



Systematic Literature Review: Analisa Sentimen Masyarakat terhadap Penerapan Peraturan ETLE

Muhammad Syiarul Amrullah¹, Syafriah Fachri Pane²

^{1,2}Sekolah Vokasi, Sarjana Terapan Teknik Informatika, Universitas Logistik dan Bisnis Internasional

¹syiarul45@gmail.com, ²syafriah.fachri@ulbi.ac.id*

Abstract

This study examines the efforts to develop a model for analyzing public sentiment regarding applying ETLE (Electronic Traffic Law Enforcement) regulations. The method used is the systematic literature review. A systematic literature review (SLR) consists of three stages: planning, conducting, and reporting. The planning stage is the determination of the SLR procedure. This stage includes preparing topics, research questions, article search criteria & inclusion and exclusion criteria. The conducting stage, namely the implementation, includes searching for articles and filtering articles. The reporting stage is the final stage of SLR. This stage includes writing the SLR results according to the article format. The explanation follows: First, hybrid is the most widely used method in developing sentiment analysis models. Apart from hybrid, several methods are used to develop sentiment analysis models, including multi-task, deep, and machine learning. Each has its advantages and disadvantages in the development of sentiment analysis models. Second, this study shows the development of a model with superior performance, namely using XGBoost as a sentiment analysis model, and the stages it goes through are preprocessing data, handling imbalanced data, and optimizing the model. Therefore, the model for analyzing public sentiment regarding the application of ETLE regulations can be an option for hybrid methods, multi-task learning, deep learning, machine learning, and the XGBoost model to obtain superior performance with preprocessing data stages, handling imbalanced data and optimization models.

Keywords: *Sentiment Analysis, Systematic Literature Review, Machine Learning, Natural Language Processing, ETLE*

Abstrak

Penelitian ini bertujuan untuk mengkaji upaya yang dilakukan dalam mengembangkan model analisa sentimen masyarakat terhadap penerapan peraturan ETLE (*Electronic Traffic Law Enforcement*). Metode yang digunakan yaitu *systematic literature review*. *Systematic literature review* (SLR) terdiri dari tiga tahap yaitu *planning*, *conducting* dan *reporting*. Tahap *planning* yaitu penetapan prosedur SLR. Tahap ini meliputi penyusunan topik, *research question*, kriteria pencarian artikel & kriteria inklusi dan eksklusi. Tahap *conducting* yaitu pelaksanaan meliputi pencarian artikel dan filterisasi artikel. Tahap *reporting* yaitu tahapan akhir dalam SLR. Tahap ini meliputi penulisan hasil SLR sesuai dengan format artikel. Adapun penjelasannya sebagai berikut: pertama, *hybrid* merupakan metode yang paling banyak digunakan dalam mengembangkan model analisa sentimen. Selain *hybrid*, terdapat beberapa metode yang digunakan dalam mengembangkan model analisa sentimen, antara lain yaitu *multi-task learning*, *deep learning* dan *machine learning*. masing-masing memiliki keunggulan dan kekurangan tersendiri dalam pengembangan model analisa sentimen. Kedua, penelitian ini menunjukkan pengembangan model dengan kinerja unggul yaitu menggunakan XGBoost sebagai model analisa sentimen dan tahapan yang dilalui yaitu *preprocessing data*, *handling imbalanced data* dan *optimization model*. Maka dari itu, model analisa sentimen masyarakat terhadap penerapan peraturan ETLE dapat menjadi pilihan untuk menggunakan metode *hybrid*, *multi-task learning*, *deep learning*, *machine learning* dan model XGBoost untuk mendapatkan kinerja unggul dengan tahapan *preprocessing data*, *handling imbalanced data* dan *optimization model*.

Kata kunci: *Analisis Sentimen, Systematic Literature Review, Machine Learning, Natural Language Processing, ETLE*

1. Pendahuluan

ETLE (*Electronic Traffic Law Enforcement*) merupakan sistem penegakan hukum dan regulasi tertib berlalu lintas dengan penggunaan teknologi untuk membuat pengendara merasa lebih aman dalam berkendara [1]. ETLE mewujudkan berlalu lintas yang aman dengan menangkap pelanggar lalu lintas dan

memberikan tilang kepada pelanggar [2]. ETLE menggunakan kamera dan alat elektronik lainnya untuk mengambil gambar dan data dari pelanggar lalu lintas [3]. Kamera dan sensor induksi magnetik digunakan untuk mendeteksi dan mengambil gambar dari pelanggar [4]. Sensor LDR (*Light Dependent Resistor*) dan sensor laser juga digunakan untuk mendeteksi pelanggaran pada lampu merah untuk pengendara yang



Lisensi

Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.

melewati *zebra cross* [5]. Hasilnya, gambar dan data pelanggar yang disimpan akan diintegrasikan dengan *Automatic Number Plate Recognition (ANPR) System* sebagai *management traffic fines*. ANPR telah terintegrasi dengan kepolisian, kejaksaan, bank, kementerian keuangan dan catatan elektronik pelanggar lalu lintas.

Implementasi berhasil merubah kebiasaan pengemudi, seperti di pangkalpinang [6]. Selain pangkalpinang, Adapun beberapa studi yang menunjukkan ETLE memiliki efektifitas dan keberhasilan tinggi, seperti: penelitian pada kota pekanbaru, polrestabes semarang dan wilayah hukum kepolisian resor bojonegoro [7]–[9]. Namun, ada beberapa studi yang menunjukkan bahwa ETLE belum efektif dikarenakan kurangnya kesadaran masyarakat dalam patuh aturan saat berkendara di jalan raya, kurangnya peralatan kamera dan *data valid* untuk deteksi pelanggar lalu lintas [10]. Adanya perbedaan hasil menunjukkan bahwa dari masing-masing studi tidak bisa digeneralisir untuk setiap daerah sehingga dapat menjadi gap pada penelitian yang dilakukan. Selain itu, terdapat gap lain yaitu, penelitian hanya mengukur efektifitas program dengan beberapa kriteria sehingga tidak mencakup aspek-aspek lainnya. Maka dari itu, artikel ini bertujuan untuk mengisi gap tersebut dengan analisa sentimen.

Analisa sentimen merupakan proses mengidentifikasi, mengekstrak, dan mengevaluasi emosi atau pendapat yang disampaikan dalam sebuah teks dengan pendekatan NLP (*Natural Language Processing*) [11]. Analisa sentimen dapat digunakan untuk menganalisis seluruh opini di dunia tanpa batasan melalui berbagai platform media sosial dan internet, hal ini menjadi alasan penting kenapa harus analisa sentimen. Perkembangan Ilmu pengetahuan dan teknologi telah merubah cara manusia dalam berkomunikasi, terhubung dan berbagi informasi. Media sosial merupakan salah satu teknologi yang paling berkembang dan banyak digunakan dalam berbagi informasi dan berkomunikasi [12]. Pertumbuhan pengguna media sosial meningkat pesat dengan 2.77 miliar pengguna media sosial aktif diseluruh dunia sejak 2019 [13]. Pertumbuhan pengguna media sosial berbanding lurus dengan peningkatan informasi yang dibagikan [14]. Twitter merupakan salah satu media sosial dengan komunikasi yang baik diantara pengguna. Twitter memberikan layanan kepada penggunanya untuk mengirim dan mengekspresikan opininya melalui media sosial [15]. Banyaknya pengguna yang mengekspresikan opininya dapat dijadikan sebagai data sentimen.

