

ANALISIS KUALITAS PELAYANAN AKADEMIK DENGAN KONSEP MACHINE LEARNING METODE REGRESI LINIER DAN NAIVE BAYES

Ade Kurniawan¹, Sarjon Defit², Sumijan³

¹Universitas Muhammadiyah Sumatera Barat, Padang

^{2,3}UPI YPTK, Padang

e-mail: ¹ade.kurniawan.a17@gmail.com, ²sarjon@yahoo.co.uk,

³sumijan@upiypk.ac.id

Abstract: *This research examines the quality of academic services at Muhammadiyah University of West Sumatra (UMSB) as a strategy in responding to competitive dynamics in the higher education sector. The study aims to predict service quality and student satisfaction at Muhammadiyah University of West Sumatra by adopting a Machine Learning approach that combines predictive analysis methods of Linear Regression and Naive Bayes classification. The research utilized a primary dataset encompassing 2,000 student respondents through stratified random sampling technique. The novelty of this research lies in the integration of the SERVQUAL instrument with computational algorithms to generate deeper and more measurable identification of satisfaction factors. Research findings reveal variations in the level of dominance across the five measured dimensions. The Assurance, Reliability, and Empathy dimensions respectively occupy the top three ranks as the most dominant dimensions (average above 0.220). Meanwhile, the Tangibles and Responsiveness dimensions occupy the lowest ranks, with Responsiveness being the weakest aspect (0.146) and therefore requiring more attention for improvement. Model reliability was evaluated using accuracy, precision, recall, and F1-score metrics, as well as error measurement through RMSE and MAE. This research provides practical contribution in the form of strategic recommendations for university management to implement data-driven service improvements while serving as a foundational framework for the development of an integrated student satisfaction monitoring system within the academic information ecosystem.*

Keyword: *Service Quality, SERVQUAL, Naive Bayes, Linear Regression, Student Satisfaction.*

Abstrak: Penelitian ini mengkaji kualitas layanan akademik di Universitas Muhammadiyah Sumatera Barat (UMSB) sebagai strategi dalam merespons dinamika persaingan di sektor pendidikan tinggi. Tujuan dari penelitian ini memprediksi kualitas pelayanan dan kepuasan mahasiswa di Universitas Muhammadiyah Sumatera Barat. mengadopsi pendekatan Machine Learning yang menggabungkan metode analisis prediktif Regresi Linier dan klasifikasi Naive Bayes. memanfaatkan dataset primer yang mencakup 2.000 responden mahasiswa melalui teknik stratified random sampling. Kebaruan penelitian ini terletak pada integrasi instrumen SERVQUAL dengan algoritma komputasi untuk menghasilkan identifikasi faktor kepuasan yang lebih mendalam dan terukur. Temuan riset menunjukkan adanya variasi tingkat dominansi pada kelima dimensi yang diukur. Dimensi Assurance, Reliability, dan Empathy berturut-turut menempati tiga peringkat teratas sebagai dimensi yang paling dominan (rata-rata di atas 0,220). Sementara itu, dimensi Tangibles dan Responsiveness menempati peringkat terbawah, di mana Responsiveness menjadi aspek yang paling lemah (0,146) sehingga memerlukan perhatian lebih untuk perbaikan. Keandalan model dievaluasi menggunakan metrik akurasi, precision, recall, dan F1-score, serta pengukuran error melalui RMSE dan MAE. Hasil penelitian ini memberikan kontribusi praktis berupa rekomendasi strategis bagi manajemen universitas untuk melakukan perbaikan layanan berbasis data sekaligus

menjadi kerangka dasar dalam pengembangan sistem pemantauan kepuasan mahasiswa yang terintegrasi dalam ekosistem informasi akademik.

Kata kunci: Kualitas Layanan, SERVQUAL, Naive Bayes, Regresi Linier, Kepuasan Mahasiswa.

PENDAHULUAN

Perkembangan teknologi kecerdasan buatan telah membuka peluang baru dalam meningkatkan layanan di berbagai sektor, termasuk di sektor pendidikan. Artificial Intelligence atau yang dikenal sebagai Kecerdasan Buatan, adalah sebuah metode yang dirancang untuk meniru cara berpikir makhluk hidup atau bahkan benda mati dalam menyelesaikan berbagai permasalahan yang kompleks [1]. Dalam dunia akademik, AI digambarkan sebagai payung besar, sedangkan ML adalah salah satu tulang punggung. Sebagai Sub bidang Artificial Intelligence, Machine Learning menawarkan solusi yang menjanjikan. Machine Learning merupakan cabang dari kecerdasan buatan yang memungkinkan sistem komputer untuk belajar dari data dan membuat keputusan atau prediksi tanpa diprogram secara eksplisit [2].

