BAB II
LANDASAN TEORI

2.1 Artificial Intelligence (AI)

Artificial Intelligence (AI) atau kecerdasan buatan merupakan salah satu cabang ilmu komputer yang berfokus pada pengembangan sistem atau mesin yang mampu meniru kemampuan berpikir manusia dalam menyelesaikan berbagai permasalahan. AI dirancang agar komputer dapat melakukan aktivitas yang umumnya membutuhkan kecerdasan manusia, seperti belajar, memahami bahasa, mengenali pola, mengambil keputusan, dan memecahkan masalah.

IBM, sebagai salah satu perusahaan teknologi global yang aktif dalam pengembangan AI, menjelaskan AI sebagai "*a field which combines computer science and robust datasets to enable problem-solving. It also encompasses sub-fields of machine learning and deep learning, which are frequently mentioned in conjunction with AI"* (IBM, 2023). Dalam penjelasan ini, AI dipahami sebagai payung besar yang mencakup berbagai pendekatan teknis untuk memungkinkan mesin menyelesaikan masalah secara otonom. Pendekatan industri ini mencerminkan pandangan praktis bahwa AI merupakan teknologi integratif yang mendukung otomatisasi berbasis data dalam skala besar. Berikut adalah tabel resume definisi AI menurut para ahli dan sumber.Perkembangan AI saat ini mengalami peningkatan yang sangat pesat seiring dengan kemajuan teknologi komputasi dan ketersediaan data dalam jumlah besar (*big* data). AI telah banyak diterapkan dalam berbagai bidang seperti kesehatan (diagnosis penyakit),

pendidikan (sistem pembelajaran adaptif), bisnis (analisis perilaku konsumen), serta pemerintahan *(smart governance).*

Dalam konteks pemerintahan, *AI* memiliki peran penting dalam meningkatkan kualitas pelayanan publik. Dengan memanfaatkan *AI,* instansi pemerintah dapat mengolah data masyarakat se cara otomatis, cepat, dan akurat, sehingga mampu meningkatkan efisiensi kerja serta kualitas pengambilan keputusan. Salah satu implementasi *AI* yang relevan adalah analisis sentimen terhadap ulasan masyarakat.

Pada penelitian ini, *AI* digunakan sebagai dasar dalam membangun sistem yang mampu menganalisis opini masyarakat terhadap pelayanan publik di DPMPTSP Provinsi Bengkulu. Dengan demikian, *AI* tidak hanya berfungsi sebagai alat bantu teknologi, tetapi juga sebagai solusi dalam meningkatkan kualitas pelayanan berbasis data

2.2 *Natural Language Processing (NLP)*

Natural Language Processing (NLP) adalah cabang dari ilmu komputer yang berfokus pada interaksi antara komputer dan bahasa manusia. Tujuannya adalah memungkinkan mesin untuk “memahami”, menganalisis, dan menghasilkan bahasa yang digunakan oleh manusia secara alami. Dalam konteks analisis sentimen, NLP berperan sebagai fondasi utama dalam mengekstraksi makna dari teks menurut Hidayah, A. K., Sunardi, D., & Sahputra, E. (2025).

Bahasa manusia memiliki struktur yang kompleks, ambigu, dan kontekstual, sehingga diperlukan teknik khusus agar komputer dapat memahaminya. Oleh karena itu, NLP memiliki peran penting dalam mengubah data teks yang tidak terstruktur menjadi data yang terstruktur dan dapat diolah lebih lanjut.

Dalam analisis sentimen, NLP memiliki peran yang sangat penting karena digunakan untuk memproses data teks yang berasal dari ulasan masyarakat. Data teks tersebut biasanya tidak terstruktur dan memiliki variasi bahasa yang sangat beragam, sehingga perlu dilakukan beberapa tahapan pengolahan sebelum dapat dianalisis.

Tahapan dalam Natural *Language Processing* meliputi: *Case Folding , Case folding* adalah proses mengubah seluruh huruf dalam teks menjadi huruf kecil *(lowercase).* Misalnya, kata “Indonesia”, “INDONESIA”, dan “indonesia” dianggap berbeda oleh komputer, sehingga perlu diseragamkan menjadi “indonesia”.

Penghapusan Simbol dan Karakter Khusus, Simbol seperti tanda baca, angka, emoji, tagar (#), *mention* (@), dan tautan *(https://...)* biasanya tidak memiliki nilai informatif dalam analisis sentimen dan harus dihapus, kecuali jika dibutuhkan untuk fitur khusus.

Tokenisasi, Tokenisasi adalah proses memecah teks menjadi unit-unit kata (token). Misalnya, kalimat “IKN terlalu mahal” diubah menjadi daftar token: [‘ikn’, ‘terlalu’, ‘mahal’]. Ini memudahkan proses analisis dan ekstraksi fitur.

Stopword Removal, Stopword adalah kata-kata umum yang tidak memiliki makna penting dalam analisis, seperti “yang”, “dan”, “di”, “ke”, dan sebagainya. Kata-kata ini dihapus agar fitur yang digunakan lebih relevan dan tidak mendominasi hasil klasifikasi.

Dengan melalui tahapan tersebut, data teks akan menjadi lebih bersih dan terstruktur sehingga dapat digunakan dalam proses klasifikasi menggunakan algoritma Naive Bayes.

