

IMPLEMENTASI ALGORITMA NAÏVE BAYES TERHADAP KLASIFIKASI KEPUASAN PENUMPANG PESAWAT

SKRIPSI

**OLEH
JIHAN SALSABILAH
NIM. 19610073**



**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025**

IMPLEMENTASI ALGORITMA NAÏVE BAYES TERHADAP KLASIFIKASI KEPUASAN PENUMPANG PESAWAT

SKRIPSI

**Diajukan Kepada
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
untuk Memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Matematika (S.Mat)**

**Oleh
Jihan Salsabilah
NIM. 19610073**

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025**

IMPLEMENTASI ALGORITMA NAÏVE BAYES TERHADAP KLASIFIKASI KEPUASAN PENUMPANG PESAWAT

SKRIPSI

Oleh
Jihan Salsabilah
NIM. 19610073

Telah Disetujui Untuk Diuji
Malang, 12 Desember 2025

Dosen Pembimbing I

Dosen Pembimbing II



Juhari, M.Si
NIPPPK.19840209 202321 1 010



Erna Herawati, M.Pd
NIPPPK. 19760723 202321 2 006

Mengetahui,
Ketua Program Studi Matematika



Dr. Fachrur Rozi, M.Si
NIP. T9800527 200801 1 012

IMPLEMENTASI ALGORITMA NAIVE BAYES TERHADAP KLASIFIKASI KEPUASAN PENUMPANG PESAWAT

SKRIPSI

Oleh
Jihan Salsabilah
NIM. 19610073

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
untuk Memperoleh Gelar Sarjana Matematika (S.Mat)

Tanggal 17 Desember 2025

Ketua Penguji : Dr. Usman Pagalay, M.Si

Anggota Penguji I : Hisyam Fahmi, M.Kom

Anggota Penguji II : Juhari, M.Si.

Anggota Penguji III : Erna Herawati, M.Pd



Mengetahui,
Ketua Program Studi Matematika



Dr. Fachrud Rozi, M.Si.
NIP. 19800527 200801 1 012

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Jihan Salsabilah
NIM : 19610073
Program Studi : Matematika
Fakultas : Sains dan Teknologi
Judul Skripsi : Implementasi Algoritma Naive Bayes Terhadap
Klasifikasi Kepuasan Penumpang Pesawat

Menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya sendiri, bukan merupakan pengambilan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan dan pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka. Apabila di kemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 17 Desember 2025

Yang membuat pernyataan,

A yellow rectangular revenue stamp with a black border. It features the number '10000' in large black digits, the word 'METERAI' above 'TEMPEL', and a serial number '62F26AND215859132' at the bottom. A black ink signature is written over the stamp.

Jihan Salsabilah

NIM. 19610073

MOTO

“Jangan khawatir akan hari esok, karena hari esok mempunyai kesusahannya sendiri. Kesusahann sehari cukuplah untuk sehari.”

KATA PENGANTAR

Dengan memanjatkan rasa syukur kehadiran Allah SWT yang telah melimpahkan rahmat, taufik dan hidayah-Nya sehingga penulis dapat menyelesaikan pengerjaan tugas akhir yang berjudul “Implementasi Algoritma Naïve Bayes terhadap Klasifikasi Penumpang Pesawat”, sebagai salah satu syarat untuk menyelesaikan Program Sarjana (S1) serta mendapatkan gelar sarjana matematika di Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang.

Dalam proses penyusunan skripsi ini, penulis menerima banyak bimbingan, masukan dan arahan dari berbagai pihak. Oleh karena itu, melalui halaman ini penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. Hj. Ilfi Nur Diana, M.Si, selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang
2. Dr. Agus Mulyono, M.Kes., selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang
3. Dr. Fachrur Rozi, M.Si selaku Ketua Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Hisyam Fahmi, M.Kom selaku Dosen Wali yang telah memberikan pengetahuan, arahan, dan nasihat kepada penulis
5. Juhari, M.Si., selaku Dosen Pembimbing 1 yang telah memberikan pengetahuan, arahan, nasihat, dan motivasi kepada penulis.
6. Erna Herawati, M.Pd., selaku Dosen Pembimbing 2 yang telah memberikan bimbingan, nasihat, dan banyak ilmu kepada penulis.
7. Seluruh dosen Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang.

8. Ayah dan ibu selaku orangtua serta keluarga tercinta yang selalu memberikan doa dan mencukupi kebutuhan yang tiada henti kepada penulis.
9. Maysa Vania Farabila selaku adik penulis yang selalu memberikan motivasi disaat lelah dan ingin menyerah.
10. Para sahabat yang bersedia meluangkan waktu dikala *stress*, Mohammad Abdullah, Moch.Afif Badjeber, Mulyo Widodo, Diashinta Cahyaning Hapsari, Cahyanita Pramestya Dewi, dan Firda Aulia.
11. Seluruh mahasiswa Program Studi Matematika angkatan 2019 yang telah memberikan bantuan, semangat, dan motivasi sampai mencapai titik ini.

Semoga Allah SWT. memberikan balasan atas segala kebaikan yang telah diberikan kepada penulis. Penulis menyadari masih terdapat kekurangan dalam penyusunan laporan penelitian ini. Oleh karena itu, penulis mengharapkan saran dan kritik yang bersifat membangun dari semua pihak.

Malang, 17 Desember 2025

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGANTAR	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN.....	iv
PERNYATAAN KEASLIAN TULISAN	v
MOTO.....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN	xii
ABSTRAK	xiii
ABSTRACT	xiv
ستخلص البحث	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	4
1.5 Batasan Masalah	4
BAB II KAJIAN TEORI	5
2.1 Data Mining.....	5
2.2 Klasifikasi.....	7
2.3 Algoritma Naïve Bayes	9
2.4 <i>Confusion matrix</i>	13
2.5 <i>Permutation Importance</i>	16
2.6 <i>K-fold Cross Validation</i>	16
2.7 Kualitas Jasa dan Pelayanan	18
2.8 Kepuasan Konsumen	18
2.9 Kajian Keislaman	19
BAB III METODE PENELITIAN	23
3.1 Jenis Penelitian.....	23
3.2 Data dan Sumber Data.....	23
3.3 Tahapan Penelitian	24
BAB IV PEMBAHASAN.....	27
4.1 Tahap Preprocessing	27
4.2 Klasifikasi.....	28
4.3 Evaluasi.....	49
4.4 Integrasi Keislaman.....	53
BAB V PENUTUP.....	55
5.1 Kesimpulan.....	55
5.2 Saran.....	55
DAFTAR PUSTAKA	56
LAMPIRAN.....	60
RIWAYAT HIDUP	69

DAFTAR TABEL

Tabel 1.1 Tabel <i>Confusion matrix</i>	14
Tabel 4.1 Hasil <i>K-fold cross validation</i>	51

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi <i>K-Fold Cross Validation</i>	17
--	----

DAFTAR LAMPIRAN

Lampiran 1 <i>Source code</i>	61
--	----

ABSTRAK

Salsabilah, Jihan. 2025. **Implementasi Algoritma Naive Bayes Terhadap Klasifikasi Kepuasan Penumpang Pesawat**. Skripsi. Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Juhari, M.Si (II) Erna Herawati, M.Pd

Kata kunci: *Naive Bayes*, kepuasan penumpang, maskapai penerbangan, klasifikasi, kualitas layanan, *machine learning*, *confusion matrix*, *k-fold cross validation*.

Transportasi udara merupakan moda transportasi yang banyak digunakan sehingga layanan menjadi faktor penting dalam menjaga kepuasan penumpang. Dengan meningkatnya jumlah penumpang setiap tahun, maskapai perlu memahami faktor-faktor yang memengaruhi kepuasan pelanggan. Penelitian ini bertujuan untuk menerapkan algoritma Naive Bayes dalam pengklasifikasian tingkat kepuasan penumpang pesawat serta mengevaluasi tingkat akurasi. Data yang digunakan adalah *dataset "Airline Passenger Satisfaction"* dari Kaggle yang berjumlah 103.904 data penumpang dengan 23 atribut terkait karakteristik dan pengalaman layanan penerbangan. Penelitian ini menggunakan pendekatan kuantitatif melalui tahapan *preprocessing*, perhitungan probabilitas prior, estimasi *likelihood* menggunakan fungsi densitas Gauss, dan proses klasifikasi dengan algoritma Naive Bayes. Evaluasi model dilakukan menggunakan *confusion matrix* dan *k-fold cross validation*. Hasil penelitian menunjukkan bahwa model menghasilkan akurasi sebesar 86% pada *confusion matrix* dan rata-rata akurasi 86.42% pada *k-fold cross validation*, yang mengindikasikan bahwa algoritma Naive Bayes memiliki performa yang baik dalam memprediksi kepuasan penumpang. Model ini diharapkan dapat memberikan kontribusi bagi pihak maskapai dalam mengidentifikasi faktor layanan yang berpengaruh serta mendukung pengambilan keputusan untuk meningkatkan kualitas pelayanan.

ABSTRACT

Salsabilah, Jihan. 2025. **Implementation of the Naive Bayes Algorithm on the Classification of Airline Passenger Satisfaction.** Thesis. Mathematics Study Program, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Supervisor: (I) Juhari, M.Si (II) Erna Herawati, M.Pd

Keywords: Naive Bayes, passenger satisfaction, airline passengers, classification, service quality, machine learning, *confusion matrix*, k-fold cross validation.

Air transportation plays a significant role in modern mobility, making service quality a crucial factor in maintaining passenger satisfaction. As the number of airline passengers continues to increase each year, airlines must understand the factors that influence customer satisfaction. This study aims to apply the Naive Bayes algorithm to classify airline passenger satisfaction and to evaluate its predictive accuracy. The dataset used in this research is the “Airline Passenger Satisfaction” dataset from Kaggle, consisting of 103.904 passenger records with 23 attributes related to demographic characteristics and flight service experiences. This study employs a quantitative approach involving data preprocessing, calculation of prior probabilities, *likelihood* estimation using the Gaussian density function, and classification using the Naive Bayes algorithm. Model performance was evaluated using a *confusion matrix* and k-fold cross validation. The results indicate that the model achieved an accuracy of 86% based on the *confusion matrix* and an average accuracy of 86.42% using k-fold cross validation, demonstrating that the Naive Bayes algorithm exhibits strong predictive capability in classifying passenger satisfaction. This model is expected to provide valuable insights for airline companies in identifying influential service factors and supporting strategic decision making to enhance service quality.

ستخلص البحث

سلسيلة، جيهان ٢٠٢٥. تطبيق خوارزمية نايف بايز (Naive Bayes) لتصنيف رضا ركاب الطائرات. بحث جامعي. قسم الرياضيات، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف: (١) جوهري، الماجستير. (٢) إيرنا هيراواتي، الماجستير.

الكلمات المفتاحية: خوارزمية نايف بايز، رضا الركاب، ركاب الطيران، التصنيف، جودة الخدمة، تعلم الآلة، مصفوفة الارتباك التحقق (k-fold).
تُعدّ وسائل النقل الجوي من أهم وسائل التنقل في العصر الحديث، الأمر الذي يجعل جودة الخدمات عاملاً حاسماً في الحفاظ على رضا الركاب. ومع الزيادة المستمرة في عدد المسافرين جواً كل عام، تبرز الحاجة إلى توظيف أساليب تحليل البيانات لفهم العوامل المؤثرة في تصنيف رضا ركاب الطائرات وتقييم (Naive Bayes) في رضا العملاء. تهدف هذه الدراسة إلى تطبيق خوارزمية نايف بايز المتاحة على منصة *Airline Passenger Satisfaction* كفاءتها التنبؤية. وقد استخدمت الدراسة مجموعة البيانات والتي تتضمن 103904 سجلات للركاب و 23 خاصية مرتبطة بالصفات الديموغرافية وتجربة السفر الجوي. تعتمد هذه، Kaggle تقدير احتمالات، (Prior Probability) الدراسة منهجاً تجريبياً كمياً يشمل معالجة البيانات، حساب احتمالات القبلية الترجيح باستخدام دالة الكثافة الغاوسية، ثم إجراء عملية التصنيف بواسطة خوارزمية نايف بايز. وتم تقييم أداء النموذج باستخدام، أظهرت النتائج أن النموذج حقق دقة بلغت 86%. (k-fold cross validation) مصفوفة الارتباك وتقنية التحقق المتقاطع. بينما بلغت الدقة المتوسطة باستخدام التحقق المتقاطع 86.42%، مما يدل على كفاءة خوارزمية نايف بايز في التنبؤ برضا الركاب وتساهم هذه النتائج في دعم شركات الطيران عبر تحديد العوامل الأكثر تأثيراً في رضا العملاء، بما يمكنها من اتخاذ قرارات استراتيجية لتحسين جودة الخدمات المقدمة.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Transportasi udara saat ini menjadi salah satu transportasi yang banyak digunakan oleh masyarakat pada saat ini. Menurut BPS, jumlah penumpang pesawat internasional pada bulan Januari hingga Februari 2024 meningkat sebesar 34.17% dan 2.06% untuk penerbangan domestik dibanding dengan kondisi pada tahun 2023 (BPS, 2024). Hal ini menyebabkan persaingan antar maskapai semakin meningkat. Kepuasan penumpang pesawat merupakan hal yang penting bagi industri penerbangan untuk pengembangan fasilitas yang dimiliki oleh maskapai.

Pertumbuhan data yang berkembang pesat menciptakan adanya kondisi di mana terdapat banyak data namun hanya didapat sedikit informasi, maka diperlukan adanya proses klasifikasi untuk pengolahan data sehingga memperoleh informasi. Hal ini menjadi fokus utama dari penelitian ini. Data kepuasan penumpang pesawat perlu untuk dikaji lebih lanjut dengan menggunakan kaidah *machine learning*. Pada penelitian ini, data bersumber pada kaggle yang berisi tentang survey kepuasan setiap pelanggan. Data ini akan diolah menggunakan teknik klasifikasi yaitu naïve bayes kemudian diklasifikasi menjadi dua kategori, yaitu “*satisfaction*” dan “*neutral or dissatisfaction*”.

Algoritma naïve bayes digunakan untuk mengkategorikan kepuasan dari penumpang pesawat berdasarkan atribut-atribut yang dimiliki. Kemudian algoritma naïve bayes digunakan untuk menghitung probabilitas posterior dari setiap kategori kepuasan berdasarkan fitur-fitur yang terdapat pada data. Probabilitas posterior

digunakan untuk mengklasifikasikan survey dari penumpang ke dalam salah satu kepuasan.

Penggunaan algoritma naïve bayes dalam klasifikasi kepuasan penumpang pesawat juga memungkinkan untuk memprediksi tingkat kepuasan di masa mendatang berdasarkan pada data historis dengan menganalisis pola dan tren dalam data kepuasan pelanggan yang terkumpul. Serta maskapai penerbangan dapat mengidentifikasi faktor-faktor yang berpengaruh terhadap kepuasan penumpang dan mengambil langkah yang tepat untuk meningkatkan kepuasan penumpang.