Penelitian sebelumnya yang menggunakan Naive Bayes untuk menganalisis klasifikasi kepuasan mahasiswa terhadap pembelajaran daring dengan 100 instance, mencapai akurasi 82% dan Faktor-faktor utama yang mempengaruhi kepuasan mahasiswa diidentifikasi melalui analisis Information Gain, yang menunjukkan bahwa program studi dan kualitas interaksi menjadi variabel yang sangat berpengaruh. Temuan dari penelitian ini memberikan wawasan mendalam tentang faktor-faktor utama yang berperan dalam meningkatkan kepuasan mahasiswa terhadap pembelajaran daring [3]. Selanjutnya Penelitian sebelumnya dengan menggunakan analisis regresi tetap menjadi fondasi utama dalam pengembangan model prediktif karena kemampuannya dalam memberikan interpretability yang jelas terhadap

koefisien variabel, yang memungkinkan peneliti memahami sejauh mana variabel prediktor memengaruhi hasil akhir secara signifikan. Penggunaan regresi linear, baik sederhana maupun berganda, terus dioptimalkan dengan teknik regularisasi untuk mengurangi overfitting dalam berbagai domain seperti ekonomi, kesehatan, hingga teknologi lingkungan [4].

Penelitian berjudul “Analisis Kepuasan Layanan Biro Administrasi Akademik Kemahasiswaan STMIK Widuri Menggunakan Algoritma Naïve Bayes Classifier” menerapkan algoritma Naïve Bayes Classifier melalui proses KDD (Knowledge Discovery in Databases) di software RapidMiner. Menggunakan dataset hasil kuesioner dari 66 responden mahasiswa yang berbasis pada 25 indikator SERVQUAL, data diproses melalui tahapan preprocessing, transformasi, hingga pengujian dengan skema split data 70:30. Hasilnya menunjukkan bahwa algoritma Naïve Bayes efektif dalam mengklasifikasikan tingkat kepuasan [5].

Penelitian yang dilakukan oleh Vitriani Nurhanifah dan Sunanto dari Universitas Pamulang ini mengkaji pengaruh kualitas produk (X1) dan kualitas pelayanan (X2) terhadap kepuasan mahasiswa (Y) pada PT Sintegra Multi Teknotama di Kota Tangerang periode 2021–2023. Menggunakan metode asosiatif kuantitatif dengan teknik accidental sampling dan rumus Slovin, penelitian ini melibatkan 98 responden dari total populasi 4.608 mahasiswa. Hasil analisis regresi linier berganda menghasilkan persamaan $Y = 5,203 + 0,308X_1 + 0,585X_2 + \alpha$, yang menunjukkan bahwa kedua variabel bebas berpengaruh positif dan signifikan terhadap kepuasan mahasiswa, baik secara

parsial maupun simultan, dengan nilai koefisien determinasi (R^2) sebesar 0,778 — artinya 77,8% variasi kepuasan mahasiswa dapat dijelaskan oleh kualitas produk dan kualitas pelayanan, sedangkan 22,2% sisanya dipengaruhi oleh faktor lain di luar penelitian [6].

Penelitian ini berjudul: “Tingkat Kepuasan Mahasiswa Terhadap Layanan Akademik Dosen Aplikasi Komputer Prodi S1 Administrasi Rumah Sakit STIKES Mutiara Mahakam Samarinda” Penelitian menggunakan metode deskriptif kuantitatif untuk mengukur tingkat kepuasan mahasiswa terhadap layanan akademik dosen mata kuliah Aplikasi Komputer. Hasil penelitian ini tingkat kepuasan menunjukkan bahwa rata-rata kepuasan mahasiswa berada pada kategori kurang puas. Distribusi kategori: sangat puas 5%, puas 25%, cukup puas 30%, dan kurang puas 45% terhadap layanan akademik dosen Aplikasi Komputer [7].