2.3 Analisis Sentimen

Menurut Asrumi, Suharijadi, D., Setiari, A. D., & Wulanda, D. P (2023) Analisis sentimen adalah bidang studi yang menganalisis opini, sentimen, evaluasi, penilaian, sikap, dan emosi masyarakat terhadap entitas seperti produk, layanan, organisasi, individu, isu, peristiwa, topik, dan atributnya. Analisis sentimes biasanya dilakukan dalam konteks data teks, seperti ulasan produk*, tweet,* posting media sosial, artikel berita, dan lainnya. Tujuan analisis sentimen adalah untuk memahami bagaimana orang merasakan atau berpikir tentang suatu topik, produk, layanan, merek, atau peristiwa.

Analisis sentimen merupakan teknik dalam data *mining* yang digunakan untuk mengidentifikasi, mengekstraksi, dan mengklasifikasikan opini atau

perasaan seseorang terhadap suatu objek, baik dalam bentuk positif, negatif, maupun netral. Analisis sentimen sering digunakan untuk memahami persepsi masyarakat terhadap suatu layanan atau produk (Liu, 2020).

Dalam konteks pelayanan publik, analisis sentimen dapat digunakan sebagai alat evaluasi untuk mengetahui tingkat kepuasan masyarakat terhadap layanan yang diberikan oleh instansi pemerintah. Dengan adanya analisis sentimen, instansi dapat mengetahui kelebihan dan kekurangan layanan yang diberikan berdasarkan opini masyarakat secara langsung.

Pada penelitian ini, analisis sentimen digunakan untuk mengklasifikasikan ulasan masyarakat terhadap pelayanan publik di DPMPTSP Provinsi Bengkulu sehingga dapat diketahui kecenderungan opini masyarakat.

Analisis sentimen biasanya dilakukan melalui beberapa tahapan, yaitu**,** pengumpulan Data,mengambil data ulasan dari berbagai sumber seperti media sosial atau *website* resmi kantor DPMPTSP yang diteliti, *Preprocessing* Data, pengolahan dan mempersiapkan data agar siap dianalisis., Klasifikasi Sentimen**,** Mengelompokkan data ke dalam kategori positif, negatif, dan netral**,** Evaluasi Hasil**,** mengukur tingkat akurasi dari model yang digunakan. Dengan analisis sentimen, instansi dapat mengetahui bagian pelayanan mana yang sudah baik dan bagian mana yang perlu diperbaiki.

2.4 Naive Bayes

Naive Bayes *Classifier* merupakan salah satu metode klasifikasi dalam *machine learning* yang menggunakan pendekatan probabilistik berdasarkan Teorema Bayes. Metode ini memiliki asumsi bahwa setiap fitur bersifat independen satu sama lain. Meskipun asumsi tersebut tergolong sederhana, Naive Bayes terbukti memiliki performa yang baik dalam klasifikasi data teks (Zhang et al., 2020).

Rumus dasar Naive Bayes adalah:

$$\boldsymbol{P(H \mid X) = \frac{P(X \mid H) \cdot P(H)}{P(X)}}$$

Penjelasan: **P(H|X)** adalah probabilitas suatu data termasuk dalam kategori tertentu, **P(X|H)** adalah probabilitas munculnya data pada kategori tersebut, **P(H)** adalah probabilitas awal dari kategori, **P(X)** adalah probabilitas keseluruhan data.

Dalam implementasinya pada analisis sentimen, Naive Bayes akan menghitung probabilitas setiap kata dalam suatu ulasan untuk menentukan apakah ulasan tersebut termasuk kategori positif, negatif, atau netral.

Metode Naive Bayes memiliki beberapa keunggulan, di antaranya proses perhitungan yang relatif cepat karena menggunakan pendekatan probabilistik yang sederhana, mudah diterapkan pada berbagai jenis permasalahan klasifikasi, serta sangat cocok digunakan untuk analisis data teks seperti analisis sentimen.

Selain itu, metode ini tidak memerlukan jumlah data pelatihan (*training data*) yang sangat besar untuk menghasilkan performa yang baik, memiliki kebutuhan komputasi yang rendah, dan mampu menangani data berdimensi tinggi, seperti hasil pembobotan TF-IDF yang umumnya memiliki ribuan fitur kata. Karakteristik tersebut menjadikan Naive Bayes sebagai salah satu algoritma yang efisien untuk klasifikasi dokumen dan teks.

Namun demikian, metode Naive Bayes juga memiliki beberapa keterbatasan. Algoritma ini mengasumsikan bahwa setiap fitur atau kata bersifat independen (tidak saling bergantung), padahal dalam kalimat sebenarnya terdapat hubungan antarkata yang memengaruhi makna suatu teks. Asumsi tersebut dapat menyebabkan model kurang mampu menangkap konteks kalimat secara utuh sehingga berpotensi menurunkan akurasi klasifikasi, terutama pada kalimat yang mengandung makna ambigu atau bergantung pada hubungan antar kata. Selain itu, performa Naive Bayes juga dapat dipengaruhi oleh ketidakseimbangan jumlah data pada setiap kelas sentimen, sehingga kelas dengan jumlah data yang lebih sedikit cenderung memiliki tingkat prediksi yang lebih rendah dibandingkan kelas yang lebih dominan.