Pada hadits yang diriwayatkan oleh Tirmidzi mengenai kejujuran dalam berbisnis dari Abu Sa'id dari Rasulullah SAW, beliau bersabda (Hasan, 2017):

عَنْ أَبِي سَعِيدٍ عَنِ النَّبِيِّ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ قَالَ التَّاجِرُ الصَّدُوقُ الْأَمِينُ مَعَ النَّبِيِّ وَالصِّدِّيقِ وَالشُّهَدَاءِ (رواه الترمذی)

“Seorang pedagang yang jujur dan dipercaya akan bersama dengan para Nabi, shiddiqun, dan pada syuhada.” (HR. Tirmidzi)

Hadist ini mengajarkan etika untuk jujur dalam berbisnis. Jujur dalam hal ini berkaitan dengan kapasitas seseorang dalam menentukan keputusan berdasarkan pertimbangan dan pemahaman yang mendalam. Hal tersebut sangat penting atau perlu ditekankan karena berbisnis bukan hanya tentang meraup keuntungan bagi pebisnis itu sendiri. Lebih dari itu, seorang pebisnis juga harus mempertimbangkan kepentingan dari pelanggan. Maka dari itu, pebisnis perlu untuk memahami klasifikasi kepuasan pelanggan pesawat guna meningkat kebaikan bagi kedua belah pihak.

Penelitian Głębockaa dan Zdrodows (Głębocka dkk., 2021) menyatakan bahwa penggunaan algoritma Naïve Bayes terbukti menjadi pengklasifikasi paling akurat pada kasus analisis masalah perawatan diri anak penyandang disabilitas.

Kemudian penelitian Anshori (Anshori, 2020) juga menyatakan bahwa algoritma naïve bayes ketika dibandingkan dengan algoritma KNN, algoritma naïve bayes lebih unggul dengan rata-rata akurasi sebesar 76,81%. Selanjutnya penelitian oleh Hariguna, Baihaqi, dan Nurwanti (Hariguna dkk., 2019) menggambarkan implementasi algoritma Naïve Bayes untuk melakukan analisis sentimen pada ulasan pelanggan. Penelitian ini menggunakan dataset berisi ulasan pelanggan yang dikategorikan sebagai positif dan negatif. Dengan menggunakan algoritma Naïve Bayes, penelitian ini berhasil mengklasifikasikan sentimen ulasan pelanggan dengan akurasi yang tinggi. Maka dari itu pada penelitian ini akan digunakan metode Naïve Bayes untuk melakukan klasifikasi terhadap kepuasan pelanggan pesawat.

1.2 Rumusan Masalah

1. Bagaimana hasil penerapan algoritma Naïve Bayes pada klasifikasi kepuasan pelanggan maskapai pesawat?
2. Berapa akurasi dari klasifikasi Naïve Bayes dalam memprediksi kepuasan pelanggan maskapai?

1.3 Tujuan Penelitian

1. Mengetahui hasil penerapan algoritma Naïve Bayes pada klasifikasi kepuasan pelanggan maskapai pesawat
2. Mengetahui berapa akurasi algoritma Naïve Bayes untuk memprediksi kepuasan pelanggan maskapai.

1.4 Manfaat Penelitian

1. Manfaat Teoritis

- a. Mampu mengetahui proses klasifikasi menggunakan algoritma Naïve Bayes terhadap klasifikasi kepuasan pelanggan pesawat
- b. Mampu merancang model klasifikasi menggunakan algoritma Naïve Bayes terhadap klasifikasi kepuasan pelanggan pesawat

2. Manfaat Praktis

Memberikan informasi dan referensi bagi pembaca terkait klasifikasi menggunakan algoritma naïve bayes terhadap kepuasan pelanggan pesawat

1.5 Batasan Masalah

1. Diasumsikan setiap variabel prediktor tidak saling berkaitan
2. Penelitian tidak membahas faktor eksternal seperti kebijakan maskapai atau aspek operasional lainnya yang tidak terdapat dalam dataset
3. Label *neutral or dissatisfaction* berada dalam satu kelompok

BAB II KAJIAN TEORI

2.1 Data Mining

Data mining merupakan proses menganalisa data yang berasal dari sudut pandang yang berbeda kemudian menyimpulkannya menjadi informasi penting yang dapat digunakan untuk menambah utilitas atau laba, meminimalisir biaya pengeluaran, atau bahkan keduanya. Secara teknis, data mining diklaim sebagai proses untuk menemukan hubungan atau pola asal ratusan atau ribuan field yang berasal dari relasional database yang besar (Lubis dkk., 2016)

Data mining juga dapat didefinisikan sebagai proses menggunakan satu atau lebih metode pembelajaran komputer atau *machine learning* untuk secara otomatis menganalisis dan mengekstraksi informasi.

Tahapan data mining dibagi menjadi enam bagian, yaitu :

1. Pembersihan data (*data cleaning*)

Sebelum proses data mining dapat dilaksanakan, data yang menjadi fokus KDD (*knowledge discovery in database*) harus dibersihkan. Proses *cleaning* mencakup antara lain menghilangkan data ganda, memeriksa data yang tidak konsisten, memeriksa data yang hilang, dan memperbaiki data yang salah seperti kesalahan cetak (tipografi). Jika terdapat permasalahan tersebut, maka data dapat diperbaiki dengan berbagai macam cara, contohnya mengubah nilai yang hilang dengan mean atau median, menggunakan *oversampling* atau *undersampling* pada data yang tidak seimbang, dan lain-lain.

2. Integrasi data (*data integration*)

Menggabungkan data dari berbagai *database* kemudian membentuk suatu *database* baru disebut integrasi data. Seringkali data yang dibutuhkan pada *data mining* tidak hanya bersumber dari satu *database*, namun juga bersumber dari banyak *database* atau file teks. Dalam melakukan identifikasi terhadap entitas-entitas unik seperti id pelanggan, alamat rumah, nomor ponsel, dan lain-lain diperlukan adanya integrasi data. Jika terjadi kesalahan pada tahap integrasi data, keputusan yang diambil dapat menjadi bencana atau menjebak. Maka, integrasi data perlu dilakukan secara detail dan teliti. Sebagai contoh jika integrasi data didasarkan pada kategori produk yang ternyata berdasarkan jenis produk maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada.

3. Seleksi data (*data selection*)

Tidak semua data yang ada pada *database* harus digunakan. Data harus dipilah terlebih dahulu sehingga hanya data yang sesuai untuk dianalisis yang akan digunakan. Contohnya dalam sebuah kasus yang meneliti faktor penyakit pada pasien pada *database* pasien, tidak perlu mengambil nama pasien, cukup dengan id pasien saja.

4. Transformasi data (*data transformation*)

Transformasi data merupakan pengubahan atau pengolahan data dari suatu struktur atau rupa ke format atau rupa lainnya. Transformasi data dilakukan untuk menambah kualitas data, penyesuaian terhadap analisis, dan memenuhi persyaratan atau dugaan tertentu dari algoritma yang digunakan. Contohnya dalam mengubah label “warna” dengan kategori “hitam”, “putih”, “abu-abu” menjadi tiga kolom terpisah.

5. Evaluasi pola (*pattern evaluation*)

Evaluasi pola atau *pattern evaluation* dilakukan untuk menemukan pola-pola yang menarik kedalam hasil *knowledge base*. Pada tahap ini, hasil *data mining* termasuk pola-pola yang memiliki ciri khas dan model prediksi dievaluasi untuk mengetahui apakah prediksi sebelumnya telah dipenuhi atau tidak.

6. Presentasi pengetahuan (*knowledge presentation*)

Presentasi pengetahuan adalah pemaparan dan visualisasi informasi tentang cara mendapatkan informasi untuk pengguna. Memformulasikan keputusan atau presentasi keputusan adalah tahap terakhir dari proses *data mining*. Tahap proses *data mining* yang terakhir adalah menyampaikan hasil *data mining* dalam bentuk yang dapat dipahami oleh semua orang, karena hal ini seringkali melibatkan orang yang tidak tahu atau tidak mengerti tentang *data mining*. Visualisasi dapat membantu dalam mengkomunikasikan hasil dari *data mining* pada orang awam.

2.2 Klasifikasi

Klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari jumlah kelas yang tersedia (Utomo dkk., 2020). Model yang dibangun berdasarkan *data train* atau data latih yang tersedia digunakan untuk mengklasifikasikan data yang baru. Klasifikasi dapat didefinisikan sebagai pekerjaan yang melakukan pelatihan atau pembelajaran terhadap fungsi target yang memetakan setiap set atribut (fitur) ke satu jumlah label kelas yang tersedia (Sanjaya dkk., 2018).

Salah satu teknik klasifikasi adalah Naïve Bayes Classifier (NBC). Naïve Bayes Classifier (NBC) merupakan salah satu metode klasifikasi dan pengklasifikasi statistik yang dapat memprediksi peluang untuk menjadi anggota kelas (Pebdika dkk., 2023). Tujuan dari klasifikasi adalah untuk mengantisipasi kelas yang dihasilkan dari item yang tidak diketahui. Menggunakan model yang telah dibangun, tahap klasifikasi dapat memprediksi mengenai label kelas pada data kemudian dapat menghasilkan hasil akhir berupa akurasi model yang telah dihasilkan.

Proses klasifikasi didasarkan pada empat komponen dasar (Nasrullah, 2021), yaitu:

1. Kelas merupakan variabel terikat dalam bentuk *categorical variabel model* yang merepresentasikan label pada objek setelah proses klasifikasi. Contohnya jenis kelamin, umur, dan lain-lain;
2. *Predictor* merupakan variabel bebas yang ditunjukkan berdasarkan karakteristik (atribut) data yang akan diklasifikasikan dan didasarkan saat klasifikasi yang telah dibuat. Contohnya seperti kepuasan pelanggan, arah dan kecepatan angin, daerah gempa, dan lain-lain;
3. *Training dataset* adalah seperangkat data yang memuat nilai-nilai dari kelas dan *predictor* yang berguna untuk mengklasifikasikan kelas yang sesuai menurut prediksi yang ada. Contohnya adalah *database* kepuasan pelanggan, kelompok pasien rumah sakit, *database* penelitian gempa, dan lain-lain;
4. *Testing dataset* merupakan sekelompok *unseen data* yang dimanfaatkan untuk proses menguji pada klasifikasi terhadap model yang dihasilkan

sebelumnya serta memungkinkan untuk melakukan proses penilaian akurasi klasifikasi terhadap hasil dari *testing dataset*.

2.3 Algoritma Naïve Bayes

Algoritma Naïve Bayes merupakan sebuah metode klasifikasi menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes (Macfud dkk., 2023). Algoritma naïve bayes dikenal sebagai teorema bayes, di mana teorema ini meramal peluang di masa depan berdasarkan peristiwa di masa lalu. Klasifikasi naïve bayes memiliki ciri utama terhadap prediksi yang sangat kuat akan independensi setiap kondisi atau peristiwa.

Cara kerja metode Naïve Bayes adalah menggunakan probabilitas atau peluang. Konsep dasar yang digunakan oleh Naïve bayes adalah Teorema Bayes, yaitu mengklasifikasi dengan mempertimbangkan nilai probabilitas $P(C|D)$, yaitu peluang kelas C jika diketahui dokumen D. Teorema Bayes didasarkan pada gagasan bahwa pengetahuan tentang probabilitas sebelumnya (prior) dapat mempengaruhi probabilitas setelahnya (posterior). Teorema ini juga dapat didefinisikan sebagai teorema yang melakukan prediksi bahwa kemungkinan di masa depan dapat terjadi berdasarkan informasi atau petunjuk baru yang ditemukan pada masa sebelumnya (Xiang & Sean, 2016)

Naïve Bayes memprediksi ada tidaknya keberadaan suatu fitur dalam suatu kelas yang memiliki keterkaitan dengan fitur yang lain. Contohnya, sesuatu yang berwarna hijau, bulat serta memiliki diameter sekitar 5 cm bisa dikategorikan sebagai buah semangka, meskipun fitur ini saling bergantung, Naïve Bayes tetap menganggap bahwa fitur tersebut tidak berpengaruh dengan lainnya.

Sebelum mengimplementasikan algoritma naïve bayes pada dataset, dataset harus dipastikan sudah diberi label untuk pelatihan model. Data ini harus terdiri fitur-fitur yang relevan dan kelas-kelas yang ingin diprediksi. Kemudian dilakukan proses *preprocessing*, tujuannya agar dapat digunakan secara efektif oleh model atau algoritma yang digunakan. Langkah yang biasanya dilakukan saat *preprocessing* yaitu *data imbalance*, *data cleaning*, dan *missing value*. Kemudian dataset dibagi menjadi dua bagian, yaitu *data train* (data uji) dan *data testing* (data test). Kemudian pada tahap klasifikasi, algoritma naïve bayes dapat diimplementasikan.

Tahapan dari proses algoritma Naïve Bayes adalah :

1. Menghitung probabilitas prior

Probabilitas prior adalah estimasi awal dari suatu peristiwa atau kondisi sebelum mempertimbangkan bukti atau informasi tambahan. Probabilitas prior menggambarkan pengetahuan awal tentang suatu kejadian. Probabilitas prior dihitung dengan cara

$$P(X) = \frac{n(X)}{N} \quad (2.1)$$

di mana:

X : Kelas dalam klasifikasi

$n(X)$: Jumlah data X

N : Total seluruh data

2. Menghitung probabilitas pada kelas kategori

$$P(C|X) = \frac{n(C)}{n(X)} \quad (2.2)$$

di mana:

C : Nilai atribut

X : Kelas dalam klasifikasi

$n(C)$: Jumlah data atribut

$n(X)$: Jumlah data kelas

3. Mencari probabilitas *likelihood*

Mencari probabilitas *likelihood* diperlukan beberapa parameter, yaitu:

a. Mean

Mean juga disebut sebagai nilai rata-rata pada suatu data. Persamaan yang digunakan untuk menghitung nilai mean adalah

$$\mu = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} \quad (2.3)$$

dengan:

μ : nilai rata – rata (mean)

x : nilai sample ke – i

n : jumlah sampel

b. Standar Deviasi

Standar deviasi biasa juga disebut dengan simpangan baku. Persamaan yang digunakan untuk menghitung standar deviasi adalah

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n-1}} \quad (2.4)$$

di mana:

σ : Standar deviasi

x_i : Nilai x ke – i

μ : Nilai rata – rata

n : Jumlah sampel

c. Fungsi densitas Gauss

Persamaan yang digunakan untuk menghasilkan densitas Gauss adalah

$$P(c_j|X) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad (2.5)$$

di mana:

c_j : Nilai atribut ke- j

X : Kelas dalam klasifikasi

$e/\text{Exp} = 2.718282$

d. Nilai *likelihood*

Perhitungan menggunakan metode naïve bayes untuk mengetahui hasil

akhir, digunakan persamaan *likelihood*

$$\begin{aligned} P(c_i|X) &= \prod_{j=1}^n P(c_j|X) \\ &= P(c_1|X) \times P(c_2|X) \times \dots \times P(c_n|X) \end{aligned} \quad (2.6)$$

di mana:

c_j : Nilai atribut dalam klasifikasi

X : Kelas dalam klasifikasi

4. Menghitung probabilitas posterior

Probabilitas posterior adalah probabilitas suatu hipotesis setelah melihat atau mempertimbangkan data. Probabilitas posterior dapat dihitung dengan cara

$$P(C|X) = \frac{P(c_j|X)P(X)}{P(C)} \quad (2.7)$$

di mana:

$P(c_j|X)$: Nilai *likelihood*

$P(X)$: Probabilitas prior

$P(C)$: Probabilitas kemunculan data pada seluruh kelas

5. Mendapat nilai maksimum *likelihood*

6. Evaluasi

Evaluasi merupakan tahap yang digunakan untuk mengukur sejauh mana model dapat memberikan prediksi yang akurat dan efektif. Evaluasi sering melibatkan *confusion matrix* yang memberikan gambaran rinci tentang performa model.

