

BAB II

TINJAUAN PUSTAKA

Penelitian-penelitian terdahulu yang sudah dilakukan untuk membuat model analisis sentimen berbasis aspek dengan data yang didapatkan dari sumber *scraping*. Penelitian-penelitian terdahulu dapat membantu peneliti untuk mendapatkan gambaran yang lebih jelas dan wawasan yang lebih luas mengenai pembuatan model yang digunakan untuk analisis sentimen dengan pemrosesan bahasa alami pada data sosial media atau khususnya terhadap sampah maupun daur ulang sampah plastik.

A. Ketersediaan *Dataset* Analisis Sentimen Berbasis Aspek

Sebagian besar penelitian terdahulu menggunakan data seperti restoran, hotel, marketplace atau e-commerce, aplikasi digital, gadget, kosmetik, dan pariwisata. Data-data pada penelitian-penelitian tersebut diberikan label sesuai dengan aspek-aspek yang dibutuhkan untuk masing-masing penelitian. Pada penelitian-penelitian dengan subyek penelitian restoran, aspek yang digunakan adalah mengenai makanan, pelayanan, tempat, harga, dan atmosfer [6], [7], [8], [9], [10]. Penelitian dengan subyek penelitian hotel [11] memiliki aspek-aspek kamar, pelayanan, lokasi, gym, dan kolam renang. Penelitian dengan subyek marketplace atau e-commerce umumnya menggunakan aspek layanan pelanggan, tampilan aplikasi, dan kualitas produk [12], [13], [14], [15]. Untuk subyek aplikasi digital aspek yang umum digunakan adalah pengalaman pengguna, promosi, dan fitur aplikasi [16], [17]. Untuk subyek lainnya ada gadget dengan aspek harga, desain, spesifikasi, citra merek [18], majalah dengan aspek produk kecantikan, kemasan, harga, dan aroma [19], serta subyek wisata dengan makanan halal, tempat ibadah, dan toilet [20].

Untuk mayoritas penelitian, label yang digunakan adalah 2 label sentimen yaitu positif dan negatif. Sentimen netral jarang digunakan, hanya pada penelitian [6], [11], [13], [15], [19]. Tanpa sentimen netral, kalimat-

kalimat yang bersifat informatif, ambigu, atau bahkan tidak memiliki sentimen yang khusus pada topik akan diklasifikasikan dengan salah.

Pada penelitian-penelitian yang terdahulu juga masih belum ditemukan adanya *dataset* secara khusus untuk daur ulang sampah plastik. Sehingga masih terdapat kekosongan *dataset* secara publik maupun dengan melakukan *web scraping* yang bisa digunakan untuk membangun model sentimen analisis berbasis aspek pada topik daur ulang sampah plastik. Hal ini menunjukkan peluang untuk riset baru yang peduli lingkungan dalam hal pengumpulan *dataset* dan pembuatan model.

B. Algoritma Penelitian Terdahulu

Penelitian-penelitian terdahulu menggunakan berbagai jenis algoritma seperti variasi-variasi pada *machine learning* dan *deep learning*. Metode *Machine learning* seperti Naive Bayes, SVM(*Support Vector Machine*), dan *Random Forest* masih banyak digunakan pada penelitian-penelitian terdahulu. Metode-metode ini, beserta dengan metode lainnya, diiringi dengan *feature extraction* yang membuat pembelajaran menjadi efektif, tetapi masih memiliki kekurangan dalam pemahaman konteks untuk kalimat. Untuk penelitian-penelitian dengan metode *machine learning*, metode Naive Bayes mendapatkan hasil terbaik akurasi 92,5% [12] dan F1-Score sebesar 76,96% [14], SVM dengan hasil terbaik akurasi 96,43% [18] dan juga hasil F1-Score 0,73 [10], dan *random forest* dengan hasil terbaik akurasi 90% [11].

Model *deep learning* sudah mulai diadopsi oleh penelitian-penelitian terdahulu, seperti model CNN, LSTM dan BiLSTM), GRU , dan model gabungan seperti CNN-BiLSTM. Pada penelitian-penelitian dengan metode *deep learning*, hasil terbaik yang didapatkan untuk CNN adalah akurasi sebesar 98,3% dengan memiliki lapisan BiLSTM di dalamnya [20]. Penelitian dengan metode LSTM secara mandiri mendapatkan hasil dengan akurasi 78,8% [6] tetapi BiLSTM mendapatkan hasil terbaik pada penggunaan CNN+BiLSTM yaitu dengan akurasi sebesar 98,3% [20]. Penelitian dengan metode GRU mendapatkan hasil terbaik dengan akurasi sebesar 97% [16] dengan *precision*

sebesar 0,85, recall sebesar 0,57, dan F1-Score sebesar 0,66 [9]. Penelitian dengan model *deep learning* umumnya mendapatkan hasil yang lebih baik daripada penelitian yang menggunakan *machine learning* konvensional.