Adapun kumpulan penelitian terkait analisa sentimen terhadap isu-isu sosial di media sosial, seperti analisa sentimen dilakukan untuk menganalisa opini masyarakat indonesia terhadap kebijakan PPKM (pemberlakuan pembatasan kegiatan masyarakat) pada media sosial twitter, analisa sentimen masyarakat terhadap vaksinasi covid dan analisa sentimen terhadap

harga kenaikan dasar listrik [16]–[18]. Berdasarkan tiga penelitian tersebut, menunjukkan bahwa analisa sentimen dapat mengisi gap terkait terbatasnya lingkup penelitian dengan analisa sentimen melalui sosial media untuk menganalisa opini masyarakat indonesia. Analisa sentimen yang dilakukan tiga penelitian sebelumnya termasuk dalam konteks *machine learning* yang mana menggunakan metode seperti *naive bayes & decision tree*. *Machine learning* merupakan cabang dari ilmu dari *artificial intelligence* dengan fokus pembelajaran komputer secara otomatis dengan data [19].

Analisa sentimen dengan *machine learning* memiliki beberapa keuntungan sebagai berikut: akurasi tinggi, penghematan waktu dan biaya dan kemampuan memproses bahasa yang kompleks [20]. Maka dari itu, penelitian ini dilakukan dalam konteks *machine learning*. Sebelum melakukan analisa sentimen terhadap masyarakat terkait ETLE perlu dilakukan *literature review* untuk menemukan metode yang tepat, menemukan gap dalam penelitian sebelumnya, memahami teori dan meningkatkan kualitas penelitian [21]. Penelitian ini memiliki kontribusi terhadap pengembangan analisa sentimen dengan eksplorasi terkait metode analisa sentimen yang banyak digunakan dan mengetahui pengembangan model analisa sentimen untuk menganalisa opini terkait penerapan peraturan ETLE, sehingga dapat mengembangkan model yang dapat mengisi gap pada model yang ada. Penelitian ini mengumpulkan data tahun publikasi, variabel, negara sumber dataset, metode, model dan evaluasi performa atau matriks (akurasi, presisi, *recall* dan F1). Penelitian ini dirancang dengan penulisan sebagai berikut: Bagian 1, menjelaskan latar belakang penelitian meliputi masalah dan solusi dalam penelitian, gap dari penelitian sebelumnya dan kontribusi dari penelitian. Bagian 2 meliputi penjelasan tahapan pada penelitian dan penerapan metode penelitian. Bagian 3 meliputi penjelasan hasil dan pembahasan penelitian.

2. Metode Penelitian

Topik pada penelitian ini adalah analisa sentimen masyarakat terhadap ETLE (*Electronic Traffic Law Enforcement*). Berikut merupakan beberapa *research question* pada penelitian ini: pertama, (RQ1) apa metode yang banyak digunakan dalam menganalisa sentimen?. Kedua, (RQ2) membangun model dengan kinerja unggul?. Berdasarkan hasil studi literatur, metode yang digunakan untuk meninjau penelitian yaitu *Systematic Literature Review (SLR)*, dikarenakan metode ini dapat melakukan pencarian literatur menyeluruh dan komprehensif terhadap studi yang relevan dengan langkah terdokumentasi dengan baik [22]. *Systematic Literature Review (SLR)* adalah suatu metode untuk secara sistematis mengidentifikasi, mengevaluasi dan meringkas hasil penelitian yang dilakukan dalam bidang tertentu [23]. Tujuannya adalah untuk memberikan bukti yang tepat dan obyektif dalam bidang tertentu [24]. SLR memiliki tiga tahap yaitu

tahap *planning, conducting & reporting*. Tahap *planning* merupakan tahapan perencanaan yang dilakukan sebelum mencari artikel, seperti menentukan *keyword, database article*, kriteria inklusi dan kriteria eksklusi [25]. Setelah tahap *planning*, selanjutnya tahap *conducting*. Tahap *conducting* merupakan tahap pelaksanaan dari SLR, dimulai dengan pencarian artikel dan *screening* artikel dengan ketentuan pada tahap *planning* [26]. Setelah tahap *conducting*, maka dilanjutkan dengan tahapan *reporting*. Tahap *reporting* merupakan tahap akhir pada SLR yang meliputi penulisan temuan penelitian secara sistematis [27]. PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) digunakan sebagai pedoman dalam melakukan SLR. PRISMA merupakan seperangkat pedoman berdasarkan bukti untuk melaporkan meninjau secara sistematis dan meta analisis [28]. PRISMA bertujuan meningkatkan transparansi dan kelengkapan laporan penelitian [29]. PRISMA memiliki empat tahap dalam meninjau literatur yaitu *identification, screening, eligibility* dan *included record studies* [30].

Identification pada PRISMA merujuk pada tahapan identifikasi penelitian yang relevan atau sumber data yang termasuk kedalam *systematic review*. Tahap ini meliputi pemilihan *database, search terms* dan *filter duplicated papers*. *Database* Google Scholar digunakan pada penelitian ini sebagai sumber pencarian artikel. Pengumpulan literatur menggunakan *tools publish or perish (POP)*, sebagai *tools* bantuan dalam mencari jurnal. *Publish or perish* merupakan perangkat lunak untuk membantu para peneliti dalam memperoleh informasi tentang kinerja publikasi mereka dan melakukan pencarian literatur. POP dapat melakukan pencarian berdasarkan rentang tahun publikasi artikel dan *keyword*. Penelitian ini menggunakan rentang waktu 2018 sampai dengan 2023 dan *keyword* yang digunakan adalah “*Sentiment Analysis*”, “*machine learning*” & “*sentiment analysis using machine learning*”.

Setelah tahap *identification*, selanjutnya tahap *screening*. tahap ini melaksanakan *screening* dari artikel yang telah dikumpulkan pada *identification*. *Screening* dilakukan berdasarkan kriteria inklusi dan eksklusi. Kriteria inklusi merupakan kriteria untuk menyeleksi artikel, yang mana jika kriteria ini dipenuhi artikel termasuk layak, sedangkan kriteria eksklusif adalah kriteria yang jika kriteria tersebut dimiliki artikel, maka artikel tersebut tidak layak [31]. adapun kriteria inklusi sebagai berikut: pertama, kriteria dalam rentang waktu 2018-2023, kedua artikel penelitian sesuai dengan topik penelitian dan ketiga, artikel memiliki SJR (*Scimago Journal Rank*) *Score* lebih dari sama dengan satu. Berikut adalah kriteria eksklusi: pertama, artikel penelitian tidak berada dalam rentang waktu 2018-2023, kedua topik penelitian artikel tidak relevan dan ketiga, SJR *Score* kurang dari satu. SJR

score dan tahun publikasi artikel dijadikan kriteria inklusi karena dapat mengurangi bias dan memastikan semua artikel ditinjau *fairly* [32].

Setelah tahap *screening*, maka dilanjutkan dengan tahap *Eligibility*. tahap *eligibility* merupakan tahap memilah *papers* lebih dalam untuk dianalisis. *Abstract, title* dan *full text* menjadi acuan dalam memilah *papers*, sehingga *papers* tersebut dapat dianalisa apakah relevan dengan topik penelitian.

