

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1. Sentimen Analisis**

Analisis sentimen adalah salah satu bidang dalam ilmu komputer yang mempelajari komputasi linguistik, pengolahan bahasa alami, dan text mining. Bidang ini bertujuan untuk menganalisis emosi, penilaian, sikap, pendapat, dan sentimen seseorang terhadap pembicara atau penulis terkait dengan suatu produk, layanan, organisasi, individu, topik publik, peristiwa, atau aktivitas tertentu. Melalui analisis ini, kita dapat memahami bagaimana perasaan dan pendapat masyarakat terhadap berbagai subjek, memberikan wawasan yang berguna untuk pengambilan keputusan dan strategi bisnis.[7]

Analisis sentimen dapat digunakan untuk memperoleh persentase sentimen positif, negatif, ataupun netral terhadap seseorang, institusi, produk, atau perusahaan dalam kondisi tertentu. Tugas utama analisis sentimen adalah mengelompokkan teks ke dalam kategori berdasarkan kalimat atau dokumen, yang kemudian akan dianalisis untuk menentukan apakah teks tersebut bersifat positif, negatif, atau netral. Melalui proses ini, kita dapat memahami persepsi publik secara lebih terperinci dan mengidentifikasi tren sentimen yang relevan untuk pengambilan keputusan strategis.

#### **Data Mining**

Data mining adalah suatu proses semi-otomatis yang melibatkan penggunaan berbagai teknik statistik, metode matematika, kecerdasan buatan, dan pembelajaran mesin dengan tujuan untuk mengeksplorasi dan menggali potensi informasi berharga yang tersembunyi dalam sejumlah besar data yang disimpan dalam sebuah basis data. Dengan menggunakan algoritma yang kompleks dan teknik analisis yang canggih, data mining memungkinkan pengguna untuk mengidentifikasi pola, hubungan, dan tren yang mungkin tidak terlihat secara langsung.

Proses data mining sendiri merupakan bagian integral dari Knowledge Discovery in Databases (KDD), yang merupakan kerangka kerja yang dirancang untuk menemukan pengetahuan baru dari data yang tersedia. Tahapan-tahapan dalam KDD meliputi pemilihan data yang relevan, pra-pengolahan data untuk mempersiapkannya agar sesuai untuk analisis, transformasi data ke dalam bentuk yang sesuai untuk pengolahan lebih lanjut, proses data mining yang melibatkan penerapan algoritma dan

teknik analisis untuk mengeksplorasi data, dan terakhir evaluasi hasil yang diperoleh dari proses tersebut.

KDD secara umum juga sering disebut sebagai penemuan pengetahuan dalam basis data. Konsep ini menekankan pentingnya dalam menemukan dan memahami informasi baru yang mungkin tidak diketahui sebelumnya, serta menerapkannya dalam pengambilan keputusan yang lebih baik dan strategis. Dengan demikian, melalui kombinasi data mining dan KDD, organisasi dapat mengoptimalkan penggunaan data mereka untuk mendapatkan wawasan yang berharga dan mendukung pengembangan bisnis serta pengambilan keputusan yang lebih efektif.[8]

### **Naïve Bayes Classifier**

Naive Bayes adalah salah satu algoritma dari machine learning. Dalam evolusi database, Naïve Bayes tergolong dalam supervised learning, yaitu jenis machine learning yang memerlukan sampel data latih yang sudah dilabeli. Supervised learning dibagi menjadi dua kategori utama, yaitu klasifikasi dan regresi. Klasifikasi digunakan ketika variabel yang dihasilkan berupa kategori, seperti warna (misalnya merah atau kuning), kondisi kesehatan (penyakit atau tidak), dan sebagainya. Sementara itu, regresi digunakan ketika variabel yang dihasilkan berupa nilai riil, seperti berat, nilai uang, dan sebagainya.

Naive Bayes termasuk dalam supervised learning klasifikasi, sebagaimana halnya dengan algoritma lain seperti Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Artificial Neural Network (ANN), Trees Gradient Boosted (TGB), dan Random Trees (RT). Sedangkan untuk regresi, contohnya adalah Decision Tree, Logistic Regression, dan Kernel Regression. Dengan beragam jenis algoritma ini, supervised learning dapat diterapkan untuk berbagai kebutuhan analisis data yang mencakup prediksi kategori atau nilai berdasarkan data latih yang telah diberikan.[9]

### **Relevansi Metode Analisis dengan Tujuan Penelitian**

Dalam konteks penelitian ini, metode Naive Bayes Classifier dalam analisis sentimen pada Twitter menjadi relevan karena kemampuannya dalam mengklasifikasikan sentimen yang berpotensi menggambarkan dukungan atau ketidaksukaan terhadap kandidat presiden 2024.