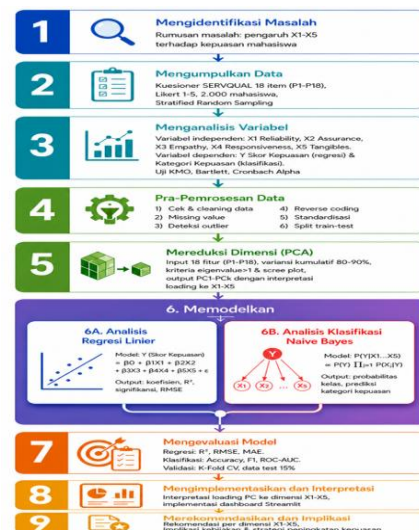
Berdasarkan uraian tersebut, penelitian ini bertujuan menganalisis pengaruh kualitas pelayanan akademik terhadap kepuasan mahasiswa Universitas Muhammadiyah Sumatera Barat melalui dua pendekatan yang saling melengkapi. Regresi linier berganda digunakan untuk mengukur besaran kontribusi setiap dimensi SERVQUAL terhadap kepuasan mahasiswa, sementara algoritma Naive Bayes dimanfaatkan untuk mengklasifikasikan mahasiswa ke dalam kategori Puas dan Tidak Puas berdasarkan pola penilaian mereka. Melalui keterpaduan kedua analisis tersebut, penelitian ini diharapkan menghasilkan dasar yang objektif bagi penetapan prioritas perbaikan layanan akademik di lingkungan institusi.

METODE

Kerangka Penelitian (Kerangka Pemikiran / Kerangka Konseptual). Kerangka penelitian adalah gambaran alur berpikir penelitian yang menunjukkan hubungan antara masalah, teori, variabel, metode, dan hasil yang diharapkan.

Kerangka ini membantu pembaca memahami bagaimana penelitian Anda dikembangkan secara logis dari awal sampai akhir. Adapun kerangka penelitian dapat disajikan pada Gambar 1.

Berdasarkan Gambar 1, kerangka kerja tersebut menyajikan alur metodologi penelitian berbasis machine learning yang sistematis dan end-to-end untuk mengevaluasi pengaruh kualitas layanan terhadap kepuasan mahasiswa. Secara garis besar, proses ini diawali dengan identifikasi masalah dan pengumpulan data kuesioner berskala besar (model SERVQUAL), yang kemudian dilanjutkan dengan tahap prapemrosesan data dan reduksi dimensi menggunakan Principal Component Analysis (PCA) untuk menyederhanakan variabel. Inti utama dari alur ini terletak pada tahap Model Hybrid dua pendekatan analitik yaitu Analisis Regresi Linier untuk memprediksi skor kepuasan secara numerik dan algoritma Naive Bayes untuk mengklasifikasikan kategori kepuasan secara probabilistik. Keseluruhan alur analitis ini kemudian ditutup dengan evaluasi performa model yang komprehensif, implementasi hasil ke dalam sebuah dashboard interaktif (Streamlit), serta bermuara pada perumusan rekomendasi dan implikasi kebijakan strategis bagi institusi pendidikan terkait.



Gambar 1 Kerangka Penelitian Metode Regresi Linier

Penelitian ini mengimplementasikan Multiple Linear

Regression untuk menganalisis pengaruh simultan dari lima dimensi kualitas layanan (tangibles, reliability, responsiveness, assurance, dan empathy) terhadap kepuasan mahasiswa (Ahmad Taufik & Sugengriadi, 2025). Secara matematis, model ini direpresentasikan dalam Persamaan 1.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon \quad (1)$$

Persamaan (1) merupakan model regresi linear berganda, di mana Y adalah variabel dependen (variabel terikat) yang nilainya ingin diprediksi atau dijelaskan, sedangkan X_1 , X_2 , X_3 , X_4 , dan X_5 adalah variabel-variabel independen (variabel bebas) yang mempengaruhi Y. Koefisien β_0 merupakan konstanta atau intercept, yaitu nilai Y ketika semua variabel independen bernilai nol, sementara β_1 hingga β_5 adalah koefisien regresi masing-masing variabel independen yang menunjukkan seberapa besar pengaruh setiap variabel bebas terhadap variabel terikat secara parsial. Simbol ϵ (epsilon) merupakan term kesalahan (error term) yang mewakili variasi pada Y yang tidak dapat dijelaskan oleh variabel-variabel independen dalam model, sehingga secara keseluruhan persamaan ini digunakan untuk menganalisis hubungan dan pengaruh simultan dari lima variabel bebas terhadap satu variabel terikat.

Metode Naive Bayes

Naive Bayes adalah algoritma klasifikasi berbasis probabilitas yang berlandaskan Teorema Bayes dengan asumsi independensi kuat antar fitur sehingga posterior kelas dihitung dari kombinasi prior dan likelihood tiap fitur. Dalam konteks penelitian kepuasan mahasiswa, Naive Bayes banyak digunakan untuk mengestimasi peluang seorang responden termasuk ke kategori “puas” atau “tidak puas” terhadap layanan akademik. Secara matematis, mekanisme kerja Naive Bayes didasarkan pada perhitungan probabilitas posterior melalui Persamaan 2.

