Jurnal Transformasi Pendidikan Modern **Vol 7, No. 3**
https://ejurnals.com/ojs/index.php/jtpm **Juli 2026**

PREDIKSI KEPUASAN MAHASISWA TERHADAP PEMBELAJARAN UNTUK PENINGKATAN LAYANAN AKADEMIK DI UNIVERSITAS KATOLIK SANTO THOMAS MENGGUNAKAN ALGORITMA NAIVE BAYES

Agnes Selaras Osabela Sarahono[1], Parasian D.P. Silitonga[2]
[1,2]Universitas Katolik Santo Thomas Medan
Email: selarassarahono@gmail.com

Abstrak: Kepuasan mahasiswa merupakan salah satu indikator penting dalam mengevaluasi kualitas pembelajaran di perguruan tinggi. Pengolahan data evaluasi kinerja dosen di Universitas Katolik Santo Thomas selama ini masih dilakukan secara deskriptif sehingga belum mampu memberikan prediksi terhadap tingkat kepuasan mahasiswa. Penelitian ini bertujuan membangun model klasifikasi tingkat kepuasan mahasiswa menggunakan algoritma Naive Bayes berdasarkan data saran atau komentar mahasiswa yang diperoleh dari Lembaga Penjaminan Mutu (LPM). Tahapan penelitian meliputi preprocessing teks berupa case folding, filtering, tokenizing, dan stemming, kemudian dilakukan pembobotan menggunakan TF-IDF, penyeimbangan data dengan SMOTE, serta proses klasifikasi menggunakan algoritma Multinomial Naive Bayes. Dataset dibagi menjadi 80% data pelatihan dan 20% data pengujian. Hasil pengujian menunjukkan bahwa model yang dibangun memperoleh nilai akurasi sebesar 95,0%, precision 92,98%, recall 100,0%, dan F1-score 96,46%. Model selanjutnya diimplementasikan dalam aplikasi berbasis web yang mampu mengklasifikasikan komentar mahasiswa, menampilkan hasil prediksi, grafik, serta laporan dalam format PDF. Hasil penelitian menunjukkan bahwa algoritma Naive Bayes dapat diterapkan dengan baik untuk mendukung evaluasi kepuasan mahasiswa terhadap kinerja dosen dan membantu pengambilan keputusan dalam upaya peningkatan kualitas layanan akademik.
Kata Kunci: Naive Bayes, Kepuasan Mahasiswa, Klasifikasi Teks, TF-IDF, SMOTE.

Abstract: *Student satisfaction is one of the important indicators used to evaluate the quality of learning in higher education institutions. The processing of lecturer performance evaluation data at Universitas Katolik Santo Thomas has so far been conducted descriptively, making it unable to predict students' satisfaction levels. This study aims to develop a student satisfaction classification model using the Naive Bayes algorithm based on students' comments obtained from the Quality Assurance Institute (LPM). The research stages include text preprocessing consisting of case folding, filtering, tokenizing, and stemming, followed by TF-IDF weighting, SMOTE for data balancing, and classification using the Multinomial Naive Bayes algorithm. The dataset was divided into 80% training data and 20% testing data. The experimental results showed that the proposed model achieved an accuracy of 95,0%, precision of 92,98%, recall of 100,0%, and an F1-score of 96,46%. The trained model was then implemented in a web-based application capable of classifying student comments, displaying prediction results, presenting graphical visualizations, and generating reports in PDF format. The findings indicate that the Naive Bayes algorithm can be effectively applied to support the evaluation of student satisfaction regarding lecturer*

performance and assist universities in improving the quality of academic services through data-driven decision-making.
Keywords: *Naive Bayes, Student Satisfaction, Text Classification, TF-IDF, SMOTE*.

