

BAB I PENDAHULUAN

1.1. Latar Belakang

Kemajuan di bidang teknologi informasi dan komunikasi telah berdampak signifikan pada banyak aspek kehidupan, khususnya dalam sektor pendidikan (Firdausy et al., 2023). E-learning kini menjadi alternatif utama bagi banyak institusi untuk menyampaikan materi pembelajaran secara efektif dan efisien. E-Learning Sunan (Sinau Temenan) telah diimplementasikan di Universitas Muria Kudus (UMK) untuk memberikan akses yang lebih baik bagi mahasiswa dalam belajar. Salah satu keunggulan platform ini adalah fleksibilitas yang ditawarkannya, memungkinkan mahasiswa untuk mengakses materi kapan saja dan di mana saja. Namun, meskipun banyak keuntungan yang ditawarkan, penting untuk memahami pengalaman pengguna dalam menggunakan e-learning. Ulasan dan feedback dari mahasiswa menjadi sumber informasi yang sangat berharga untuk mengevaluasi kepuasan dan efektivitas platform (Arofah et al., 2024). Mahasiswa sering kali memberikan pendapat mengenai aspek-aspek seperti kemudahan penggunaan, kualitas materi, dan interaksi dengan instruktur. Namun, dengan meningkatnya volume ulasan, analisis manual terhadap feedback tersebut menjadi tantangan tersendiri. Proses ini tidak hanya memakan waktu, tetapi juga berpotensi mengabaikan pola-pola penting dalam opini pengguna. Di sinilah sentiment analysis atau analisis sentimen memegang peranan penting. Analisis sentimen memungkinkan kita untuk secara otomatis mengidentifikasi, mengekstrak, dan mengelompokkan opini pengguna terhadap e-learning berdasarkan kategori sentimen, seperti positif, negatif, atau netral. Analisis ini memberikan pemahaman lebih dalam mengenai pengalaman pengguna yang tidak hanya berhenti pada tingkat individu, tetapi juga menggambarkan kecenderungan umum persepsi mahasiswa secara keseluruhan terhadap platform tersebut. Dengan demikian, analisis sentimen dapat membantu pengelola *e-learning* dalam mengidentifikasi aspek-aspek yang perlu diperbaiki atau dikembangkan lebih lanjut.

Sebagai metode dalam data mining, analisis sentimen memanfaatkan berbagai teknik pemrosesan bahasa alami untuk mengevaluasi opini pengguna

dalam skala besar (Apriani et al., 2024). Algoritma Naive Bayes merupakan salah satu metode yang umum digunakan untuk klasifikasi teks, termasuk analisis sentimen, karena algoritma ini mengasumsikan independensi antar fitur yang menjadikannya efisien dalam memproses data besar. Menurut (Muktafin et al., 2020), penggunaan analisis sentimen secara otomatis memungkinkan pengelola platform untuk memahami persepsi pengguna dengan lebih baik, yang pada gilirannya dapat meningkatkan pengalaman belajar secara keseluruhan. Sementara itu, penelitian lain oleh (Putri et al., 2023) menunjukkan bahwa Naïve Bayes memiliki performa yang baik dalam analisis sentimen, sehingga menjadi pilihan yang sesuai untuk penelitian ini. Berdasarkan hal tersebut, penelitian ini bertujuan untuk menerapkan teknik data mining dalam proses analisis sentimen terhadap kuesioner yang dikumpulkan dari mahasiswa UMK pengguna E-Learning Sunan. Diharapkan, hasil penelitian ini dapat memberikan wawasan yang lebih mendalam mengenai persepsi pengguna serta membantu pengelola dalam meningkatkan kualitas layanan dan pengalaman belajar. Untuk menangani permasalahan ini, penulis mengajukan penelitian dengan judul “Implementasi Data Mining dalam Analisis Sentimen Pengguna E-Learning Sunan (Sinau Temenan) Menggunakan Algoritma Naïve Bayes Classifier”.

