

BAB II

TINJAUAN PUSTAKA

Penelitian oleh Budaya dan Suniantara [9] bertujuan untuk mengevaluasi efektivitas algoritma *machine learning* dalam analisis sentimen terhadap ulasan Google mengenai pusat kesehatan masyarakat di Bali [9]. Metode yang digunakan melibatkan representasi teks dengan *Term Frequency-Inverse Document Frequency* (TF-IDF) dan penanganan ketidakseimbangan data menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE). Dataset terdiri dari 1.834 ulasan publik yang telah melalui proses *preprocessing* dan pelabelan ulang oleh dua pakar. Lima algoritma digunakan dalam klasifikasi, yaitu Logistic Regression, Support Vector Machine (SVM), Naive Bayes, Random Forest, dan XGBoost. Hasil evaluasi metrik menunjukkan bahwa Logistic Regression dan Naive Bayes memberikan performa terbaik dengan *accuracy* 0,89, serta nilai Random Forest memiliki *precision* tinggi pada kelas negatif (0.90), namun *recall*-nya rendah (0.79) menunjukkan banyak sentimen negatif yang tidak terdeteksi. Namun, penelitian ini masih terbatas pada pendekatan *single-label classification* dan belum mengeksplorasi hubungan antar label dalam satu dokumen, yang membuka peluang penelitian lanjutan dengan pendekatan *multi-label classification*.

Penelitian oleh Mishra dan Singh [10] bertujuan untuk mengatasi ketidakseimbangan dalam *multi-label learning* melalui metode *Feature Construction and SMOTE-based Imbalance Handling* (FCSMI). Metode ini menggabungkan konstruksi fitur berbasis jarak antar instansi minoritas dan penerapan SMOTE untuk menghasilkan data sintetis yang seimbang dalam ruang fitur baru. Dataset yang digunakan berasal dari enam *benchmark datasets* seperti Emotions, Yeast, Scene, Enron, Llog-F, dan RCV1S1. Algoritma klasifikasi dasar yang digunakan adalah Logistic Regression, dengan hasil terbaik ditunjukkan oleh FCSMI dalam metrik *AUC* mengungguli metode lain pada empat dari enam dataset, serta menempati peringkat dua tertinggi pada satu dataset lainnya. Secara numerik, FCSMI mencapai skor *AUC* tertinggi pada dataset Yeast (0.8294), Scene (0.9383), Llog-F (0.8647), dan RCV1S1 (0.8708), serta menempati posisi kedua pada dataset

Enron (0.8750) dan keempat pada dataset Emotions (0.8176). Meskipun hasilnya unggul, penelitian ini belum menjangkau penerapan pada domain spesifik seperti analisis sentimen berbasis aspek dan belum mengeksplorasi hubungan antar label secara mendalam, sehingga membuka celah untuk penelitian selanjutnya.

Saputra, Hama Aziz, dan Dimililer [11] bertujuan meningkatkan akurasi analisis sentimen pada media sosial dengan menerapkan algoritma *Extreme Gradient Boosting* (XGBoost) berbasis *ensemble learning*. Metode yang dikembangkan, bernama *SentiXGBoost*, menggabungkan enam jenis fitur (termasuk TF-IDF dan *opinion lexicon*) dan memanfaatkan berbagai *base classifiers* seperti *Naïve Bayes*, Random Forest, Logistic Regression, dan *Stochastic Gradient Descent*, yang kemudian dikombinasikan menggunakan XGBoost sebagai *meta-classifier*. Dataset utama yang digunakan adalah SemEval-2017 Task 4 Subtask B, yang berisi unggahan Twitter dengan jumlah data besar dan telah dipisahkan menjadi data latih dan uji. Hasil evaluasi menunjukkan performa terbaik dengan akurasi sebesar 90.8%, *precision* 92.7%, *recall* 98.1%, dan *F1-score* 94.0%, menjadikan *SentiXGBoost* sebagai metode unggul dibandingkan metode *ensemble* lainnya. Namun, pendekatan *multi-label classification* belum digunakan dan tidak ada analisis mendalam mengenai distribusi label yang tidak seimbang atau keterkaitan antar aspek sentimen, yang membuka peluang penelitian lanjutan.

Aurangzeb, Ayub, dan Alhussein [12] mengembangkan pendekatan *aspect-based multi-label classification* menggunakan model Support Vector Machine (SVM) yang dioptimalkan dengan *Genetic Algorithm (GA)* dalam kerangka *Evolutionary Ensembler (EEn)*. Tujuan utama penelitian ini adalah meningkatkan akurasi dan keragaman dari pembelajar *multi-label* dalam klasifikasi sentimen berbasis aspek. Metode yang digunakan mencakup *word2vec* untuk ekstraksi fitur dan optimasi multi-objektif berbasis evolusioner untuk menggabungkan akurasi serta diversifikasi model. Dataset yang digunakan terdiri dari delapan sumber data *multi-label* nyata, seperti ulasan medis, hotel, film, dan emosi, yang diperoleh dari Kaggle. Algoritma utama yang digunakan adalah SVM dengan pengoptimalan *GA*, serta dibandingkan dengan *Naïve Bayes*, *CNN*, *RAkEL*, dan *ECC*. Hasil evaluasi metrik menunjukkan bahwa model EEn menghasilkan nilai *accuracy* tertinggi

mencapai 95,10% pada dataset otomotif, *precision* hingga 94,32% pada dataset hotel, *recall* mencapai 94,00% pada dataset film, dan *F1-score* tertinggi sebesar 94,00% juga diperoleh pada dataset emosi. Namun demikian, pendekatan ini belum mengeksplorasi pengaruh langsung antar label atau hubungan semantik antar aspek dalam satu dokumen, yang dapat menjadi ruang pengembangan selanjutnya.

