

ANALISIS SENTIMEN OPINI MAHASISWA PADA KUALITAS LAYANAN AKADEMIK BERBASIS *PYTHON TEXTBLOB* DENGAN PENGUJIAN ALGORITMA *RANDOM FOREST* (STUDI KASUS: UNIVERSITAS KAHURIPAN KEDIRI)

Winda Indriani¹, Candra Adipradana²

^{1,2}Universitas Kahuripan Kediri

Email: winda.indriani@students.kahuripan.ac.id

Abstrak

Pendidikan tinggi memainkan peran penting dalam membentuk potensi sumber daya manusia. Kualitas layanan akademik di kampus sangat mempengaruhi pencapaian akademik dan kepuasan mahasiswa. Universitas Kahuripan Kediri bertanggung jawab untuk memberikan layanan akademik terbaik, dan pemahaman terhadap sentimen mahasiswa dapat membantu meningkatkan lingkungan akademik dan citra lembaga. Penelitian ini menggunakan teknologi analisis sentimen dengan Python TextBlob dan metode Random Forest untuk mengeksplorasi opini mahasiswa mengenai layanan akademik. Hasil menunjukkan akurasi 73% untuk keseluruhan data dan 81% untuk data uji dengan perhitungan manual, serta hasil yang konsisten dengan algoritma Random Forest menggunakan RapidMiner. Namun, peningkatan jumlah dataset tidak selalu meningkatkan akurasi model, dengan akurasi menurun dari 81% pada 9 dataset menjadi 65% pada 200 dataset. Hal ini menunjukkan bahwa penambahan data percobaan tidak selalu meningkatkan performa model dan dapat menurunkan akurasi. Hasil analisis sentiment menunjukkan bahwa 63.33% data memiliki sentimen positif, 26.67% sentimen netral, dan 10.00% sentimen negatif. Mayoritas ulasan memberikan tanggapan yang baik, sementara sebagian kecil memberikan tanggapan yang kurang baik. Secara keseluruhan, mayoritas sentimen dalam data ini cenderung positif.

Kata Kunci: *Data Mining, Analisis Sentimen, Random Forest, TextBlob*

Abstract

Higher education plays an important role in shaping human resource potential. The quality of academic services on campus greatly influences academic achievement and student satisfaction. Kahuripan University Kediri is responsible for providing the best academic services, and understanding student sentiment can help improve the academic environment and image of the institution. This research uses sentiment analysis technology with Python TextBlob and the Random Forest method to explore student opinions regarding academic services. The results show an accuracy of 73% for all data and 81% for test data with manual calculations, as well as consistent results with the Random Forest algorithm using RapidMiner. However, increasing the number of datasets does not always increase model accuracy, with accuracy decreasing from 81% on 9 datasets to 65% on 200 datasets. This shows that adding experimental data does not always improve model performance and can reduce accuracy. The results of sentiment analysis show that 63.33% of the data has positive sentiment, 26.67% neutral sentiment, and 10.00% negative sentiment. The majority of reviews provide good responses, while a small number provide unfavorable responses. Overall, the majority of sentiment in this data tends to be positive.