1.2. Perumusan Masalah

Berdasarkan Latar Belakang sebelumnya, penulis mengidentifikasi permasalahan, yaitu bagaimana penerapan algoritma *Naive Bayes Classifier* dalam mengklasifikasikan sentimen ulasan mahasiswa mengenai Sunan *E-learning* sebagai media pembelajaran dan seberapa akurat model *Naive Bayes Classifier* dalam mengklasifikasikan sentimen ulasan tersebut.

1.3. Batasan Masalah

Penelitian ini dibatasi pada beberapa aspek berikut:

- a. Data yang digunakan adalah data hasil kuisisioner yang dibuat dan disebarluaskan sendiri oleh penulis kepada seluruh mahasiswa UMK.
- b. Data yang diolah dalam penelitian ini berjumlah sekitar 387 responden.

- c. Metode yang digunakan untuk analisis sentimen adalah *Naive Bayes Classifier*.
- d. Sentimen yang dianalisis hanya dibatasi pada tiga kategori utama: positif dan negatif.
- e. *Tools* yang akan digunakan dalam penelitian ini adalah *Google Colaboratory* dan *Jupyter Notebook* untuk proses pengolahan dan analisis data serta implementasi *interface* sistem.

1.4. Tujuan

Penelitian ini berfokus pada penerapan serta evaluasi metode *Naive Bayes Classifier* dalam mengelompokkan sentimen mahasiswa terkait penggunaan *E-learning* Sunan sebagai media pembelajaran. Selain itu, penelitian ini juga bertujuan untuk mengetahui seberapa akurat model ini dalam pengklasifikasian sentimen.

1.5. Manfaat

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

- a. Bagi Peneliti dan Akademisi: Memberikan kontribusi pada pengembangan ilmu pengetahuan, terutama terkait penerapan *machine learning* untuk analisis sentimen pada *e-learning*.
- b. Bagi Praktisi *E-learning*: Memberikan wawasan mengenai persepsi pengguna untuk meningkatkan kualitas layanan.
- c. Bagi Mahasiswa Lain: Menjadi referensi dan inspirasi dalam pengembangan penelitian di bidang data mining dan *machine learning* untuk analisis data teks.

1.6. Metode Penelitian

1.6.1 Metode Pengumpulan Data

Dalam rangka untuk mencapai tujuan penelitian, penulis memerlukan metodologi yang tepat dan terarah. Oleh karena itu, metodologi dan langkah-langkah yang diambil dalam teknik pengumpulan data yang digunakan oleh penulis dalam penelitian ini dengan cara:

a. Sumber Data Primer

Data primer merupakan informasi yang didapatkan secara langsung dari objek penelitian atau sumber data awal, sumber data ini mencakup:

1. Observasi

Pengumpulan data melalui pengamatan dan pencatatan terhadap peristiwa yang terjadi secara langsung, misalnya mengamati proses pendaftaran pelanggan sampai pembayaran tagihan oleh pelanggan.

2. Wawancara

Pengumpulan data melalui tatap muka dan tanya jawab langsung dengan sumber data atau pihak-pihak yang berkepentingan yang berhubungan dengan penelitian. Misalnya wawancara dengan bagian pengelola dan administrasi mengenai pembukuan transaksi dengan pelanggan.

3. Kuesioner

Kuesioner merupakan metode utama untuk mengumpulkan data opini dan sentimen mahasiswa terhadap platform Sunan. Kuesioner disebarakan secara daring kepada mahasiswa UMK, dengan fokus pada pertanyaan mengenai kepuasan, kemudahan penggunaan, serta kualitas materi pembelajaran yang disediakan.

b. Sumber Data Sekunder

Sumber Data sekunder merupakan informasi yang didapatkan dari sumber informasi yang sudah ada, seperti jurnal, buku, dan laporan penelitian sebelumnya. Data sekunder yang peneliti peroleh diperoleh melalui pendekatan studi pustaka. Pendekatan studi pustaka adalah teknik yang digunakan untuk menghimpun informasi dan data dengan

merujuk pada literatur seperti buku, jurnal, serta sumber online yang relevan dengan topik penelitian. Dalam metode ini, penulis mengumpulkan berbagai referensi dari sumber-sumber yang relevan untuk kemudian digunakan sebagai acuan atau landasan dalam penulisan tugas akhir.

1.6.2 Metode *Machine Learning*

Pada bagian ini, dijelaskan metode *machine learning* yang digunakan untuk penelitian ini, yang berfokus pada klasifikasi sentimen menggunakan algoritma *Naive Bayes*. Penelitian ini menggunakan algoritma *Naive Bayes* untuk melakukan klasifikasi sentimen pada teks. Algoritma ini dipilih karena kesederhanaannya dan kemampuannya yang baik dalam menangani data teks, khususnya dalam masalah klasifikasi biner (positif, negatif, dan netral) yang umum digunakan dalam analisis sentimen. Algoritma *Naive Bayes* bekerja dengan prinsip probabilistik berdasarkan teorema Bayes, dan sering digunakan dalam pengklasifikasian teks.

Untuk mempersiapkan data teks agar dapat diproses oleh model *Naive Bayes*, digunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk pembobotan kata. Metode TF-IDF akan memberikan bobot lebih tinggi pada kata-kata yang lebih jarang muncul di seluruh dokumen tetapi sering muncul dalam satu dokumen tertentu, yang menunjukkan relevansi kata tersebut terhadap sentimen yang ingin diklasifikasikan. Langkah-langkah perancangan sistem meliputi:

a. *Preprocessing Data*

Penghapusan tanda baca, angka, dan kata-kata yang tidak relevan. Tokenisasi teks untuk memecah teks menjadi kata-kata. Sedangkan *stemming* untuk mereduksi kata menjadi bentuk dasarnya.

b. Pembobotan Kata dengan TF-IDF

Menerapkan teknik TF-IDF untuk menghasilkan representasi numerik dari teks yang dapat diproses oleh model *Naive Bayes*.

1.6.3 Pengembangan Sistem

Pengembangan sistem dilakukan menggunakan bahasa pemrograman *Python*, yang menyediakan pustaka-pustaka yang berguna untuk pengolahan bahasa alami *Natural Language Processing* (NLP). Beberapa pustaka yang digunakan antara lain:

- a. *Scikit-learn*: Untuk implementasi model *Naive Bayes*, serta evaluasi model dengan metrik seperti *accuracy*, *precision*, *recall*, dan *F1-score*.
- b. NLTK atau *spaCy*: Untuk pemrosesan teks, termasuk *case folding*, *filtering*, *correction slang word*, *tokenizing* dan *stemming*.
- c. *TfidfVectorizer* dari *Scikit-learn*: Untuk pembobotan kata menggunakan metode TF-IDF.

1.6.4 Pengujian Sistem

Setelah model dilatih, dilakukan uji performa untuk mengevaluasi seberapa baik model dalam mengklasifikasikan sentimen. Beberapa metrik evaluasi yang digunakan dalam penelitian ini antara lain:

- a. *Acuration*: Untuk mengukur proporsi prediksi yang benar dibandingkan dengan total prediksi yang dibuat.
- b. *Precision*: Mengukur seberapa tepat prediksi positif yang dibuat oleh model.
- c. *Recall*: Mengukur seberapa banyak sentimen positif yang berhasil diidentifikasi oleh model.
- d. *F1-Score*: Merupakan rata-rata harmonis antara presisi dan recall, memberikan gambaran umum tentang keseimbangan keduanya.
- e. *Confusion Matrix*: Untuk memberikan gambaran visual dari performa model dengan menunjukkan jumlah prediksi yang benar dan salah untuk setiap kelas.

1.6.5 Kerangka Pemikiran

Adapun kerangka penelitian yang akan dilakukan dalam implementasi proses klasifikasi tersebut adalah sebagai berikut:

