Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 03, 2024, DOI 10.54209/infosains.v14i03
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Implementation Of Naïve Bayes Algorithm For Children's Clothing Recommendation

Yudi Sutaryana[1], Lalu Puji Indra Kharisma[2*]
[1*,2]STMIK Syaikh Zainuddin Nahdlatul Wathan Anjani, Lombok Timur, NTB

Article Info
Keywords:
Naïve Bayes,
Recommendations,
Product Purchase

ABSTRACT
Many buying and selling activities are carried out online, which is very easy for buyers. One of themn the purchases of children's clothing, when buying clothes children for in choosing a garment, is sometimes very confusing there are many interesting choices. Apart from this, several factors are taken into account such as body type, material, color and price. From these factors, a system of recommendations necessary when purchasing children's clothing, in those online. The Naïve Bayes algorithm is a simple probabilistic classification which a set of probabilities by adding the frequencies. Naïve Bayes is supposed to provide accurate determining recommendations regarding clothing for women such as style, price, rating, size, neckline, length of sleeve, size, material, fabric type, decoration, pattern type, recommendation.

Corresponding Author:
Lalu Puji Indra Kharisma
STMIK Syaikh Zainuddin Nahdlatul Wathan Anjani, Lombok Timur, NTB
lalupujii@gmail.com

INTRODUCTION

Parents, when choosing their children's clothes, sometimes think that the clothing choices they choose are not attractive. For example, when buying clothes, parents need advice or other people's opinions regarding the clothes they want to choose, starting from price, material, size, and so on. Things like this often make parents confused before choosing clothes for their children.

As we know, nowadays many sellers sell online, and that means buyers no longer need to come directly to the shop. Because now you can buy clothes online. However, the problem is, because sales are carried out online, this means that the items to be purchased cannot be seen directly, but only by looking at photos or videos. And this is what makes parents who want to buy clothes online confused, because they don't see directly the quality and quantity of the products being sold, whether the stock of the goods being sold is available or not, and what is more important is whether the price of the clothes is appropriate or not. the quality. Parents' ability to choose children's clothes can be influenced by several factors such as model, color, price and so on. Currently there are many types and types of children's clothing that can attract their attention, examples of clothing images ranging from images of apples, pears, rectangles, images of clocks, and so on (Mariana, Setiawan, & Irawan, 2015).

Apart from that, people also have to look at several descriptions of the clothes and reviews of the clothes they want to buy before choosing clothes for children, starting from

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 03, 2024, DOI 10.54209/infosains.v14i03
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

what material the clothes are made of, what model they are, color, price, pattern and ratings from previous buyers. I have already bought clothes at this shop. All of this is very influential as a reference for recommendations when purchasing clothes online. Therefore, we really need a system that can help determine recommendations for choosing children's clothes. Decision support systems are interactive information systems that provide information (Mahendra et al., 2023), modeling and data manipulation. This system is used to assist in decision making (Suryadi & Harahap, 2018)

In this study, the Naive Bayes method was used to determine recommendations for children's clothing. Naive Bayes is a method for classification that calculates probabilistically by adding up the frequencies and combinations of values from a given dataset to determine the probability of an outcome (Medkhar, Bote, & Deshmukh, 2013) (Septiani et al., 2024)

Apart from that, according to (Saleh, 2015) Naive Bayes is a simple probabilistic classification that calculates a set of probabilities by adding up the frequencies and combinations of values from a given dataset. The Naive Bayes algorithm is a data mining method which is included in the ten most popular data mining classifications among other algorithms (Saputra, Taufik, Ramdhani, Oktapian, & Marsusanti, 2018).

Data mining itself means analyzing data from different perspectives and summarizing it into useful information (Medkhar et al., 2013). Data mining has several stages, namely data cleaning, data integration, data transformation, data mining techniques, pattern evaluation, knowledge presentation (Peling, Arnawan, Arthawan, & Janardana, 2017) (Suryadana & Sarasvananda, 2024).

There are 6 stages of data mining with the following explanation (Meilani & Susanti, 2015) is Data Cleaning: The process of eliminating inconsistent data or irrelevant data. Data Integration Combining or combining data from several sources (Dewi et al., 2021). Data Selection: Selection of data from a set of operational data before obtaining knowledge discovery information in the database. Data Transformation Data is changed or combined into a format suitable for processing in data mining. Data MiningThe main process when methods are applied to find knowledge or information from data. Pattern Evaluation To identify interesting patterns in the knowledge based found. Knowledge presentation: Knowledge of the methods used to obtain the knowledge obtained by the user

METHODS

Metode Naïve Bayes merupakan metode yang memanfaatkan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes. Naïve 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).

Probabilitas Prior

Equation of prior probability:

$$P(H) = \frac{Nj}{N} \quad (1)$$

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 03, 2024, DOI 10.54209/infosains.v14i03
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Where:
Nj : The amount of data in a class
N : Total amount of data

Probabilitas Posterior
Equation of Bayes' theorem (Saputra et al., 2018):

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (2)$$

Where:
X : Data with unknown class
H : A data hypothesis is a specific class
P(H|X) : Probability of Hypothesis H based on condition X (posterior probability)
P (H) : Probability of hypothesis H (prior probability)
P(X|H) : Probability of X based on the conditions in hypothesis H
P(X) : X probability

Class determination is done by comparing the probability value of a sample being in one class with the probability value of a sample being in another class. To determine the suitable class of a sample is done by comparing the posterior value for each class, and taking the class with the highest posterior value.

RESULTS AND DISCUSSION

The data used in this research was obtained from the UCI Machine Learning Repository. The amount of Dresses_attribute_sales data is 500 data, in determining recommendations using attributes consisting of 12 attributes and 1 attribute used for the recommendation class. The attributes used are style, price, rating, size, season, neckline, sleevelength, waiseline, material, fabric type, decoration, pattern type and recommendation. From the dresses_attribute_sales data, there are 290 clothing data that are not recommended for purchase and there are 210 clothing data that are recommended for purchase.