Keywords: Data Mining, Sentiment Analysis, Random Forest, TextBlob

Pendahuluan

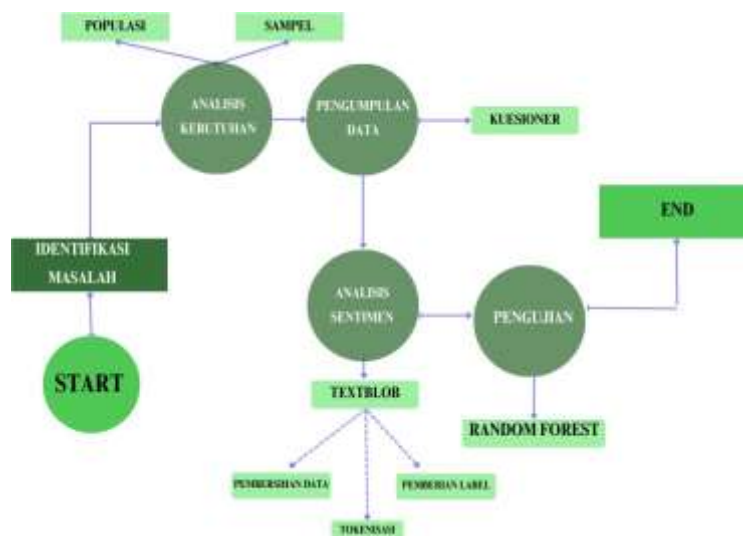
Pendidikan tinggi merupakan fondasi utama dalam pembentukan potensi dan kualitas sumber daya manusia, dengan institusi pendidikan tinggi memainkan peran penting dalam pengalaman belajar mahasiswa. Kualitas layanan akademik di kampus berdampak signifikan pada pencapaian akademik dan kepuasan mahasiswa. Era digital saat ini memungkinkan mahasiswa mengekspresikan opini mereka mengenai layanan akademik melalui media sosial dan forum online. Universitas Kahuripan Kediri, sebagai salah satu perguruan tinggi di Kabupaten Kediri, bertanggung jawab memberikan layanan akademik terbaik. Memahami sentimen mahasiswa membantu universitas merespon lebih efektif terhadap masukan dan kebutuhan mereka, menciptakan lingkungan akademik yang positif, dan meningkatkan citra lembaga.

Penggunaan teknologi analisis sentimen menjadi penting. Metode ini dapat menggunakan Python TextBlob untuk mengekstrak sentimen positif, negatif, atau netral dari teks. Pengklasifikasian diusulkan dengan dataset untuk melakukan proses data latih dan model yang diperoleh (Latifah 2020). Dengan adanya seleksi fitur tentu Random Forest dapat bekerja pada big data dengan parameter yang kompleks secara efektif (Fatmasari et al. 2022). Penelitian sebelumnya telah mengaplikasikan analisis sentimen pada berbagai konteks, termasuk pendidikan. Namun, penelitian terkait analisis sentimen opini mahasiswa tentang kualitas layanan akademik di Universitas Kahuripan Kediri masih terbatas. Penelitian ini bertujuan

mengeksplorasi opini dan sentimen mahasiswa tentang kualitas layanan akademik menggunakan teknik TextBlob dan metode Random Forest, diharapkan dapat memberikan wawasan mendalam dan membantu institusi pendidikan tinggi meningkatkan layanan akademik mereka secara efisien.

Metode Penelitian

Metode yang digunakan dalam penelitian ini adalah kuantitatif deskriptif



1. Identifikasi Masalah

Penelitian ini berfokus pada kualitas layanan akademik di Universitas Kahuripan Kediri, yang mempengaruhi pencapaian akademik, kepuasan mahasiswa, serta citra universitas. Di era digital, mahasiswa menggunakan media sosial untuk menyampaikan opini mereka, sehingga penting bagi universitas untuk menganalisis dan merespons masukan ini. Namun, analisis sentimen mahasiswa mengenai layanan akademik masih terbatas, menghambat perbaikan yang tepat sasaran. Penelitian ini menggunakan metode TextBlob untuk analisis sentimen dan membandingkannya dengan metode Random Forest untuk menilai keakuratan.

2. Analisis Kebutuhan

a. Populasi

Populasi adalah wilayah generalisasi yang terdiri atas obyek/subyek yang mempunyai kualitas dan karakteristik tertentu yang ditetapkan oleh peneliti untuk dipelajari dan kemudian ditarik kesimpulannya (Erri, Lestari, and Asymar 2021).

Populasi dalam penelitian ini adalah mahasiswa aktif di Universitas Kahuripan Kediri

b. Sampel

Sampel secara sederhana diartikan sebagai bagian dari populasi yang menjadi sumber data yang sebenarnya dalam suatu penelitian (Amin, Garancang, and Abunawas 2023). Dengan kata lain, sampel adalah sebagian dari populasi untuk mewakili seluruh populasi. Dalam penelitian ini, sampel akan terdiri dari sejumlah mahasiswa yang dipilih secara acak atau dengan Teknik random sampling lainnya dari populasi mahasiswa Universitas Kahuripan Kediri.

3. Pengumpulan Data

Data dikumpulkan menggunakan kuesioner berbasis Google Form, yang mencakup informasi demografis dan pertanyaan terbuka mengenai pengalaman mahasiswa dengan layanan akademik. Kuesioner dikirim melalui grup WhatsApp mahasiswa dan diberikan waktu pengumpulan sekitar dua bulan.