PENDAHULUAN

Kepuasan mahasiswa merupakan salah satu indikator penting dalam menilai kualitas proses pembelajaran di perguruan tinggi. Tingkat kepuasan mahasiswa dapat mencerminkan kualitas kinerja dosen dalam menyampaikan materi, berkomunikasi, serta menciptakan suasana belajar yang efektif. Oleh karena itu, hasil evaluasi kepuasan mahasiswa dapat dijadikan sebagai dasar dalam meningkatkan mutu pembelajaran dan layanan akademik. Di Universitas Katolik Santo Thomas, evaluasi kinerja dosen dilakukan setiap semester melalui kuesioner yang dikelola oleh Lembaga Penjaminan Mutu (LPM). Namun, pengolahan hasil evaluasi tersebut masih bersifat deskriptif sehingga hanya menyajikan rekapitulasi data dan belum mampu memberikan prediksi terhadap tingkat kepuasan mahasiswa pada data baru.

Perkembangan teknik *machine learning* memungkinkan data evaluasi mahasiswa dimanfaatkan untuk membangun model prediksi yang dapat mengklasifikasikan tingkat kepuasan secara otomatis. Salah satu algoritma klasifikasi yang banyak digunakan adalah Naive Bayes, karena memiliki proses komputasi yang sederhana, cepat, dan mampu memberikan performa yang baik pada klasifikasi data teks. Dalam penelitian ini, data komentar atau saran mahasiswa digunakan sebagai fitur masukan yang diproses melalui tahapan *case folding*, *filtering*, *tokenizing*, dan *stemming*. Selanjutnya, data diubah menjadi representasi numerik menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF) sebelum dilakukan proses klasifikasi menggunakan algoritma Naive Bayes. Untuk mengatasi ketidakseimbangan jumlah data antar kelas, diterapkan metode Synthetic Minority Over-sampling Technique (SMOTE) sehingga model dapat belajar secara lebih optimal.

Beberapa penelitian sebelumnya menunjukkan bahwa algoritma Naive Bayes memiliki kinerja yang baik dalam klasifikasi teks, termasuk analisis sentimen dan klasifikasi kepuasan pengguna. Meskipun demikian, penelitian mengenai prediksi tingkat kepuasan mahasiswa terhadap kinerja dosen yang diimplementasikan dalam sistem berbasis web masih relatif terbatas, khususnya pada lingkungan Universitas Katolik Santo Thomas. Oleh karena itu, penelitian ini mengembangkan sistem klasifikasi berbasis web yang mampu mengolah komentar mahasiswa

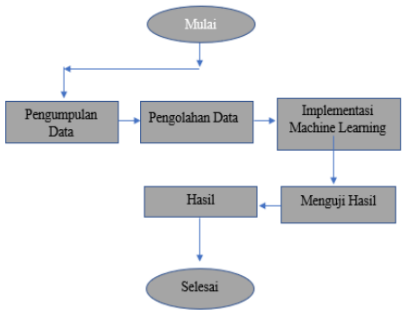
secara otomatis, menampilkan hasil prediksi, visualisasi grafik, serta menghasilkan laporan dalam format PDF untuk mendukung proses evaluasi pembelajaran.

Tujuan penelitian ini adalah membangun model klasifikasi tingkat kepuasan mahasiswa menggunakan algoritma Naive Bayes berdasarkan data evaluasi kinerja dosen yang diperoleh dari LPM Universitas Katolik Santo Thomas, serta mengimplementasikan model tersebut ke dalam aplikasi berbasis web. Hasil penelitian diharapkan dapat membantu pihak universitas dalam melakukan evaluasi pembelajaran secara lebih cepat, objektif, dan mendukung pengambilan keputusan berbasis data untuk meningkatkan kualitas layanan akademik

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode klasifikasi (*classification*) menggunakan algoritma Naive Bayes untuk memprediksi tingkat kepuasan mahasiswa terhadap kinerja dosen berdasarkan data hasil evaluasi pembelajaran. Data penelitian diperoleh dari Lembaga Penjaminan Mutu (LPM) Universitas Katolik Santo Thomas yang berasal dari hasil kuesioner evaluasi kinerja dosen. Sistem yang dibangun bertujuan mengklasifikasikan komentar atau saran mahasiswa ke dalam dua kategori, yaitu Puas dan Tidak Puas, sehingga hasilnya dapat dimanfaatkan sebagai bahan evaluasi dalam meningkatkan kualitas layanan akademik.