Tarekegn, Giacobini, dan Michalak [13] melakukan studi literatur sistematis yang mengulas berbagai pendekatan dalam menangani masalah *imbalanced multi-label classification* (MLC). Tujuan utama penelitian ini adalah untuk mengidentifikasi dan menganalisis strategi penanganan ketidakseimbangan kelas dalam konteks *multi-label*, seperti ketimpangan antar label dan antar label-set. Metode yang dibahas meliputi *resampling*, *classifier adaptation*, *ensemble methods*, dan *cost-sensitive learning*. Karena merupakan kajian literatur, dataset yang digunakan bersifat referensial dari berbagai penelitian terdahulu, serta tidak menerapkan satu algoritma spesifik, melainkan menganalisis puluhan metode. Evaluasi metrik yang dibahas mencakup yaitu *micro-F1*. Nilai *micro-F1* tertinggi dicapai oleh metode gabungan MLeNN + REMEDIAL dengan skor sebesar 0.638 pada dataset Yeast, 0.534 pada Enron, dan 0.759 pada Scene. *Research gap* yang diidentifikasi adalah perlunya pengembangan pendekatan hibrida yang menggabungkan teknik *resampling* dan *label correlation*, serta perlunya pengklasifikasi *multi-label* yang *imbalance-aware* tanpa ketergantungan pada *resampling*.

Penelitian oleh Tao dan Fang [18] bertujuan untuk meningkatkan performa analisis sentimen *multi-label* dengan mengintegrasikan pendekatan *transfer learning* pada metode *Aspect-Based Sentiment Analysis* (ABSA) dan memperkenalkan pendekatan baru bernama *Aspect Enhanced Sentiment Analysis* (AESAs). Metode yang digunakan mencakup ekstensi dua model *transfer learning* terkini, yaitu *BERT* dan *XLNet*, dengan penyesuaian fungsi aktivasi dan *loss function* agar mendukung klasifikasi *multi-label*. Dataset yang digunakan terdiri dari tiga domain berbeda: Yelp untuk ulasan restoran (8.172 kalimat), WineMag (80.638 ulasan anggur), dan Rotten Tomatoes (48.755 ulasan film), masing-masing digunakan untuk mengevaluasi ABSA dan AESA. Algoritma yang digunakan

meliputi *BERT* dan *XLNet*, yang dibandingkan dengan 27 model *baseline* termasuk *LSTM*, *SVC + CC*, dan *RF + LP*. Hasil terbaik diperoleh oleh *XLNet* pada Yelp dengan *subset accuracy* 66,61% dan *micro F1-score* 0,77, serta oleh *BERT* pada Wine dan Movie dengan *macro F1-score* hingga 0,95 dan *micro F1-score* hingga 0,97. Namun, penelitian ini masih memiliki celah, seperti performa terbatas pada input tingkat kalimat dan belum dievaluasinya efisiensi pada skenario data terbatas secara lebih luas.

Penelitian oleh Van-Ho Nguyen et al. [19] bertujuan mendeteksi emosi pelanggan dalam layanan hotel dan pariwisata menggunakan model klasifikasi *multi-label* berbasis *ensemble learning*. Metode yang diterapkan mencakup transformasi masalah menjadi *multi-class classification*, *feature selection* berbasis *variance threshold*, dan *term frequency inverse document frequency* (TF-IDF), dengan proses pelabelan manual untuk meningkatkan kualitas data. Dataset yang digunakan terdiri atas 415.998 ulasan pelanggan hotel di Vietnam yang diambil dari situs Agoda, Booking.com, dan Traveloka. Model utama yang digunakan adalah *stacking ensemble* dengan pendekatan *label powerset* (LP), yang mencapai *exact match ratio* sebesar 88%, *precision* 91%, *recall* 87%, dan *F1-score* 89%, mengungguli model *Convolutional Neural Network* (CNN) dan *Long Short-Term Memory* (LSTM). Namun, penelitian ini belum mempertimbangkan penerapan pada data waktu nyata (*real-time data*) serta eksplorasi lebih lanjut terhadap persebaran geografis dan variasi semantik dalam bahasa Vietnam, yang menjadi celah untuk penelitian berikutnya.