2.4 *Confusion matrix*

Confusion matrix merupakan alat evaluasi yang digunakan dalam pemodelan statistik dan *machine learning* untuk mengukur kinerja suatu model klasifikasi (Ainurrohman, 2021). *Confusion matrix* memberikan gambaran rinci tentang sejauh mana model mampu memprediksi kelas dengan benar dan seberapa sering kesalahan prediksi terjadi.

Confusion matrix biasanya digunakan dalam konteks klasifikasi biner, di mana ada dua kelas yang diinginkan untuk diprediksi, yaitu kelas positif dan kelas negatif. Namun, konsep ini dapat diperluas ke dalam masalah klasifikasi dengan lebih dari dua kelas.

Dalam *confusion matrix* terdapat empat komponen utama (Erinsyah dkk., 2023), yaitu:

1. True positive (TP)

True positive meruju pada jumlah kasus di mana model memprediksi dengan benar bahwa suatu observasi adalah positif dan memang benar positif.

2. True negative (TN)

True negative mengacu pada jumlah kasus di mana model memprediksi dengan benar bahwa suatu observasi adalah negatif dan memang benar negatif dalam realitas.

3. False positive (FP)

False positif mengacu pada jumlah kasus di mana model salah memprediksi bahwa suatu observasi adalah positif, padahal sebenarnya negatif.

4. False negative (FN)

False negatif merujuk pada jumlah kasus di mana model salah memprediksi bahwa suatu observasi adalah negatif, padahal sebenarnya positif.

Tabel 1.1 Tabel *Confusion matrix*

	POSITIVE	NEGATIVE
POSITIVE	TP (True Positive)	FP (False Positive)
NEGATIVE	FN (False Negative)	TN (True Negative)

Konsep ini terkait dengan evaluasi kinerja model klasifikasi. Dengan menghitung TP, TN, FP, dan FN maka selanjutnya dapat menghitung berbagai

metrik evaluasi yang memberikan wawasan tentang performa model seperti akurasi, presisi, dan *recall*.

Dalam data mining akurasi, presisi, dan *recall* adalah metrik evaluasi yang digunakan untuk mengukur kinerja model klasifikasi. Setiap metrik ini menunjukkan wawasan atau seberapa baik model klasifikasi dapat meramal label atau kelas dengan benar

1. Akurasi (*accuracy*)

Akurasi digunakan untuk mengukur sejauh mana model klasifikasi dapat memprediksi dengan benar seluruh kasus dalam dataset (Romadloni dkk., 2022). Akurasi merupakan perbandingan antara jumlah prediksi benar (TP dan TN) dengan total jumlah prediksi (TP, TN, FP, dan FN)

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.8)$$

Akurasi memberikan gambaran umum tentang kinerja model, tetapi dapat menjadi tidak akurat jika distribusi kelas tidak seimbang dalam dataset. Akurasi yang tinggi pada dataset yang tidak seimbang mungkin mengarah pada hasil yang buruk.

2. Presisi (*precision*)

Presisi mengukur sejauh mana prediksi positif yang dibuat oleh model benar (Putra, 2023). Presisi mengukur kemampuan model dalam menghindari memberikan hasil positif palsu.

$$Presisi = \frac{TP}{TP+FP} \quad (2.9)$$

Presisi berguna saat ingin memastikan bahwa model memberikan prediksi positif, prediksi tersebut cenderung akurat.

3. *Recall*

Recall mengukur sejauh mana model dapat mengidentifikasi semua kasus positif yang sebenarnya dalam dataset (Sholikhah dkk., 2023). *Recall* mengukur kemampuan model untuk tidak melewatkan observasi yang sebenarnya positif.

$$Recall = \frac{TP}{TP+FN} \quad (2.10)$$

Recall penting saat menghindari *false negatif* (FN) adalah prioritas.

2.5 *Permutation Importance*

Permutation Importance adalah metode yang digunakan untuk mengevaluasi seberapa penting setiap variabel atau fitur dalam model (Ramirez dkk., 2022). Metode ini bekerja dengan cara mengacak nilai-nilai fitur tertentu pada data kemudian melihat perubahan performa model ketika fitur tersebut diacak. Jika fitur atau variabel yang diacak sangat penting, maka nilainya akan menurunkan performa model secara signifikan. Sebaliknya, jika fitur tidak penting, performa model akan tetap stabil meskipun nilai fitur tersebut diacak.

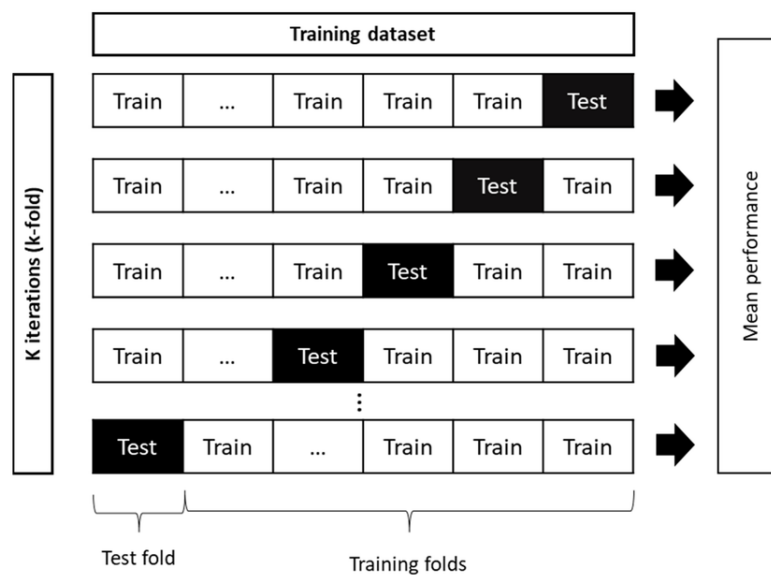
Teknik ini sangat berguna karena tidak bergantung pada jenis model yang digunakan, sehingga dapat diterapkan pada hampir semua model *machine learning*. Sehingga dapat membantu dalam penelitian untuk menentukan fitur mana yang paling relevan atau berpengaruh dalam model yang diterapkan.

2.6 *K-fold Cross Validation*

Cross validation adalah teknik untuk memvalidasi keakuratan sebuah model yang dibangun berdasarkan data set tertentu (Dwiki dkk., 2021). Keakuratan sebuah

model ketika dijalankan dapat diketahui dengan cara ini. Salah satu teknik dari *cross validation* adalah *k-fold cross validation*.

K-fold cross validation merupakan salah satu teknik *cross validation* yang membagi *data set* secara acak kemudian dikelompokkan sebanyak nilai K yang digunakan. Banyaknya nilai K akan ditentukan oleh peneliti. Teknik ini merupakan salah satu cara untuk mengetahui berapa besar persentase keberhasilan suatu sistem dalam melakukan proses perulangan dengan mengacak *data train* dan *data test* sehingga sistem akan lebih teruji. Berikut ilustrasi alur kerja dari *k-fold cross validation*



Gambar 2. 1 Ilustrasi *K-Fold Cross Validation*

Pada proses ini terjadi pengkombinasian antara data yang digunakan sebagai *data train* dan *data test* dengan tujuan untuk mengetahui akurasi dari masing-masing kombinasi yang akan menentukan performa dari sistem prediksi yang dibuat.

2.7 Kualitas Jasa dan Pelayanan

Salah satu unsur yang terpenting dalam mewujudkan kepuasan konsumen adalah pelayanan. Perusahaan sudah seharusnya memberikan pelayanan yang berkualitas agar harapan konsumen tercapai. Kualitas pelayanan merupakan upaya pemenuhan kebutuhan yang diiringi dengan keinginan konsumen serta ketepatan cara penyampainnya agar dapat memenuhi harapan dan kepuasan pelanggan tersebut (Santoso dkk., 2019). Kualitas pelayanan adalah kunci keberhasilan dalam persaingan yang semakin ketat di dunia bisnis. Kualitas pelayanan terbaik adalah ketika semua sumber daya manusia suatu perusahaan berfokus pada kepuasan pelanggan.

2.8 Kepuasan Konsumen

Kepuasan konsumen dapat didefinisikan sebagai evaluasi kesadaran dan kecintaan konsumen terhadap barang atau jasa yang telah disampaikan kepada produsen tertentu (Husna, 2011). Secara sederhana kepuasan dapat didefinisikan sebagai upaya untuk mencapai sesuatu atau membuat sesuatu memadai. Kepuasan dipengaruhi oleh tingkat harapan atas kualitas jasa. Apabila produk atau jasa memenuhi atau melampaui harapan konsumen, konsumen akan merasa puas yang dapat menghasilkan loyalitas.

Kepuasan konsumen yang rendah akan berdampak besar pada bisnis. Pelanggan yang tidak puas cenderung meninggalkan bisnis untuk mencari perusahaan yang lebih baik. Hal ini dapat mengakibatkan berkurangnya atau bahkan hilangnya pendapatan. Kepuasan konsumen merupakan elemen penting dalam strategi terciptanya bisnis yang sehat dan berkelanjutan.

2.9 Kajian Keislaman

Dalam Islam, berbisnis sangat dianjurkan oleh Rasulullah SAW karena Rasulullah SAW merupakan seorang pedagang dan kegiatan ini juga merupakan salah satu kegiatan yang sah untuk mencari nafkah, memenuhi kebutuhan hidup, serta hal lain yang dapat membantu banyak orang. Rasulullah SAW juga mendorong umat Islam agar tidak bergantung pada orang lain, dengan berbisnis seseorang akan menjadi lebih mandiri secara ekonomi dan dapat meningkatkan kualitas hidup.

Memiliki maskapai penerbangan merupakan salah satu contoh kegiatan dalam berbisnis. Meningkatkan kepuasan pelanggan merupakan salah satu cara penerapan perilaku atau etika yang baik dalam berbisnis. Dalam Al-Qur'an pada surah Al-Baqarah ayat 188 dikatakan bahwa (Kemenag, 2019):

وَلَا تَأْكُلُوا أَمْوَالَكُمْ بَيْنَكُمْ بِالْبَاطِلِ وَتُدْنُوا بِهَا إِلَى الْحُكَّامِ لِتَأْكُلُوا فَرِيقًا مِّنْ أَمْوَالِ النَّاسِ بِالْإِثْمِ وَأَنْتُمْ تَعْلَمُونَ
(البقرة ٢ : ١٨٨)

“Dan janganlah kamu memakan harta orang lain di antara kamu dengan jalan yang bathil dan janganlah kamu membawa urusan harta itu kepada hakim, supaya kamu memakan sebagian harta orang lain itu dengan yang batil sedangkan kamu mengetahui.” (Q.S. Al-Baqarah, 2:188)

Dalam tafsir Jalalain, Ibnu Abi Hatim dan Sa'id bin Jubair menyatakan bahwa asbab An-nuzul dari ayat ini adalah seperti yang dikemukakan oleh Ibnu Abi Hatim: "Umru-ul Qais bin 'Abis dan Abdan bin Asywa Al-Hadrami terlibat dalam salah satu pertikaian mengenai tanah mereka, sehingga Umru-ul Qais hendak mengucapkan sumpahnya dalam hal itu." Maka turunlah ayat "Dan janganlah sebahagian kamu memakan harta sebahagian yang lain di antara kamu turun padanya."

Ayat ini mencakup segala sesuatu yang bertentangan dengan syariah Islam, seperti mengambil harta orang lain, memanipulasi dalam perdagangan, melakukan praktek riba, dan hal-hal lain yang dilarang oleh Islam. Ayat ini memberikan arahan kepada umat muslim untuk berlaku jujur dan adil. Termasuk dalam menjalankan bisnis, jujur dan adil merupakan sifat yang sudah seharusnya dimiliki oleh pebisnis. Contohnya dalam fasilitas yang harus diberikan untuk mendapatkan kepuasan pelanggan. Pemilik maskapai harus memberikan fasilitas yang tepat atau sesuai dengan apa yang sudah dibayar oleh pelanggan. Jika pemilik maskapai tidak memberikan fasilitas yang sesuai dengan apa yang sudah dibayar, maka pemilik maskapai sudah memakan harta orang lain yaitu pelanggan. Perbuatan ini sangat tidak disukai oleh Allah SWT. Fasilitas ini harus diberikan secara adil dan jujur agar tidak terjadi kerugian antara kedua belah pihak baik pemilik maskapai pesawat maupun pelanggan.

Pada hadits riwayat Thabrani dalam al-Mu'jam al-Awsat No. 897 Nabi SAW bersabda (Maududin dkk., 2021):

عَنْ عَائِشَةَ، أَنَّ رَسُولَ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ قَالَ: إِنْ اللَّهُ عَزَّ وَجَلَّ يُحِبُّ إِذَا عَمِلَ أَحَدُكُمْ عَمَلًا أَنْ يُتَّقِنَهُ
(رواه الطبران)

“Sesungguhnya Allah sangat mencintai orang yang jika melakukan suatu pekerjaan, dilakukan dengan itqan (tepat, terarah, jelas, dan tuntas) .” (HR. Thabrani)

Hadits ini juga mengajarkan kita sebagai umat muslim untuk melakukan segala pekerjaan dengan terencana dan terprogram dengan baik. Mengklasifikasi kepuasan pelanggan menggunakan algoritma naïve bayes merupakan salah satu cara bagi pebisnis untuk mengetahui apa rencana dan program yang dapat dilakukan

agar pelanggan semakin merasa puas dan tidak dirugikan. Hal ini dapat bermanfaat bagi kedua belah pihak, baik pebisnis maupun pelanggan.

Kemudian pada surah Az-Zalzalah ayat 7 sampai 8 dikatakan bahwa (Kemenag, 2019):

فَمَنْ يَعْمَلْ مِثْقَالَ ذَرَّةٍ خَيْرًا يَرَهُ (٧) وَمَنْ يَعْمَلْ مِثْقَالَ ذَرَّةٍ شَرًّا يَرَهُ (الزلزلة ٩٩: ٧-٨)

“Siapa yang mengerjakan kebaikan seberat zarrah, dia akan melihat (balasan)-nya (7) Siapa yang mengerjakan kejahatan seberat zarrah, dia akan melihat (balasan)-nya (8)” (Q.S Az-Zalzalah 99:7-8)

Ayat ini mengajarkan umat muslim untuk memperhatikan hal-hal kecil dalam berbisnis. Dalam hal ini, fasilitas sekecil apapun yang bisa diberikan harus bisa didapatkan oleh pelanggan karena semuanya akan dipertanggung jawabkan atau dihisab. Contohnya dalam memberikan pelayanan terhadap pelanggan, hendaklah pegawai murah senyum dan mengucapkan perkataan yang baik tanpa menyinggung perasaan pelanggan.

