupakan pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi. Dengan Naïve Bayes diharapkan dapat memberikan keputusan yang akurat dalam menentukan rekomendasi pakaian wanita dengan atribut yang sudah ditentukan seperti *style, price, rating, size, season, neckline, sleeve length, waiseline, material, fabric type, decoration, pattern type, recommendation*.

Abstract

Many activities buying and selling are carried out online, which of is very easy to buyer. One of them in the purchases of women's clothing, in choosing a clothing sometime feels very confusing because there are many interesting choice. Apart of there are several factors that considered such as body type, material, model, color and price. From the factors is needed system for recommendations when purchasing women's clothing, especially those online. Naïve Bayes Algorithm is simple probabilitc classification that calculates set of probabilities by summing frequencies. The Naïve Bayes is expected to provide accurate decisions in determining recommendations for women's clothing such as style, price, rating, size, season, neckline, sleeve length, waiseline, material, fabric type, decoration, pattern type, recommendation.

Keywords: *Naïve Bayes, Rekomendasi*

1. Pendahuluan

Sebagian dari wanita terkadang dalam memilih pakaian sangat membingungkan. Seperti pada saat melakukan pembelian pakaian pasti membutuhkan masukan atau pendapat dari orang lain mengenai baju yang akan dipilih, dimulai dari model, harga, bahan, ukuran dan sebagainya yang sesuai. Hal seperti ini sering membuat sebagian orang menjadi kesulitan dan membingungkan untuk memilih sebuah pakaian.

Sudah banyak penjualan yang dilakukan secara *online*, tentunya sangat memudahkan bagi para pembeli untuk tidak datang lagi ke sebuah toko. Tetapi untuk membeli pakaian secara *online* harus melihat terlebih dahulu kualitas dan kuantitas dari barang yang dijual. Seperti

stok dari barang yang dijual masih tersedia atau tidak dan yang lebih penting lagi yaitu kualitas dari pakaiannya.

Kecocokan seseorang dalam memilih sebuah pakaian dapat dipengaruhi oleh beberapa faktor seperti tipe tubuh yang dimiliki. Tipe tubuh ini banyak macamnya seperti tipe apel, pear, persegi, dan jam pasir (Mariana, Setiawan, & Irawan, 2015).

Selain itu, harus melihat deskripsi dari pakaian dan ulasan dari pakaian yang akan dibeli seperti bahan dari pakaian, model, warna, harga, corak dan rating dari para pembeli yang sudah membeli pakaian tersebut. Itu semua sangat dibutuhkan sebagai rekomendasi pada saat melakukan pembelian pakaian secara *online*. Untuk itu dibutuhkan suatu sistem yang dapat membantu dalam menentukan rekomendasi pakaian.

Sistem pendukung keputusan merupakan sistem informasi interaktif yang menyediakan informasi, pemodelan dan pemanipulasian data. Sistem ini digunakan untuk membantu dalam pengambilan keputusan (Suryadi & Harahap, 2018)

Pada penelitian ini menggunakan metode Naive Bayes dalam menentukan rekomendasi pakaian wanita. Naive bayes merupakan metode untuk klasifikasi yang menghitung probabilitas dengan dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan untuk menentukan probabilitas hasil (Medkhar, Bote, & Deshmukh, 2013)

Selain itu, menurut (Saleh, 2015) Naive Bayes merupakan sebuah pengklasifikasian probabilitas sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan.

Algoritma Naive Bayes merupakan metode data mining yang termasuk kedalam sepuluh klasifikasi data mining yang paling populer diantara algoritma-algoritma lainnya (Saputra, Taufik, Ramdhani, Oktapian, & Marsusanti, 2018)

Data mining sendiri memiliki arti menganalisis data dari perspektif yang berbeda dan meringkasnya menjadi informasi yang berguna (Medkhar et al., 2013)

Data mining memiliki beberapa tahapan yaitu data *cleaning*, data *integration*, data *transformation*, teknik data mining, *patern evaluation*, *knowledge presentation* (Peling, Arnawan, Arthawan, & Janardana, 2017).