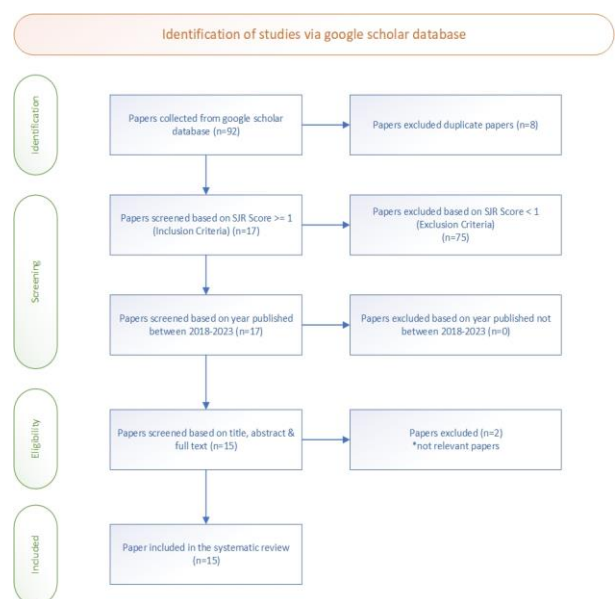
Setelah tahap *eligibility*, maka dilanjutkan dengan tahapan *included*. Tahapan *included* merupakan tahap akhir pada SLR. *Papers* yang telah melalui setiap tahap pada PRISMA menjadi hasil *final* pada protokol *included*.

3. Hasil dan Pembahasan

Temuan-temuan artikel yang telah disintesis dan dianalisis akan disajikan secara sistematis dan terfokus, termasuk kesimpulan yang diperoleh dari hasil analisis pada bagian ini.

3.1. Hasil

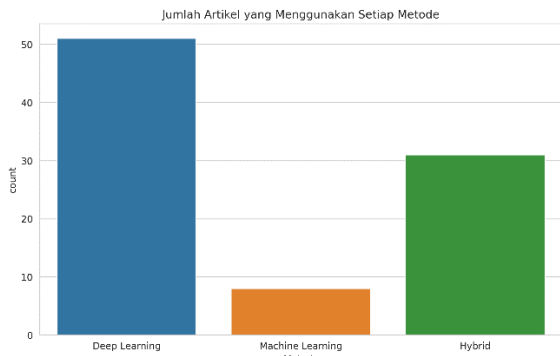
Gambar 1, merupakan alur PRISMA pada penelitian. Berdasarkan Gambar 1, sebanyak 100 artikel dikumpulkan, namun 92 artikel lolos tahap identifikasi dan 8 artikel dikecualikan karena sudah termasuk dalam 92 artikel tersebut. Selanjutnya, 92 artikel pada tahap identifikasi dilakukan filterisasi dengan kriteria inklusi dan eksklusi, dengan hasilnya 17 artikel yang dilanjutkan ke tahap eligibilitas dan 75 artikel dikecualikan. Pada tahap eligibilitas, terdapat 15 artikel yang masuk dalam tinjauan sistematis dan 2 artikel dikecualikan.



Gambar 1. Prisma Diagram

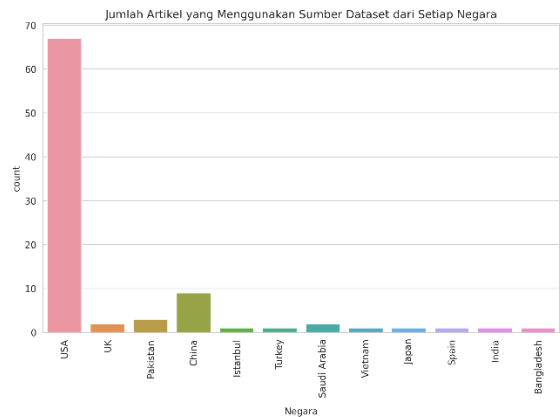
100 artikel yang telah dikumpulkan pada tahap *identification*. Berdasarkan Gambar 2, terdapat 50

artikel dengan metode *deep learning*, 8 artikel dengan metode *machine learning*, 1 artikel dengan metode *multi-task Learning* dan 33 artikel dengan metode *hybrid*.



Gambar 2. Grafik jumlah artikel berdasarkan metode yang digunakan

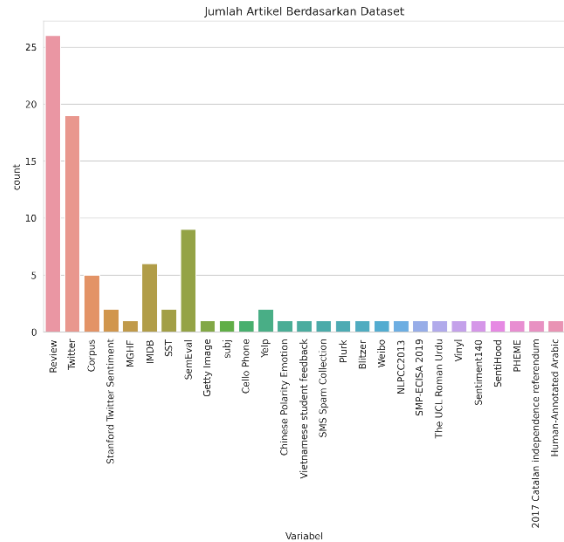
Berdasarkan Gambar 3, terdapat 68 artikel yang bernegara di USA, 9 artikel bernegara di China, 3 artikel bernegara di Pakistan, 2 artikel bernegara di UK dan Arab Saudi, 1 artikel masing-masing bernegara di Turki, Istanbul, Vietnam, Jepang, Spanyol, India dan *Bangladesh*.



Gambar 3. Grafik jumlah artikel berdasarkan negara

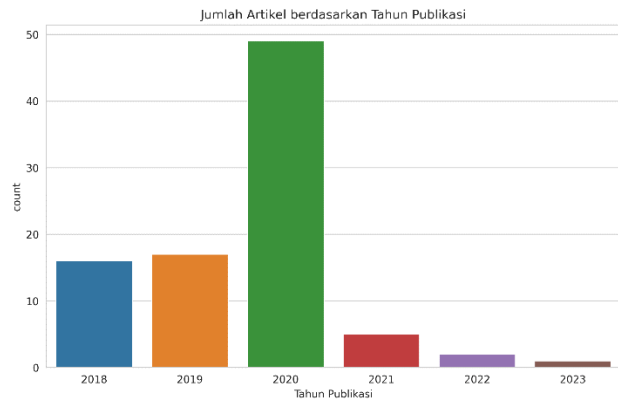
Berdasarkan Gambar 4, 28 artikel dengan variabel *review*, 19 artikel dengan *twitter*, 9 artikel dengan *SemEval*, 5 artikel dengan *Corpus*, 6 artikel dengan *IMDB Movie Reviews*, 2 artikel dengan variabel *SST (Stanford Sentiment Treebank)*, *Stanford Twitter Sentiment* dan *Yelp*, 1 artikel dengan variabel *MGHF*, *Getty Image*, *Subj (Subjectivity dataset)*, *CelloPhone*, *Chinese polarity emotion*, *Vietnamese student feedback*, *plurk*, *SMS Spam Collection*, *blitzer*, *weibo*, *NLPCC2013 (International Conference on Natural*

Language Processing and Chinese Computing 2013), *Chinese implicit sentiment*, *The UCL (University College London) Roman Urdu*, *Vinyl*, *Sentiment140*, *Sentihood*, *PHEME (Computing Veracity of Information in Social Media)*, *2017 catalan independence referendum* dan *Human-Annotated Arabic*.