Tahapan penelitian dimulai dari proses pengumpulan data, dilanjutkan dengan preprocessing teks, pembobotan menggunakan TF-IDF, penyeimbangan data menggunakan SMOTE, pembagian data menjadi data pelatihan dan data pengujian, pelatihan model menggunakan algoritma Naive Bayes, evaluasi model, hingga implementasi model ke dalam aplikasi berbasis web. Alur penelitian ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian

A. Pengumpulan data

Data yang digunakan dalam penelitian ini merupakan data hasil evaluasi kinerja dosen yang diperoleh dari Lembaga Penjaminan Mutu (LPM) Universitas Katolik Santo Thomas. Dataset terdiri atas data komentar atau saran mahasiswa terhadap proses pembelajaran yang diberikan oleh dosen. Sebelum digunakan dalam proses klasifikasi, data terlebih dahulu dilakukan proses seleksi untuk menghilangkan data yang tidak lengkap maupun data yang tidak relevan dengan tujuan penelitian. Selanjutnya dilakukan proses pelabelan sehingga setiap data memiliki kelas **Puas** atau **Tidak Puas** yang akan digunakan sebagai target klasifikasi.

B. Preprocessing Data

Tahap preprocessing bertujuan untuk membersihkan data teks sehingga lebih mudah diproses oleh algoritma klasifikasi. Tahapan preprocessing yang diterapkan dalam penelitian ini meliputi:

1. **Case Folding**

   Case folding merupakan proses mengubah seluruh karakter huruf menjadi huruf kecil (*lowercase*) agar tidak terjadi perbedaan antara huruf kapital dan huruf kecil. Sebagai contoh, kata **"Dosen"**, **"DOSEN"**, dan **"dosen"** akan dianggap sebagai satu kata yang sama.

2. **Filtering**

   Filtering merupakan proses menghapus karakter yang tidak diperlukan, seperti angka, tanda baca, simbol, serta kata-kata umum (*stopword*) yang tidak memberikan informasi penting dalam proses klasifikasi. Tahapan ini bertujuan mengurangi jumlah kata yang tidak berpengaruh terhadap hasil prediksi.

3. **Tokenizing**

   Tokenizing merupakan proses memecah kalimat menjadi kumpulan kata (*token*). Setiap kata yang dihasilkan akan digunakan sebagai fitur dalam proses pembentukan model klasifikasi menggunakan algoritma Naive Bayes.

4. Stemming

Stemming dilakukan untuk mengubah kata berimbuhan menjadi kata dasar. Pada penelitian ini digunakan pustaka **Sastrawi** untuk proses stemming bahasa Indonesia sehingga kata-kata yang memiliki makna sama tetapi bentuk berbeda dapat diseragamkan.

C. Pembobotan TF-IDF

Setelah melalui preprocessing, data teks diubah menjadi bentuk numerik menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF). Metode ini memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam dokumen serta tingkat kemunculannya pada seluruh dokumen. Dengan demikian, kata-kata yang memiliki informasi penting akan memperoleh bobot yang lebih besar dibandingkan kata-kata yang sering muncul pada seluruh dokumen.

D. Penyeimbangan Data Menggunakan SMOTE

Distribusi kelas pada dataset penelitian tidak seimbang sehingga dapat mempengaruhi kinerja algoritma klasifikasi. Oleh karena itu diterapkan metode Synthetic Minority Over-sampling Technique (SMOTE) pada data pelatihan. Metode ini menghasilkan data sintetis pada kelas minoritas sehingga jumlah data antar kelas menjadi lebih seimbang. Penerapan SMOTE diharapkan mampu meningkatkan kemampuan model dalam mengenali data pada kelas minoritas dan mengurangi bias terhadap kelas mayoritas.