Kemudian pada surah Al-Furqan ayat 67 dikatakan bahwa (Kemenag, 2019):

وَالَّذِينَ إِذَا أَنْفَقُوا لَمْ يُسْرِفُوا وَلَمْ يَقْتُرُوا وَكَانَ بَيْنَ ذَلِكَ قَوَامًا (الفرقان ٦٧)

“Dan, orang-orang yang apabila berinfak tidak berlebihan dan tidak (pula) kikir. (Infak mereka) adalah pertengahan antara keduanya.”

Ayat ini mengajarkan umat muslim untuk memberikan sesuatu dengan tepat, tidak kekurangan ataupun berlebihan. Dalam berbisnis, umat muslim dihimbau untuk menjaga keadilan dan keseimbangan. Tidak boleh terlalu boros atau royal dalam memberikan fasilitas, namun juga tidak boleh terlalu kikir atau pelit yang dapat merugikan orang lain maupun diri sendiri. Sehingga diperlukan adanya survei terhadap kepuasan pelanggan kemudian mengklasifikasi

menggunakan algoritma naïve bayes agar tercipta fasilitas yang tepat tanpa merugikan siapapun.

BAB III METODE PENELITIAN

3.1 Jenis Penelitian

Penelitian ini merupakan penelitian eksperimen yang menggunakan pendekatan kuantitatif yang harus memenuhi persyaratan dalam proses pengujian sebab dan akibat. Penggunaan metode ini juga sejalan dengan rumusan masalah yaitu mengetahui akurasi dari klasifikasi naïve bayes dalam memprediksi kepuasan pelanggan.

3.2 Data dan Sumber Data

Data yang digunakan adalah data set yang berasal dari Kaggle Repository tahun 2020 (KLEIN, 2020). Data set berisi survei kepuasan penumpang maskapai penerbangan internasional sebanyak 103904 data penumpang. Data ini memuat 23 atribut terkait karakteristik penumpang dan pengalaman mereka selama menggunakan jasa penerbangan. Atribut pertama adalah *gender* yang menunjukkan jenis kelamin penumpang, serta *customer type* yang membedakan antara pelanggan loyal dan tidak loyal. Atribut *age* atau usia penumpang yang dapat memengaruhi ekspektasi terhadap layanan. Selain itu, terdapat atribut *type of travel* yang membedakan tujuan perjalanan bisnis atau pribadi, serta *class* yang menunjukkan kelas penerbangan (*economy*, *economy plus*, dan *business*). Atribut *flight distance* menggambarkan jarak tempuh penerbangan dalam kilometer, sedangkan *departure delay in minutes* dan *arrival delay in minutes* mengukur keterlambatan keberangkatan maupun kedatangan yang secara langsung memengaruhi tingkat kepuasan.

Di samping itu, dataset ini juga mencakup sejumlah atribut yang berkaitan dengan kualitas layanan maskapai. Atribut tersebut dinilai menggunakan skala 0 sampai 5 oleh penumpang, meliputi *inflight wifi service*, *departure/arrival time convenient*, *ease of online booking*, dan *gate location*. Aspek kenyamanan penerbangan dicerminkan melalui atribut *food and drink*, *seat comfort*, *leg room service*, serta *inflight entertainment*. Faktor lain yang tidak kalah penting adalah online boarding dan check-in service yang berhubungan dengan kemudahan proses administrasi. Dari sisi pelayanan awak kabin, terdapat atribut *on-board service* dan *inflight service* yang menggambarkan kualitas interaksi serta pelayanan selama penerbangan. Kemudian atribut *baggage handling* dan *cleanliness* memberikan informasi mengenai penanganan bagasi serta kebersihan kabin yang turut memengaruhi persepsi penumpang terhadap maskapai.

Keseluruhan atribut digunakan sebagai variabel independen dalam penelitian, sementara variabel dependet atau target adalah *satisfaction*, yaitu label kepuasan penumpang yang terbagi dalam dua kategori, yaitu *satisfied* (puas) dan *neutral or dissatisfied* (netral atau tidak puas).

3.3 Tahapan Penelitian

Tahapan penelitian yang dilakukan dalam mengklasifikasi kepuasan penumpang pesawat dengan menggunakan algoritma naïve bayes adalah sebagai berikut:

1. Mengunduh dan mengamati dataset *Airline Passenger Satisfaction* yang berasal dari website Kaggle. Data ini harus terdiri dari fitur-fitur yang relevan dengan kelas yang ingin diklasifikasi.

2. Penerapan Algoritma Naïve Bayes

a. Preprocessing

Preprocessing yang dilakukan pada dataset kepuasan pelanggan pesawat adalah melakukan pemeriksaan data, yaitu pemeriksaan *data imbalance* dan *missing value*. Pemeriksaan *data imbalance* adalah proses untuk mengevaluasi sejauh mana distribusi kelas atau label target dalam dataset bersifat seimbang atau tidak seimbang. Jika data yang dihasilkan tidak seimbang atau *imbalance* maka dapat mempengaruhi kinerja model klasifikasi. Data yang tidak seimbang atau imbalance akan diatasi dengan cara *oversampling* atau *undersampling*. *Oversampling* merupakan penambahan lebih banyak data pada kelas mayoritas. *Undersampling* merupakan pengurangan atau mengeliminasi data pada kelas mayoritas. *Missing value* merupakan proses untuk mengidentifikasi apakah terdapat nilai yang hilang atau tidak lengkap dalam dataset. Jika terdapat data yang hilang atau tidak lengkap juga dapat mempengaruhi kinerja model klasifikasi. Maka, jika terdapat *missing value* atau data yang hilang dilakukan pengisian nilai data yang hilang dengan nilai rata-rata atau mean yang terdapat pada kolom.

b. Klasifikasi

Dataset yang dihasilkan pada tahap *preprocessing* kemudian diklasifikasi kepuasan pelanggan antara *satisfied* dan *neutral or dissatisfied* menggunakan algoritma naïve bayes. Proses klasifikasi dilakukan dengan menghitung probabilitas prior terlebih dahulu,

kemudian menghitung probabilitas kelas pada tiap kategori. Setelah didapat probabilitas prior dan probabilitas kelas kategori, kemudian menghitung nilai *mean* dan *standar deviasi* pada data dengan kategori numerik yang akan digunakan untuk perhitungan naïve bayes dengan fungsi densitas Gauss. Kemudian menghitung probabilitas posterior dan nilai *likelihood* maksimal sehingga didapatkan suatu keputusan *satisfied* atau *neutral or dissatisfied*.

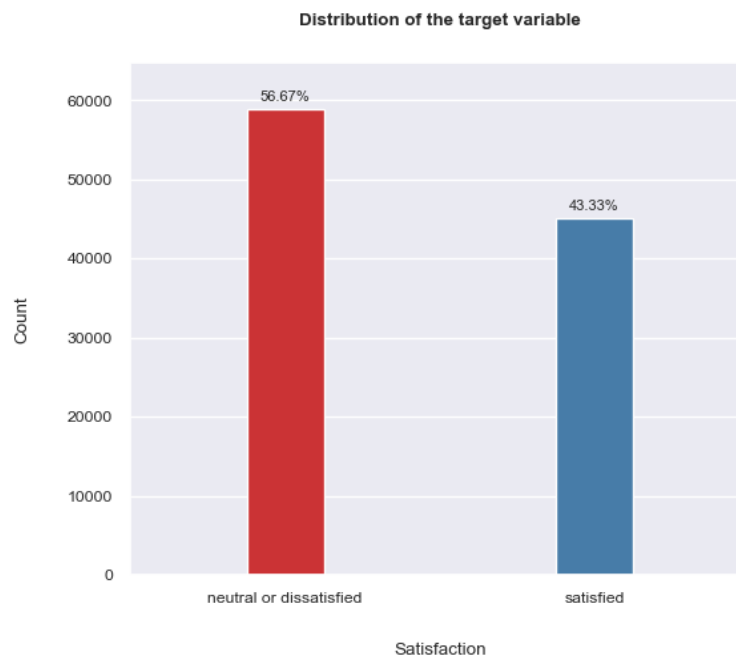
3. Evaluasi Performa

Akurasi merupakan bagian dari evaluasi dalam tahap penerapan algoritma naïve bayes pada klasifikasi kepuasan pelanggan. Selain mengetahui berapa besar akurasi yang dihasilkan, pada tahap evaluasi dilakukan analisis *presisi* dan *recall* pada data model klasifikasi kepuasan pelanggan pesawat menggunakan algoritma naïve bayes. Kemudian evaluasi juga menggunakan *k-fold cross validation* agar semua data mendapat peran sebagai *data train* dan *data test* sehingga akurasi lebih valid dan teliti. Hasil akurasi tersebut dapat menentukan apakah model klasifikasi kepuasan pelanggan pesawat sudah baik atau belum. Kemudian akurasi dievaluasi menggunakan *permutation importance* untuk menentukan fitur atau variabel yang terpenting pada model.

BAB IV PEMBAHASAN

4.1 Tahap Preprocessing

Tahap preprocessing merupakan tahap penting yang harus dilakukan sebelum mengklasifikasi data. Langkah pertama yang dilakukan dalam tahap *preprocessing* adalah pengecekan *data imbalance*. Jika data yang dihasilkan tidak seimbang atau *imbalanced*, maka algoritma tidak dapat memaksimalkan akurasi dan banyak kesalahan yang akan terjadi. Algoritma akan cenderung fokus pada nilai variabel tertinggi dan mengabaikan variabel yang rendah. Suatu dataset dikatakan tidak seimbang apabila kelas minoritas memiliki proporsi kurang dari 20% dari total data. Pada data penelitian ini hasil yang didapat adalah data yang seimbang antara *neutral or dissatisfied* dan *satisfied*. *Neutral or dissatisfied* sebanyak 56.67% dan *satisfied* sebanyak 43.33%. Sehingga algoritma akan dapat bekerja dengan baik dalam dataset ini tanpa mengabaikan variabel manapun.



Gambar 4. 1 *Data Distribution*

Langkah selanjutnya yaitu melakukan *data cleaning* pada dataset. *Data cleaning* dilakukan untuk membersihkan data dari *noise*, duplikasi, menghapus, hingga melengkapi data yang hilang sehingga memudahkan model untuk mengklasifikasi data.

Pada dataset ini terdapat *missing value* pada variabel “*Arrival Delay in Minutes*” yaitu sebanyak 310 data. Nilai yang hilang atau *missing value* yang tidak diatasi akan mengakibatkan kesalahan atau kegagalan dalam hasil analisis serta menghasilkan kesimpulan yang tidak sesuai dengan yang diharapkan. *Missing value* pada penelitian ini diatasi dengan menggunakan nilai rata-rata atau *mean* yang dihasilkan dari kolom “*Arrival Delay in Minutes*”. Data yang hilang pada atribut *Arrival Delay in Minutes* menjadi nol atau tidak terdapat data yang hilang sehingga diharapkan mendapatkan hasil kesimpulan yang sesuai dengan yang diharapkan.

4.2 Klasifikasi

Klasifikasi merupakan proses yang dilakukan untuk menentukan apakah penumpang pesawat merasa *satisfied* (puas) atau *neutral or dissatisfied* (netral atau tidak puas). Metode yang digunakan dalam proses klasifikasi adalah algoritma naïve bayes. Kemudian dataset dibagi menjadi *data training* dan *data testing*. Pada penelitian ini dibagi menjadi 75% *data training* dan 25% *data testing*. Pembagian ini dipilih karena proporsi tersebut dianggap mampu memberikan keseimbangan antara kebutuhan model untuk belajar dari data dalam jumlah yang besar serta kebutuhan evaluasi yang representatif. Dengan porsi 75%, model memperoleh cakupan *data training* yang memadai sehingga mampu

mengenali pola dan hubungan antar variabel secara lebih stabil. Sementara itu, alokasi 25% sebagai *data testing* tetap cukup besar untuk memberikan penilaian yang objektif mengenai performa model dalam mengklasifikasi data yang belum pernah dilihat sebelumnya. Sehingga diperoleh *data training* sebanyak 77928 data dan *data testing* sebanyak 25976 data.

Data training digunakan untuk melatih model klasifikasi naïve bayes tentang bagaimana atribut-atribut dalam data berkaitan dengan kategori atau kelas. *Data testing* digunakan untuk menguji sejauh mana model tersebut dapat melakukan prediksi yang benar serta membantu pengukuran kinerja dan akurasi model klasifikasi naïve bayes.

Berikut langkah-langkah untuk mengklasifikasi menggunakan algoritma naïve bayes:

4.2.1 Menghitung Probabilitas Prior

Langkah awal dalam mengklasifikasi data adalah menghitung probabilitas pada setiap kategori atau *prior* pada setiap kelas. Pada penelitian ini kategori dibagi menjadi 2, yaitu kategori *satisfied* dan *neutral or dissatisfied*.

$$P(\text{satisfaction}) = \frac{45025}{103904} = 0.4333$$

$$P(\text{neutral or dissatisfied}) = \frac{58879}{103904} = 0.5666$$

Nilai ini menunjukkan bahwa dari keseluruhan data, proporsi penumpang yang termasuk kategori *neutral or dissatisfied* lebih tinggi dibandingkan penumpang yang *satisfied*.

4.2.2 Menentukan Probabilitas Kelas Kategori

Setelah menghitung probabilitas *prior* dari setiap kelas, selanjutnya menghitung probabilitas pada setiap kategori.

1. Kelas *Gender*

Terdapat dua kategori pada kelas *Gender* atau jenis kelamin, yaitu *Female* (perempuan) dan *Male* (laki-laki).

a. *Female*

$$P(\text{female}|\text{satisfied}) = \frac{22534}{45025} = 0.5004$$

$$P(\text{female}|\text{neutral or dissatisfied}) = \frac{30193}{58879} = 0.5127$$

b. *Male*

$$P(\text{male}|\text{satisfied}) = \frac{22491}{45025} = 0.4995$$

$$P(\text{male}|\text{neutral or dissatisfied}) = \frac{28686}{58879} = 0.4872$$

Pada atribut *gender* menunjukkan bahwa penumpang *female* lebih dominan pada kelas *neutral or dissatisfied*, sedangkan penumpang *male* lebih dominan pada kelas *satisfied*. . Meskipun selisih probabilitasnya tidak terlalu besar, atribut ini memberikan indikasi bahwa *gender* atau jenis kelamin memiliki pengaruh terhadap kecenderungan tingkat kepuasan penumpang.