Gambar 4. Grafik jumlah artikel variabel artikel

Berdasarkan Gambar 5, 16 artikel publikasi tahun 2018, 16 artikel publikasi tahun 2019, 48 artikel publikasi tahun 2020, 4 artikel publikasi tahun 2021 dan 2 artikel publikasi tahun 2022. Artikel yang *included* dianalisa dan diekstrak informasinya lalu digabungkan. Tabel 1. merupakan hasil ekstraksi informasi dari artikel *included*.



Gambar 5. Grafik jumlah artikel berdasarkan tahun publikasi

Tabel 1. Rangkuman Artikel Terpilih

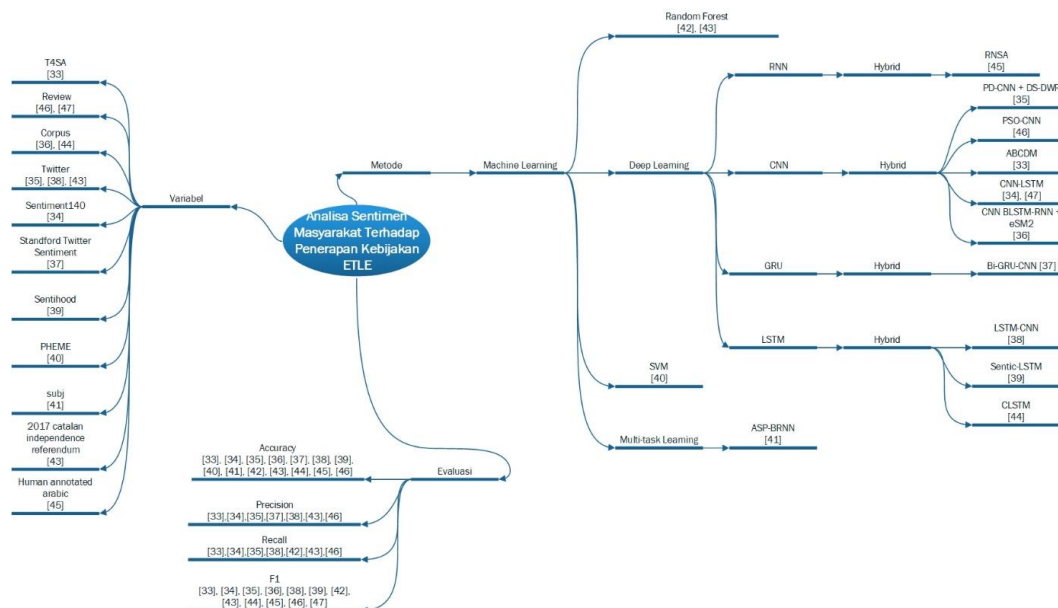
No	Variabel	Negara	Model	Metode	Akurasi	Presisi	Recall	F1
[33]	T4SA	USA	ABCDM (<i>Attention-based Bidirectional CNN-RNN Deep Model</i>)	Hybrid	96.87%	99.08%	99.04%	96.94%
[34]	Sentiment140	USA	CNN-LSTM (<i>Convolutional Neural Network - Long Short-Term Memory</i>)	Hybrid	84.56%	83.90%	84.50%	84.10%

[35]	Twitter	USA	PD-CNN (parallel dilated convolution neural network) + DS-DWR (domain-specific distributed word representation)	Hybrid	76.29%	75.93%	75.64%	75.18%
[36]	Corpus	USA	CNN BLSTM-RNN (Convolutional Neural Network - Bidirectional Long Short-Term Memory Recurrent Neural Network)	Hybrid	93.00%			95.00%
[37]	Stanford Twitter Sentiment	USA	Bi-GRU-CNN (Bidirectional Gated Recurrent Unit - Convolutional Neural Network)	Deep Learning	90.59%			
[38]	Twitter	UK	LSTM-CNN (Long Short-Term Memory - Convolutional Neural Network)	Hybrid	92.66%	88.32%	89.47%	90.71%
[39]	SentiHood	Japan	Sentic LSTM (Long Short-Term Memory)	Hybrid	89.32%			78.49%
[40]	PHEME	USA	SVM (Support Vector Machine)	Machine Learning	89%			
[41]	subj	USA	ASP-BRNN (Attention-based Separate Pooling Bidirectional Recurrent Neural Network)	Multi-task learning	95.70%			
[42]	2017 Catalan independence referendum	Spain	Random Forest	Machine Learning	85.80%	92.22%	89.50%	90.80%
[43]	Twitter	India	Random Forest	Machine Learning	96.62%	99.00%	99.00%	98.00%
[44]	Corpus	Bangladesh	CLSTM (Convolutional Long Short-Term Memory)	Deep Learning	85.80%			86.00%
[45]	Human-Annotated Arabic	Saudi Arabia	RNSA (Recurrent Neural Sentiment Analysis)	Hybrid	82.60%			69.98%
[46]	Review	USA	PSO-CNN (Particle Swarm Optimization - Convolutional Neural Network)	Hybrid	88.52%	94.30%	85.63%	86.03%
[47]	Review	USA	CNN-LSTM (Convolutional Neural Network - Long Short-Term Memory)	Hybrid				89.44%

3.2. Pembahasan

Pembahasan akan menjelaskan hasil dari informasi yang telah diekstraksi sehingga tujuan penelitian ini dapat tercapai. Tujuan dari penelitian ini yaitu untuk memperoleh deskripsi tentang analisa sentimen untuk

menjawab dua pertanyaan penelitian terkait (RQ1) apa metode yang banyak digunakan dalam analisa sentimen dan model apa yang memiliki kinerja unggul (Akurasi, Presisi, *Recall* dan *F1 Score*). Adapun pembahasannya sebagai berikut:



Gambar 6. Taksonomi Studi Literatur

3.2.1 (RQ1) apa metode yang banyak digunakan dalam menganalisa sentimen?

Dalam mengembangkan model analisa sentimen, terdapat beberapa metode yang digunakan. Berdasarkan Gambar 6, artikel yang terkumpul menjelaskan penggunaan metode seperti *deep learning*, *machine learning*, *multi-task learning* dan *hybrid*. *hybrid* menjadi metode populer digunakan dalam analisa sentimen. Selanjutnya ialah metode *deep learning*, *multi-task learning* dan *machine learning*. Hal tersebut dapat dilihat dari kuantitas pada metode *deep learning* dengan 2 artikel, metode *machine learning* dengan 3 artikel, metode *multi-task learning* dengan 1 artikel dan metode *hybrid* dengan 9 artikel. Berikut merupakan penjelasan dari masing-masing metode dalam mengembangkan model analisa sentimen.

Salah satu metode yang digunakan untuk mengembangkan model analisa sentimen yaitu *deep learning* [48]. *deep learning* merupakan subbidang dari *machine learning* yang terinspirasi oleh otak manusia dengan penggunaan jaringan saraf tiruan dalam memproses data [49]. Berdasarkan artikel yang telah dikaji, beberapa studi dengan metode *deep learning*, menunjukkan pengembangan model analisa sentimen dengan efektif, terutama penggunaan data twitter sebagai variabel [37]. Data dengan penggunaan banyak bahasa memiliki banyak tantangan dalam proses analisa sentimen. Tantangan dalam penggunaan data dengan banyak bahasa seperti dataset tidak memiliki label yang sesuai disetiap datanya, lalu alat *preprocessing* kurang memadai pada beberapa bahasa. Studi yang menggunakan dataset dengan bahasa bengali mengalami tantangan tersebut. Namun, tantangan tersebut dapat diatasi dengan pembuatan dataset *multi class* dari beberapa media sosial seperti twitter, facebook, media sosial lainnya. Lalu, tindakan *preprocessing* dengan menghapus *emoticon*, simbol, karakter khusus, *stemming*, penghapusan *stop words* dan ekstraksi fitur dengan TF-IDF untuk meningkatkan kinerja algoritma klasifikasi teks, hasilnya akurasi mencapai 85.8% dan F1 86% menggunakan model CLSTM [44].