Tahapan data mining ada 6 dengan penjelasan sebagai berikut (Meilani & Susanti, 2015):

- a. *Data Cleaning*
Proses menghilangkan data yang tidak konsisten atau data yang tidak relevan
- b. *Data Integration*
Penggabungan atau mengkombinasikan sebuah data dari beberapa sumber.
- c. *Data Selection*
Pemilihan data dari sekumpulan data operasional sebelum mendapatkan informasi *knowledge discovery in database*.
- d. *Data Transformation*
Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining.
- e. *Data Mining*

Proses utama saat metode diterapkan untuk menemukan pengetahuan atau informasi dari data.

- f. *Pattern Evaluation*
Untuk mengidentifikasi pola-pola menarik kedalam *knowledge based* yang ditemukan.
- g. *Knowledge presentation*
Pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna

2. Metode Penelitian

Metode Naive Bayes merupakan metode yang memanfaatkan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes. Naive Bayes merupakan metode pengklasifikasian yang sangat sederhana dengan mengasumsikan klasifikasi atribut. Dengan metode Naive Bayes terlebih dahulu mencari Nilai Probabilitas dan *likelihood* maksimum dari setiap atribut untuk masing-masing kelas (Karthika & Sairam, 2015).

2.1 Probabilitas Prior

Persamaan dari probabilitas *prior*:

$$P(H) = \frac{N_j}{N}$$

Dimana:

N_j : Jumlah data pada suatu class

N : Jumlah total data

2.2 Probabilitas Posterior

Persamaan dari teorema bayes (Saputra et al., 2018):

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

Dimana:

X : Data dengan class yang belum diketahui

H : Hipotesis data merupakan suatu class spesifik

$P(H|X)$: Probabilitas Hipotesis H berdasarkan kondisi X (posterior probabilitas)

$P(H)$: Probabilitas hipotesis H (prior probabilitas)

$P(X|H)$: probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$: Probabilitas X

Penentuan *class* dilakukan dengan cara membandingkan nilai probabilitas suatu sampel berada di *class* yang satu dengan nilai probabilitas suatu sampel berada di *class* yang lain.

Untuk menentukan *class* yang cocok dari suatu sampel dilakukan dengan cara membandingkan nilai posterior untuk masing-masing *class*, dan mengambil *class* dengan nilai posterior yang tertinggi.

3. Hasil dan Pembahasan

Data yang digunakan dalam penelitian ini diperoleh dari *UCI Machine Learning Repository*. Jumlah Data *Dresses_attribute_sales* sebanyak 500 data, dalam penentuan rekomendasi menggunakan atribut yang terdiri dari 12 atribut dan 1 atribut yang digunakan untuk *class* rekomendasi. Atribut yang digunakan yaitu *style, price, rating, size, season, neckline, sleeve length, waiseline, matrial, fabric type, decoration, pattern type dan recommendation*. Dari data *dresses_attribute_sales*, ada 290 data gaun yang tidak rekomended untuk dibeli dan ada 210 data gaun yang rekomended untuk dibeli.

3.1 Probabilitas Prior

Nilai probabilitas prior dihitung berdasarkan data masa lalu. Total semua data 500 dengan *class* yang berbeda, 290 dengan *class* 0 yang artinya tidak rekomended, dan 210 dengan *class* 1 yang artinya rekomended. Nilai probabilitas prior digunakan untuk menghitung nilai probabilitas prior dari setiap atribut.

Tabel 1. Atribut

No	Nama Atribut	Status Atribut
1	Style	Diketahui
2	Price	Diketahui
3	Rating	Diketahui
4	Size	Diketahui
5	Season	Diketahui
6	Neckline	Diketahui
7	Sleeve length	Diketahui
8	Waiseline	Diketahui
9	Material	Diketahui
10	Fabric type	Diketahui
11	Decoration	Diketahui
12	Pattern type	Diketahui
13	Recommendation	Dicari

Tabel 2. Nilai Atribut

Nama Atribut	Nilai Atribut
Style	Bohemian, Brief, Casual, Cute, Fashion, Flare, Novelty, OL, Party, Sexy, Vintage, Work
Price	Average, High, Low,