4. Analisis Sentimen

- a. Pengumpulan Data: Data opini mahasiswa dikumpulkan dari survey dan media sosial.
- b. Pra-pemrosesan Data: Data dibersihkan, dilakukan tokenisasi, dan penghapusan kata-kata yang tidak relevan.
- c. Analisis Sentimen: Menggunakan TextBlob untuk menentukan sentimen (positif, negatif, atau netral) dari setiap opini.
- d. Pelatihan Model: Data dibagi menjadi set pelatihan dan set pengujian. Model Random Forest dilatih menggunakan data latih.

5. Pengujian

Implementasi model Random Forest dilakukan dengan menggunakan RapidMiner untuk menganalisis sentimen opini mahasiswa. Pengujian ini mencakup pemrosesan data, konfigurasi model, pelatihan model, dan evaluasi kinerja model. Hasil pengujian dibandingkan dengan hasil dari TextBlob untuk menentukan keakuratan dan efektivitas masing-masing metode.

Pada Tahap ini dilakukan perhitungan data sampel menggunakan Algoritma Random Forest untuk menganalisis sentiment opini mahasiswa. Untuk menghitung nilai gini index, gini (S_i) didefinisikan sebagai :

$$\text{Gini}(S_i) = 1 - \sum_{i=0}^{c-1} (P_i)^2 \quad \text{persamaan ke(3)}$$

Jika S_i dataset dibagi menjadi dua himpunan bagian T_1 dan T_2 dengan ukuran N_1 dan N_2 masing-masing, index pada split data berisi contoh-contoh dari kelas n index $gini(S_i)$ didefinisikan sebagai berikut:

$$Gini_{split} = \sum_{i=0}^{c-1} \left(\frac{N_i}{N} \right) gini(S_i) \quad \text{persamaan ke(4)}$$

6. Uji Kompetensi Formula

Uji kompetensi dilakukan untuk mengevaluasi efektivitas metode analisis sentimen dan memastikan akurasi serta keandalan hasil penelitian. Proses ini mencakup pengumpulan dan pra-pemrosesan data, analisis sentimen, pengujian algoritma, dan evaluasi hasil.

Dengan pendekatan ini, penelitian diharapkan dapat memberikan gambaran yang komprehensif mengenai sentimen mahasiswa terhadap kualitas layanan akademik di Universitas Kahuripan Kediri dan membantu universitas dalam melakukan perbaikan yang diperlukan.

Hasil Penelitian

1. Identifikasi Masalah

Penelitian ini mengidentifikasi beberapa masalah berdasarkan hasil survei:

- a. Tahapan Analisis Sentimen: Menggunakan TextBlob, tahapan meliputi pengumpulan data opini mahasiswa melalui survei, pra-pemrosesan data, klasifikasi sentimen (positif, negatif, atau netral), dan validasi hasil klasifikasi menggunakan dataset berlabel untuk mengukur akurasi.
- b. Hasil Analisis Sentimen:

Sebagian besar opini mahasiswa bersifat positif, dengan pujian terhadap pelayanan akademik, kualitas pembelajaran, fasilitas, dan kualitas sarana dan prasarana. Opini negatif menyoroti masalah seperti keterlambatan dalam pelayanan dan kurangnya ketersediaan dosen. Opini netral sering berisi kritik konstruktif dan saran perbaikan.
- c. Algoritma Pengujian: Menggunakan algoritma Random Forest untuk menguji akurasi hasil sentimen.

2. Analisis Kebutuhan

Hasil penelitian analisis kebutuhan berdasarkan populasi dan sampel mahasiswa Universitas Kahuripan Kediri:

- Populasi: 450 mahasiswa yang terdaftar dan aktif dalam berbagai program studi di

Universitas Kahuripan Kediri.

- Sampel: 100 mahasiswa yang dipilih secara acak, namun dalam praktiknya hanya berhasil mengumpulkan 30 data sampel.

3. Pengumpulan Data

Pengumpulan data dilakukan melalui kuesioner digital yang mencerminkan opini mahasiswa tentang layanan akademik. Jawaban yang diterima dikumpulkan secara digital untuk memudahkan analisis selanjutnya.