Selain *deep learning*, *machine learning* juga merupakan salah satu metode yang digunakan dalam pengembangan model analisa sentimen [50]. *Machine learning* digunakan dalam analisa sentimen untuk mengidentifikasi, mengekstrak opini dari data teks [48]. Berdasarkan artikel yang telah dikaji, *machine learning* dapat memiliki hasil lebih baik dari model *deep learning*, hasil tersebut merupakan dari studi yang telah dilakukan oleh peneliti terkait analisa sentimen dengan data twitter, namun dari hasilnya yang tinggi studi ini memiliki beberapa kekurangan atau gap seperti: kurangnya teknik menyimpan konten twitter secara otomatis dan pertimbangan fitur konteks visual yang masih kurang [42]. Beberapa studi lainnya, memiliki kekurangan atau gap yang sama dengan penelitian sebelumnya [40], [43].

Setelah *deep learning* dan *machine learning*, ada sebuah paradigma yang memanfaatkan informasi yang berhubungan kedalam beberapa tugas untuk meningkatkan performa generalisasi semua *task* disebut *multi-task machine learning* [51]. Berdasarkan hasil kajian literatur, *multi-task machine learning* dapat meningkatkan kemampuan generalisasi dari model untuk setiap tugasnya dengan penggunaan arsitektur *neural network* yang terdiri dari lapisan bersama dan lapisan khusus [52]. Lapisan bersama digunakan untuk mengekstrak fitur-fitur umum dari data yang digunakan dalam semua tugas, dan lapisan khusus digunakan untuk mengekstrak fitur-fitur dari penggunaan data disetiap *task* [53]. Berikut adalah studi yang menggunakan pendekatan *multi-task learning* dalam menganalisa sentimen, penelitian dengan tujuan mengembangkan model analisa sentimen dengan pendekatan *multi-task learning* [41]. Penelitian ini memiliki beberapa permasalahan yang menjadi fokus pada penelitian ini, yaitu: mengatasi keterbatasan model yang sudah ada dan improvisasi kinerja model terhadap tugas-tugas analisa sentimen. Usulan 3 model dari penelitian ini, yaitu: *Unified Pooling RNN (UP-RNN)*, *Unified Pooling Bidirectional RNN (UP-BRNN)*, *Attention-based Separate Pooling Bidirectional Recurrent Neural Network (ASP-BRNN)* dan *Separate Pooling Bidirectional RNN (SP-BRNN)*. Keempat model tersebut, akan diuji dengan delapan *task* analisa sentimen. Hasilnya, ASP-BRNN memiliki hasil paling unggul dari ketiga model lainnya. ASP-BRNN memiliki dua improvisasi dalam penggunaan fitur. Pertama, fitur bersama tidak berikan tugas ketika menangkap fitur khusus. Kedua, penggunaan mekanisme *attention* untuk pembobotan setiap fitur bersama. Penggunaan mekanisme *attention* berdasarkan konteks vektorisasi kata.

Beberapa metode memiliki kelebihan masing-masing dalam memproses data. Namun, dari masing-masing model juga memiliki kekurangan atau gap pada model analisa sentimen yang telah dikembangkan, seperti: membuat ruang dengan dimensi tinggi pada *layer*, memperlakukan fitur yang berbeda sama pentingnya, masalah *vanishing*, kurangnya pertimbangan konteks visual & *exploding gradients* pada model dan masalah *co-occurrence* dari ketergantungan terhadap input teks panjang dan pendek [33]. Adanya gap dalam penelitian, membuat banyak peneliti melakukan studi untuk mengisi gap tersebut. Berdasarkan artikel yang telah dikaji, banyak studi melakukan penggabungan beberapa metode dalam satu model analisa sentimen. Metode tersebut dinamakan *Hybrid*.

Metode *Hybrid* merupakan kolaborasi dari pendekatan yang berbeda untuk meningkatkan akurasi. Analisa sentimen dengan metode *Hybrid* menggabungkan pendekatan seperti *lexical-based methods*, *machine learning-based methods*, & *deep learning-based*

methods. Hasilnya, akurasi menjadi lebih akurat dari penggunaan satu pendekatan. Adapun penjelasan terkait kumpulan studi dengan *improvement* menggunakan metode *hybrid*: studi dengan *proposed* model ABCDM (*Attention-based Bidirectional CNN-RNN Deep Model*), berhasil mengisi kekurangan atau gap dalam penelitian. Adapun pencapaian pada studi ini adalah ketergantungan jangka panjang terhadap data input dengan penggunaan layer ekstraksi fitur yang terdiri dari Bi LSTM dan GRU, mengatasi masalah *co-occurrence* dengan *layer* yang terdiri dari CNN dengan *kernel size* yang bervariasi, penggunaan *attention layer* untuk memberikan bobot disetiap kata berdasarkan pentingnya dan penggunaan teknik *stack generalization* untuk meningkatkan performa model, hasilnya model ABCDM memiliki performa maksimal dari akurasi dan *F1 Score* [33]. Disisi lain, studi dengan model PD-CNN (*Parallel Dilated Convolution Neural Network*) + DS-DWR (*Domain-Specific Distributed Word Representation*), berhasil mengatasi beberapa masalah pada penelitian dengan metode *hybrid* seperti masalah kompleksitas dan penghabisan waktu pada teknik pengolahan fitur [35]. Selanjutnya, studi dengan model CNN BLSTM-RNN (*Convolutional Neural Network Bidirectional Long Short-Term Memory Recurrent Neural Network*), berhasil menambahkan fungsi rekomendasi dengan pendekatan *content-based* dan *collaborative filtering* pada model analisa sentimen. Hasilnya, sistem ini dapat mendeteksi depresi dan *stress* dan dapat memberikan pesan yang menenangkan, membuat senang dan memberikan peringatan sesuai rekomendasi sistem [36]. Beberapa studi lainnya, fokus pada peningkatan akurasi menggunakan pendekatan optimasi, integrasi pengetahuan umum kedalam jaringan saraf; *AffectiveSpace*, penggunaan ontologi dan teknik ekstraksi fitur seperti; *word embedding* dan leksikon sentimen [36], [37], [39], [45], [46]. Maka dari itu, metode *hybrid* menjadi metode yang populer digunakan dalam studi sebagai peningkatan model analisa sentimen dengan satu metode atau pendekatan. Diikuti dengan *multi-task learning*, *deep learning* dan *machine learning*.

3.2.2 (RQ2) membangun model dengan kinerja unggul?

