

BAB II TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Tabel 2. 1 Penelitian Terdahulu

No.	Penulis	Judul	Metode	Hasil
1	Michelle Graciela, Rikky & Hafiz Irsyad (2024)	Klasifikasi Opini Masyarakat Terhadap Naturalisasi Pemain Sepak Bola Menggunakan KNN dan SMOTE	KNN tanpa SMOTE vs. KNN dengan SMOTE	Hasil penelitian menunjukkan KNN tanpa SMOTE: 98%, KNN dengan SMOTE: menurun
2	Leno Dwi Cahya & tim (2023)	Improving Multi-label Classification Performance on Imbalanced Datasets Through SMOTE Technique and Data Augmentation Using IndoBERT Model	IndoBert (SMOTE vs. Augmentasi Data)	Hasil penelitian menunjukkan SMOTE: 82%, Augmentasi Data: 78%.

2.2 Analisis Sentimen

Analisis sentimen adalah proses memahami, ekstraksi dan olah data teks yang tidak terstruktur yang berguna sebagai teknik mengambil informasi sentimen, termasuk sebuah pendapat atau opini. Teknik ini dapat digunakan untuk menilai suatu kalimat dan kecenderungannya terhadap suatu topik, apakah bersifat negatif, positif, atau netral. Analisis sentimen dapat diterapkan di berbagai bidang seperti politik, ekonomi, hukum, sosial, dan lainnya. Terutama dari media sosial Twitter,

yang membuka peluang bagi para peneliti untuk mempelajari sentimen pengguna dalam memberikan komentar terhadap suatu topik pembahasan [17].

2.3 NLP

Natural Language Processing (NLP) merupakan cabang ilmu komputer yang memanfaatkan pengetahuan dalam linguistik komputasional untuk memahami dan mengembangkan cara komputer berinteraksi dengan bahasa manusia secara alami [18]. NLP memungkinkan komputer berinteraksi dengan manusia melalui bahasa alami, baik dalam bentuk ucapan maupun teks. Tujuan utamanya adalah memahami, menganalisis, dan menghasilkan teks dalam bahasa manusia, serta mengembangkan sistem yang mampu memproses dan berkomunikasi dengan manusia dengan cara yang menyerupai komunikasi antar manusia [19].

2.4 IndoBert

BERT (Bidirectional Encoder Representations from Transformers) adalah model bahasa buatan Google tahun 2018 yang menggunakan arsitektur Transformer untuk memahami konteks kata secara dua arah (bidirectional). Berbeda dengan model sebelumnya yang hanya satu arah, BERT dapat memahami kata berdasarkan konteks sebelum dan sesudahnya. Versi BERT Base memiliki 110 juta parameter, sedangkan BERT Large mencapai 340 juta parameter dengan konfigurasi yang lebih kompleks untuk pemahaman bahasa yang lebih dalam.

Untuk bahasa Indonesia, dikembangkan versi khusus bernama IndoBERT, salah satunya adalah indobenchmark/IndoBERT-base-p1, yang telah dioptimalkan untuk pemrosesan bahasa Indonesia. Model ini banyak digunakan dalam berbagai tugas pemrosesan bahasa alami, termasuk analisis sentimen. Dalam konteks tersebut, IndoBERT mampu mengklasifikasikan sentimen suatu teks ke dalam tiga kategori, yaitu positif, negatif, dan netral [20].