

Systematic Literature Review: Klasifikasi Emosi pada Data Teksual

Putu Widyantara Artanta Wibawa^[1], Cokorda Rai Adi Pramatha^[2]

Program Studi Informatika^{[1], [2]}
Fakultas MIPA, Universitas Udayana
Jimbaran, Bali, Indonesia

putuwaw973@gmail.com^[1], cokorda@unud.ac.id^[2]

Abstract— Emotions are a person's response to an event. Emotions can be expressed verbally or nonverbally. Over time people can express their emotions through social media. Considering that emotion is a reflection of society's response, it is important to classify emotions in society to find out the community's response as information for consideration in decision-making. This study is aimed to identify and analyze the datasets, methods, and evaluation metrics that are being used in the classification of emotional texts in textual data from research data from 2013 to 2022. Based on the inclusion and exclusion design in selecting literature, a total of 50 kinds of literature were used in extracting and synthesizing data. Analysis of the data shows that out of 50 pieces of literature, there are 36 works of literature that use public datasets while 14 kinds of literature use private datasets. In the method of developing models for classifying, the SVM and Naive Bayes models are the most popular among the other models. In evaluating the model, the F-measure or F1-score metric is the most widely used metric compared to other metrics.

Keywords— *systematic literature review, emotion classification, text data*

Abstrak—Emosi adalah respon seseorang terhadap suatu peristiwa. Emosi dapat diekspresikan secara verbal maupun non-verbal. Seiring dengan perkembangan zaman, orang dapat mengekspresikan emosinya melalui media sosial. Mengingat emosi adalah cerminan dari respon seseorang, maka penting untuk melakukan klasifikasi terhadap emosi dalam masyarakat untuk mengetahui respon dari masyarakat sehingga dapat digunakan sebagai bahan pertimbangan dalam pengambilan keputusan. Penelitian ini bertujuan untuk mengidentifikasi serta menganalisis dataset, metode, dan metrik evaluasi yang digunakan dalam penelitian mengenai klasifikasi emosi pada data tekstual dari tahun 2013 sampai tahun 2022. Berdasarkan desain inklusi dan eksklusi dalam pemilihan literatur, didapatkan sebanyak 50 literatur yang akan digunakan dalam ekstraksi dan sintesis data. Hasil analisis menunjukkan bahwa dari 50 literatur, 36 literatur menggunakan dataset publik dan 14 literatur menggunakan dataset privat. Pada metode dalam pengembangan model untuk melakukan klasifikasi, model SVM dan Naive Bayes merupakan model yang paling populer diantara model lainnya. Dalam melakukan evaluasi model, metrik F-measure atau F1-score adalah metrik yang paling banyak digunakan dibandingkan dengan metrik lainnya.

Kata Kunci—*systematic literature review, klasifikasi emosi,*

data tekstual

I. PENDAHULUAN

Emosi didefinisikan sebagai respon seseorang terhadap suatu peristiwa atau situasi secara sadar yang terjadi pada durasi tertentu [1]. Emosi dapat diekspresikan menggunakan komunikasi verbal maupun secara non-verbal [2]. Secara verbal, emosi dapat diekspresikan secara lisan maupun menggunakan tulisan. Sementara secara non-verbal, emosi dapat diekspresikan melalui bahasa tubuh seperti gerakan tangan dan raut wajah. Berdasarkan representasinya, emosi terbagi ke dalam dunia jenis yaitu dimensional dan kategorikal. Emosi dimensional merepresentasikan emosi ke dalam sebuah ruang spasial, sementara emosi kategorikal menempatkan emosi ke dalam beberapa kelas atau kategori yang berbeda [3].