Selain metode dalam analisa sentimen, algoritma dan proses penelitian termasuk dalam bagian penting dalam pengembangan model analisa sentimen. Algoritma dan proses yang tepat, dapat menghasilkan model dengan kinerja unggul. Analisa sentimen memiliki beberapa proses yang perlu dilakukan seperti *preprocessing data*, *handling imbalanced data* dan *optimization model*. *Preprocessing data* merupakan proses mengubah data mentah menjadi data dengan format yang dengan mudah digunakan disegala bidang [54]. *Preprocessing data* analisa sentimen memiliki beberapa tahapan seperti: *cleaning text*, *tokenization*, *text normalization*, *spell checking*, *dimensionality reduction*. *Preprocessing data* perlu dilakukan karena dapat

mencegah model dari *overfitting* [55]. *Overfitting* merupakan suatu kondisi dari model analisa sentimen menjadi terlalu kompleks sehingga hasil prediksi tidak dapat digeneralisasi dengan baik [56]. Selain *preprocessing data*, tahapan selanjutnya yaitu *handling imbalanced data*. *Imbalance data* merupakan sebuah kondisi dimana antar kelas pada dataset tidak memiliki jumlah yang seimbang [57]. *Imbalance data* juga dapat menyebabkan model menjadi *overfitting* [58]. Selain *handling imbalanced data*, tahapan selanjutnya yaitu *optimization model*. *Optimization model* merupakan proses menemukan *hyperparameter* terbaik untuk model, sehingga dapat meminimalisir kesalahan model [59]. Peneliti telah melakukan *review* dari algoritma dan proses setiap studi pada artikel terpilih dan menuliskannya pada Tabel 2.

Berdasarkan Tabel 2, banyak model tidak melalui semua proses atau tahapan analisa sentimen. Hal tersebut dapat menyebabkan model menjadi *overfitting*, jika model menjadi *overfitting* walaupun memiliki kinerja unggul, model tersebut tidak dapat menerima data baru dan akurasi yang tidak akurat. Maka dari itu peneliti membuat *proposed model* dan menentukan tahapan yang akan digunakan sebagai analisa sentimen terhadap penerapan peraturan ETLE. *Proposed model* yang digunakan peneliti yaitu XGBoost. XGBoost merupakan algoritma *machine learning* yang digunakan untuk masalah optimasi pada pengaplikasian *machine learning* [60]. XGBoost menjadi *proposed model* dikarenakan studi yang telah dilakukan menyatakan bahwa XGBoost kompatibel dalam melakukan analisa sentimen [61].

4. Kesimpulan

Systematic Literature Review yang dilakukan memberikan informasi untuk mengembangkan model analisa sentimen masyarakat terhadap penerapan peraturan ETLE. Penelitian ini memiliki kontribusi terhadap pengembangan analisa sentimen dengan eksplorasi terkait metode analisa sentimen yang banyak digunakan dan mengetahui pengembangan model analisa sentimen untuk menganalisa opini terkait penerapan peraturan ETLE. Berdasarkan kajian 15 artikel, pertama, *hybrid* merupakan metode yang paling banyak digunakan dalam mengembangkan model analisa sentimen. Selain *hybrid*, terdapat beberapa metode yang digunakan dalam mengembangkan model analisa sentimen, antara lain yaitu *multi-task learning*, *deep learning* dan *machine learning*. Masing-masing memiliki keunggulan dan kekurangan tersendiri dalam pengembangan model analisa sentimen. Kedua, penelitian ini menunjukkan membangun model dengan kinerja unggul yaitu menggunakan XGBoost sebagai model analisa sentimen dan tahapan yang dilalui yaitu *preprocessing data*, *handling imbalanced data* dan *optimization model*. Maka dari itu, model analisa sentimen masyarakat terhadap penerapan peraturan ETLE dapat menjadi pilihan untuk menggunakan

metode *hybrid, multi-task learning, deep learning, preprocessing data, handling imbalanced data* dan *machine learning* dan model XGBoost untuk *optimization model*.
 mendapatkan kinerja unggul dengan tahapan

Tabel 2. Membangun Model dengan Kinerja Unggul

Penelitian	Model	Preprocessing data					handling imbalanced data	Optimization model
		Cleaning Text	Tokenization	Text Normalization	Spell Checking	Dimensionality reduction		
[33]	ABCDM	✓	✓			✓	✓	✓
[34]	CNN-LSTM							
[35]	PD-CNN + DS-DWR	✓	✓	✓			✓	✓
[36]	BLSTM-RNN						✓	
[37]	Bi-GRU-CNN	✓		✓	✓			✓
[38]	LSTM-CNN		✓					✓
[39]	Sentic LSTM	✓	✓					
[40]	SVM	✓						
[41]	ASP-BRNN	✓		✓		✓		✓
[42]	Random Forest	✓	✓	✓				
[43]	Random Forest	✓	✓	✓				
[44]	CLSTM	✓		✓	✓	✓	✓	✓
[45]	RNSA	✓	✓	✓				✓
[46]	PSO-CNN	✓	✓	✓		✓		✓
[47]	CNN-LSTM	✓	✓	✓		✓		✓
Proposed Model	XGBoost	✓	✓	✓	✓	✓	✓	✓

Daftar Rujukan

[1] A. S. Nugroho, "Electronic Traffic Law Enforcement (E-Tle) Mobile Sebagai Difusi Inovasi, Interoperabilitas Menuju E-Tle Nasional (Studi Implementasi E-Tle Mobile Di Wilayah Jawa Tengah)," *Jurnal Ilmu Kepolisian*, 2022.

[2] A. N. Leonita, I. Islah, And H. Hisbah, "Penegakan Hukum Terhadap Pelanggaran Lalu Lintas Di Kota Jambi Melalui Tilang Elektronik Atau Electronic Traffic Law Enforcement (Etle)," *Jurnal Ilmiah Universitas Batanghari Jambi*, 2022.

[3] A. A. Rahman, R. H. Yoga, And M. D. Atmadja, "Sepeda Motor Untuk Data Investigasi Kecelakaan Lalu Lintas Berbasis Modul Gsm," 2021.

[4] D. M. Nugraha And S. D. Kurniawan, "Prototype Sistem Pengawasan Pada Lampu Lalu Lintas Di Persimpangan Jalan Menggunakan Rasperry Pi Untuk Mencatat Pelanggaran Lampu Merah," 2016.

[5] A. Israk, R. Satra, And F. Fattah, "Perancangan Sistem Pendeteksi Pelanggaran Lampu Lalu Lintas Menggunakan Rasperry Pi 3 Berbasis Internet Of Things," *Buletin Sistem Informasi Dan Teknologi Islam*, 2021.

[6] R. Safitri, M. Fahri, And R. Arlianda, "Perilaku Berkendara Dampak Penerapan Electronic Traffic Law Enforcement (Etle) Pada Simpang Bersinyal Di Kota Pangkalpinang," *Bentang : Jurnal Teoritis Dan Terapan Bidang Rekayasa Sipil*, 2023.

[7] E. Syafitri And D. Mashur, "Efektivitas Implementasi Program Electronic Traffic Law Envoorcement (Etle) Nasional Dalam Peningkatan Pelayanan Publik Di Kota Pekanbaru," *Cross-Border*, Vol. 5, No. 2, Pp. 1322–1337, 2022.

[8] M. Sari And A. Saputra, "Implementasi Pemberlakuan E-Tilang Terhadap Pelanggaran Lalu Lintas Oleh Polrestabes Semarang," *Jurnal Komunikasi Hukum (Jkh)*, Vol. 9, No. 1, Pp. 901–917, 2023.

[9] Y. Armala And M. Yasir, "Implementasi Electronic Traffic Law Enforcement (Etle) Di Wilayah Hukum Kepolisian Resor Bojonegoro," *Justitiable-Jurnal Hukum*, Vol. 5, No. 1, Pp. 32–44, 2022.