2. Kelas *Customer Type*

Terdapat dua kategori pada kelas *Customer Type* atau jenis tipe pelanggan, yaitu *Loyal* (setia) dan *Disloyal* (tidak setia).

a. Loyal

$$P(\text{loyal}|\text{satisfied}) = \frac{40533}{45025} = 0.9002$$

$$P(\text{loyal}|\text{neutral or dissatisfied}) = \frac{44390}{58879} = 0.7539$$

b. Disloyal

$$P(\text{disloyal}|\text{satisfied}) = \frac{4492}{45025} = 0.0997$$

$$P(\text{disloyal}|\text{neutral or dissatisfied}) = \frac{14489}{58879} = 0.2460$$

Pada atribut *Customer Type*, penumpang *Loyal* memiliki probabilitas yang jauh lebih tinggi untuk berada pada kelas *Satisfied*, sedangkan penumpang *Disloyal* lebih banyak muncul pada kelas *Neutral or Dissatisfied*. Hal ini menunjukkan bahwa tingkat loyalitas merupakan faktor kuat yang berkaitan dengan kepuasan penumpang.

3. Kelas Type of Travel

Terdapat dua kategori pada kelas *Type of Travel* atau tipe perjalanan, yaitu *Personal Travel* (perjalanan pribadi) dan *Business travel* (perjalanan bisnis).

a. Personal travel

$$P(\text{personal travel}|\text{satisfied}) = \frac{3279}{45025} = 0.0728$$

$$P(\text{personal travel}|\text{neutral or dissatisfied}) = \frac{28970}{58879}$$

$$= 0.4920$$

b. Business travel

$$P(\text{business travel}|\text{satisfied}) = \frac{41746}{45025} = 0.9271$$

$$P(\text{business travel}|\text{neutral or dissatisfied}) = \frac{29909}{45025} \\ = 0.6642$$

Atribut *Type of Travel* memperlihatkan perbedaan yang sangat jelas, di mana penumpang yang berada pada *Business Travel* lebih dominan pada kelas *Satisfied*, sedangkan penumpang dengan *Personal Travel* lebih sering ditemukan pada kelas *Neutral or Dissatisfied*. Dengan demikian, tujuan perjalanan memiliki pengaruh signifikan terhadap tingkat kepuasan.

4. Kelas Class

Terdapat tiga kategori pada kelas *Class* atau kelas, yaitu *Eco* (ekonomi), *business* (bisnis), dan *Eco plus* (ekonomi plus).

a. Eco

$$P(\text{eco}|\text{satisfied}) = \frac{8701}{45025} = 0.1932$$

$$P(\text{eco}|\text{neutral or dissatisfied}) = \frac{38044}{58879} = 0.6461$$

b. Business

$$P(\text{business}|\text{satisfied}) = \frac{34480}{45025} = 0.7657$$

$$P(\text{business}|\text{neutral or dissatisfied}) = \frac{15185}{58879} = 0.2579$$

c. Eco plus

$$P(\text{eco plus}|\text{satisfied}) = \frac{1844}{45025} = 0.0409$$

$$P(\text{eco plus}|\text{neutral or dissatisfied}) = \frac{5650}{58879} = 0.0959$$

Atribut *Class* menunjukkan bahwa penumpang *Business Class* sangat dominan pada kelas *Satisfied*, sementara penumpang *Eco Class* dan *Eco Plus* lebih banyak muncul pada kelas *Neutral or Dissatisfied*. Hal ini mengindikasikan bahwa kategori kelas penerbangan merupakan salah satu faktor yang paling membedakan antara penumpang puas dan tidak puas.

4.2.3 Menghitung nilai mean dan standar deviasi kelas numerik

Setelah menghitung probabilitas kelas pada setiap kategori, selanjutnya akan menghitung nilai mean dan standar deviasi pada setiap kelas numerik untuk digunakan pada perhitungan fungsi densitas Gauss.

1. Age

a. Mean

$$\mu_{\text{satisfied}} = \frac{1879820}{45025} = 41.750$$

$$\mu_{\text{neutral or dissatisfied}} = \frac{2211889}{58879} = 37.566$$

b. Standar deviasi

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1}}$$

$$\sigma_{\text{satisfied}} = 12.7676$$

$$\sigma_{\text{neutral or dissatisfied}} = 16.4596$$

2. *Flight distance*

a. Mean

$$\mu_{satisfied} = \frac{68894565}{45025} = 1530.1402$$

$$\mu_{neutral\ or\ dissatisfied} = \frac{54693879}{58879} = 928.9199$$

b. Standar deviasi

$$\sigma_{satisfied} = 1128.114$$

$$\sigma_{neutral\ or\ dissatisfied} = 790.4456$$

3. *Inflight wifi service*

a. Mean

$$\mu_{satisfied} = \frac{142337}{45025} = 3.1611$$

$$\mu_{neutral\ or\ dissatisfied} = \frac{141288}{58879} = 2.3996$$

b. Standar deviasi

$$\sigma_{satisfied} = 1.5886$$

$$\sigma_{neutral\ or\ dissatisfied} = 0.9643$$

4. *Departure/arrival time convenient*

a. Mean

$$\mu_{satisfied} = \frac{133738}{45025} = 2.9703$$

$$\mu_{neutral\ or\ dissatisfied} = \frac{184239}{58879} = 3.1291$$

b. Standar deviasi

$$\sigma_{satisfied} = 1.5521$$

$$\sigma_{neutral\ or\ dissatisfied} = 1.5003$$

5. *Ease of online booking*

a. Mean

$$\mu_{satisfied} = \frac{136497}{45025} = 3.0315$$

$$\mu_{neutral\ or\ dissatisfied} = \frac{249956}{58879} = 2.5468$$

b. Standar deviasi

$$\sigma_{satisfied} = 1.5752$$

$$\sigma_{neutral\ or\ dissatisfied} = 1.2058$$

6. *Gate location*

a. Mean

$$\mu_{satisfied} = \frac{134079}{45025} = 2.9778$$

$$\mu_{neutral\ or\ dissatisfied} = \frac{175231}{58879} = 2.9761$$

b. Standar deviasi

$$\sigma_{satisfied} = 1.3742$$

$$\sigma_{neutral\ or\ dissatisfied} = 1.1984$$

7. *Food and drink*

a. Mean

$$\mu_{satisfied} = \frac{158547}{45025} = 3.5213$$

$$\mu_{neutral\ or\ dissatisfied} = \frac{174167}{58879} = 2.9580$$

b. Standar deviasi

$$\sigma_{satisfied} = 1.2361$$

$$\sigma_{neutral\ or\ dissatisfied} = 1.3465$$

8. *Online boarding*

a. Mean

$$\mu_{satisfied} = \frac{181337}{45025} = 4.0274$$

$$\mu_{neutral\ or\ dissatisfied} = \frac{156390}{58879} = 2.6561$$

b. Standar deviasi

$$\sigma_{satisfied} = 1.1915$$

$$\sigma_{neutral\ or\ dissatisfied} = 1.1458$$

9. *Seat comfort*

a. Mean

$$\mu_{satisfied} = \frac{178593}{45025} = 3.9665$$

$$\mu_{neutral\ or\ dissatisfied} = \frac{178774}{58879} = 3.0362$$

b. Standar deviasi

$$\sigma_{satisfied} = 1.1420$$

$$\sigma_{neutral\ or\ dissatisfied} = 1.3031$$

10. *Inflight entertainment*

a. Mean

$$\mu_{satisfied} = \frac{178521}{45025} = 3.9649$$

$$\mu_{neutral\ or\ dissatisfied} = \frac{170405}{58879} = 2.8941$$

b. Standar deviasi

$$\sigma_{satisfied} = 1.0768$$

$$\sigma_{neutral\ or\ dissatisfied} = 1.3236$$

11. *Onboard service*

a. Mean

$$\mu_{satisfied} = \frac{173676}{45025} = 3.8573$$

$$\mu_{neutral\ or\ dissatisfied} = \frac{177765}{58879} = 3.0191$$

b. Standar deviasi

$$\sigma_{satisfied} = 1.1271$$

$$\sigma_{neutral\ or\ dissatisfied} = 1.2857$$

12. *Leg room service*

a. Mean

$$\mu_{satisfied} = \frac{172092}{45025} = 3.8221$$

$$\mu_{neutral\ or\ dissatisfied} = \frac{176096}{58879} = 2.9908$$

b. Standar deviasi

$$\sigma_{satisfied} = 1.1755$$

$$\sigma_{neutral\ or\ dissatisfied} = 1.3031$$

13. *Baggage handling*

a. Mean

$$\mu_{satisfied} = \frac{178587}{45025} = 3.8221$$

$$\mu_{netral\ or\ dissatisfied} = \frac{198775}{58879} = 3.3759$$

b. Standar deviasi

$$\sigma_{satisfied} = 1.0995$$

$$\sigma_{neutral\ or\ dissatisfied} = 1.1769$$

14. *Checkin service*

a. Mean

$$\mu_{satisfied} = \frac{164163}{45025} = 3.6460$$

$$\mu_{dissatisfied} = \frac{179166}{58879} = 3.0429$$

b. Standar deviasi

$$\sigma_{satisfied} = 1.1587$$

$$\sigma_{neutral\ or\ dissatisfied} = 1.2811$$

15. *Inflight service*

a. Mean

$$\mu_{satisfied} = \frac{178725}{45025} = 3.9694$$

$$\mu_{neutral\ or\ dissatisfied} = \frac{199530}{58879} = 3.3888$$

b. Standar deviasi

$$\sigma_{satisfied} = 1.0914$$

$$\sigma_{neutral\ or\ dissatisfied} = 1.1755$$

16. *Cleanliness*

a. Mean

$$\mu_{satisfied} = \frac{168589}{45025} = 3.7443$$

$$\mu_{neutral\ or\ dissatisfied} = \frac{172876}{58879} = 2.9361$$

b. Standar deviasi

$$\sigma_{satisfied} = 1.1422$$

$$\sigma_{neutral\ or\ dissatisfied} = 1.3259$$

17. *Departure delay in minutes*

a. Mean

$$\mu_{satisfied} = \frac{567679}{45025} = 12.6080$$

$$\mu_{neutral\ or\ dissatisfied} = \frac{971723}{58879} = 16.5037$$

b. Standar deviasi

$$\sigma_{satisfied} = 35.3822$$

$$\sigma_{neutral\ or\ dissatisfied} = 40.1915$$

18. *Arrival delay in minutes*

a. Mean

$$\mu_{satisfied} = \frac{567085}{45025} = 12.5948$$

$$\mu_{neutral\ or\ dissatisfied} = \frac{1005335}{58879} = 17.0745$$

b. Standar deviasi

$$\sigma_{satisfied} = 35.9167$$

$$\sigma_{neutral\ or\ dissatisfied} = 40.5083$$

Pada atribut numerik, nilai mean dan standar deviasi digunakan untuk menggambarkan pola penyebaran data pada masing-masing kelas. Nilai mean menunjukkan rata-rata nilai atribut dalam kelas tertentu, sedangkan standar deviasi menunjukkan seberapa besar variasi atau penyebaran nilai tersebut dari rata-ratanya. Mean yang berbeda pada kedua kelas menunjukkan adanya perbedaan karakteristik rata-rata antara penumpang *Satisfied* dan *Neutral or*

Dissatisfied. Sementara itu, standar deviasi yang lebih besar menandakan bahwa nilai atribut pada kelas tersebut lebih bervariasi.

4.2.4 Perhitungan Naïve Bayes dengan Fungsi Densitas Gauss

Setelah diperoleh hasil mean dan standar deviasi pada setiap atribut, tahap selanjutnya adalah menghitung metode naïve bayes dengan rumus densitas Gauss pada data numerik. Untuk mengklasifikasi apakah penumpang *satisfied* (puas) atau *neutral or dissatisfied* (netral atau tidak puas), diberi contoh jika diketahui pelanggan perempuan berumur 26 tahun dengan type pelanggan loyal, type travel dan kelas bisnis, flight distance 900, inflight wifi 4, departure or arrive 4, ease of online booking 5, gate location 3, food and drink 3, online boarding 5, seat comfort 5, inflight entertainment 4, on board 4, leg room 4, baggage 3, checkin service 5, inflight service 4, cleanliness 3, departure delay 3, dan arrive delay 3.

1. Age

$$P(\text{age}|\text{satisfied}) = \frac{1}{\sqrt{2\pi} (12.7676)} e^{\frac{-(26-41.750)^2}{2(12.7676)^2}} = 0.05216$$

$$P(\text{age}|\text{neutral or dissatisfied}) = \frac{1}{\sqrt{2\pi} (16.4596)} e^{\frac{-(26-37.566)^2}{2(16.4596)^2}}$$

$$= 0.07682$$

2. Flight distance

$$P(\text{flightdistance}|\text{satisfied}) = \frac{1}{\sqrt{2\pi} (1128.114)} e^{\frac{-(400-1530.1402)^2}{2(1128.114)^2}}$$

$$= 0.00719$$

$P(\text{flightdistance}|\text{neutral or dissatisfied})$

$$= \frac{1}{\sqrt{2\pi}(790.4456)} e^{\frac{-(400-928.9199)^2}{2(790.4456)^2}} = 0.01134$$

3. *Inflight wifi*

$$P(\text{inflightwifi}|\text{satisfied}) = \frac{1}{\sqrt{2\pi}(1.5886)} e^{\frac{-(4-3.1612)^2}{2(1.5886)^2}} = 0.27533$$

$P(\text{inflightwifi}|\text{neutral or dissatisfied})$

$$= \frac{1}{\sqrt{2\pi}(0.9643)} e^{\frac{-(4-2.3996)^2}{2(0.9643)^2}}$$

$$= 0.10249$$

4. *Departure or arrive*

$$P(\text{departureorarrive}|\text{satisfied}) = \frac{1}{\sqrt{2\pi}(1.5521)} e^{\frac{-(4-2.9703)^2}{2(1.5521)^2}}$$

$$= 0.25696$$

$P(\text{departureorarrive}|\text{neutral or dissatisfied})$

$$= \frac{1}{\sqrt{2\pi}(1.5003)} e^{\frac{-(4-31.291)^2}{3(1.5003)^2}} = 0.27520$$

5. *Ease of online booking*

$$P(\text{easeofonlinebooking}|\text{satisfied}) = \frac{1}{\sqrt{2\pi}(1.5752)} e^{\frac{-(5-3.0315)^2}{2(1.5752)^2}}$$

$$= 0.14558$$

$P(\text{easeofonlinebooking}|\text{neutral or dissatisfied})$

$$= \frac{1}{\sqrt{2\pi}(1.2058)} e^{\frac{-(5-2.5468)^2}{2(1.2058)^2}} = 0.04586$$

6. Gate location

$$P(\text{gatelocation}|\text{satisfied}) = \frac{1}{\sqrt{2\pi}(1.3742)} e^{\frac{-(3-2.9778)^2}{2(1.3742)^2}} = 0.34027$$

$$P(\text{gatelocation}|\text{neutral or dissatisfied})$$

$$= \frac{1}{\sqrt{2\pi}(1.2058)} e^{\frac{-(3-3.0315)^2}{2(1.2058)^2}} = 0.36318$$