E. Pembagian Data

Dataset dibagi menjadi dua bagian menggunakan metode train-test split, yaitu sebesar 80% sebagai data pelatihan dan 20% sebagai data pengujian. Pembagian dilakukan secara acak menggunakan parameter random_state = 42 dan stratify agar distribusi kelas tetap proporsional pada data pelatihan maupun data pengujian.

F. Pelatihan Model Naive Bayes

Model klasifikasi dibangun menggunakan algoritma Multinomial Naive Bayes. Algoritma ini menghitung probabilitas setiap kata terhadap masing-masing kelas berdasarkan Teorema Bayes. Selanjutnya model akan menentukan kelas dengan nilai probabilitas posterior terbesar

sebagai hasil prediksi. Setelah proses pelatihan selesai, model disimpan menggunakan pustaka Joblib agar dapat digunakan kembali pada aplikasi berbasis web.

G. Evaluasi Model

Evaluasi model dilakukan menggunakan data pengujian yang tidak ikut digunakan pada proses pelatihan. Kinerja model diukur menggunakan confusion matrix untuk memperoleh nilai accuracy, precision, recall, dan F1-score. Keempat metrik tersebut digunakan untuk mengetahui tingkat keberhasilan model dalam mengklasifikasikan tingkat kepuasan mahasiswa terhadap kinerja dosen.

H. Implementasi Sistem

Model yang telah memperoleh hasil evaluasi kemudian diimplementasikan ke dalam aplikasi berbasis web menggunakan framework Flask. Sistem menyediakan beberapa fitur utama, yaitu pengelolaan data, proses klasifikasi komentar mahasiswa, visualisasi hasil prediksi dalam bentuk grafik, serta pembuatan laporan hasil prediksi dalam format PDF. Implementasi ini bertujuan agar proses evaluasi kepuasan mahasiswa dapat dilakukan secara lebih cepat, efisien, dan mendukung pengambilan keputusan berbasis data di Universitas Katolik Santo Thomas.

HASIL DAN PEMBAHASAN

1. Distribusi Data Latih dan Data Uji

Setelah proses pengumpulan, pembersihan data, dan pelabelan selesai dilakukan, dataset kemudian dibagi menjadi dua bagian, yaitu data training dan data testing. Pembagian data dilakukan menggunakan metode *train-test split* dengan perbandingan 80% sebagai data training dan 20% sebagai data testing. Data training digunakan untuk melatih model Naive Bayes agar dapat mempelajari pola hubungan antara komentar mahasiswa dengan kelas kepuasan, sedangkan data testing digunakan untuk menguji kemampuan model dalam mengklasifikasikan data yang belum pernah dipelajari sebelumnya. Berdasarkan proses pembagian tersebut diperoleh 30.323 data training 24.258 dan data testing 6.065. Berikut gambar data latih dan data uji

No	Tahun	Nama Mata Kuliah	Saran	Kelas
1	2023	Algoritma dan Pemrograman	Dosen sangat baik dan menjelaskan materi dengan jelas	Puas
2	2023	Struktur Data	Cara mengajar dosen cukup baik namun perlu lebih interaktif	Puas
3	2024	Basis Data	Materi sulit dipahami dan dosen kurang responsif terhadap mahasiswa	Tidak Puas
4	2025	Sistem Operasi	Dosen mengajar dengan baik dan materi cukup mudah dipahami	Puas
5	2025	Pemrograman Web	Cara mengajar membosankan dan penjelasan tidak mudah dipahami	Tidak Puas

Gambar 2 Data latih

No	Tahun	Mata Kuliah	Saran	Kelas Asli
1	2023	Pemrograman Web	Dosen mengajar dengan sangat baik dan interaktif	Puas
2	2024	Interaksi Manusia dan Komputer	Penjelasan dosen cukup jelas	Puas
3	2024	Etika Profesi TI	Penjelasan materi kurang jelas	Tidak Puas
4	2025	Pengolahan Citra Digital	Cara mengajar tidak baik dan membosankan	Tidak Puas
5	2023	Kecerdasan Buatan	Dosen menjelaskan materi dengan sangat jelas dan mudah dipahami	Puas

Gambar 3 Data uji

2. Hasil Evaluasi Model

Setelah proses pelatihan model selesai dilakukan, tahap berikutnya adalah evaluasi model menggunakan data testing sebanyak 6.065 data. Evaluasi bertujuan untuk mengetahui tingkat kinerja algoritma Naive Bayes dalam mengklasifikasikan tingkat kepuasan mahasiswa terhadap kinerja dosen berdasarkan komentar mahasiswa.