[10] Y. Yuliantoro And A. Sulchan, "The Effectiveness Against Traffic Violations With Electronic Traffic Law Enforcement (Etle)," *Law Development Journal*, 2021.

[11] V. V. Kumar, K. M. K. Raghunath, V. Muthukumaran, R. B. Joseph, I. S. Beschi, And A. K. Uday, "Aspect Based Sentiment Analysis And Smart Classification In Uncertain Feedback Pool," *International Journal Of System Assurance Engineering And Management*, Vol. 13, No. 1, Pp. 252–262, 2022, Doi: 10.1007/S13198-021-01379-2.

[12] M. Ewing, L. R. Men, And J. O'neil, "Using Social Media To Engage Employees: Insights From Internal Communication Managers," *International Journal Of Strategic Communication*, Vol. 13, No. 2, Pp. 110–132, Mar. 2019, Doi: 10.1080/1553118x.2019.1575830.

[13] T. R. Soomro And M. Hussain, "Social Media-Related Cybercrimes And Techniques For Their Prevention.," *Appl. Comput. Syst.*, Vol. 24, No. 1, Pp. 9–17, 2019.

[14] C. M. Pulido, B. Villarejo-Carballido, G. Redondo-Sama, And A. Gómez, "Covid-19 Infodemic: More Retweets For Science-Based Information On Coronavirus Than For False Information," *International Sociology*, Vol. 35, No. 4, Pp. 377–392, 2020.

[15] A. Karami, V. Shah, R. Vaezi, And A. Bansal, "Twitter Speaks: A Case Of National Disaster Situational Awareness," *J InfSci*, Vol. 46, No. 3, Pp. 313–324, 2020.

- [16] T. Krisdiyanto, "Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PpkM Pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers," *Jurnal Coreit: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, 2021.
- [17] A. Harun And D. P. Ananda, "Analisa Sentimen Opini Publik Tentang Vaksinasi Covid-19 Di Indonesia Menggunakan Naïve Bayes Dan Decission Tree," *Malcom: Indonesian Journal Of Machine Learning And Computer Science*, 2021.
- [18] A. Kusuma And A. Nugroho, "Analisa Sentimen Pada Twitter Terhadap Kenaikan Tarif Dasar Listrik Dengan Metode Naïve Bayes," *Jurnal Ilmiah Teknologi Informasi Asia*, 2021.
- [19] M. I. Wardah And S. D. Putra, "Implementasi Machine Learning Untuk Rekomendasi Film Di Imdb Menggunakan Collaborative Filtering Berdasarkan Analisa Sentimen Imdb," *Jurnal Manajemen Informatika Jayakarta*, 2022.
- [20] S. F. Pane And J. Ramdan, "Pemodelan Machine Learning : Analisis Sentimen Masyarakat Terhadap Kebijakan PpkM Menggunakan Data Twitter," *Jurnal Sistem Cerdas*, 2022.
- [21] N. Nurfauziah And A. Putra, "Systematic Literature Review: Etnomatematika Pada Rumah Adat," *Jurnal Riset Pembelajaran Matematika*, 2022.
- [22] M. Shin And J. S. Haberl, "Thermal Zoning For Building Hvac Design And Energy Simulation: A Literature Review," *Energy Build*, Vol. 203, P. 109429, 2019, Doi: <https://doi.org/10.1016/j.enbuild.2019.109429>.
- [23] M. E. M. Abo *Et Al.*, "A Multi-Criteria Approach For Arabic Dialect Sentiment Analysis For Online Reviews: Exploiting Optimal Machine Learning Algorithm Selection," *Sustainability*, Vol. 13, No. 18, 2021, Doi: 10.3390/Su131810018.
- [24] E. C. Blalock, Y. Fan, And X. Lyu, "A Systematic Literature Review Of Chinese Entrepreneurship: Utilizing Feminist Theory With Implications For Public Policy," *Entrepreneurship & Regional Development*, Vol. 35, No. 5–6, Pp. 482–510, May 2023, Doi: 10.1080/08985626.2023.2184873.
- [25] P. Zare, C. Pettit, S. Z. Leao, And O. Gudes, "Digital Bicycling Planning: A Systematic Literature Review Of Data-Driven Approaches," *Sustainability*, 2022.
- [26] W. Mengist, T. Soromessa, And G. Legese, "Method For Conducting Systematic Literature Review And Meta-Analysis For Environmental Science Research," *Methodsx*, Vol. 7, P. 100777, 2020.
- [27] I. Nurhas, "A Brief Overview Of The Process Of Conducting A Systematic Literature Review (Slr).," 2021.
- [28] M. L. Rethlefsen *Et Al.*, "Prisma-S: An Extension To The Prisma Statement For Reporting Literature Searches In Systematic Reviews," *Syst Rev*, Vol. 10, No. 1, P. 39, 2021, Doi: 10.1186/S13643-020-01542-Z.
- [29] A. Schnieder mann, "Shaping The Qualities, Values And Standards Of Science. How Reporting Guidelines Improve The Transparency Of Biomedical Research," *Front Res Metr Anal*, P. 34, 2022.
- [30] M. Ashiq, S. U. Rehman, M. Safdar, And H. Ali, "Academic Library Leadership In The Dawn Of The New Millennium: A Systematic Literature Review," *The Journal Of Academic Librarianship*, Vol. 47, No. 3, P. 102355, 2021, Doi: <https://doi.org/10.1016/j.acalib.2021.102355>.
- [31] S. Nabilah, H. Pujiastuti, And S. Syamsuri, "Systematic Literature Review : Literasi Numerasi Dalam Pembelajaran Matematika, Jenjang, Materi, Model Dan Media Pembelajaran," *Jiip - Jurnal Ilmiah Ilmu Pendidikan*, 2023.
- [32] J. P. Romanelli, P. Meli, R. P. Naves, M. C. Alves, And R. R. Rodrigues, "Reliability Of Evidence-Review Methods In Restoration Ecology," *Conservation Biology*, Vol. 35, No. 1, Pp. 142–154, Feb. 2021, Doi: <https://doi.org/10.1111/Cobi.13661>.
- [33] M. E. Basiri, S. Nemati, M. Abdar, E. Cambria, And U. R. Acharya, "Abcdm: An Attention-Based Bidirectional Cnn-Rnn Deep Model For Sentiment Analysis," *Future Generation Computer Systems*, Vol. 115, Pp. 279–294, 2021, Doi: <https://doi.org/10.1016/j.future.2020.08.005>.
- [34] S. Yoo, J. Song, And O. Jeong, "Social Media Contents Based Sentiment Analysis And Prediction System," *Expert Syst Appl*, Vol. 105, Pp. 102–111, 2018, Doi: <https://doi.org/10.1016/j.eswa.2018.03.055>.
- [35] M. Alam, F. Abid, C. Guangpei, And L. V Yunrong, "Social Media Sentiment Analysis Through Parallel Dilated Convolutional Neural Network For Smart City Applications," *Comput Commun*, Vol. 154, Pp. 129–137, 2020, Doi: <https://doi.org/10.1016/j.comcom.2020.02.044>.
- [36] R. L. Rosa, G. M. Schwartz, W. V. Ruggiero, And D. Z. Rodríguez, "A Knowledge-Based Recommendation System That Includes Sentiment Analysis And Deep Learning," *Ieee Trans Industr Inform*, Vol. 15, No. 4, Pp. 2124–2135, 2019, Doi: 10.1109/Tii.2018.2867174.
- [37] F. Abid, M. Alam, M. Yasir, And C. Li, "Sentiment Analysis Through Recurrent Variants Latterly On Convolutional Neural Network Of Twitter," *Future Generation Computer Systems*, Vol. 95, Pp. 292–308, 2019, Doi: <https://doi.org/10.1016/j.future.2018.12.018>.
- [38] S. Ahmad, M. Z. Asghar, F. M. Alotaibi, And I. Awan, "Detection And Classification Of Social Media-Based Extremist Affiliations Using Sentiment Analysis Techniques," *Human-Centric Computing And Information Sciences*, Vol. 9, No. 1, P. 24, 2019, Doi: 10.1186/S13673-019-0185-6.
- [39] Y. Ma, H. Peng, T. Khan, E. Cambria, And A. Hussain, "Sentic Lstm: A Hybrid Network For Targeted Aspect-Based Sentiment Analysis," *Cognit Comput*, Vol. 10, No. 4, Pp. 639–650, 2018, Doi: 10.1007/S12559-018-9549-X.
- [40] O. Ajao, D. Bhowmik, And S. Zargari, "Sentiment Aware Fake News Detection On Online Social Networks," In *Icassp 2019 - 2019 Ieee International Conference On Acoustics, Speech And Signal Processing (Icassp)*, 2019, Pp. 2507–2511. Doi: 10.1109/Icassp.2019.8683170.
- [41] Y. Cai, Q. Huang, Z. Lin, J. Xu, Z. Chen, And Q. Li, "Recurrent Neural Network With Pooling Operation And Attention Mechanism For Sentiment Analysis: A Multi-Task Learning Approach," *Knowl Based Syst*, Vol. 203, P. 105856, 2020, Doi: <https://doi.org/10.1016/j.knsys.2020.105856>.
- [42] G. A. Ruz, P. A. Henríquez, And A. Mascareño, "Sentiment Analysis Of Twitter Data During Critical Events Through Bayesian Networks Classifiers," *Future Generation Computer Systems*, Vol. 106, Pp. 92–104, 2020, Doi: <https://doi.org/10.1016/j.future.2020.01.005>.
- [43] A. S. Neogi, K. A. Garg, R. K. Mishra, And Y. K. Dwivedi, "Sentiment Analysis And Classification Of Indian Farmers' Protest Using Twitter Data," *International Journal Of Information Management Data Insights*, Vol. 1, No. 2, P. 100019, 2021, Doi: <https://doi.org/10.1016/j.jime.2021.100019>.
- [44] R. Haque, N. Islam, M. Tasneem, And A. K. Das, "Multi-Class Sentiment Classification On Bengali Social Media Comments Using Machine Learning," *International Journal Of Cognitive Computing In Engineering*, Vol. 4, Pp. 21–35, 2023, Doi: <https://doi.org/10.1016/J.Ijcc.2023.01.001>.
- [45] A. Abdi, S. M. Shamsuddin, S. Hasan, And J. Piran, "Deep Learning-Based Sentiment Classification Of Evaluative Text Based On Multi-Feature Fusion," *Inf*