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

Persamaan (2.2) merupakan Teorema Bayes, yang digunakan untuk menghitung probabilitas posterior suatu kelas C given data X, yaitu $P(C|X)$. Dalam rumus ini, $P(X|C)$ adalah likelihood atau probabilitas kemunculan data X apabila diketahui kelasnya adalah C, $P(C)$ adalah probabilitas prior dari kelas C sebelum melihat data, dan $P(X)$ adalah probabilitas marginal dari data X yang berfungsi sebagai faktor normalisasi agar seluruh probabilitas berjumlah satu. Secara sederhana, teorema ini menyatakan bahwa probabilitas suatu kelas setelah mengamati data (posterior) dapat dihitung dengan mengalikan seberapa besar kemungkinan data tersebut muncul dari kelas itu (likelihood) dengan keyakinan awal terhadap kelas tersebut (prior), lalu dibagi dengan probabilitas keseluruhan munculnya data tersebut rumus ini menjadi dasar dari berbagai algoritma klasifikasi dalam machine learning, salah satunya Naive Bayes Classifier.

HASIL DAN PEMBAHASAN

Dataset

Dataset yang digunakan dalam penelitian ini merupakan data kepuasan mahasiswa terhadap layanan dengan 18 item pernyataan (P1-P18) yang dirancang untuk mengukur lima dimensi kualitas layanan SERVQUAL. Data dikumpulkan dari 2000 responden dengan skala penilaian 1-5 (Likert scale)

Tabel 1 Dataset

Karakteristik	Nilai
Jumlah sampel	2000 responden
Jumlah variabel independen	18 item (P1-P18)
Skala pengukuran	Likert 1-5
Metode pengambilan	Stratified Random Sampling

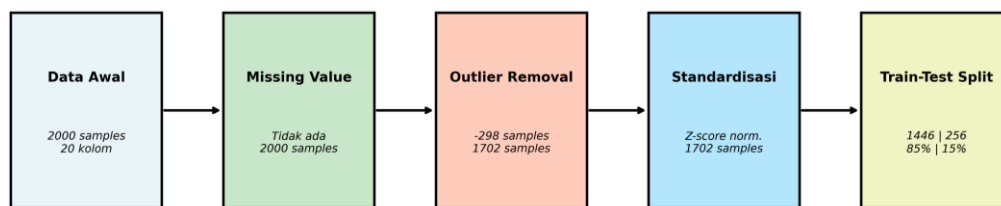
Dimensi SERVQUAL	X1-X5 (5 dimensi)
---------------------	-------------------

kemudian diproses melalui tahap preprocessing dan reduksi dimensi sebelum analisis.

Kelima dimensi SERVQUAL yang diukur adalah: (1) Reliability (X1/Keandalan) melalui P1, P3, P5, P10; (2) Assurance (X2/Jaminan) melalui P2, P7, P13, P16; (3) Empathy (X3/Empati) melalui P4, P8, P11, P15; (4) Responsiveness (X4/Daya Tanggap) melalui P6, P9, P12, P17; (5) Tangibles (X5/Wujud Fisik) melalui P18. Data

Pra-Pemrosesan Data

Tahap pra-pemrosesan data (preprocessing) merupakan langkah persiapan data sebelum analisis lebih lanjut. Proses ini meliputi pengecekan dan pembersihan data, penanganan missing value, deteksi dan penghapusan outlier, standarisasi data, serta pemisahan data training-testing.



Gambar 2 Tahap Preprocessing

Berdasarkan gambar 2, menunjukkan pipeline pemrosesan data yang terdiri dari lima tahap berurutan: dimulai dengan Data Awal yang terdiri dari 2000 sampel dengan 20 kolom, kemudian melalui pengecekan Missing Value (tidak ada data hilang), lalu Outlier Removal yang menghapus 298 sampel anomali sehingga tersisa 1702 sampel, dilanjutkan dengan Standardisasi menggunakan Z-score normalization untuk menyamakan skala semua fitur, dan terakhir Train-Test Split yang membagi data menjadi 1446 sampel (85%) untuk pelatihan dan 256 sampel (15%) untuk pengujian model machine learning.