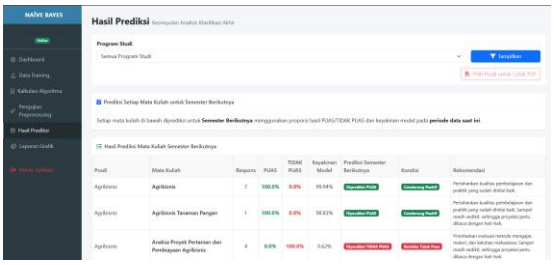
Pengukuran performa model dilakukan menggunakan beberapa metrik evaluasi, yaitu Accuracy, Precision, Recall, dan F1-Score. Accuracy digunakan untuk mengukur tingkat ketepatan keseluruhan model dalam melakukan klasifikasi. Precision digunakan untuk mengetahui tingkat ketepatan prediksi pada kelas yang diprediksi positif. Recall digunakan untuk mengukur kemampuan model dalam mengenali seluruh data pada kelas yang sebenarnya. Sedangkan F1-Score merupakan nilai harmonis antara precision dan recall sehingga memberikan gambaran performa model secara keseluruhan. Berikut contoh gambar hasil model.

```
Laporan Klasifikasi Model:
              precision    recall  f1-score   support

        PUAS       0.96      1.00      0.98      5230
  TIDAK PUAS       1.00      0.76      0.86       835

    accuracy                           0.97      6065
   macro avg       0.98      0.88      0.92      6065
weighted avg       0.97      0.97      0.97      6065
```

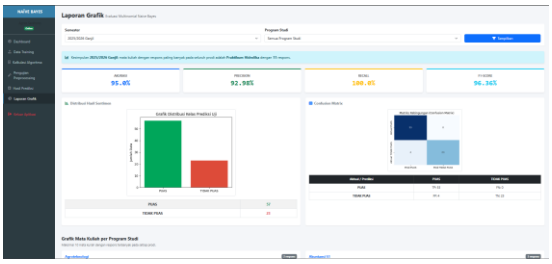
Gambar 4. Hasil Model

3. Hasil Klasifikasi Data Uji

Model Naive Bayes yang telah dilatih digunakan untuk mengklasifikasikan 30.323 data uji berdasarkan komentar mahasiswa terhadap kinerja dosen. Sebelum proses klasifikasi, data terlebih dahulu melalui tahapan preprocessing, pembobotan TF-IDF, kemudian Hasil pengujian menunjukkan bahwa model mampu mengklasifikasikan data ke dalam kategori Puas dan Tidak Puas dengan nilai akurasi sebesar 95,0%, precision 92,98%, recall 100,0%, dan F1-score 96,46%. Hasil klasifikasi ditampilkan melalui aplikasi berbasis web dalam bentuk hasil prediksi, grafik, dan laporan PDF, sebagaimana ditunjukkan pada Gambar 5 dan Gambar 6.



Gambar 5. Hasil Prediksi



Gambar 6. Hasil Grafik.

KESIMPULAN

Berdasarkan hasil penelitian, algoritma Naive Bayes berhasil diterapkan untuk mengklasifikasikan tingkat kepuasan mahasiswa terhadap kinerja dosen berdasarkan data komentar mahasiswa di Universitas Katolik Santo Thomas. Melalui tahapan preprocessing, pembobotan TF-IDF, dan penyeimbangan data menggunakan SMOTE, model memperoleh performa yang baik dengan nilai akurasi sebesar 95,0%, precision 92,98%, recall 100,0%, dan F1-score 96,46%. Model yang dikembangkan juga berhasil diimplementasikan ke dalam aplikasi berbasis web sehingga dapat membantu Lembaga Penjaminan Mutu (LPM) dalam mengolah hasil evaluasi pembelajaran dan mendukung pengambilan keputusan untuk meningkatkan kualitas pembelajaran.