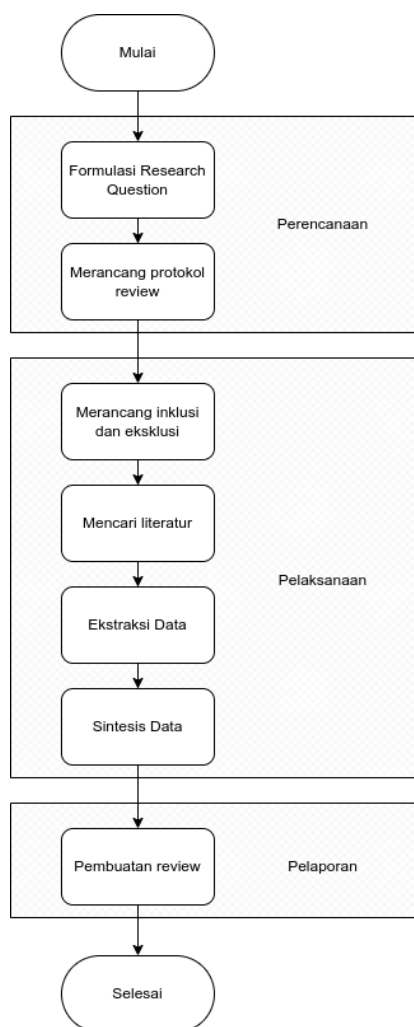
Dengan kemajuan teknologi saat ini, seseorang dapat dengan mudah mengekspresikan emosi yang dimilikinya secara verbal melalui tulisan pada media sosial. Untuk dapat mengetahui emosi tersebut perlu dilakukan klasifikasi emosi. Klasifikasi emosi adalah proses untuk memetakan atau mengkategorikan dokumen ke dalam sekumpulan emosi yang telah ditentukan sebelumnya [4]. Mengingat emosi adalah respon seseorang terhadap suatu peristiwa, klasifikasi emosi menjadi sangat penting untuk dilakukan di berbagai sektor kehidupan. Pada sektor bisnis, respon dari pengguna sangat penting sebagai bahan pertimbangan dalam pengambilan keputusan untuk meningkatkan kualitas produk yang dihasilkan. Selain itu, dalam sektor pendidikan klasifikasi emosi juga dapat membantu guru untuk meningkatkan metode pembelajaran berdasarkan respon dari siswa [5].

Saat ini, klasifikasi emosi pada data tekstual adalah salah satu cabang dari NLP (Natural Language Processing) yang berkembang paling cepat [6]. Perkembangan yang cepat ini juga menyebabkan tersebarnya dataset, metode, serta skema evaluasi yang digunakan klasifikasi emosi, sehingga mempersulit untuk mendapatkan gambaran menyeluruh dari klasifikasi emosi. Oleh karena itu, diperlukan sebuah *literature review* yang bertujuan untuk mengidentifikasi dan menganalisis dataset, metode, serta skema evaluasi dari klasifikasi emosi khususnya dalam rentang waktu sepuluh tahun terakhir dari tahun 2013 sampai tahun 2022.

II. METODE PENELITIAN

A. Desain Penelitian

Pada penelitian ini, metode *literature review* yang digunakan adalah *systematic literature review* atau tinjauan pustaka sistematis. *Systematic literature review* atau *systematic review* merupakan cara untuk mengidentifikasi, mengevaluasi, menginterpretasikan semua penelitian yang tersedia terkait dengan pertanyaan penelitian tertentu [7]. Terdapat tiga tahapan utama dalam melakukan *systematic literature review*, ketiga tahap tersebut adalah perencanaan, pelaksanaan, dan pelaporan. Seluruh tahapan utama dalam *systematic literature review* ditunjukkan pada Gambar 1.



Gambar 1. Tahapan *Systematic Literature Review*

B. Pertanyaan Penelitian

Penentuan pertanyaan penelitian (research question) adalah bagian terpenting dalam tahap perencanaan karena akan memandu keseluruhan proses dari *systematic literature review* [8]. Pertanyaan penelitian dibangun menggunakan kriteria PICOC (Population, Intervention, Comparison, Outcome, Context) [7]. Tabel I menunjukkan formulasi pertanyaan penelitian yang dibangun dari PICOC.