Penelitian oleh Salim [20] bertujuan untuk menganalisis opini publik di Twitter mengenai pencemaran udara di Jakarta dengan membandingkan algoritma Support Vector Machine (SVM) dan Naive Bayes. Penelitian ini menggunakan pendekatan *lexicon-based* dengan tahapan *text preprocessing* serta validasi melalui *k-fold cross validation* dalam kerangka metode SEMMA. Dataset dikumpulkan melalui *crawling* dari Twitter dengan kata kunci terkait polusi udara Jakarta sepanjang tahun 2023, meskipun jumlah tweet tidak dijelaskan secara eksplisit. Algoritma yang digunakan adalah SVM dan Naive Bayes untuk klasifikasi sentimen positif dan negatif. Hasil evaluasi menunjukkan bahwa SVM memperoleh akurasi sebesar

91%, *precision* 92%, *recall* 90%, dan *F1-score* 91%, sementara Naive Bayes menghasilkan akurasi 85%, *precision* 83%, *recall* 80%, dan *F1-score* 81%, yang membuktikan bahwa SVM lebih unggul dalam klasifikasi sentimen terkait isu ini. Namun, pendekatan *multi-label classification* tidak diterapkan dalam studi ini, dan hubungan antar label dalam opini publik belum dieksplorasi, sehingga menyisakan ruang untuk pengembangan lebih lanjut.

Pada penelitian sebelumnya, mayoritas penelitian berfokus pada pendekatan *single-label classification* atau menerapkan *multi-label classification* pada domain umum seperti ulasan produk, layanan, dan emosi. Penelitian ini secara khusus mengangkat isu lokal, yakni sentimen masyarakat terhadap pembangunan pagar laut di Indonesia, yang bersumber dari data Twitter. Penelitian ini menggabungkan pendekatan *aspect-based sentiment analysis* dengan dua skema klasifikasi, yaitu *single-label* dan *multi-label*, untuk mengungkap opini publik berdasarkan aspek sosial, ekonomi, dan lingkungan. Selain itu, penelitian ini turut membandingkan performa tiga algoritma Support Vector Machine (SVM), Random Forest, dan XGBoost baik dalam kondisi dengan maupun tanpa penerapan teknik penyeimbangan data seperti SMOTE dan MLSMOTE. Dengan fokus pada isu kebijakan publik yang spesifik dan pendekatan evaluasi model yang komprehensif, penelitian ini diharapkan dapat memberikan kontribusi metodologis dan aplikatif dalam pengembangan sistem klasifikasi sentimen yang akurat dan relevan secara kontekstual.

Tabel 2. 1 Tabel Penelitian Terdahulu

Penulis	Metode	Algoritma	Dataset	Metrik Evaluasi	Tipe Label
Budaya & Suniantara	TF-IDF, SMOTE	Logistic Regression, SVM, Naive Bayes, RF, XGBoost	<i>Google Reviews (1.834 ulasan)</i>	<i>Accuracy, Precision, Recall</i>	<i>Single-label</i>
Mishra & Singh	FCSMI (<i>Feature Construction</i> + SMOTE)	Logistic Regression	<i>Emotions, Yeast, Scene, Enron, Llog-F, RCVIS1</i>	<i>AUC</i>	<i>Multi-label</i>
Saputra, Hama Aziz, & Dimililer	SentiXGBoost (6 fitur termasuk TF-IDF)	NB, RF, LR, SGD + XGBoost (meta-classifier)	<i>SemEval-2017 Task 4B (Twitter)</i>	<i>Accuracy, F1-score</i>	<i>Single-label</i>
Aurangzeb, Ayub, & Alhussein	Word2Vec, Genetic Algorithm	SVM + GA, Naive Bayes, CNN, RAKEL, ECC	<i>Kaggle (Hotel, Film, Emosi, Medis)</i>	<i>Accuracy, F1-score</i>	<i>Multi-label</i>
Tarekegn, Giacobini, & Michalak	Literatur Review (Resampling, Adaptation, Ensemble)	-	-	<i>F1-score</i>	<i>Multi-label</i>
Tao & Fang	AESA, Transfer Learning (ABSA to AESA)	BERT, XLNet, LSTM, SVC+CC, RF+LP	<i>Yelp (8.172), WineMag (80.638), Rotten Tomatoes (48.755)</i>	<i>Macro F1, Micro F1</i>	<i>Multi-label</i>
Nguyen et al.	TF-IDF, LP, Feature Selection, Ensemble	LP Ensemble, CNN, LSTM	<i>Agoda, Booking.com, Traveloka (415.998 ulasan)</i>	<i>Exact Match, F1-score</i>	<i>Multi-label</i>
Salim	Lexicon-based, SEMMA, k-fold CV	SVM, Naive Bayes	Twitter (<i>Polusi Jakarta</i>)	<i>Accuracy, F1-score</i>	<i>Single-label</i>
Martinus	TF-IDF, Binary	SVM, Random Forest,	Twitter (<i>Pagar Laut</i>)	<i>Precision, Recall, F1-</i>	<i>Single-label dan</i>

Bagas Billy Valentino Putra	Relevance, SMOTE & non-SMOTE	XGBoost		Score, Confusion Matrix	<i>Multi-label</i>
-----------------------------------	---------------------------------	---------	--	----------------------------	---------------------------