DAFTAR PUSTAKA

Anggraini, R., dkk. (2020). *Sistem Basis Data*. Jakarta: Informatika.

Astuti, Y. (2009). *Analisis dan Perancangan Sistem Informasi*. Yogyakarta: Andi.

Dewayani, S., & Wasino. (2021). *Perancangan Basis Data*. Semarang: UNNES Press.

Duggan, J., et al. (1970). *Database Management Systems*. New York: McGraw-Hill.

Faridi. (2015). *Pemrograman Menggunakan Sublime Text*. Jakarta: Elex Media Komputindo.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3), 37–54.

Felita, M., dkk. (2020). Penggunaan Sublime Text dalam Pengembangan Perangkat Lunak. *Jurnal Teknologi Informasi*, 5(2), 45–52.

Fildes, R. (1992). The Evaluation of Forecasting Methods. *International Journal of Forecasting*, 8(1), 19–27.

Gorunescu, F. (2011). *Data Mining: Concepts, Models and Techniques*. Berlin: Springer.

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). San Francisco: Morgan Kaufmann.

Istiqomah. (2022). *Manajemen Basis Data MySQL*. Yogyakarta: Deepublish.

Kotler, P., & Keller, K. L. (2016). *Marketing Management* (15th ed.). New Jersey: Pearson Education.

Larose, D. T. (2014). *Discovering Knowledge in Data: An Introduction to Data Mining*. New Jersey: Wiley.

Larose, D. T. (2015). *Data Mining Methods and Models*. New Jersey: Wiley.

Londol, S., & S. (2022). *Perancangan Sistem Basis Data*. Jakarta: Informatika.

Mitchell, T. M. (1997). *Machine Learning*. New York: McGraw-Hill.

Nugroho, A. (2010). *Rekayasa Perangkat Lunak Menggunakan UML dan Java*. Yogyakarta: Andi.

Patil, T. R., & Sherekar, S. S. (2013). Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification. *International Journal of Computer Science and Applications*, 6(2), 256–261.

Putri, A. (2019). Pemodelan Sistem Menggunakan Activity Diagram. *Jurnal Sistem Informasi*, 7(1), 15–22.

Ramdhany, A. (2024). *Perancangan Class Diagram dalam Sistem Informasi*. Bandung: Informatika.

Rish, I. (2001). An Empirical Study of the Naive Bayes Classifier. *Proceedings of IJCAI Workshop on Empirical Methods in AI*.

Russell, S., & Norvig, P. (2016). *Artificial Intelligence: A Modern Approach* (3rd ed.). New Jersey: Pearson Education.

Sallaby, A., & Kanedi, I. (2020). Analisis Use Case Diagram pada Sistem Informasi. *Jurnal Media Informatika*, 4(1), 23–30.

Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, 3(3), 210–229.

Sonata, F. (2019). Unified Modeling Language (UML) dalam Perancangan Sistem Informasi. *Jurnal Ilmu Komputer*, 6(2), 1–9.

Sukamto, R. A. (2018). *Rekayasa Perangkat Lunak Terstruktur dan Berorientasi Objek*. Bandung: Informatika.

Supono, & Putratama, V. (2016). *Pemrograman Web dengan PHP dan MySQL*. Jakarta: Elex Media Komputindo.

Turban, E., Sharda, R., & Delen, D. (2018). *Decision Support and Business Intelligence Systems*. New Jersey: Pearson.

Yoga, I. M., dkk. (2021). Pemodelan Activity Diagram pada Sistem Informasi. *Jurnal Teknologi dan Sistem Informasi*, 9(2), 60–68.

Zhang, H. (2004). The Optimality of Naive Bayes. *Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference*