

## BAB 1. PENDAHULUAN

### 1.1 Latar Belakang

Perkembangan zaman saat ini membuat penggunaan media sosial, terutama *Instagram* dan *Twitter*, telah menjadi sangat umum di kalangan penggemar sepakbola. Melalui *platform* ini, penggemar dapat dengan mudah berbagi opini, merayakan kemenangan, dan juga menyuarakan kekecewaan terhadap klub favorit mereka, seperti *Manchester United*. Media sosial telah mengubah cara penggemar berinteraksi dengan klub, memungkinkan pertukaran informasi dan sentimen secara langsung dan global. Selain itu, media sosial juga menjadi indikator penting untuk melihat bagaimana publik merespons suatu klub, yang dapat mempengaruhi reputasi dan nilai komersialnya (Jara et al., 2024). Tren ini juga diamati di Indonesia, dimana media sosial menjadi alat standar untuk komunikasi bisnis dan pemasaran (Nurmaelinda & Rusydi, 2023). Selain itu, peran media sosial dalam memperkuat hubungan pengguna dan memfasilitasi penyebaran informasi disoroti dalam konteks perpustakaan digital (A. Lubis, 2022). Perkembangan ini mendorong kebutuhan akan metode klasifikasi yang efektif untuk mengolah dan menganalisis data teks secara otomatis.

*Naive Bayes Classifier* merupakan salah satu algoritma *machine learning* yang paling populer dan efektif untuk klasifikasi teks, khususnya dalam analisis sentimen. Algoritma ini didasarkan pada teorema *Bayes* dengan asumsi independensi yang kuat antara fitur-fitur yang digunakan. Keunggulan utama *Naive Bayes* terletak pada kemampuannya untuk menangani *dataset* yang besar dengan efisien, memiliki kompleksitas komputasi yang rendah, dan memberikan hasil yang akurat meskipun dengan asumsi independensi yang sederhana (Surya Sayogo et al., 2024). Karakteristik ini menjadikan *Naive Bayes* sangat cocok untuk aplikasi *real-time* dan sistem dengan *resource* terbatas.

Klasifikasi teks menggunakan *machine learning* telah menjadi area penelitian yang sangat aktif dalam beberapa dekade terakhir. Berbagai algoritma telah dikembangkan dan diuji untuk tugas klasifikasi teks, mulai dari metode

tradisional hingga *deep learning*. *K-Nearest Neighbors* (KNN) mengklasifikasikan data berdasarkan jarak terpendek ke objek data dan menggunakan mayoritas kategori dari tetangga terdekatnya (Cholil et al., 2021). *Support Vector Machine* (SVM) menggunakan fungsi *hyperplane* untuk membentuk wilayah-wilayah pemisah antara kelas-kelas yang ada (Fikri et al., 2020). *Recurrent Neural Network* (RNN) memanfaatkan arsitektur jaringan saraf dengan keadaan tersembunyi dan *loop* umpan balik untuk memproses urutan data (Tarkus et al., 2020). *Naïve Bayes* adalah salah satu algoritma pembelajaran induktif yang paling efektif dan efisien untuk *machine learning* dan *data mining* (Syarli & Muin, 2020). Metode *Naive Bayes* tetap menjadi pilihan utama peneliti karena kesederhanaan implementasi dan performa yang konsisten.

Perbandingan empiris menunjukkan superioritas *Naive Bayes* dalam berbagai konteks aplikasi. Penelitian dari Kurnia et al., (2024) membuktikan bahwa *Naive Bayes* memiliki akurasi dan presisi yang lebih tinggi dibandingkan *CNN* dalam mengidentifikasi berita hoax politik Indonesia. Mustriyanto et al., (2022) mendemonstrasikan efektivitas *Naive Bayes* yang lebih baik dibandingkan *Decision Tree* dalam analisis sentimen pengguna layanan PLN, dengan akurasi data *testing* mencapai 80% dibandingkan 56% yang dicapai oleh *Decision Tree*. Puteri et al., (2023) juga memvalidasi keunggulan *Naive Bayes* dibandingkan KNN dalam konteks diagnosa penyakit diabetes. Fikri et al., (2020) mengkonfirmasi bahwa *Naive Bayes* lebih unggul dibandingkan SVM dalam melakukan analisis sentimen.

Karakteristik data teks media sosial memiliki tantangan khusus dalam analisis sentimen yang memerlukan pendekatan algoritma yang kokoh. Data teks dari platform media sosial sering mengandung *noise*, singkatan, *emoji*, dan variasi bahasa informal yang dapat mempengaruhi akurasi klasifikasi (Dwiyansaputra et al., 2025). *Naïve Bayes* adalah metode probabilistik sederhana yang diterapkan pada teorema bayes dengan kecepatan pemrosesan yang tinggi dalam data besar (Iqbal et al., 2023). Kemampuan algoritma *Naive Bayes* bekerja berdasarkan prinsip probabilitas *Bayes* dan asumsi bahwa setiap fitur dalam data bersifat independen satu sama lain, membuatnya cocok untuk menangani kompleksitas linguistik dalam

teks media sosial, termasuk penggunaan bahasa gaul, typo, dan struktur kalimat yang tidak baku (Haq et al., 2024).

Proses *preprocessing* dan *feature extraction* merupakan komponen krusial dalam implementasi *Naive Bayes Classifier* untuk analisis sentimen. Tahapan *preprocessing* meliputi *tokenization*, *stemming*, *stopword removal*, dan normalisasi teks yang dapat meningkatkan performa klasifikasi secara signifikan (Khairunnisa et al., 2021). *Feature extraction* menggunakan teknik seperti *Term Frequency-Inverse Document Frequency* (TF-IDF) atau *Bag of Words* memungkinkan *Naive Bayes* untuk mengidentifikasi pola sentimen dengan lebih akurat (Panjaitan, 2024). Optimalisasi parameter dan pemilihan fitur yang tepat dapat meningkatkan *precision* dan *recall* algoritma *Naive Bayes* hingga mencapai tingkat akurasi yang kompetitif dengan metode *machine learning* yang lebih kompleks (Panjaitan, 2024).

Evaluasi performa *Naive Bayes Classifier* dalam analisis sentimen memerlukan pendekatan yang komprehensif menggunakan berbagai metrik evaluasi. Metrik standar seperti akurasi, presisi, *recall*, dan *F1-score* memberikan gambaran menyeluruh tentang kemampuan algoritma dalam mengklasifikasikan sentimen (Siti Aisah et al., 2024). *Cross-validation* dan pembagian data *training-testing* yang tepat menjadi faktor penting untuk memastikan generalisasi model yang optimal (Azmi et al., 2023).

Konteks data media sosial Indonesia menyediakan lingkungan yang ideal untuk pengembangan dan pengujian *Naive Bayes Classifier*. Menurut laporan Asosiasi Pengguna Jasa Internet Indonesia (APJII), sebanyak 79,5% dari total 221 juta pengguna aktif media sosial mengaksesnya setiap hari, menghasilkan volume data teks yang sangat besar untuk analisis sentimen (Haryanto, 2024). Mayoritas pengguna adalah Generasi Z (34,40%) dan generasi milenial (30,62%), yang cenderung menggunakan bahasa informal dan ekspresif dalam media sosial, memberikan tantangan menarik untuk algoritma klasifikasi teks (Alfiyani et al., 2025).

Implementasi *Naive Bayes Classifier* untuk analisis sentimen media sosial memiliki nilai praktis yang signifikan. Algoritma ini dapat mengklasifikasikan teks

ke dalam kategori sentimen positif, negatif, atau netral dengan tingkat akurasi yang tinggi. Kemampuan *Naive Bayes* untuk menangani *noise* dan variasi bahasa dalam teks media sosial menjadikannya pilihan yang tepat untuk aplikasi *real-world*. Penelitian ini menggunakan metode *Naive Bayes Classifier* untuk menganalisis sentimen pada data *Twitter* dan *Instagram*, dengan fokus pada optimalisasi *preprocessing* teks dan *feature engineering* untuk meningkatkan performa klasifikasi.

Penelitian ini penting dilakukan karena kebutuhan akan sistem analisis sentimen yang akurat dan efisien semakin meningkat di era digital. Pengembangan model *Naive Bayes* yang dioptimalkan dapat memberikan kontribusi signifikan untuk pemahaman opini publik secara otomatis. Implementasi algoritma ini dalam bentuk sistem *web* menggunakan *flask* akan memungkinkan akses yang mudah untuk analisis sentimen *real-time*, yang dapat dimanfaatkan untuk berbagai keperluan bisnis dan penelitian.

Penelitian ini menawarkan solusi berupa implementasi *Naive Bayes Classifier* yang dioptimalkan untuk mengklasifikasikan sentimen positif, negatif dan netral dari data media sosial. Pendekatan kuantitatif digunakan dengan data yang dikumpulkan melalui *crawling* dan *web scraping* dari platform *Instagram* dan *Twitter*. Proses pengumpulan data dilakukan menggunakan *library python* untuk *Twitter* dan *extension* pihak ketiga dari *google chrome* untuk *Instagram*. Model *Naive Bayes* akan dilatih dan divalidasi menggunakan pembagian data *training* dan *testing*, dengan evaluasi performa menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*. Hasil akhir akan diimplementasikan dalam bentuk aplikasi *web* menggunakan *Flask* dengan pengujian menggunakan metode *black box testing* untuk memastikan fungsionalitas sistem yang optimal.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang ini, permasalahan penelitian dapat dirumuskan sebagai berikut:

1. Bagaimana metode *Naïve Bayes* digunakan untuk mencapai tujuan klasifikasi sentimen (positif, negatif, dan netral) pada data media sosial Instagram dan Twitter terkait Manchester United?
2. Bagaimana tingkat akurasi dan performa algoritma *Naïve Bayes* dalam mengklasifikasikan sentimen terhadap *Manchester United* di media sosial *Instagram* dan *Twitter*?

### 1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Menerapkan metode *Naïve Bayes* untuk mengklasifikasikan sentimen menjadi kategori positif, negatif, dan netral.
2. Menganalisis dan mengevaluasi performa algoritma *Naïve Bayes* dalam klasifikasi sentimen pada media sosial.

### 1.4 Batasan Masalah

Batasan masalah dari penelitian ini adalah:

1. Platform media sosial yang digunakan adalah *Instagram* dan *Twitter*.
2. Data yang digunakan dalam penelitian ini adalah *tweet* dengan topik *Manchester United* dan komentar *Instagram* dengan topik *Manchester United*
3. Data yang digunakan dibatasi dengan hanya penggunaan bahasa Indonesia.
4. Data yang digunakan dalam penelitian ini merupakan data *tweet* dan komentar dengan klasifikasi positif, negatif dan netral.
5. Jumlah data yang diambil adalah 5000 data teks diantaranya 2500 *tweet* dan 2500 Komentar *Instagram*

### 1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. Hasil penelitian ini diharapkan dapat digunakan sebagai referensi dalam memahami analisis sentimen terhadap Klub *Manchester United* melalui data *Instagram* dan *Twitter*. Hal ini diharapkan dapat memberikan

kontribusi pada pengetahuan akademis terkait analisis sentimen di kalangan mahasiswa, khususnya di Politeknik Negeri Jember.

2. Penelitian ini diharapkan dapat memberikan kontribusi yang bermanfaat bagi manajemen Klub *Manchester United* dalam memahami sentimen dan persepsi penggemar di media sosial, khususnya *Instagram* dan *Twitter*. Sebagai pemahaman yang lebih baik tentang sentimen penggemar, klub dapat meningkatkan strategi komunikasi dan pemasaran mereka, serta mengoptimalkan pengelolaan citra klub secara keseluruhan. Selain itu, penelitian ini juga dapat menjadi panduan bagi klub-klub lain dalam menerapkan analisis sentimen pada media sosial untuk meningkatkan interaksi dan keterlibatan penggemar.