- Process Manag*, Vol. 56, No. 4, Pp. 1245–1259, 2019, Doi: <https://doi.org/10.1016/j.ipm.2019.02.018>.
- [46] R. Kumar, H. S. Pannu, And A. K. Malhi, “Aspect-Based Sentiment Analysis Using Deep Networks And Stochastic Optimization,” *Neural Comput Appl*, Vol. 32, No. 8, Pp. 3221–3235, 2020, Doi: 10.1007/S00521-019-04105-Z.
- [47] C. Colón-Ruiz And I. Segura-Bedmar, “Comparing Deep Learning Architectures For Sentiment Analysis On Drug Reviews,” *J Biomed Inform*, Vol. 110, P. 103539, 2020, Doi: <https://doi.org/10.1016/j.jbi.2020.103539>.
- [48] M. A. Rahman, H. Budianto, And E. I. Setiawan, “Aspect Based Sentimen Analysis Opini Publik Pada Instagram Dengan Convolutional Neural Network,” *Journal Of Intelligent System And Computation*, 2019.
- [49] H. Bunyamin And Meyliana, “Classical And Deep Learning Time Series Prediction Techniques In The Case Of Indonesian Economic Growth,” *Iop Conf Ser Mater Sci Eng*, Vol. 1077, 2021.
- [50] B. Hakim, “Analisa Sentimen Data Text Preprocessing Pada Data Mining Dengan Menggunakan Machine Learning,” *Jbase - Journal Of Business And Audit Information Systems*, 2021.
- [51] Y. Zhang And Q. Yang, “A Survey On Multi-Task Learning,” *Ieee Trans Knowl Data Eng*, Vol. 34, Pp. 5586–5609, 2017.
- [52] Y. Zhang And Q. Yang, “An Overview Of Multi-Task Learning,” *Natl Sci Rev*, Vol. 5, Pp. 30–43, 2018.
- [53] A. Gesmundo And J. Dean, “An Evolutionary Approach To Dynamic Introduction Of Tasks In Large-Scale Multitask Learning Systems,” *Arxiv*, Vol. Abs/2205.12755, 2022.
- [54] C. Fan, M. Chen, X. Wang, J. Wang, And B. Huang, “A Review On Data Preprocessing Techniques Toward Efficient And Reliable Knowledge Discovery From Building Operational Data,” *Front Energy Res*, Vol. 9, 2021, Doi: 10.3389/Fenrg.2021.652801.
- [55] K. Yu, L. Tan, L. Lin, X. Cheng, Z. Yi, And T. Sato, “Deep-Learning-Empowered Breast Cancer Auxiliary Diagnosis For 5g Remote E-Health,” *Ieee Wirel Commun*, Vol. 28, No. 3, Pp. 54–61, 2021.
- [56] S. Biswas And H. Rajan, “Fair Preprocessing: Towards Understanding Compositional Fairness Of Data Transformers In Machine Learning Pipeline,” *Proceedings Of The 29th Acm Joint Meeting On European Software Engineering Conference And Symposium On The Foundations Of Software Engineering*, 2021.
- [57] N. Lu And T. Yin, “Transferable Common Feature Space Mining For Fault Diagnosis With Imbalanced Data,” *Mech Syst Signal Process*, Vol. 156, P. 107645, 2021.
- [58] N. Malave And A. V. Nimkar, “A Survey On Effects Of Class Imbalance In Data Pre-Processing Stage Of Classification Problem,” *International Journal Of Computational Systems Engineering*, Vol. 6, No. 2, Pp. 63–75, 2020.
- [59] D. Kuhn, P. M. Esfahani, V. A. Nguyen, And S. Shafieezadeh-Abadeh, “Wasserstein Distributionally Robust Optimization: Theory And Applications In Machine Learning,” *Arxiv*, Vol. Abs/1908.08729, 2019.
- [60] N.-T. Ngo *Et Al.*, “Proposing A Hybrid Metaheuristic Optimization Algorithm And Machine Learning Model For Energy Use Forecast In Non-Residential Buildings,” *Sci Rep*, Vol. 12, 2022.
- [61] B. Albadani, R. Shi, And J. Dong, “A Novel Machine Learning Approach For Sentiment Analysis On Twitter Incorporating The Universal Language Model Fine-Tuning And Svm,” *Applied System Innovation*, 2022.