	Medium, Very-High
Rating	0 – 5
Size	Free, L, M, S, Small, XL
Season	Autumn, Autumn, Spring, Summer, Winter
Neckline	Backless, Boat-neck, Bowneck, Halter, Mandarin-collor, Null, o-neck, Open, Peterpan-collor, Ruffled, Scoop, Slash-neck, Sqare-collor, Sweetheart, Turndowncollor, V-Neck
Sleeve length	Butterfly, Caps-leeves, Full, Half, Halfsleeves, Null, Petal, Short, Sleeveless, Threequarter, Thressqatar, Turndowncollor, Urndowncollor
Waiseline	Dropped, Empire, Natural, Null, Princess
Material	Acrylic, Cashmere, Chiffonfabric, Cotton, Knitting, lace, linen, lycra, Microfiber, Milksilk, Mix, Modal, Model, Null, Nylon, Other, Polyster, Rayon, Shiffon, Silk, Sill, Spandex, Viscos, Wool.
Fabric type	Batik, Broadcloth, chiffon, Corduroy, Dobby, Flannel, Jersey, Knitted, Knitting, lace, null, Organza, Other, Poplin, Satin, Shiffon, Terry, Tulle, Woolen
Decoration	Applique, Beading, Bow, Button
Pattern type	Animal, dot, geometric, null, patchwork, plaid, print, solid, striped
Recommendation	Ya, Tidak

Diketahui data tes untuk menghitung nilai probabilitas di setiap atribut. Berikut adalah data tes yang ada dapat dilihat pada tabel 3.3

Tabel 3. Data Tes

Dress_Id	919930954
Style	Casual
Price	Low
Rating	4,4
Size	Free

Season	Summer
Neckline	V-Neck
SleeveLength	Short
Waiseline	Empire
Material	Cotton
FabricType	Chiffon
Decoration	Lace
PatternType	Solid
Recommendation	?

Decorat ion	Lace (Ya)	32	210	0,15
	Lace (Tidak)	38	290	0,13
Pattern Type	Solid (Ya)	79	210	0,38
	Solid (Tidak)	124	290	0,43

Tabel 4. Nilai Probabilitas Prior Atribut

Atribut	Nilai Atribut	Jumlah Data	P (X C)	
Style	Casual (Ya)	89	210	0,42
	Casual (Tidak)	143	290	0,49
Price	Low (Ya)	72	210	0,34
	Low (Tidak)	102	290	0,35
Rating	4,4 (Ya)	6	210	0,03
	4,4 (Tidak)	21	290	0,07
Size	Free (Ya)	77	210	0,37
	Free (Tidak)	96	290	0,33
Season	Summ er (Ya)	54	210	0,26
	Summ er (Tidak)	106	290	0,37
Necklin e	V- Neck (Ya)	60	210	0,29
	V- Neck (Tidak)	64	290	0,22
SleeveL ength	Short (Ya)	27	210	0,13
	Short (Tidak)	69	290	0,24
Waiseli ne	Empire (Ya)	47	210	0,22
	Empire (Tidak)	57	290	0,20
Material	Cotton (Ya)	64	210	0,30
	Cotton (Tidak)	88	290	0,30
Fabric Type	Chiffon (Ya)	57	210	0,27
	Chiffon (Tidak)	78	290	0,27

3.2 Probabilitas Posterior

Perhitungan pada probabilitas posterior digunakan untuk menentukan *class* terhadap data baru yang akan diklasifikasikan. Berikut adalah perhitungan probabilitas posterior dari data tes pada tabel 3.3:

$$\begin{aligned}
 P(X|Recommendation) &= \\
 &P(Casual|Recommendation) * \\
 &P(Low|Recommendation) * \\
 &P(4,4|Recommendation) * \\
 &P(Free|Recommendation) * \\
 &P(Summer|Recommendation) * \\
 &P(V-Neck|Recommendation) * \\
 &P(Short|Recommendation) * \\
 &P(Empire|Recommendation) * \\
 &P(Cotton|Recommendation) * \\
 &P(Chiffon|Recommendation) * \\
 &P(Lace|Recommendation) * \\
 &P(Solid|Recommendation) * \\
 &= 0,42*0,34*0,03*0,37*0,26*0,29*0,13* \\
 &0,22*0,30*0,27*0,15*0,38*210 \\
 &= 0,00000320
 \end{aligned}$$

