

ANALISIS KLASIFIKASI KEPUASAN PENUMPANG MASKAPAI PENERBANGAN MENGUNAKAN ALGORITMA NAÏVE BAYES

Agung Wijayanto, Januario Freitas Araujo Bernardo, Sapto Pamungkas
Program Studi Magister Teknik Informatika
Univesitas Amikom Yogyakarta
Email: agung.1179@students.amikom.ac.id, januario.1200@students.amikom.ac.id,
sapto.1169@students.amikom.ac.id

Abstrak

Dalam dunia transportasi khususnya penerbangan banyak sekali hal menarik yang bisa dianalisa maupun diperbarui, contohnya dalam menilai kepuasan penumpang maskapai sebagai evaluasi maupun mengetahui seberapa besar kepuasan pelanggan dalam menggunakan transportasi penerbangan tersebut. Karena alasan inilah sehingga peneliti akan melakukan penelitian mengenai hal tersebut di atas dengan menggunakan algoritma Naïve Bayes, dimana pengukuran tersebut diukur dengan 16 variabel pengukur dan 1 variabel respon. Digunakan dataset sebanyak 129.880 record untuk selanjutnya dari dataset tersebut akan dibagi menjadi data training dan data testing, dimana pembagian data tersebut akan dibuat dalam 4 kondisi yaitu 90%, 85%, 80% dan 75% untuk masing data training dan sisanya adalah data testing sehingga dapat dilihat dari keempat pembagian ini manakah yang memiliki tingkat akurasi yang paling tinggi. Dari proses analisa menggunakan algoritma Naïve Bayes yang dilakukan dengan aplikasi KNime, didapatkan bahwa pembagian data training 90% dan Data Testing 10% adalah hasil akurasi tertinggi dengan persentasi akurasi sebesar 81,466%. Dengan hasil ini maka 16 variabel pengukur yang ditetapkan peneliti dapat dijadikan sebagai sebuah acuan untuk mengukur tingkat kepuasan penumpang maskapai penerbangan.

Kata kunci—Klasifikasi, Tingkat Kepuasan, Algoritma, Naïve Bayes

Abstract

In the world of transportation, especially aviation, there are many interesting things that can be analyzed and updated, for example in assessing airline passenger satisfaction as an evaluation and knowing how much customer satisfaction is in using the flight transportation. It is for this reason that researchers will conduct research on the above by using the Naïve Bayes algorithm where the measurement is measured by 16 measuring variables and 1 response variable. The dataset used as many as 129,880 records, then the dataset will be divided into training data and testing data, where the data distribution will be made in 4 conditions, namely 90%, 85%, 80% and 75% for each training data and the rest is testing data so that It can be seen from the four divisions which one has the highest level of accuracy. From the analysis process using the Naïve Bayes algorithm which was carried out with the KNime application, it was found that the sharing of training data was 90% and Data Testing 10% was the highest accuracy result with an accuracy percentage of 81.466%. With these results, the 16 measuring variables defined by the researcher can be used as a reference for measuring the level of satisfaction of airline passengers.

Keywords— Classification, Satisfaction Level, Algorithm, Naïve Bayes

I. PENDAHULUAN

Seiring dengan perkembangan zaman di era Industri 4.0 setiap maskapai memiliki cara atau standar kualitas pelayanan yang berbeda-beda. Semakin baik cara atau standar kualitas pelayanan sebuah maskapai, maka kepuasan penumpang akan meningkat dan minat calon penumpang maskapai penerbangan akan semakin tinggi untuk menggunakan jasanya. Jika minat calon penumpang sebuah maskapai terus meningkat, maka standar kualitas dari pelayanan

jasa maskapai penerbangan harus mampu menghadapi beberapa pesaing yang semakin kompetitif. Tetapi jika pelayanan maskapai penerbangan dinilai buruk dapat mengakibatkan menurunnya minat penumpang. Analisis kepuasan adalah proses menganalisa hasil ulasan dari pengguna layanan untuk menentukan sebuah opini atau perasaan terhadap suatu layanan secara keseluruhan. Tujuannya sendiri adalah untuk memastikan perilaku atau opini dari seorang yang mengulas dan memberi pendapat dengan memperhatikan suatu topik

dan aspek tertentu[1]. Pada penelitian ini proses analisis kepuasan penumpang dengan menentukan beberapa variabel dilihat dari berbagai sisi layanan penerbangan, beberapa kriteria / variabel yang menjadi tolak ukur adalah : layanan *onboard*, ketepatan waktu kedatangan dan keberangkatan, kebersihan, bagasi, dan lainnya. Beberapa teknik yang dapat digunakan dalam proses analisis kepuasan penumpang penerbangan salah satunya dengan menggunakan Algoritma Naïve Bayes. Algoritma ini termasuk algoritma klasifikasi yang cukup populer dan merupakan sepuluh algoritma terbaik didalam bidang data mining. Algoritma Naive Bayes menggunakan teori probabilitas dari cabang ilmu matematika yang digunakan untuk mencari sebuah peluang paling besar dari semua kemungkinan klasifikasi yang didapatkan, dengan melihat frekuensi dari setiap hasil klasifikasi dari data training [2]. Dalam klasifikasi

algoritma Naïve Bayes ini, hasil yang diberikan akan ditampilkan sebagai hasil kepuasan penumpang dalam bentuk pernyataan puas dan netral/tidak puas serta menghitung nilai akurasi dan menganalisa apakah variabel-variabel yang ditentukan peneliti dapat menjadi acuan untuk menilai tingkat kepuasan penumpang maskapai penerbangan.

2. METODE PENELITIAN

2.1 Seleksi Variabel Penelitian

Penelitian ini dilakukan dengan menggunakan data yang diambil dari situs www.kaggle.com pada *Airline Passenger Satisfaction* dengan data berjumlah 129.880 record dan terdiri dari 16 variabel pengukur dan 1 variabel respon. Berikut adalah variabel dari data yang akan digunakan.

Tabel 1. Variabel Penelitian

Variabel	Informasi Data
<i>Satisfaction</i>	Variabel Respon Kepuasan
<i>Gender</i>	Jenis kelamin penumpang
<i>Customer Type</i>	Jenis pelanggan
<i>Age</i>	Usia penumpang sebenarnya
<i>Type of Travel</i>	Tujuan penerbangan penumpang
<i>Class</i>	Kelas perjalanan di pesawat penumpang
<i>Flight Distance</i>	Jarak penerbangan dari perjalanan
<i>Departure/Arrival time convenient</i>	Tingkat kepuasan waktu Keberangkatan / Kedatangan
<i>Food and drink</i>	Tingkat kepuasan makanan dan minuman
<i>Seat comfort</i>	Tingkat kepuasan kenyamanan kursi
<i>On-board service</i>	Tingkat kepuasan layanan On-board
<i>Baggage handling</i>	Tingkat kepuasan penanganan bagasi
<i>Checkin service</i>	Tingkat kepuasan layanan Check-in
<i>Inflight service</i>	Tingkat kepuasan layanan dalam pesawat
<i>Cleanliness</i>	Tingkat Kepuasan Kebersihan
<i>Departure Delay in Minutes</i>	Menit penundaan saat keberangkatan
<i>Arrival Delay in Minutes</i>	Menit penundaan saat Kedatangan

Dari variabel di atas, data akan dibuat dalam bentuk excel atau menggunakan format CSV agar bisa diolah dan dianalisa dalam aplikasi Knime atau sejenisnya.

2.2 Preprocessing Data

Sebelum data digunakan untuk tahapan analisa klasifikasi, dataset harus dilakukan data

preprocessing yang nantinya meliputi : pembersihan data yang tidak lengkap (missing value), transformasi data, dan konversi data agar dapat digunakan pada tahap klasifikasi.

2.3 Split Data Training dan Testing

Pada tahap ini akan dilakukan pembagian dataset menjadi 2 bagian yaitu data training dan data

test dengan 4 kali pembagian persentase yang berbeda untuk membandingkan hasil akurasi

yang didapat sehingga hasilnya akan optimal.

Tabel 2. Persentase Split Dataset

No	Persentase Split	
	Data Training	Data Test
1	90%	10%
2	85%	15%
3	80%	20%
4	75%	25%

2.4 Klasifikasi Naïve Bayes

Data yang telah dipisah menjadi data training dan data test, selanjutnya dilakukan klasifikasi dengan menggunakan algoritma Naïve Bayes. Teorema Bayes dirumuskan sebagai berikut[9] :

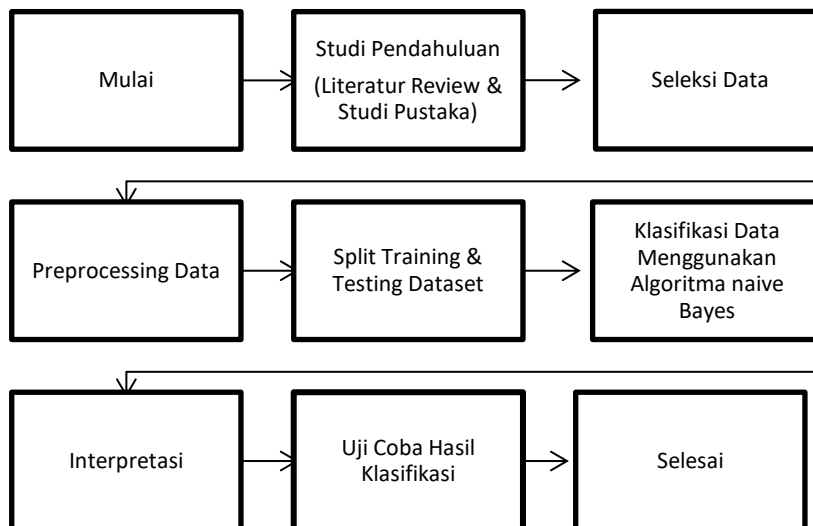
$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

Dimana Y merupakan label kelas yang menjadi target klasifikasi dan X merupakan variabelnya. Teorema Bayes dapat diartikan sebagai pembelajaran berdasarkan data latih untuk

pembangunan model dari setiap kombinasi Y dengan semua fitur X sehingga kita akan mendapatkan informasi peluang perolehan kelas Y berdasarkan variabel – variabel X yang diamati[10].

Dari proses klasifikasi di atas, selanjutnya dilakukan proses validasi data dimana Validasi adalah proses mengevaluasi dan menginterpretasikan keakuratan hasil model prediksi dari proses klasifikasi[11]

2.5 Alur Penelitian



Gambar 1. Alur Penelitian

3. HASIL DAN PEMBAHASAN

3.1 Tahap Preprocessing Dataset

Setelah mendapatkan dataset sebanyak 129.880 record, maka tahap selanjutnya adalah melakukan persiapan data untuk selanjutnya akan dilakukan analisa klasifikasi dengan cara pembersihan data yang tidak lengkap atau data tersebut kosong. Hasil dari pembersihan data tersebut didapatkan sebanyak 393 data yang

kosong yang berada pada variabel Arrival Delay in Minutes. Jadi total dataset yang sebelumnya sebanyak 129.880 berkurang menjadi 129.487 setelah dikurangi dari 393 data yang kosong . Dari hasil pembersihan data yang kosong, didapatkan informasi dataset yang akan digunakan pada masing-masing variabel:

Tabel 3. Variabel dan klasifikasi data penelitian

Variabel	Klasifikasi Data
Satisfaction	Satisfied, Neutral or Dissatisfied
Gender	Male, Female
Customer Type	Loyal Customer, Disloyal Customer
Age	7-85 Tahun
Type of Travel	Personal Travel, Business travel
Class	Business, Eco, Eco Plus
Flight Distance	31 - 4983km
Departure/Arrival time convenient	Skala 0 – 5
Food and drink	Skala 0 – 5
Seat comfort	Skala 0 – 5
On-board service	Skala 0 – 5
Baggage handling	Skala 0 – 5
Checkin service	Skala 0 – 5
Inflight service	Skala 0 – 5
Cleanliness	Skala 0 – 5
Departure Delay in Minutes	0 -1592 Menit
Arrival Delay in Minutes	0-1584 Menit

3.2 Klasifikasi Data Menggunakan Algoritma Naive Bayes Tahap Preprocessing Dataset

Pada tahap klasifikasi ini akan dilakukan pencarian nilai accuracy, precision dan recall

dengan menggunakan algoritma Naive Bayes dengan pembagian data seperti yang dijelaskan di atas, dimana hasilnya adalah sebagai berikut:

Tabel 4. hasil Split Dataset

No	Persentase Split		Total Data	
	Data Training	Data Test	Data Training	Data Test
1	90%	10%	116.538	12.949
2	85%	15%	110.063	19.424
3	80%	20%	103.589	25.898
4	75%	25%	97.115	32.372

Tabel di atas menunjukkan besarnya data yang akan dibagi menjadi data Training dan dataset dari total dataset sebesar 129.487. Selanjutnya adalah melatih semua data training menggunakan algoritma Naive Bayes sehingga

hasil prediksi pada data test dapat diukur akurasi serta melihat seberapa besar data yang terklasifikasi benar dan salah dari data test tersebut. Hasilnya adalah sebagai berikut:

Tabel 5. hasil Klasifikasi Algoritma Naive Bayes pada DataTest

No	Persentase Split		Data Terklasifikasi Benar	Data Terklasifikasi Salah	Accuracy	Error
	Data Training	Data Test				
1	90%	10%	10.549	2.400	81,466%	18,534%
2	85%	15%	15.802	3.622	81,353%	18,647%

3	80%	20%	21.082	4.816	81,404%	18,596%
4	75%	25%	26.352	6.020	81,404%	18,596%

Dari tabel di atas didapatkan hasil bahwa dengan banyaknya data training yang gunakan sebesar 90% dan 10 % data test maka hasil akurasi sebesar 81,466% dan merupakan nilai tertinggi untuk akurasi yang didapat. Tetapi hasil selanjutnya cukup berbeda dimana dengan pembagian lebih sedikit di data training dan data test yaitu pada pembagian 80%-20% didapatkan tingkat akurasi yang lebih tinggi dibandingkan

dengan pembagian dataset 85%-15% seperti yang terlihat pada tabel di atas. Dikarenakan nilai akurasi yang didapat adalah berdasarkan pembagian dataset sebesar 90%-10%, maka statistik akurasi klasifikasi kepuasan pengguna jasa layanan maskapai penerbangan akan digunakan pembagian terbaik dengan hasil sebagai berikut:

Tabel 6. hasil Analisis Klasifikasi pada Split data 90%-10%

Klasifikasi	Recall	Precision	F-Mean
Satisfied,	0.75	0.81	0.779
Neutral or Dissatisfied	0.865	0.818	0.841

3.3 Uji Coba Hasil Klasifikasi

Setelah mendapatkan nilai akurasi yang dinilai paling tinggi, maka tahap terakhir adalah melakukan uji coba pada hasil akurasi tertinggi tersebut dengan menggunakan dataset baru :

untuk mengukur tingkat kepuasan pengguna maskapai penerbangan. Disini akan dicoba menggunakan 2 dataset baru yaitu sebagai berikut

Tabel 7. Dataset untuk Ujicoba Algoritma Naïve Bayes

Variabel Uji	Jawaban Dataset	
	Dataset 1	Dataset 2
Gender	Male	Male
Customer Type	Loyal Customer,	Disloyal Customer,
Age	23 Tahun	45 Tahun
Type of Travel	Personal Travel	Personal Travel
Class	Eco Plus	Business
Flight Distance	521 km	1233 Km
Departure/Arrival time convenient	2	4
Food and drink	3	5
Seat comfort	3	4
On-board service	3	4
Baggage handling	3	5
Checkin service	2	5
Inflight service	3	3
Cleanliness	4	4
Departure Delay in Minutes	120 Menit	15 Menit
Arrival Delay in Minutes	30 Menit	10 Menit

Hasil dari ujicoba kedua dataset di atas adalah sebagai berikut:

Tabel 8. Hasil Ujicoba Algoritma Naïve Bayes

Dataset	Klasifikasi Kepuasan
1	Neutral or Dissatisfied
2	Satisfied,

Dari kedua tabel di atas, didapatkan bahwa tingkat kepuasan seorang penumpang maskapai penerbangan dapat diukur secara garis besar bahwa maskapai penerbangan wajib meminimalisir keterlambatan penerbangan dengan memberikan kepuasan dari layanan dalam penerbangan terlebih untuk penerbangan dengan jarak yang cukup jauh.

Dari hasil penelitian ini, diharapkan dapat menjadi sebuah referensi untuk mengukur tingkat kepuasan seorang penumpang pada maskapai penerbangan guna terus meningkatkan layanan penerbangan di masa yang akan datang.

4. KESIMPULAN

Dari hasil penelitian yang telah dilakukan pada data training dan data set didapatkan hasil bahwa dengan banyaknya data training yang digunakan sebesar 90% dan 10 % data test maka hasil akurasi sebesar 81,466% dan merupakan nilai tertinggi untuk akurasi yang didapat. Tetapi hasil selanjutnya cukup berbeda

dimana dengan pembagian lebih sedikit di data training dan data test yaitu pada pembagian 80%-20% didapatkan tingkat akurasi yang lebih tinggi yaitu 81,404% dibandingkan dengan pembagian dataset 85%-15% yaitu 81,353%. Dikarenakan nilai akurasi yang didapat adalah berdasarkan pembagian dataset sebesar 90%-10%, maka statistik akurasi klasifikasi kepuasan pengguna jasa layanan maskapai penerbangan akan digunakan pembagian terbaik dengan hasil bahwa maskapai penerbangan wajib meminimalisir keterlambatan penerbangan dengan memberikan kepuasan dari layanan dalam penerbangan terlebih untuk penerbangan dengan jarak yang cukup jauh.

5. SARAN

Agar penelitian ini bisa ditingkatkan dan mendapatkan hasil yang lebih baik maka disarankan untuk mempertimbangkan dataset lain dalam penelitian untuk akurasi yang akan dihasilkan oleh Naïve Bayes.

DAFTAR PUSTAKA

- [1] A. B. P. Negara, H. Muhardi, and I. M. Putri, "Analisis Sentimen Maskapai Penerbangan Menggunakan Metode Naive Bayes dan Seleksi Fitur Information Gain," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 3, p. 599, 2020, doi: 10.25126/jtiik.2020711947.
- [2] S. A. Aaputra, Didi Rosiyadi, Windu Gata, and Syepri Maulana Husain, "Sentiment Analysis Analisis Sentimen E-Wallet Pada Google Play Menggunakan Algoritma Naive Bayes Berbasis Particle Swarm Optimization," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 3, pp. 377–382, 2019, doi: 10.29207/resti.v3i3.1118.
- [3] T. H. Apandi and C. A. Sugianto, "Algoritma Naive Bayes untuk Prediksi Kepuasan Pelayanan Perekaman e-KTP (Naive Bayes Algorithm for Satisfaction Prediction of e-ID," *JUITA (Jurnal Inform. UMP)*, vol. 7, no. November, pp. 125–128, 2019.
- [4] M. Hasan, "Menggunakan Algoritma Naive Bayes Berbasis," vol. 9, pp. 317–324, 2017.
- [5] W. D. Septiani, "Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis," *None*, vol. 13, no. 1, pp. 76–84, 2017, doi: 10.33480/pilar.v13i1.149.
- [6] A. Prabhat and V. Khullar, "Sentiment classification on big data using Naïve bayes and logistic regression," *2017 Int. Conf. Comput. Commun. Informatics, ICCCI 2017*, no. January 2017, 2017, doi: 10.1109/ICCCI.2017.8117734.
- [7] A. D. Poernomo and S. Suharjito, "Indonesian online travel agent

- sentiment analysis using machine learning methods,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 14, no. 1, p. 113, 2019, doi: 10.11591/ijeecs.v14.i1.pp113-117.
- [8] J. Singh, G. Singh, and R. Singh, “Optimization of sentiment analysis using machine learning classifiers,” *Human-centric Comput. Inf. Sci.*, vol. 7, no. 1, 2017, doi: 10.1186/s13673-017-0116-3.
- [9] Y. T. Samuel and K. DEwi, “Penggunaan Metode NAÏVE BAYES Dalam Mengukur Tingkat Kepuasan Pengguna Terhadap Online System Universitas Advent Indonesia,” *TelKa*, vol. 9, no. 02, pp. 147–153, 2019, doi: 10.36342/teika.v9i02.2162.
- [10] H. Oktavianto and R. P. Handri, “Analisis Klasifikasi Kanker Payudara Menggunakan Algoritma Naive Bayes,” *INFORMAL Informatics J.*, vol. 4, no. 3, p. 117, 2020, doi: 10.19184/isj.v4i3.14170.
- [11] N. Normah, “Naïve Bayes Algorithm For Sentiment Analysis Windows Phone Store Application Reviews,” *Sinkron*, vol. 3, no. 2, p. 13, 2019, doi: 10.33395/sinkron.v3i2.242.