Sesuai dengan penelitian-penelitian terdahulu yang dikumpulkan untuk kajian literatur, dapat disimpulkan bahwa masih ada peluang untuk penelitian lebih lanjut terhadap pembuatan *dataset* dan model pembelajaran mesin pada topik analisis sentimen berbasis aspek. *Dataset* yang digunakan oleh penelitian-penelitian terdahulu menggunakan teknik *web scraping* seperti yang akan diterapkan pada penelitian ini untuk melakukan pengumpulan data. Ada juga penelitian yang dapat menggunakan *dataset* yang sudah tersedia untuk memperbaiki performansi dari algoritma yang digunakan. Penelitian-penelitian terdahulu juga masih belum ada yang mengangkat subyek penelitian lingkungan sehingga membuka kesempatan penelitian akibat kebutuhan Indonesia akan pengelolaan lebih lanjut atas isu-isu lingkungan.

Penelitian terdahulu rata-rata melakukan pelabelan aspek dengan dua label, yaitu positif dan negatif *dataset* yang akan dibuat pada penelitian akan memiliki label positif, negatif, dan netral. Algoritma yang digunakan untuk penelitian terdahulu masih ada yang menggunakan teknik pembelajaran mesin klasik seperti *Naive Bayes*, *Logistic Regression*, dan *Random Forest*.

Pada penelitian ini, seperti beberapa penelitian yang terdahulu, akan digunakan arsitektur *Deep learning* dalam bentuk *Feed Forward Neural Network* sebagai pembandingan untuk *neural network* yang tidak dapat melakukan pemrosesan data secara urut, LSTM dan BiLSTM yang dipilih karena kemampuan algoritma dalam mengingat data sekuensial sehingga cocok untuk teks dan data sekuensial lainnya dalam analisis sentimen berbasis aspek. Metode evaluasi dengan menggunakan metrik akurasi, *precision*, *recall*, dan F1-Score juga diadopsi dari penelitian-penelitian terdahulu karena dapat merepresentasikan kinerja dari model secara akurat. Tabel 2.1 merupakan sebuah perbandingan penelitian-penelitian terdahulu dengan penelitian yang dilakukan.

Tabel 2.1 Tabel Perbandingan Tinjauan Pustaka

No	Judul Penelitian	Label	Subyek Penelitian	Aspek	Metode	Hasil
1	<i>Aspect-Based Sentiment Analysis on Indonesian Restaurant Review Using a Combination of Convolutional Neural Network and Contextualized Word Embedding</i> [8]	Positif dan negatif	Ulasan restoran	<i>Food, service, price, dan place</i>	CNN dan contextualized word embedding	F1 Score Sentiment: 0,91 F1 Score Aspect: 0,86
2	<i>Aspect-Based Sentiment Analysis of Online Marketplace Reviews Using Convolutional Neural Network</i> [14]	Positif dan negatif	Online marketplace	<i>Accuracy, quality, service, packaging, price, dan delivery</i>	CNN vs Naïve Bayes	F1 Score (CNN): 92,02%, F1 Score (NB): 76,96%
3	Analisis Sentimen berbasis Aspek Ulasan Pelanggan Restoran Menggunakan LSTM dengan Adam Optimizer [6]	Negatif, setengah negatif, netral, setengah positif, dan positif	Ulasan restoran	Makanan, tempat, pelayanan, dan harga	LSTM	Accuracy: 78,7%
4	Analisis Sentimen berdasarkan Aspek pada Ulasan Restoran Berbahasa Indonesia Menggunakan Convolutional Neural Network (CNN) dan Gated Recurrent Unit (GRU) [9].	Positif dan negatif	Ulasan restoran	Makanan, harga, tempat, dan pelayanan	CNN dan GRU	Precision 0,85, recall 0,57, dan F1-score 0,66