$$\begin{aligned}
 P(X|NonRecommendation) &= \\
 &P(Casual|NonRecommendation) * \\
 &P(Low|NonRecommendation) * \\
 &P(4,4|NonRecommendation) * \\
 &P(Free|NonRecommendation) * \\
 &P(Summer|NonRecommendation) * \\
 &P(V-Neck|NonRecommendation) * \\
 &P(Short|NonRecommendation) * \\
 &P(Empire|NonRecommendation) * \\
 &P(Cotton|NonRecommendation) * \\
 &P(Chiffon|NonRecommendation) * \\
 &P(Lace|NonRecommendation) * \\
 &P(Solid|NonRecommendation) * \\
 &= 0,49*0,35*0,07*0,33*0,37*0,22*0,24* \\
 &0,20*0,30*0,27*0,13*0,43*290 \\
 &= 0,00002080
 \end{aligned}$$

4. Kesimpulan

Dapat ditarik kesimpulan dari penelitian ini sebagai berikut:

- Melihat hasil dari probabilitas posterior bahwa $P(X|Recommendation)$ lebih kecil di bandingkan hasil $P(X|NonRecommendation)$, maka dapat

diketahui bahwa data tes ini termasuk data NonRecommendation.

- b. Naïve Bayes merupakan metode yang dapat membantu dalam menyelesaikan masalah klasifikasi seperti rekomendasi pakaian wanita.
- c. Naïve Bayes sangat baik digunakan untuk mendukung penunjang keputusan.
- d. Metode Naïve Bayes merupakan metode klasifikasi dengan teorema bayes yang dikenal dengan ilmu probabilitas.

Referensi

- Karthika, S., & Sairam, N. (2015). A Naïve Bayesian Classifier for Educational Qualification. *Indian Journal of Science and Technology*, 8(16), 1–5. <http://doi.org/10.17485/ijst/2015/v8i16/62055>
- Mariana, P., Setiawan, H., & Irawan, P. L. T. (2015). Implementasi Augmented Reality Untuk Visualisasi Pakaian Wanita. *SMATIKA Jurnal*, 5(2), 51–57.
- Medkhar, D. S., Bote, M. P., & Deshmukh, S. D. (2013). Heart Disease Prediction System Using Naive Bayes. *International Journal Of Enhanced Research In Science Technology & Engineering*, 2(3), 1–5. <http://doi.org/10.1093/infdis/139.2.232>
- Meilani, B. D., & Susanti, N. (2015). Aplikasi Data Mining Untuk Menghasilkan Pola Kelulusan Siswa Dengan Metode Naive Bayes. *Jurnal Ilmiah NERO*, 1(3), 182–189.
- Peling, I. B. A., Arnawan, I. N., Arthawan, I. P. A., & Janardana, I. (2017). Implementation of Data Mining To Predict Period of Students Study Using Naive Bayes Algorithm. *International Journal of Engineering and Emerging Technology*, 2(1), 53–57.
- Saleh, A. (2015). Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *CITEC Journal*, 2(3), 207–217. <http://doi.org/10.24076/citec.2015v2i3.49>
- Saputra, R. A., Taufik, A. R., Ramdhani, L. S., Oktapian, R., & Marsusanti, E. (2018). Sistem Pendukung Keputusan Dalam Menentukan Metode Kontrasepsi Menggunakan Algoritma Naive Bayes. *SNIT 2018*, 106–111.
- Suryadi, A., & Harahap, E. (2018). Sistem Rekomendasi Penerimaan Mahasiswa Baru Menggunakan Naive Bayes Classifier Di Institut Pendidikan Indonesia. *JOUTICA*, 3(2), 171–182.