Prior Probability

Prior probability values are calculated based on past data. The total of all data is 500 with different classes, 290 with class 0 which means not recommended, and 210 with class 1 which means recommended. Prior probability values are used to calculate the prior probability values of each attribute.

Table 1. Attributes

No	Attribute Name	Attribute Status
1	Style	Is known
2	Price	Is known
3	Ratings	Is known
4	Size	Is known
5	Seasons	Is known
6	Necklines	Is known

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 03, 2024, DOI 10.54209/infosains.v14i03
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

No	Attribute Name	Attribute Status
7	Sleeve length	Is known
8	Waiseline	Is known
9	Material	Is known
10	Fabrictype	Is known
11	Decoration	Is known
12	Patterntype	Is known
13	Recommendations	Searching for

Tabel 2. Attribute Values

Attribute Name	Attribute Value
Style	bohemians, Briefs, Casual, Cute, Fashion, Flare, Novelty, OL, Party, Sexy, Vintage, Work
Price	Average, High, Low, Medium, Very-High
Ratings	0 – 5
Size	Free, L, M, S, Small, XL
Seasons	Automn, Autumn, Spring, Summer, Winter
Necklines	Backless, Boat-neck, Bowneck, Halter, Mandarin-collor, Null, oneck, Open, Peterpancollor, Ruffled, Scoop, Slash-neck, Sqarecollor, Sweetheart, Turndowncolour, V-Neck
Sleeve length	Butterfly, Caps-leeves, Full, Half, Halfsleeves, Null, Petal, Short, sleeveless, threequarters, Thressqatar, turndown color, Urndowncolour
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.
Fabrictype	Batik, Broadcloth, chiffon, Corduroy, Dobby, Flannel, Jersey, Knitted, Knitting, lace, null, Organza, Other, Poplin, Satin, Shiffon, Terry, Tulle, Woolen
Decoration	Appliques, Beading, Bows, Buttons
Patterntype	Animal, dot, geometric, null, patchwork, plaid, print, solid, striped
Recommendations	Yes No

Know the test data to calculate the probability value for each attribute. The following is the existing test data which can be seen in table 3.

Table 3. Test Data

Dress_Id	919930954
Style	Casual
Price	Low

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 03, 2024, DOI 10.54209/infosains.v14i03
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Dress_Id	919930954
Rating	4,4
Size	Free

Table 4. Prior Probability Values of Attributes

Attribute	Attribute Value		Amount of data	P(X|C)
Style	Casual (Yes)	89	210	0.42
	Casual (No)	143	290	0.49
Price	Low (Yes)	72	210	0.34
	Low (No)	102	290	0.35
Ratings	4.4 (Yes)	6	210	0.03
	4.4 (No)	21	290	0.07
Size	Free (Yes)	77	210	0.37
	Free (No)	96	290	0.33
Seasons	Summer (Yeah)	54	210	0.26
	Summer (No)	106	290	0.37
Necklin e	VNeck (Yes)	60	210	0.29
	VNeck (No)	64	290	0.22
SleeveLength	Short (Yes)	27	210	0.13
	Short (No)	69	290	0.24
Waiseli ne	Empire (Yeah)	47	210	0.22
	Empire (No)	57	290	0.20
Material	Cotton (Yes)	64	210	0.30
	Cotton (No)	88	290	0.30
Fabric Type	Chiffon (Yeah)	57	210	0.27
	Chiffon (No)	78	290	0.27

Calculations on posterior probabilities are used to determine the class of new data that will be classified. The following is the posterior probability calculation from the test data in table 3:

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) *

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 03, 2024, DOI 10.54209/infosains.v14i03
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

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

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

CONCLUSION

Conclusions can be drawn from this research as follows: Seeing the results of the posterior probability that P(X|Recommendation) is smaller than the results of P(X|NonRecommendation), it can be interpreted that the data in this test includes NonRecommendation data. Naïve Bayes is a method that can help in solving classification problems such as recommending children's clothing. Naïve Bayes is very well used to support decision support. The Naïve Bayes method is a classification method using Bayes' theorem which is known as probability science.

REFERENCE

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 Anak. *SMATIKA Jurnal2*, *5*(2), 51–57.

Dewi, I. G. A. M. P., Parwita, W. G. S., & Setiawan, I. M. D. (2021). Algoritma Decision Tree untuk Klasifikasi Calon Debitur LPD Desa Adat Anggungan. *Jurnal Krisnadana*, *1*(1), 23–36.

Mahendra, G. S., Hariyono, R. C. S., Purnawati, N. W., Hatta, H. R., Sudipa, I. G. I., Hamali, S., Sarjono, H., & Meilani, B. D. (2023). *BUKU AJAR SISTEM PENDUKUNG KEPUTUSAN*. PT. Sonpedia Publishing Indonesia.

Septiani, A. T. D., Kuncoro, A. P., Subarkah, P., & Riyanto, R. (2024). Perbandingan Kinerja Metode Naïve Bayes Classifier dan K-Nearest Neighbor pada Analisis Sentimen Ulasan Mobile Banking Jenius. *Jurnal Krisnadana*, *3*(2), 67–77.

Suryadana, K., & Sarasvananda, I. B. G. (2024). Streamlining Inventory Forecasting with Weighted Moving Average Method at Parta Trading Companies. *Jurnal Galaksi*, *1*(1), 12–21.

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.

Meilani, B. D., & Susanti, N. (2015). Aplikasi Data Mining Untuk Menghasilkan Pola Kelulusan Siswa Ddengan 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/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.