TABEL I FORMULA RESEARCH QUESTION BERDASARKAN PICOC

PICOC	Formulasi Pertanyaan
Population	NLP, klasifikasi teks, machine learning, deep learning
Intervention	klasifikasi emosi, klasifikasi, model, metode, dataset, metrik evaluasi
Comparison	-
Outcome	prediksi akurasi dari klasifikasi emosi, metode klasifikasi emosi
Context	dataset berukuran besar dan kecil, penelitian pada bidang industri dan akademik

Berdasarkan formulasi dari Tabel I, adapun pertanyaan penelitian yang digunakan dalam *systematic literature review* ini ditunjukkan pada Tabel II.

TABEL II RESEARCH QUESTION

	Research Question
RQ1	Apakah jenis dataset yang paling banyak digunakan dalam klasifikasi emosi pada data tekstual?
RQ2	Apakah metode yang paling banyak digunakan dalam klasifikasi emosi pada data tekstual?
RQ3	Apakah metode terbaik ketika digunakan dalam klasifikasi emosi pada data tekstual?
RQ4	Apakah metrik evaluasi yang paling banyak digunakan dalam klasifikasi emosi pada data tekstual?

C. Strategi Pencarian

Pada tahapan ini akan dilakukan pemilihan perpustakaan digital, merancang *string* pencarian, melakukan pencarian, dan mengambil literatur yang sesuai dengan *string* pencarian. Sebelum melakukan pencarian, akan dilakukan pemilihan perpustakaan digital. Perpustakaan atau *database* digital harus dipilih untuk meningkatkan kemungkinan dalam menemukan literatur yang relevan. Pencarian akan dilakukan pada perpustakaan digital yang paling populer akan dicari untuk menemukan literatur dengan cakupan yang luas [9]. Adapun daftar perpustakaan digital yang akan digunakan dalam *systematic literature review* ini ditunjukkan pada Tabel III.

TABEL III. TABEL PERPUSTAKAAN DIGITAL

Nama	Situs Web
Google Scholar	https://scholar.google.com/
IEEE Xplore	https://ieeexplore.ieee.org/
ACM Digital Library	https://dl.acm.org/
ScienceDirect	https://www.sciencedirect.com/
Springer	https://www.springer.com/

Setelah penentuan perpustakaan digital, selanjutnya adalah merancang *string* pencarian, adapun *string* pencarian yang digunakan adalah:

(emotion) AND (detection OR classification OR recognition OR analysis) AND (textual or text)

Proses pencarian dilakukan dengan mengambil sampel sebanyak 50 artikel teratas pada setiap perpustakaan digital yang ada, sehingga akan didapatkan 250 artikel.

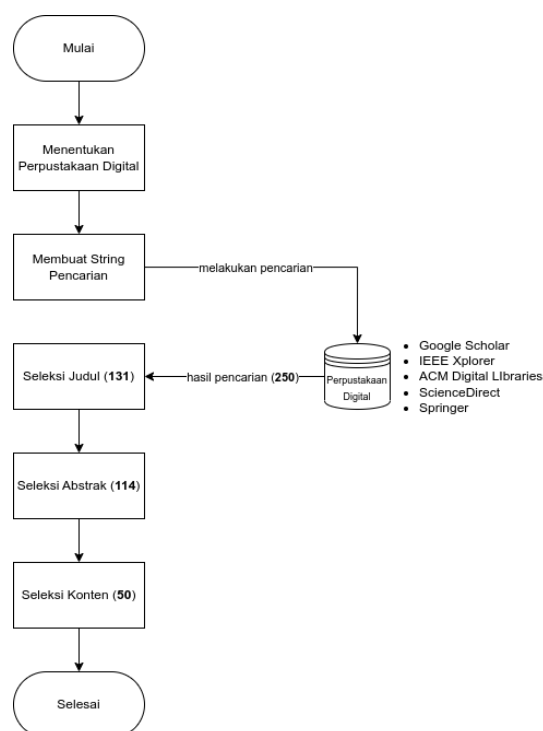
D. Kriteria Pemilihan

Setelah mendapatkan daftar referensi, selanjutnya adalah melakukan seleksi terhadap setiap literatur untuk menentukan apakah literatur tersebut harus disertakan untuk ekstraksi dan analisis data [10]. Untuk melakukan seleksi, akan digunakan kriteria inklusi dan eksklusi yang ditunjukkan pada Tabel IV.