4. Analisis Sentimen

Analisis sentimen dilakukan dengan Python menggunakan TextBlob, melalui beberapa langkah preprocessing data:

a. Pembersihan Data:

- Mengimpor pustaka Pandas, TextBlob, dan NLTK.
- Membaca file CSV
- Membersihkan teks dan mengubah huruf menjadi kecil.

b. Labeling:

- Membaca dataset.
- Mendefinisikan daftar kata positif dan negatif.
- Memberikan label sentimen berdasarkan kata-kata dalam teks dan polaritas sentimen.

c. Pembagian Model Data Training dan Testing:

- Analisis data dilakukan terhadap dua set data: data training (70%) dan data testing (30%).

5. Pengujian

Tahap pengujian meliputi perhitungan manual dan penggunaan RapidMiner:

- Perhitungan Manual: Menghitung Gini Index dan akurasi untuk data keseluruhan dan data testing.
- RapidMiner: Menggunakan Algoritma Random Forest, mendapatkan akurasi 73% untuk keseluruhan data dan 80% untuk data testing.

6. Uji Kompetensi Formula

Uji kompetensi dilakukan dengan data dummy untuk menghitung indeks dan mengukur akurasi menggunakan Gini Split, serta membandingkan hasilnya dengan alat analitik RapidMiner:

- Data Dummy dengan 50 Dataset: Akurasi 72%
- Data Dummy dengan 100 Dataset: Akurasi 70%
- Data Dummy dengan 200 Dataset: Akurasi 65%

7. Distribusi Presentase Sentimen

Hasil analisis sentiment pada penelitian ini menunjukkan bahwa sebagian besar data, yaitu sebesar 63.33%, memiliki sentimen positif. Ini berarti mayoritas ulasan atau pendapat yang dianalisis memberikan tanggapan yang baik atau menguntungkan. Selain itu, 26.67% dari data menunjukkan sentimen netral, yang menandakan bahwa ulasan atau pendapat tersebut tidak menunjukkan kecenderungan yang kuat ke arah positif maupun negatif. Sementara itu, hanya 10.00% dari data yang memiliki sentimen negatif, menunjukkan bahwa sebagian kecil ulasan atau pendapat memberikan tanggapan yang kurang baik atau tidak menguntungkan. Secara keseluruhan, hasil ini memberikan gambaran menyeluruh mengenai sentimen yang terkait dengan data yang diteliti, dengan mayoritas menunjukkan kecenderungan positif.

Simpulan

Berdasarkan penelitian yang dilakukan kesimpulan yang diperoleh, yaitu:

1. Metode TextBlob dapat digunakan secara efektif untuk mengklasifikasikan sentimen dalam analisis opini mahasiswa terhadap layanan akademik.
2. Sebagian besar opini mahasiswa bersifat positif dengan presentase sebesar 63,33%, namun ada juga kritik yang dapat digunakan untuk peningkatan layanan.
3. Metode Random Forest menunjukkan akurasi bervariasi dalam mengukur opini mahasiswa di Universitas Kahuripan Kediri, dengan 73% untuk keseluruhan data dan 80% untuk data uji. Penambahan dataset kadang menurunkan akurasi dari 81% menjadi 65%, sehingga perlu berhati-hati dalam menambah data.

Daftar Pustaka

- Amin, Nur Fadilah, Sabaruddin Garancang, and Kamaluddin Abunawas. 2023. "Konsep Umum Populasi Dan Sampel Dalam Penelitian." *Jurnal Pilar* 14(1): 15–31.
- Erri, Dirgahayu, Ajeng Puji Lestari, and Hasta Herlan Asymar. 2021. "Pengaruh Gaya Kepemimpinan Terhadap Kinerja Karyawan Pada PT MELZER GLOBAL SEJAHTERA JAKARTA." *Jurnal Inovasi Penelitian* 1(10): 1–208.
- Fatmasari, Rhini et al. 2022. "Analisis Sentimen Dalam Pengkategorian Komentar Youtube Terhadap Layanan Akademik Dan Non-Akademik Universitas Terbuka Untuk Prediksi Kepuasan." *Building of Informatics, Technology and Science (BITS)* 4(2): 395–404.

Latifah, Safitri Linawati; Siti Nurdiani; Kartika Handayani; 2020. “Prediksi Prestasi Akademik Mahasiswa Menggunakan Random Forest Dan C.45.” VIII(1): 47–52.