7. Food and drink

$$P(\text{foodanddrink}|\text{satisfied}) = \frac{1}{\sqrt{2\pi}(1.2361)} e^{\frac{-(3-3.5213)^2}{2(1.2361)^2}}$$

$$= 0.32829$$

$$P(\text{foodanddrink}|\text{neutral or dissatisfied})$$

$$= \frac{1}{\sqrt{2\pi}(1.3465)} e^{\frac{-(3-2.9580)^2}{2(1.3465)^2}} = 0.34363$$

8. Online boarding

$$P(\text{onlineboarding}|\text{satisfied}) = \frac{1}{\sqrt{2\pi}(1.1915)} e^{\frac{-(5-4.0274)^2}{2(1.1915)^2}}$$

$$= 0.26192$$

$$P(\text{onlineboarding}|\text{neutral or dissatisfied})$$

$$= \frac{1}{\sqrt{2\pi}(1.1458)} e^{\frac{-(5-2.6561)^2}{2(1.1458)^2}} = 0.04599$$

9. Seat comfort

$$P(\text{seatcomfort}|\text{satisfied}) = \frac{1}{\sqrt{2\pi}(1.1420)} e^{\frac{-(5-3.9665)^2}{2(1.1458)^2}} = 0.24787$$

$$P(\text{seatcomfort}|\text{neutral or dissatisfied})$$

$$= \frac{1}{\sqrt{2\pi}(1.3031)} e^{\frac{-(5-3.0362)^2}{2(1.3031)^2}} = 0.11226$$

10. Inflight entertainment

$$P(\text{inflightentertainment}|\text{satisfied}) = \frac{1}{\sqrt{2\pi}(1.0768)} e^{\frac{-(4-3.9649)^2}{2(1.0768)^2}}$$

$$= 0.38424$$

$$P(\text{inflightentertainment}|\text{neutral or dissatisfied})$$

$$= \frac{1}{\sqrt{2\pi}(1.3236)} e^{\frac{-(4-2.8941)^2}{2(1.3236)^2}} = 0.24459$$

11. On board

$$P(\text{onboard}|\text{satisfied}) = \frac{1}{\sqrt{2\pi}(1.1271)} e^{\frac{-(4-3.8573)^2}{2(1.1271)^2}} = 0.37277$$

$$P(\text{onboard}|\text{neutral or dissatisfied}) = \frac{1}{\sqrt{2\pi}(1.2857)} e^{\frac{-(4-3.0191)^2}{2(1.2857)^2}}$$

$$= 0.26299$$

12. Leg room

$$P(\text{legroom}|\text{satisfied}) = \frac{1}{\sqrt{2\pi}(1.1755)} e^{\frac{-(4-3.8221)^2}{2(1.1755)^2}} = 0.36376$$

$$P(\text{legroom}|\text{neutral or dissatisfied}) = \frac{1}{\sqrt{2\pi}(1.3031)} e^{\frac{-(4-2.9908)^2}{2(1.3031)^2}}$$

$$= 0.25892$$

13. Baggage handling

$$P(\text{baggagehandling}|\text{satisfied}) = \frac{1}{\sqrt{2\pi}(1.0995)} e^{\frac{-(3-3.9663)^2}{2(1.0995)^2}}$$

$$= 0.25857$$

$$P(\text{baggagehandling}|\text{neutral or dissatisfied})$$

$$= \frac{1}{\sqrt{2\pi}(1.1769)} e^{\frac{-(3-3.3759)^2}{2(1.1769)^2}} = 0.14191$$

14. Checkin service

$$P(\text{checkin service}|\text{satisfied}) = \frac{1}{\sqrt{2\pi}(1.1587)} e^{\frac{-(5-3.6460)^2}{2(1.1587)^2}}$$

$$= 0.18724$$

$$P(\text{checkin service}|\text{neutral or dissatisfied})$$

$$= \frac{1}{\sqrt{2\pi}(1.2811)} e^{\frac{-(5-3.0429)^2}{2(1.2811)^2}} = 0.10973$$

15. Inflight service

$$P(\text{inflight service}|\text{satisfied}) = \frac{1}{\sqrt{2\pi}(1.0914)} e^{\frac{-(4-3.9694)^2}{2(1.0914)^2}}$$

$$= 0.38172$$

$$P(\text{inflight service}|\text{neutral or dissatisfied})$$

$$= \frac{1}{\sqrt{2\pi}(1.1755)} e^{\frac{-(4-3.9694)^2}{2(1.1755)^2}} = 0.32143$$

16. Cleanliness

$$P(\text{cleanliness}|\text{satisfied}) = \frac{1}{\sqrt{2\pi}(1.1422)} e^{\frac{-(3-3.7443)^2}{2(1.1422)^2}} = 0.30187$$

$$P(\text{cleanliness}|\text{neutral or dissatisfied})$$

$$= \frac{1}{\sqrt{2\pi}(1.3259)} e^{\frac{-(3-2.9361)^2}{2(1.3259)^2}} = 0.34605$$

17. Departure delay

$$P(\text{departure delay}|\text{satisfied}) = \frac{1}{\sqrt{2\pi}(35.3822)} e^{\frac{-(3-12.6080)^2}{2(35.3822)^2}}$$

$$= 0.06464$$

$$P(\text{departure delay}|\text{neutral or dissatisfied}) =$$

$$\frac{1}{\sqrt{2\pi}(40.1915)} e^{\frac{-(3-6.5037)^2}{2(40.1915)^2}} = 0.05947$$

18. Arrival delay

$$P(\text{arrivaldelay}|\text{satisfied}) = \frac{1}{\sqrt{2\pi}(35.9167)} e^{\frac{-(3-12.5974)^2}{2(35.9167)^2}}$$

$$= 0.06427$$

$$P(\text{arrivaldelay}|\text{neutral or dissatisfied})$$

$$= \frac{1}{\sqrt{2\pi}(40.5083)} e^{\frac{-(3-17.0745)^2}{2(40.5083)^2}} = 0.05900$$

Kemudian menggunakan perhitungan metode naïve bayes dengan rumus *likelihood*:

$$P(C|X) = \prod_{j=1}^n P(c_j|X)$$

$$P(C|\text{Satisfied})$$

$$\begin{aligned} &= P(0.5004|\text{Satisfied}) \times P(0.9002|\text{Satisfied}) \\ &\times P(0.9271|\text{Satisfied}) \times P(0.7657|\text{Satisfied}) \\ &\times P(0.05216|\text{Satisfied}) \times P(0.00719|\text{Satisfied}) \\ &\times P(0.27533|\text{Satisfied}) \times P(0.25696|\text{Satisfied}) \\ &\times P(0.14558|\text{Satisfied}) \times P(0.34027|\text{Satisfied}) \\ &\times P(0.32829|\text{Satisfied}) \times P(0.26192|\text{Satisfied}) \\ &\times P(0.24787|\text{Satisfied}) \times P(0.38424|\text{Satisfied}) \\ &\times P(0.37277|\text{Satisfied}) \times P(0.36376|\text{Satisfied}) \\ &\times P(0.25857|\text{Satisfied}) \times P(0.18724|\text{Satisfied}) \\ &\times P(0.38172|\text{Satisfied}) \times P(0.30187|\text{Satisfied}) \\ &\times P(0.06464|\text{Satisfied}) \times P(0.06427|\text{Satisfied}) \\ &\times P(0.4333|\text{Satisfied}) = 4.68707 \times 10^{-15} \end{aligned}$$

$$\begin{aligned}
&P(C|\textit{Neutral or dissatisfied}) \\
&= P(0.5127|\textit{Neutral or dissatisfied}) \\
&\times P(0.7539|\textit{Neutral or dissatisfied}) \\
&\times P(0.9271|\textit{Neutral or dissatisfied}) \\
&\times P(0.7657|\textit{Neutral or dissatisfied}) \\
&\times P(0.07682|\textit{Neutral or dissatisfied}) \\
&\times P(0.11343|\textit{Neutral or dissatisfied}) \\
&\times P(0.10249|\textit{Neutral or dissatisfied}) \\
&\times P(0.27520|\textit{Neutral or dissatisfied}) \\
&\times P(0.04586|\textit{Neutral or dissatisfied}) \\
&\times P(0.36318|\textit{Neutral or dissatisfied}) \\
&\times P(0.34363|\textit{Neutral or dissatisfied}) \\
&\times P(0.04599|\textit{Neutral or dissatisfied}) \\
&\times P(0.11226|\textit{Neutral or dissatisfied}) \\
&\times P(0.24459|\textit{Neutral or dissatisfied}) \\
&\times P(0.26299|\textit{Neutral or dissatisfied}) \\
&\times P(0.25892|\textit{Neutral or dissatisfied}) \\
&\times P(0.14191|\textit{Neutral or dissatisfied}) \\
&\times P(0.10973|\textit{Neutral or dissatisfied}) \\
&\times P(0.32143|\textit{Neutral or dissatisfied}) \\
&\times P(0.34605|\textit{Neutral or dissatisfied}) \\
&\times P(0.05947|\textit{Neutral or dissatisfied}) \\
&\times P(0.05900|\textit{Neutral or dissatisfied}) \\
&\times P(0.5666|\textit{Neutral or dissatisfied}) = 1.14278 \times 10^{-16}
\end{aligned}$$

Setelah seluruh *likelihood* setiap atribut dihitung menggunakan fungsi densitas Gauss dan dikalikan sesuai asumsi independensi pada naïve bayes, diperoleh nilai *likelihood* total untuk masing-masing kelas. Pada data uji yang digunakan, diperoleh dua nilai *likelihood* yaitu *likelihood* kelas *satisfied* sebesar 4.68707×10^{-15} dan *likelihood* kelas *neutral or dissatisfied* sebesar 1.14278×10^{-16} .

Kedua angka ini menggambarkan tingkat kesesuaian data uji terhadap pola distribusi pada masing-masing kelas. Nilai *likelihood* yang lebih besar menunjukkan bahwa data tersebut lebih menyerupai karakteristik fitur pada kelas tersebut. Karena pada data ini nilai *likelihood* pada kelas *satisfied* lebih besar, maka data uji lebih sesuai dengan karakteristik kelas tersebut.

4.2.5 Menghitung Probabilitas Posterior

Setelah memperoleh nilai *likelihood* dan probabilitas prior untuk setiap kelas, tahap selanjutnya adalah menghitung probabilitas posterior. Probabilitas posterior menunjukkan seberapa besar kemungkinan data uji termasuk ke dalam masing-masing kelas setelah mempertimbangkan seluruh informasi atribut yang dimilikinya.

$$P(C|satisfied) = \frac{(4.68707 \times 10^{-15})(0.4333)}{(4.68707 \times 10^{-15} + 1.14278 \times 10^{-16})} = 0.976199$$

$$P(C|Neutral or dissatisfied)$$

$$= \frac{(1.14278 \times 10^{-16})(0.5666)}{(4.68707 \times 10^{-15} + 1.14278 \times 10^{-16})} = 0.0238012$$

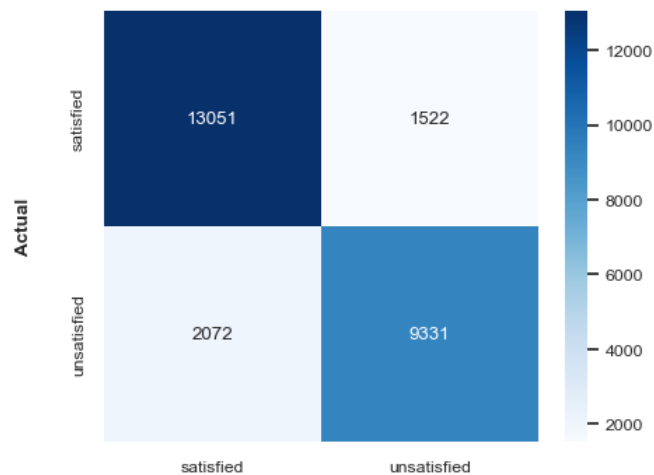
Nilai posterior tersebut menunjukkan bahwa data uji memiliki kecenderungan yang sangat kuat untuk masuk ke dalam kelas *satisfied*. Dengan kata lain, setelah mempertimbangkan seluruh fitur penumpang dan membandingkannya dengan pola pada masing-masing kelas, model ini menilai bahwa peluang penumpang tersebut termasuk kategori puas mencapai sekitar 97% jauh lebih tinggi dibandingkan peluang menjadi tidak puas yang hanya sekitar 2.38%.

4.2.6 Menghitung Nilai *Likelihood* Maksimum

Nilai *likelihood* maksimum menunjukkan kelas yang memiliki tingkat kesesuaian paling tinggi dengan data uji. Untuk mengklasifikasikan jenis penumpang pesawat dapat dikatakan puas atau netral atau tidak puas dapat dilihat dari nilai akhir probabilitas yang mendekati 1 atau bernilai sama dengan 1. Dari hasil perhitungan probabilitas posterior diketahui bahwa hasil akhir yang diperoleh untuk setiap probabilitas, yaitu *Satisfied* = 0.976199 dan *Neutral or dissatisfied* = 0.0238012. Sehingga dapat dikategorikan bahwa penumpang pesawat yang berjenis kelamin perempuan berumur 26 tahun dengan *type* pelanggan loyal, *type travel* dan kelas bisnis, *flight distance* 900, *inflight wifi* 4, *departure or arrive* 4, *ease of online booking* 5, *gate location* 3, *food and drink* 3, *online boarding* 5, *seat comfort* 5, *inflight entertainment* 4, *on board* 4, *leg room* 4, *baggage* 3, *checkin service* 5, *inflight service* 4, *cleanliness* 3, *departure delay* 3, dan *arrive delay* 3 merupakan pelanggan yang *satisfied* atau puas.

4.3 Evaluasi

Evaluasi yang dilakukan adalah dengan cara menghitung akurasi, presisi, dan recall, pada data testing sebanyak 25976 data yang didapat dari *confusion matriks*.



Gambar 4. 2 *Confusion matriks*

Berdasarkan *confusion matriks* yang dihasilkan, terdapat sebanyak 13051 sampel yang benar-benar puas dan diprediksi benar sebagai puas yang disebut dengan *true positive (TP)*. Kemudian terdapat 1522 sampel yang sebenarnya tidak puas tetapi diprediksi sebagai puas yang disebut dengan *false positive (FP)*. Kemudian terdapat 2072 sampel yang sebenarnya puas tetapi diprediksi sebagai tidak puas yang disebut dengan *false negative (FN)*. Kemudian terdapat 9331 sampel yang benar-benar tidak puas dan diprediksi dengan benar sebagai tidak puas yang disebut dengan *true negative (TN)*. Maka terdapat 15123 pelanggan yang puas dan 10853 pelanggan yang netral atau tidak puas.

Berdasarkan *confusion matriks* yang didapat, maka dihasilkan akurasi, presisi, dan recall.

$$Akurasi = \frac{13051 + 9331}{25976} \times 100\% = 86\%$$

Nilai akurasi menunjukkan bahwa pelanggan yang benar diprediksi merasa puas maupun yang merasa netral atau tidak puas dari keseluruhan pelanggan adalah sebesar 86%.

$$presisi\ satisfied = \frac{13051}{13051 + 1522} \times 100\% = 89\%$$

Presisi pada kelas satisfied adalah 89%. Ini berarti bahwa dari semua prediksi yang dibuat oleh model sebagai “satisfied” (puas), 89% diantaranya benar-benar puas. Dengan kata lain, ketika model memprediksi bahwa seorang pelanggan puas, maka ada kemungkinan sekitar 89% bahwa prediksi tersebut benar.

$$presisi\ neutral\ or\ dissatisfied = \frac{9331}{9331 + 2072} \times 100\% = 81\%$$

Presisi pada kelas neutral or dissatisfied adalah 81%. Ini berarti bahwa dari semua prediksi yang dibuat oleh model sebagai “neutral or dissatisfied”, 81% diantaranya benar-benar tidak puas. Dengan kata lain, ketika model memprediksi bahwa seorang pelanggan tidak puas, ada kemungkinan sekitar 81% bahwa prediksi tersebut benar.

$$recall\ satisfied = \frac{13051}{13051 + 2072} \times 100\% = 86\%$$

Recall pada kelas satisfied adalah 86%. Ini berarti bahwa dari semua pelanggan yang sebenarnya puas, model berhasil memprediksi 86% diantaranya sebagai “satisfied”. Dengan kata lain, jika ada pelanggan yang benar-benar puas, model dapat mendeteksinya dengan kemungkinan 86%

$$recall\ neutral\ or\ dissatisfied = \frac{9331}{9331 + 1522} \times 100\% = 85\%$$

Recall pada kelas neutral or dissatisfied adalah 85%. Ini berarti bahwa dari semua pelanggan yang sebenarnya tidak puas, model berhasil memprediksi 85% diantaranya sebagai “unsatisfied”. Dengan kata lain, jika ada pelanggan yang benar-benar tidak puas, model dapat mendeteksinya dengan kemungkinan 85%.

Kemudian menggunakan *k-fold cross validation* untuk mengetahui seberapa valid nilai akurasi yang dihasilkan. Nilai *k* yang ditentukan adalah 5, maka data akan terbagi menjadi 5 (lima) bagian pada setiap proses dan diacak antara *data train* dan *data test*. Sehingga pada *k-fold cross validation*, semua data mendapat peran menjadi *data train* dan *data test*. Hasil akurasi yang didapat pada setiap *k* ditunjukkan pada Tabel 4.1.

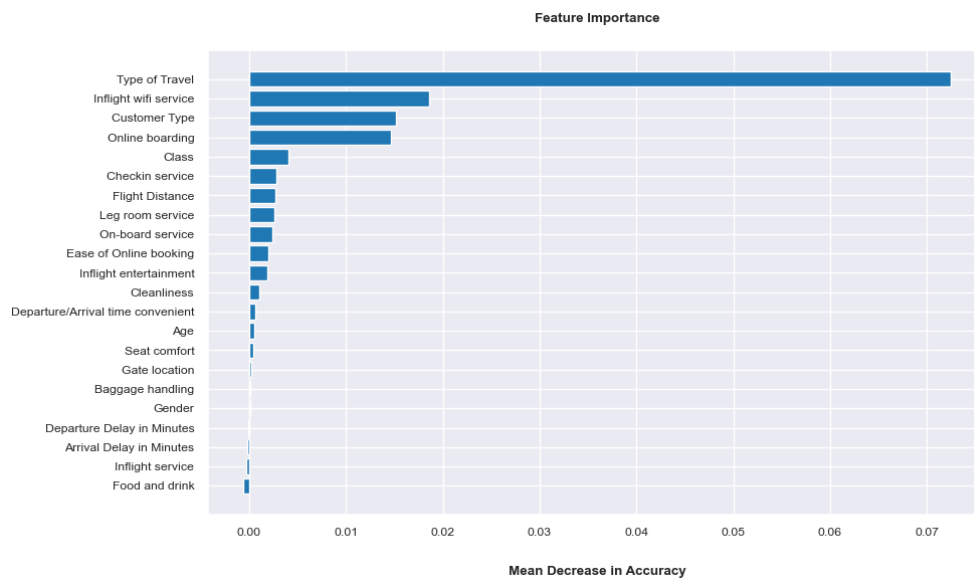
Tabel 4.1 Hasil *K-fold cross validation*

Fold ke-	Akurasi
1	86.12%
2	86.24%
3	86.19%
4	86.38%
5	87.18%
Rata-rata	86.42%

Hasil nilai akurasi pada tiap fold menunjukkan bahwa setiap *fold* memiliki nilai akurasi yang berbeda-beda yaitu dengan nilai 86.12% hingga 87.18% dengan rata-rata 86.42%. Berdasarkan tabel, dapat diketahui bahwa nilai akurasi terbaik berada pada fold ke-5 dengan nilai akurasi sebesar 87.18%. Hasil akurasi pada tiap fold juga menunjukkan bahwa hasil nilai akurasi adalah konsisten dan

menghasilkan nilai yang tidak berbeda jauh pada setiap pengulangan atau fold, sehingga dapat diinterpretasikan bahwa model maupun data dalam penelitian dikatakan reliabel.

Kemudian menggunakan *permutation importance* untuk mengukur variabel atau fitur yang paling berpengaruh dalam model. *Permutation importance* menunjukkan seberapa besar penurunan akurasi model jika fitur tersebut dihilangkan secara acak. Variabel atau fitur yang paling berpengaruh pada model ini berdasarkan *permutation importance* adalah sebagai berikut.



Gambar 4.3 Hasil *feature importance*

Berdasarkan Gambar 4.3, variabel yang paling berpengaruh pada model ini merupakan variabel *type of travel* dan variabel *inflight wifi service*. Ini menunjukkan bahwa fitur tersebut memberikan kontribusi paling signifikan terhadap akurasi model, dengan nilai *feature importance* 0.075. Di sisi lain, variabel *food and drink* adalah fitur yang paling tidak penting karena memiliki nilai *permutation importance* terendah. Ini menunjukkan bahwa fitur tersebut memberikan kontribusi paling sedikit terhadap akurasi model.

4.4 Integrasi Keislaman Terhadap Implementasi Algoritma Naïve Bayes

Berdasarkan pembahasan sebelumnya telah diperoleh proses dan hasil dari implementasi algoritma naïve bayes terhadap klasifikasi kepuasan penumpang pesawat. Hasil yang ditunjukkan merupakan faktor-faktor layanan yang berkontribusi terhadap ketidakpuasan maupun kepuasan pelanggan. Hal ini dilakukan untuk menjaga hak pembeli atau pengguna jasa dalam bertransaksi.

Prinsip kejujuran dan keadilan dalam bermualah merupakan salah satu mekanisme dari khiyar aib, yaitu hak pembeli atau pengguna jasa untuk membatalkan akad apabila ditemukan cacat yang tidak dijelaskan sebelumnya. Prinsip ini menekankan pentingnya transparansi dan perlindungan hak konsumen. Dalam Al-Qur'an Surah An-Nisa ayat 29 dikatakan bahwa:

يَا أَيُّهَا الَّذِينَ آمَنُوا لَا تَأْكُلُوا أَمْوَالَكُمْ بَيْنَكُمْ بِالْبَاطِلِ إِلَّا أَنْ تَكُونَ تِجَارَةً عَنْ تَرَاضٍ مِّنْكُمْ وَلَا تَقْتُلُوا أَنْفُسَكُمْ إِنَّ اللَّهَ كَانَ بِكُمْ رَحِيمًا (النِّسَاء ٢٩)

“Wahai orang-orang yang beriman, janganlah kamu memakan harta sesamamu dengan cara yang batil (tidak benar), kecuali berupa perniagaan atas dasar suka sama suka di antara kamu. Jangan kamu membunuh dirimu. Sesungguhnya Allah adalah Maha Penyayang kepadamu”

Ayat ini mengingatkan agar transaksi dilakukan atas dasar kejujuran dan kerelaan, bukan karena penipuan atau informasi yang disembunyikan. Dalam konteks layanan transportasi dan janji pelayanan yang tidak sesuai dengan kenyataan dapat dikategorikan sebagai tindakan yang batil, karena merugikan pengguna jasa.

Penggunaan algoritma naïve bayes dalam menganalisis tingkat kepuasan penumpang memberikan manfaat besar dalam menjaga nilai-nilai syariah tersebut. Algoritma ini dapat mengidentifikasi pola keluhan, ketidaksesuaian layanan, serta potensi cacat tersembunyi dalam sistem pelayanan yang tidak terdeteksi secara

manual. Dengan demikian, pihak penyedia layanan dapat melakukan perbaikan secara lebih tepat sasaran dan bertanggung jawab.

Dalam Sunan Ibnu Majah no. 2246 yang diriwayatkan melalui sanad Muhammad bin Bashshar, dari Wahb bin Jarir, dikatakan bahwa Rasulullah SAW bersabda:

لَا يَحِلُّ لِأَحَدٍ أَنْ يَبِيعَ سَلْعَةً إِلَّا بَيَّنَّ مَا يُعْيِبُهَا، وَلَا يَحِلُّ لِمَنْ عَلِمَ ذَلِكَ إِلَّا بَيَّنَّهُ (حديث رواه ابن ماجه)
“Tidak halal bagi seseorang menjual suatu barang kecuali jika ia menjelaskan cacat yang ada padanya”

Hadist ini mendukung prinsip bahwa segala kekurangan atau aib dalam layanan harus diinformasikan kepada pelanggan. Ketika hal ini tidak dilakukan, maka pelanggan berhak menuntut atau membatalkan transaksi. Dengan demikian, integrasi antara implementasi algoritma naïve bayes dan nilai-nilai keislaman memberikan peluang besar untuk menciptakan sistem layanan yang transparan dalam menginformasikan kualitas, adil dalam menanggapi keluhan, dan proaktif dalam memperbaiki kekurangan. Hal ini sesuai dengan prinsip keadilan (‘adl), kejujuran (sidq), dan amanah (amanah) dalam Islam, yang merupakan fondasi utama transaksi dalam muamalah.

BAB V PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan pada penelitian implementasi algoritma naïve bayes terhadap kepuasan penumpang pesawat, terbukti bahwa algoritma naïve bayes dapat digunakan untuk mengklasifikasi kepuasan penumpang pesawat. Klasifikasi yang dihasilkan dapat menjadi suatu keputusan apakah pelanggan termasuk dalam klasifikasi puas atau tidak puas sehingga dapat memberikan tolak ukur untuk maskapai dalam pengembangan fasilitas perusahaan.

Pada penelitian ini data mining dan algoritma naïve bayes dapat menampilkan hasil prediksi penumpang pesawat dengan menggunakan bahasa *python* diperoleh *confusion matriks* yang menghasilkan akurasi sebesar 86%. Kemudian didapat nilai rata-rata akurasi menggunakan *k-fold cross validation* sebesar 86.42%. Sehingga dapat dinyatakan bahwa performa model naïve bayes yang diterapkan dalam penelitian ini sudah baik karena memiliki nilai lebih dari 50%

5.2 Saran

Pada penelitian ini penulis hanya mengklasifikasi menggunakan metode naïve bayes. Penulis menyarankan hal-hal berikut untuk meningkatkan dan memperbaiki penelitian selanjutnya:

1. Menerapkan metode klasifikasi yang lain pada dataset
2. Penelitian dikembangkan dengan cara menggabungkan algoritma lain agar bisa membandingkan hasil dari masing-masing metode

DAFTAR PUSTAKA

- Ainurrohmah. (2021). Akurasi Algoritma Klasifikasi pada Software Rapidminer dan Weka. *Prosiding Seminar Nasional Matematika*, 4, 493–499.
<https://journal.unnes.ac.id/sju/index.php/prisma/>
- Anshori, M. A. I. (2020). *Perbandingan Metode Naive Bayes Classifier dengan K-Nearest Neighbor (KNN) untuk Klasifikasi Kategori Abstrak Skripsi*.
- BPS. (2024, April). *Berita Resmi Statistik*. Badan Pusat Statistik.
- ChengXiang, Z., & Sean, M. (2016). *Text Data Management and Analysis: A practical Introduction to Information Retrieval and Text Mining*.
- Dwiki, A., Putra, A., & Juanita, S. (2021). *Analisis Sentimen Pada Ulasan Pengguna Aplikasi Bibit Dan Bareksa Dengan Algoritma KNN*. 8(2).
<http://jurnal.mdp.ac.id>
- Erinsyah, M. F., Karenina, V., & Wibowo, S. D. (2023). Klasifikasi Rentang Usia Dan Gender Dengan Deteksi Suara Menggunakan Metode Deep Learning Algoritma Cnn (Convolutional Neural Network). *Jurnal Sistem Komputer*, 12(2), 195–202. <https://doi.org/10.34010/komputika.v12i2.10516>
- Głębocka, D. A., & Zdrodowska, M. (2021). Analysis children with disabilities self-care problems based on selected data mining techniques. *Procedia Computer Science*, 192, 2854–2862.
<https://doi.org/10.1016/j.procs.2021.09.056>
- Hariguna, T., Maulana Baihaqi, W., & Nurwanti, A. (2019). Sentiment Analysis of Product Reviews as A Customer Recommendation Using the Naive Bayes

- Classifier Algorithm. *International Journal of Informatics and Information Systems*, 2(2), 48–55.
- Hasan, F. N. (2017). *Keutamaan Pedagang yang Jujur dan Amanah*.
- Husna, A. P. A. (2011). Analisis Pengaruh Persepsi Harga, Citra, dan Kualitas Layanan Terhadap Kepuasan Pelanggan dalam Membangun Minat Mereferensikan (Studi Penelitian pada Pelanggan CV. Tria Jasa Boga Semarang). *Jurnal Sains Pemasaran Indonesia*, X, 1–32.
- Kemenag. (2019). *Qur'an Kemenag*.
- KLEIN, T. (2020). *Airline Passenger Satisfaction*.
<https://www.kaggle.com/code/frixinglife/airline-passenger-satisfaction#Reading-data>
- Lubis, R., & Ginanjar Mabrur, A. (2016). Penerapan Data Mining Untuk Memprediksi Kriteria Nasabah Kredit. *Universitas Komputer Indonsia*.
- Maududin, I. A., Mansur, A. T., & Supraha, W. (2021). MANAJEMEN PENDIDIKAN ISLAM PERIODE MADINAH PERPEKTIF AL MUBARAKFURI. *Islamic Management: Jurnal Manajemen Pendidikan Islam*, 4(01), 87. <https://doi.org/10.30868/im.v4i01.1059>
- Nasrullah, A. H. (2021). *Implementasi Algoritma Decision Tree Untuk Klasifikasi Produk Lairs*. 7(2). <http://ejournal.fikom-unasman.ac.id>
- Pebdika, A., Herdiana, R., & Solihudin, D. (2023). KLASIFIKASI MENGGUNAKAN METODE NAIVE BAYES UNTUK MENENTUKAN CALON PENERIMA PIP. Dalam *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Nomor 1).