TABEL IV TABEL KRITERIA INKLUSI DAN EKSKLUSI

Inklusi	Eksklusi
Literatur mengembangkan model atau metode komputasi untuk melakukan klasifikasi emosi	Literatur tidak mengembangkan model atau hanya melakukan survei atau studi literatur
Menggunakan dataset data tekstual semi terstruktur atau tidak terstruktur	Menggunakan dataset data tekstual terstruktur atau diluar data tekstual
Literatur memiliki judul serta abstrak yang sesuai dengan string pencarian	Literatur memiliki judul serta abstrak yang tidak sesuai dengan string pencarian
Literatur ditulis dalam bahasa Inggris	Literatur tidak ditulis dalam bahasa Inggris
Literatur menggunakan metrik untuk evaluasi model	Literatur tidak menggunakan metrik untuk evaluasi model

Tahapan seleksi akan dilakukan dalam tiga tahap, yaitu seleksi terhadap judul, seleksi terhadap abstrak dan kata kunci, serta seleksi terhadap konten atau isi dari artikel.



Gambar 2. Tahapan Pencarian dan Seleksi Literatur

Setelah dilakukan seleksi pada judul artikel, artikel yang awalnya berjumlah 250 artikel berkurang menjadi 131 artikel. Kemudian dilakukan seleksi terhadap abstrak dan kata kunci pada artikel, sehingga didapatkan 114 artikel.

Tahap terakhir dalam seleksi adalah seleksi pada konten yang menghasilkan 50 artikel. Dengan demikian, terdapat 50 literatur yang akan digunakan untuk tahap ekstraksi serta sintesis data. Seluruh tahapan dalam pencarian serta seleksi artikel ditunjukkan pada Gambar 2.

E. Ekstraksi Data

Literatur yang sudah melewati proses inklusi dan eksklusi akan diekstraksi untuk mendapatkan data yang nantinya digunakan untuk menjawab pertanyaan penelitian. Tujuan dari tahap ini adalah merancang formulir ekstraksi data untuk secara akurat merekam informasi yang diperoleh peneliti dari studi utama [7]. Ekstraksi data dilakukan terhadap beberapa kategori yang diidentifikasi berdasarkan pertanyaan penelitian. Terdapat tiga kategori yang digunakan untuk menjawab pertanyaan penelitian yang ditunjukkan pada Tabel V.

TABEL V TABEL KATEGORI EKSTRAKSI DATA

Kategori	Pertanyaan Penelitian
Dataset Klasifikasi Emosi	RQ1
Metode Klasifikasi Emosi	RQ2, RQ3
Metrik Evaluasi Klasifikasi Emosi	RQ4

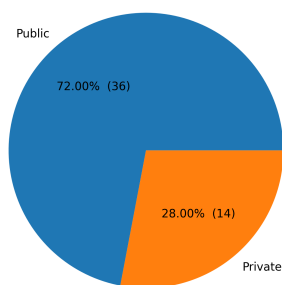
F. Sintesis Data

Sintesis data melibatkan proses penyusunan dan peringkasan hasil literatur utama yang telah diekstraksi [7]. Tujuan dari sintesis data adalah untuk mengumpulkan bukti-bukti dari literatur utama untuk menjawab pertanyaan penelitian [9]. Tipe sintesis data yang digunakan dalam *systematic literature review* ini adalah sintesis naratif (deskriptif). Data akan ditabulasikan dengan cara yang konsisten dengan pertanyaan penelitian. Beberapa jenis visualisasi seperti diagram batang dan diagram lingkaran juga digunakan dalam penyajian data.