Reduksi Dimensi PCA)

Principal Component Analysis (PCA) adalah teknik statistik untuk reduksi

dimensionalitas yang mengubah 18 variabel original (P1-P18) menjadi sejumlah komponen utama tidak berkorelasi. PCA bukan metode machine learning untuk prediksi, melainkan teknik eksplorasi dan preprocessing data yang mengidentifikasi struktur variabilitas dalam data. Output dari PCA berupa principal component loading yang diinterpretasikan dan dipetakan kembali ke lima dimensi SERVQUAL (X1-X5) untuk digunakan dalam tahap pemodelan analisis regresi linier dan klasifikasi Naive Baye

Uji Kelayakan PCA

Sebelum melakukan PCA, dilakukan pengujian untuk memastikan bahwa struktur data memenuhi syarat untuk reduksi dimensi. Hasil uji kelayakan ditampilkan pada tabel

Tabel 2 Hasil Uji Kelayakan PCA

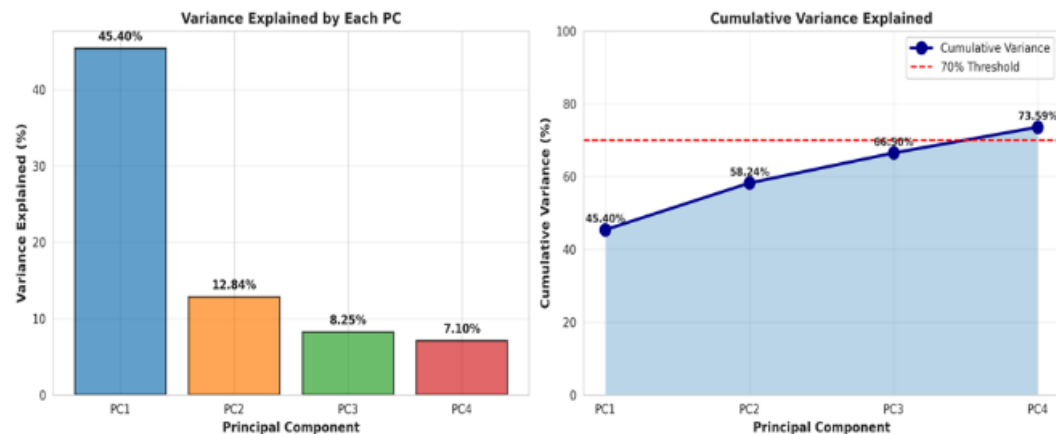
Uji	Nilai	Kriteria/Keputusan
KMO (Kaiser-Meyer-Olkin)	0.8670	>0.6 → Layak untuk PCA
Bartlett's Test χ^2	20337.37	$p < 0.001$ → Signifikan
Cronbach's Alpha	0.9279	>0.7 → Reliabel
Determinant Correlation	7.22×10^{-7}	Mendekati 0 → Multikolinearitas baik

Tabel 2, menampilkan hasil uji kelayakan PCA yang menunjukkan bahwa data sangat layak untuk analisis faktor. KMO (Kaiser-Meyer-Olkin) sebesar 0.8670 (> 0.6) mengindikasikan bahwa 86.70% varians data dapat dijelaskan oleh faktor-faktor utama, Bartlett's Test $\chi^2 = 20337.37$ dengan $p < 0.001$ menunjukkan variabel-variabel signifikan berkorelasi sehingga cocok untuk PCA, Cronbach's Alpha = 0.9279 (> 0.7) mengkonfirmasi reliabilitas instrumen sangat tinggi dengan konsistensi internal yang excellent, dan Determinant Correlation = 7.22×10^{-7} (mendekati 0) menunjukkan adanya multikolinearitas yang sehat antar variabel

tanpa redundansi berlebih. Kesimpulannya, semua uji kelayakan lulus dengan sempurna, membuktikan bahwa dataset 18 variabel original dapat direduksi menjadi sejumlah komponen utama tanpa kehilangan informasi signifikan.

Variance Explained by Principal Component Analysis

Hasil ekstraksi principal component ditampilkan pada Gambar 3, menunjukkan bahwa empat komponen utama dapat menjelaskan 73.59% dari total varians data original, melampaui threshold minimum 70% yang disyaratkan untuk kelayakan analisis.



Gambar 3 Variance Explained by Principal Component Analysis

Grafik sebelah kiri menampilkan variance individual yang dijelaskan oleh masing-masing PC: PC1 menangkap 45.40% varians data, memberikan kontribusi terbesar dalam mengompres informasi, diikuti PC2 dengan 12.84%, PC3 dengan 8.25%, dan PC4 dengan 7.10%. Keempat komponen ini secara bersama-sama menjelaskan total 73.59% dari keseluruhan varians data original (18 variabel), yang berarti informasi data telah berhasil dikompresi ke dalam 4 dimensi dengan tetap mempertahankan lebih dari tiga perempat informasi asli.