- Putra, A. E. (2023). *Analisis Sentimen Tanggapan Masyarakat Kepada Video Youtube Mengenai Resesi 2023 Menggunakan Metode Naive Bayes dan K-Nearest Neighbor (KNN)*. Universitas Islam Negeri Syarif Hidayatullah.
- Putra Sanjaya, A., Indriani, N., & Si, S. (2018). *Penerapan Support Vector Mechine dengan Optimasi Particle Swarm Optimization Untuk Mendeteksi Penyakit Malaria*.
- Ramirez, S. G., Hales, R. C., Williams, G. P., & Jones, N. L. (2022). Extending SC-PDSI-PM with neural network regression using GLDAS data and Permutation Feature Importance. *Environmental Modelling and Software*, 157. <https://doi.org/10.1016/j.envsoft.2022.105475>
- Romadloni, P. L., Kusuma, B. A., & Baihaqi, W. M. (2022). Komparasi Metode Pembelajaran Mesin Untuk Implementasi Pengambilan Keputusan dalam Menentukan Promosi jabatan Karyawan. *Universitas Amikom Purwokerto*.
- Santoso, J. B., Kunci, K., Produk, K., Pelayanan, K., Harga, K., Konsumen, D., & Loyalitas, K. (2019). *Pengaruh Kualitas Produk, Kualitas Pelayanan, dan Harga Terhadap Kepuasan dan Loyalitas Konsumen (Studi Pada Konsumen Geprek Bensu Rawamangun)* (Vol. 16, Nomor 01).
- Sholikhah, I. I., Tri, A., Harjanta, J., & Latifah, K. (2023). *Machine Learning Untuk Deteksi Berita Hoax Menggunakan BERT*. <https://binus.ac.id/malang/2022/05/crisp-dm-cross-industry-standard-process-for-data-mining/>
- Utomo, D. P., & Mesran, M. (2020). Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung. *JURNAL*

MEDIA INFORMATIKA BUDIDARMA, 4(2), 437.

<https://doi.org/10.30865/mib.v4i2.2080>

Zainal Macfud, A., Pandu Kusuma, A., Dwi Puspitasari, W., Balitar Blitar Jl Majapahit No, I., Sananwetan, K., Blitar, K., & Timur, J. (2023). Analisis Algoritma Naive Bayes (NBC) Pada Klasifikasi Tingkat Minat Barang di Toko Violet Cell. Dalam *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Nomor 1).

LAMPIRAN

Lampiran 1 *Source code*

```
import os

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.model_selection import cross_val_score

from sklearn.Naive_bayes import GaussianNB

from sklearn.feature_selection import SelectKBest, f_classif

from sklearn.model_selection import train_test_split

from matplotlib import rcParams

rcParams['axes.titleweight'] = 'bold'

rcParams['axes.labelweight'] = 'bold'

rcParams['axes.labelpad'] = 20

rcParams['axes.titlepad'] = 20

sns.set_style('darkgrid')

sns.set_context('notebook', font_scale=.8)

# directory

dir = os.getcwd()

folder = 'D:\Skripshit\Python Airline\data'

# read train data

path = os.path.join(dir, folder)
```

```

train = pd.read_csv(os.path.join(path, 'train.csv'))

test = pd.read_csv(os.path.join(path, 'test.csv'))

# drop Unnamed: 0, and Id columns

train.drop(['Unnamed: 0', 'id'], axis=1, inplace=True)

test.drop(['Unnamed: 0', 'id'], axis=1, inplace=True)

# display data

train.head()

# display data information

train.info()

# plot target variable

fig, ax = plt.subplots(figsize=(6, 5))

sns.countplot(x='satisfaction', data=train, ax=ax, width=.25, palette='Set1')

ax.set_title('Distribution of the target variable')

ax.set_xlabel('Satisfaction')

ax.set_ylabel('Count')

# max value & total count

max = np.max(train['satisfaction'].value_counts())

total = train['satisfaction'].value_counts().sum()

for x, y in enumerate(train['satisfaction'].value_counts()):

    v = '%.2f%%' % (y / total * 100)

    ax.annotate(v, xy=(x, y), xytext=(0, 3), textcoords='offset points', ha='center',

va='bottom', fontsize=8)

ax.set_ylim(0, max * 1.1)

plt.show()

```

```

#data train

train.isnull().sum()

#fill missing values

mean = train['Arrival Delay in Minutes'].mean()

train['Arrival Delay in Minutes'].fillna(mean, inplace=True)

#setelah diisi dengan mean

train.isnull().sum()

#missing values in categorical column

for col in train.columns:

    if train[col].dtype == 'object':

        print(col, train[col].unique())

#plot satisfaction by customer type and gender

fig, ax = plt.subplots(1, 2, figsize=(12, 5), sharey=True)

sns.countplot(x='Customer Type', hue='satisfaction', data=train, ax=ax[0],

width=.5, palette='Set1')

ax[0].set_title('Satisfaction by customer type')

ax[0].set_xlabel('Satisfaction')

ax[0].set_ylabel('Count')

sns.countplot(x='Gender', hue='satisfaction', data=train, ax=ax[1], width=.5,

palette='Set1')

ax[1].set_title('Satisfaction by customer gender')

ax[1].set_xlabel('Satisfaction')

ax[1].set_ylabel('Count')

ax[0].legend_.remove()

```

```

ax[1].legend(loc='center left', frameon=False, title='Satisfaction',
bbox_to_anchor=(1, .5))

plt.tight_layout()

plt.show()

#plot satisfaction by type of travel and class

fig, ax = plt.subplots(1, 2, figsize=(12,5), sharey=True)

sns.countplot(x='Type of Travel', hue='satisfaction', data=train, ax=ax[0],
width=.5, palette='Set1')

ax[0].set_title('Satisfaction by type of travel')
ax[0].set_xlabel('Satisfaction')
ax[0].set_ylabel('Count')

sns.countplot(x='Class', hue='satisfaction', data=train, ax=ax[1], width=.75,
palette='Set1')

ax[1].set_title('Satisfaction by class')
ax[1].set_xlabel('Satisfaction')
ax[1].set_ylabel('Count')

ax[0].legend_.remove()

ax[1].legend(loc='center left', frameon=False, title='Satisfaction',
bbox_to_anchor=(1, .5))

plt.tight_layout()

plt.show()

fig, ax = plt.subplots(1, 2, figsize=(12,5))

colors = sns.color_palette()

sns.histplot(train['Age'], ax=ax[0], bins=20, color=colors[0], kde=True)

```

```

ax[0].set_title('Distribution of Age')

ax[0].set_xlabel('Age')

ax[0].set_ylabel('Density')

sns.histplot(test['Flight Distance'], ax=ax[1], bins=20, color=colors[1], kde=True)

ax[1].set_title('Distribution of flight distance')

ax[1].set_xlabel('Flight Distance')

ax[1].set_ylabel('Density')

plt.tight_layout()

plt.show()

#all numerical columns

numeric = train.select_dtypes(include=np.number).columns

numeric = numeric.drop(['Age', 'Flight Distance', 'Departure Delay in Minutes',
'Arrival Delay in Minutes'])

col = 2

row = len(numeric) // col

for i in range(row):

    fig, ax = plt.subplots(1, 2, figsize=(12,5))

    for j in range(col):

        data = train[numeric[i * col + j]]

        sns.countplot(x=data, hue='satisfaction', data=train, ax=ax[j], palette='Set1')

        ax[j].set_title('Distribution of ' + ' ' + numeric[i * col+j])

        ax[j].set_xlabel(numeric[i * col + j])

        ax[j].set_ylabel('Density')

```

```

ax[1].legend(loc='center left', frameon=False, title='Class', bbox_to_anchor=(1,
.5))

plt.tight_layout()

plt.show()

#average rating for each numeric

avg = train[numeric].mean().sort_values(ascending=False)

fig, ax = plt.subplots(figsize=(8, 8))

sns.barplot(x=avg.values, y=avg.index, ax=ax, palette='tab10', orient='h')

ax.set_title('Average rating')

ax.set_xlabel('Feature')

ax.set_ylabel('Average rating')

for x, y in enumerate(avg.values):

    ax.annotate('{:.2f}'.format(y), xy=(y, x), va='center', ha='left', fontsize=8,
xytext=(5, 0), textcoords='offset points')

ax.set_xlim(0, np.max(avg.values) * 1.1)

plt.tight_layout()

plt.show()

from sklearn.preprocessing import LabelEncoder

# encode catagorical data

encoder = LabelEncoder()

# for train data

for item in train.columns:

    if train[item].dtype == 'object':

        train[item] = encoder.fit_transform(train[item])

```



```
# for test data

for item in test.columns:

    if test[item].dtype == 'object':

train.describe()

X_train = train.drop('satisfaction', axis=1)

y_train = train['satisfaction']

X_test = test.drop('satisfaction', axis=1)

y_test = test['satisfaction']

y_train = y_train.to_numpy()

y_test = y_test.to_numpy()

from sklearn.preprocessing import MinMaxScaler

scaller = MinMaxScaler()

X_train = scaller.fit_transform(X_train)

X_test = scaller.transform(X_test)

from sklearn.metrics import accuracy_score

from sklearn.metrics import roc_auc_score

from sklearn.metrics import classification_report

from sklearn.metrics import confusion_matrix

from sklearn.Naive_bayes import GaussianNB

model = GaussianNB()

model.fit(X_train, y_train)

y_pred = model.predict(X_test)

accuracy = accuracy_score(y_test, y_pred)

roc_auc = roc_auc_score(y_test, y_pred)
```

```

print('%-10s %.2f % ('Accuracy', accuracy))

print('%-10s %.2f % ('ROC UAC', roc_auc))

print()

print(classification_report(y_test, y_pred))

fig, ax = plt.subplots(figsize=(5, 4))

sns.heatmap(confusion_matrix(y_test, y_pred), annot=True, fmt='d',
cmap='Blues', ax=ax)

ax.set_title('Confusion matrix')

ax.set_xlabel('Predicted')

ax.set_ylabel('Actual')

ax.xaxis.set_ticklabels(['satisfied', 'unsatisfied'])

ax.yaxis.set_ticklabels(['satisfied', 'unsatisfied'])

plt.show()

F = 5

scores = cross_val_score(model, X_train, y_train, cv=F, scoring='accuracy')

for fold, score in enumerate(scores):

    print(f'Fold {fold + 1}: Accuracy = {score * 100:.2f} %')

mean_accuracy = np.mean(scores)*100

std_accuracy = np.std(scores) * 100

print(f'Mean Accuracy = {mean_accuracy:.2f} %')

print(f'Standar Deviation = {std_accuracy:.2f} %')

from sklearn.inspection import permutation_importance

import pandas as pd

```

```
feature_names = train.drop("satisfaction", axis=1).columns

#ndarray to dataframe

X_train = pd.DataFrame(X_train, columns=feature_names)

X_test = pd.DataFrame(X_test, columns=feature_names)

result = permutation_importance(model, X_train, y_train, n_repeats=10,
                                random_state=42)

sorted_idx = result.importances_mean.argsort()

plt.figure(figsize=(10, 6))

plt.barh(range(len(sorted_idx)), result.importances_mean[sorted_idx],
         align='center')

plt.yticks(range(len(sorted_idx)), X_train.columns[sorted_idx])

plt.title("Feature Importance")

plt.xlabel("Mean Decrease in Accuracy")

plt.show()
```

RIWAYAT HIDUP



Jihan Salsabilah, lahir di Bengkulu pada 29 Desember 2001.

Penulis telah menempuh pendidikan mulai dari TK Kemala Bhayangkari dan lulus pada tahun 2008, dilanjutkan menempuh sekolah dasar di SDN 09 Bengkulu Utara dan lulus

pada tahun 2014. Kemudian penulis melanjutkan pendidikan sekolah menengah pertama di SMPN 01 Bengkulu Utara dan lulus pada tahun 2016. Selanjutnya menempuh pendidikan sekolah menengah atas di SMAN 01 Bengkulu Utara dan lulus pada tahun 2019. Pada tahun yang sama, penulis melanjutkan pendidikan di Universitas Islam Negeri Maulana Malik Ibrahim Malang pada Program Studi Matematika, Fakultas Sains dan Teknologi.

Selama menempuh pendidikan di perguruan tinggi, penulis aktif mengikuti beberapa kegiatan baik di dalam lingkup kampus maupun luar kampus. Penulis bergabung dalam organisasi mahasiswa yaitu Himpunan Mahasiswa Jurusan (HMJ) “Integral” Matematika selama satu periode sebagai anggota divisi kewirausahaan.



KEMENTERIAN AGAMA RI
UNIVERSITAS ISLAM NEGERI
MAULANA MALIK IBRAHIM MALANG
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Gajayana No.50 Dinoyo Malang Telp. / Fax. (0341)558933

BUKTI KONSULTASI SKRIPSI

Nama : Jihan Salsabilah
NIM : 19610073
Fakultas / Jurusan : Sains dan Teknologi / Matematika
Judul Skripsi : Implementasi Algoritma Naive Bayes Terhadap
Klasifikasi Kepuasan Penumpang pesawat
Pembimbing I : Juhari, M.Si
Pembimbing II : Erna Herawati, M.Pd

No	Tanggal	Hal	Tanda Tangan
1.	15 Februari 2025	Konsultasi Bab I, II, dan III	1. JH
2.	10 Maret 2025	Konsultasi Bab I, II, dan III	2. JH
3.	15 Maret 2025	Konsultasi Bab I, II, dan III	3. JH
4.	20 Maret 2025	ACC Bab I, II, dan III	4. JH
5.	2 April 2025	Konsultasi Kajian Agama Bab I dan II	5. JH
6.	8 April 2025	Konsultasi Kajian Agama Bab I dan II	6. JH
7.	1 Mei 2025	Konsultasi Kajian Agama Bab I dan II	7. JH
8.	15 Mei 2025	ACC Kajian Agama Bab I dan II	8. JH
9.	10 Juni 2025	ACC Seminar Proposal	9. JH
10.	8 Juli 2025	Konsultasi Revisi Seminar Proposal	10. JH
11.	5 Agustus 2025	Konsultasi Bab IV	11. JH
12.	20 Agustus 2025	Konsultasi Bab IV	12. JH
13.	10 September 2025	Konsultasi Bab IV dan V	13. JH
14.	8 Oktober 2025	ACC Bab IV dan V	14. JH
15.	15 Oktober 2025	Konsultasi Kajian Agama Bab IV	15. JH



KEMENTERIAN AGAMA RI
UNIVERSITAS ISLAM NEGERI
MAULANA MALIK IBRAHIM MALANG
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Gajayana No.50 Dinoyo Malang Telp. / Fax. (0341)558933

16.	27 November 2025	Konsultasi Kajian Agama Bab IV	16.
17.	28 November 2025	ACC Kajian Agama Bab IV	17.
18.	28 November 2025	ACC Seminar Hasil	18.
19.	4 Desember 2025	Konsultasi Revisi Seminar Hasil	19.
20.	11 Desember 2025	ACC Sidang Skripsi	20.
21.	17 Desember 2025	ACC Keseluruhan	21.

Malang, 17 Desember 2025

Mengetahui,

Ketua Program Studi Matematika



Dr. Fachrur Rozi, M.Si.

NIP. 19800527 200801 1 012