No	Judul Penelitian	Label	Subyek Penelitian	Aspek	Metode	Hasil
5	Analisis Sentimen berbasis Aspek pada Ulasan Pelanggan Restoran Bakso President Malang dengan Metode <i>Naïve Bayes Classifier</i> [7].	Positif dan negatif	Ulasan restoran	Makanan, layanan, dan atmosfer	<i>Multinomial Naïve Bayes</i>	<i>Accuracy</i> : 76–88%
6	Analisis Sentimen berbasis Aspek terhadap Ulasan Hotel Tentrem Yogyakarta menggunakan Algoritma <i>Random Forest Classifier</i> [11].	Positif, netral, dan negatif	Ulasan hotel	Pelayanan, kamar, gym, lokasi, dan kolam renang	<i>Random Forest</i>	<i>Accuracy</i> (aspek kamar): 90%
7	Analisis Sentimen berbasis Aspek pada Review Female Daily menggunakan TF-IDF dan <i>Naïve Bayes</i> [19].	Positif, netral, dan negatif	Majalah	Harga, kemasan, produk, dan aroma	<i>Naïve Bayes</i>	F1 Score: 62,81%
8	Analisis Sentimen berbasis Aspek <i>Marketing Mix</i> terhadap Ulasan Aplikasi Dompot Digital (Studi Kasus: Aplikasi Linkaja pada Twitter) [15].	Positif, netral, dan negatif	Aplikasi	Produk, harga, promosi, dan tempat	SVM	<i>Accuracy</i> : 45–77,9%
9	Perbandingan <i>Naïve Bayes</i> , SVM, dan k-NN untuk Analisis Sentimen Gadget berbasis Aspek [18].	Positif dan negatif	Gadget	Desain, harga, spesifikasi, dan citra merek	<i>Naïve Bayes</i> , SVM, k-NN	SVM Average. <i>Accuracy</i> : 96,43%
10	Analisis Sentimen berbasis Aspek pada Aplikasi Tokopedia menggunakan LDA dan <i>Naïve Bayes</i> [12].	Positif dan negatif	Aplikasi Marketplace	<i>Helpful</i> , pelayanan, pengalaman belanja, dan tampilan	<i>Naïve Bayes</i>	<i>Accuracy</i> : 92,5%, AUC: 0,95

No	Judul Penelitian	Label	Subyek Penelitian	Aspek	Metode	Hasil
11	BLTSM untuk Analisis Sentimen berbasis Aspek pada Aplikasi Belanja Online dengan Cicilan [13].	Positif, netral, dan negatif	Aplikasi Marketplace	<i>Customer service</i> , tampilan aplikasi, dan produk yang ditawarkan	BiLSTM	<i>Accuracy sentiment</i> 81%, <i>Accuracy aspect</i> 78%
12	Analisis Sentimen berbasis Aspek terhadap Data Ulasan Rumah Makan menggunakan Metode <i>Support Vector Machine</i> (SVM) [10].	Positif dan negatif	Restoran	<i>Ambience, drink quality, food quality, restaurant</i> , dan <i>service</i>	SVM	F1-Score: 0,73 <i>Accuracy</i> : 88%
13	<i>Aspect-Based Sentiment Analysis</i> pada Aplikasi Pelacakan Kasus Covid-19 (Studi Kasus: Pedulilindungi) [16].	Positif dan negatif	Aplikasi	Aplikasi <i>user interface, user experience</i> , kode OTP, kartu vaksin, dan akses layanan	CNN, GRU, dan TCN	<i>Accuracy</i> : 97,02
14	Analisis Sentimen berbasis Aspek pada Wisata Halal dengan Metode <i>Deep Learning</i> [20].	Positif dan negatif	Tempat wisata	<i>Halal food, mosque</i> , dan toilet	BiLSTM, CNN	CNN <i>Accuracy (aspect)</i> : 98,3%, CNN <i>Accuracy (sentiment)</i> : 93,96%
15	Analisis Sentimen berbasis Aspek pada Ulasan Aplikasi Shopee menggunakan Algoritma <i>Naïve Bayes</i> [17].	Positif dan negatif	Aplikasi	Video tutorial, <i>user account</i> , dan <i>customer service</i>	<i>Complement Naïve Bayes</i>	<i>Accuracy</i> : 92,5%

No	Judul Penelitian	Label	Subyek Penelitian	Aspek	Metode	Hasil
16	Pembangunan <i>Dataset</i> dan Model Klasifikasi Sentimen BerbasisAspek PadaTopik Daur Ulang Plastik	Positif, negatif, dan netral	Lingkungan , Daur ulang sampah	<i>recyclability of plastic in general, recyclability of PET plastic, recycling process, dan future of recycling</i>	FFNN, LSTM, dan BiLSTM	BiLSTM <i>average accuracy:</i> 0,84437, <i>precision:</i> 0,81117, <i>recall:</i> 0,68083, F1 Score: 0,73449