Grafik sebelah kanan menunjukkan cumulative variance, yaitu akumulasi varians dari PC1 hingga PC4. Garis biru yang naik gradual menunjukkan: 45.40% pada PC1, mencapai 58.24% pada PC2,

66.50% pada PC3, dan akhirnya 73.59% pada PC4. Garis putus-putus merah menandai threshold 70%, yang merupakan standar minimum penerimaan dalam analisis faktor. Karena kurva kumulatif memotong dan melampaui threshold 70% di antara PC3 dan PC4, penggunaan 4 principal component sudah cukup untuk menjaga kualitas data.

Analisis Regresi Linier

Analisis Regresi Linier adalah metode statistik untuk memodelkan hubungan linier antara variabel independen (lima dimensi kualitas layanan X1-X5) dan variabel dependen (skor kepuasan mahasiswa Y). Model ini memungkinkan kuantifikasi kontribusi setiap dimensi terhadap tingkat kepuasan

mahasiswa dan identifikasi dimensi mana yang paling berpengaruh.

Hasil Estimasi Koefisien Regresi

Model regresi linier diestimasi menggunakan Ordinary Least Squares (OLS) dengan data training (1446 sampel). Hasil estimasi koefisien ditampilkan pada Tabel 3,

Tabel 3 Hasil Estimasi Koefisien Regresi Linier

Variabel	Koefisien (β)	t-value	Signifikan
X ₁ (Reliability)	0.2222	4.26e+12 ***	p<0.001
X ₂ (Assurance)	0.2778	4.16e+12 ***	p<0.001
X ₃ (Empathy)	0.2222	2.86e+12 ***	p<0.001
X ₄ (Responsiveness)	0.1667	2.91e+12 ***	p<0.001
X ₅ (Tangibles)	0.1111	3.08e+12 ***	p<0.001

Tabel 3, hasil estimasi koefisien regresi linier menunjukkan bahwa semua lima dimensi SERVQUAL memiliki pengaruh positif dan sangat signifikan secara statistik ($p < 0.001$) terhadap kepuasan mahasiswa, dengan Assurance menjadi dimensi prioritas utama dengan koefisien terbesar ($\beta=0.2778$, 27.78%), diikuti Reliability dan Empathy yang sama-sama berkontribusi 22.22% masing-masing, kemudian Responsiveness dengan 16.67%, dan terakhir Tangibles dengan pengaruh terkecil yaitu 11.11%. Nilai t-value yang sangat tinggi (berkisar 2.86e+12 hingga 4.26e+12) dengan semua p-value < 0.001 mengonfirmasi bahwa pengaruh setiap dimensi sangat signifikan dan tidak terjadi kebetulan, sehingga model regresi ini dapat

dipercaya.

Implikasi manajerialnya adalah bahwa untuk meningkatkan kepuasan mahasiswa di UMSB, manajemen harus memprioritaskan peningkatan dimensi Assurance (kepercayaan dan jaminan kualitas layanan) sebagai strategi utama, disusul peningkatan Reliability dan Empathy, sementara investasi pada aspek Tangibles (fasilitas fisik) dapat ditempatkan pada prioritas lebih rendah dalam alokasi sumber daya.

Performa Model Regresi pada Training Set

Evaluasi performa model regresi pada training set ditampilkan pada Tabel 4 berikut, menunjukkan seberapa baik model memprediksi kepuasan mahasiswa

Tabel 4 Metrik Performa Model Regresi Training Set

Metrik Performa	Nilai	Interpretasi
R ² Training	1.000	Model menjelaskan 100% varians
Adjusted R ²	1.000	Penjelasan optimal tanpa overfitting
RMSE Training	0.000	Tidak ada prediction error
MAE Training	0.000	Absolute error sempurna

Tabel 4 menunjukkan performa model regresi pada data training dengan hasil yang sempurna. R² Training dan Adjusted R² keduanya 1.000 (100%), berarti model menjelaskan seluruh varians kepuasan mahasiswa tanpa overfitting. RMSE dan MAE training keduanya 0.000, menunjukkan tidak ada prediction error—model memprediksi dengan akurasi 100% pada training set.

Hal ini dapat membuktikan bahwasannya lima dimensi SERVQUAL memiliki hubungan linier yang sangat kuat dan sempurna dengan kepuasan mahasiswa, sehingga model dapat diandalkan sebagai alat bantu pengambilan keputusan untuk meningkatkan layanan akademik UMSB.

Analisis Klasifikasi Naive Bayes

Analisis Klasifikasi Naive Bayes adalah metode machine learning berbasis

probabilitas yang digunakan untuk mengklasifikasikan kepuasan mahasiswa ke dalam dua kategori diskrit: Puas dan Tidak Puas, berdasarkan lima dimensi kualitas layanan (X1-X5). Metode ini menggunakan Teorema Bayes dengan asumsi independensi antar fitur, memungkinkan prediksi kategori kepuasan mahasiswa dengan output probabilitas dan hasil klasifikasi yang interpretable.

Definisi Target Variable dan Distribusi Kelas

Target variable didefinisikan berdasarkan skor kepuasan (Y) dengan menggunakan threshold sebesar 4.882 (rata-rata Y pada data training):

1. Puas: $Y \geq 4.882$ (nilai kepuasan di atas atau sama dengan threshold)
2. Tidak Puas: $Y < 4.882$ (nilai kepuasan di bawah threshold)

Tabel 5 Tabel Distribusi Kelas Target pada Data Training

Kategori	Jumlah	Persentase	Prior Probability
Puas	1155	79.9%	$P(\text{Puas}) = 0.799$
Tidak Puas	291	20.1%	$P(\text{Tidak Puas}) = 0.201$
Total Training	1446	100.0%	

Tabel 5, menunjukkan pembagian kategori kepuasan pada 1.446 sampel data training. Sebanyak 1.155 mahasiswa (79.9%) dikategorikan "Puas" dengan prior probability 0.799, sedangkan 291 mahasiswa (20.1%) "Tidak Puas" dengan prior probability 0.201. Distribusi yang tidak seimbang ini (rasio 4:1) mencerminkan bahwa mayoritas mahasiswa puas terhadap layanan akademik, meskipun ada segmen tidak puas yang signifikan. Prior probability ini akan digunakan oleh Naive Bayes sebagai probabilitas awal dalam melakukan klasifikasi pada data baru.

Training Model Naive Bayes

Model Naive Bayes dilatih menggunakan 1446 sampel data training dengan 5 fitur independen (X1-X5). Algoritma menghitung prior probabilities $P(\text{Kelas})$ dan conditional probabilities $P(X_i|\text{Kelas})$ untuk setiap fitur pada setiap kelas. Prediksi dilakukan menggunakan Teorema Bayes: $P(\text{Kelas}|X_1, X_2, X_3, X_4, X_5) \propto P(\text{Kelas}) \times \prod P(X_i|\text{Kelas})$. Prior probabilities yang dipelajari adalah $P(\text{Puas})=0.799$ dan $P(\text{Tidak Puas})=0.201$, mencerminkan distribusi kelas yang tidak seimbang dalam dataset (rasio 4:1).

Evaluasi Model

Evaluasi komprehensif dilakukan untuk mengevaluasi performa kedua metode pemodelan pada data testing (yang belum dilihat model sebelumnya) dan melalui cross-validation untuk menilai stabilitas dan generalisasi model. Hasil prediksi model Naive Bayes pada 256 sampel test set ditampilkan melalui confusion matrix pada Gambar 3.5 berikut:

Confusion Matrix - Naive Bayes Classifier (Test Set)
n=256 samples

True Label	Puas	196	11
	Tidak Puas	0	49
		Puas	Tidak Puas
		Predicted Label	

Gambar 7 Confusion Matix

Confusion Matrix ini menunjukkan performa model Naive Bayes dalam memprediksi kepuasan terhadap layanan akademik mahasiswa dari 256 sampel. Model berhasil mengidentifikasi 196 layanan akademik mahasiswa yang memuaskan dengan benar, 49 layanan akademik mahasiswa yang tidak memuaskan dengan benar, namun melakukan kesalahan hanya 11 kali

(menganggap layanan akademik mahasiswa yang tidak memuaskan sebagai memuaskan). Dengan total 245 prediksi yang benar, model ini mencapai akurasi

95.7%, yang menunjukkan performa sangat baik dalam memprediksi kepuasan terhadap layanan akademik mahasiswa.

Tabel 6 Metrik Klasifikasi Naive Bayes pada Test Set

Metrik Klasifikasi	Nilai	Interpretasi
Accuracy	95.70%	245/256 prediksi benar
Precision	100.00%	Semua prediksi 'Tidak Puas' benar
Recall (Sensitivity)	94.69%	49/52 sampel 'Tidak Puas' terdeteksi
F1-Score	97.27%	Harmonic mean precision-recall
AUC-ROC	0.9992	Excellent discrimination ability

Tabel 6, menunjukkan berModel Naive Bayes menunjukkan performa klasifikasi yang luar biasa dengan accuracy 95.70% (245 dari 256 prediksi benar), precision 100% yang berarti semua prediksi "Tidak Puas" akurat tanpa false positive, recall 94.69% yang menangkap 49 dari 52 sampel "Tidak Puas" yang sebenarnya, F1-Score 97.27% sebagai keseimbangan sempurna antara precision dan recall, dan AUC-ROC 0.9992 yang menunjukkan kemampuan diskriminasi model sangat excellent dalam membedakan kedua kelas. Secara keseluruhan, model ini adalah classifier yang sangat handal dan dapat dipercaya untuk memprediksi kepuasan mahasiswa dengan tingkat kesalahan yang sangat minimal.

SIMPULAN

Evaluasi komprehensif menunjukkan bahwa kedua metode pemodelan menghasilkan performa yang sangat baik dan konsisten, dengan Analisis Regresi Linier mencapai $R^2=1.0$ pada training set yang menunjukkan model dapat menjelaskan seluruh variasi skor kepuasan melalui lima dimensi SERVQUAL, sementara Naive Bayes Classifier mencapai accuracy 95.70%, precision 100%, recall 94.69%, F1-Score 97.27%, dan AUC 0.9992 pada test set, sedangkan K-Fold Cross-Validation memastikan konsistensi model dengan standar deviasi $\pm 2.00\%$ yang mengindikasikan generalisasi yang baik.

Semua hasil konsisten mengidentifikasi Assurance sebagai dimensi prioritas utama untuk meningkatkan kepuasan layanan akademik mahasiswa, dan kedua model siap untuk diimplementasikan dalam aplikasi praktis. Pendekatan kombinasi yang menggabungkan exploratory analysis dengan PCA, explanatory dengan regresi linier, dan predictive dengan Naive Bayes memberikan pemahaman holistik tentang struktur data, kontribusi dimensi, dan kemampuan prediksi kategori kepuasan layanan akademik mahasiswa secara menyeluruh.

DAFTAR PUSTAKA

- A. Pangestu, D. Hamdhana, and R. Suwanda, "Student Academic Consultation Chatbot Using Meta AI Large Language Models and Retrieval-Augmented Generation," *Journal of Artificial Intelligence and Software Engineering*, vol. 6, no. 1, pp. 26–36, 2026, doi: 10.30811/jaise.v6i1.8875.
- L. Nur Halimah, S. Riyadi, A. Fatahillah Jurjani, A. Prayogi, and S. Dwi Laksana, "IMPLEMENTASI PENGGUNAAN MACHINE LEARNING DALAM PEMBELAJARAN: SUATU TELAAH DESKRIPTIF," 2025.
- A. Susanti, F. Azima, and F. Surya Nugraha, "SEMINAR NASIONAL AMIKOM SURAKARTA (SEMNASA) 2024 ANALISIS KLASIFIKASI KEPUASAN

-
- MAHASISWA TERHADAP PEMBELAJARAN DARING MENGGUNAKAN ALGORITMA NAIVE BAYES,” 2024.
- D. Poni Egistin, M. Yahdi Rauza, R. Has Ramadhan, and S. Ramadani, “Analisis regresi linier sederhana dan penerapannya,” 2025. [Online]. Available: <https://creativecommons.org/licenses/by/4.0/>
- A. Kepuasan *et al.*, “Nanda Kharisma KATA KUNCI,” 2025.
- V. Nurhanifah, “PENGARUH KUALITAS PRODUK DAN PELAYANAN TERHADAP KEPUASAN PELANGGAN PADA PT SINTEGRA MULTI TEKNOTAMA DI KOTA TANGERANG PERIODE 2021-2023,” 2024. [Online]. Available: <https://academicajournal.org/index.php/jeac>
- [P. Purwaningsih *et al.*, “Tingkat Kepuasan Mahasiswa Terhadap Layanan Akademik Dosen Aplikasi Komputer Prodi S1 Administrasi Rumah Sakit STIKES Mutiara Mahakam Samarinda,” *Jurnal Pengabdian kepada Masyarakat Nusantara*, vol. 5, no. 1, pp. 340–345, Jan. 2024, doi: 10.55338/jpkmn.v5i1.2483.