III. HASIL DAN PEMBAHASAN

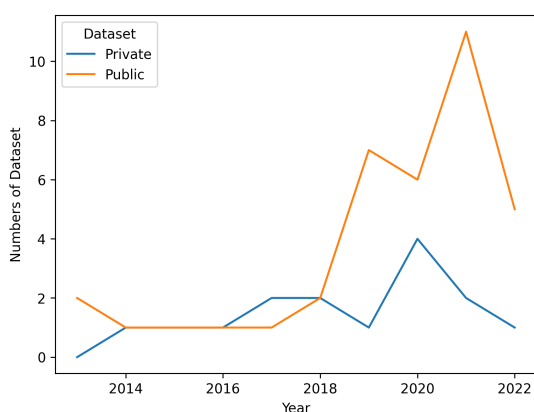
A. Dataset Terbanyak Digunakan dalam Klasifikasi Emosi pada Data Tekstual

Dataset yang digunakan dalam *systematic literature review* ini adalah dataset tekstual baik terstruktur maupun tidak terstruktur. Dari 50 literatur, terdapat 36 literatur yang menggunakan dataset yang bersifat publik sementara 14 literatur menggunakan dataset yang bersifat privat. Dataset yang bersifat publik sebagian besar menggunakan dataset ISEAR (International Survey on Emotion Antecedents and Reactions) maupun SemEval (Semantic Evaluation). Sementara untuk dataset privat umumnya didapatkan dari teks pada media sosial. Gambar 3 menunjukkan perbandingan jumlah dataset publik dan privat yang digunakan dari 50 literatur yang digunakan pada *systematic literature review* ini.



Gambar 3. Grafik Distribusi Dataset

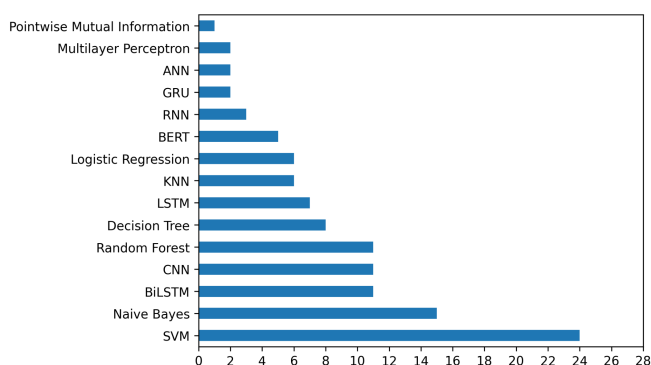
Pada Gambar 4 menunjukkan bagaimana tren penggunaan dataset publik dan privat dari tahun ke tahun, mulai dari tahun 2013 sampai tahun 2022. Dari Gambar 4 terlihat bahwa penggunaan dataset publik cenderung mengalami peningkatan dibandingkan dengan dataset privat. Hal ini dimungkinkan mengingat tiap tahunnya semakin banyak studi yang dipublikasikan, sehingga ada semakin banyak data yang bisa untuk digunakan.



Gambar 4. Grafik Tren Penggunaan Dataset

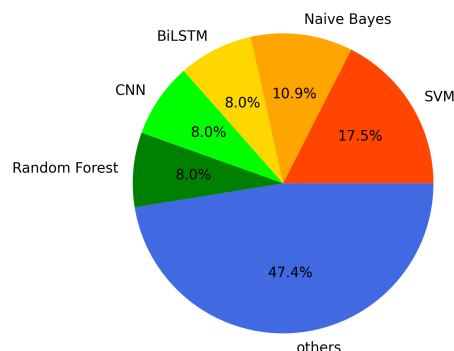
B. Metode Terbanyak Digunakan dalam Klasifikasi Emosi pada Data Teksual

Dari 50 literatur yang digunakan, terdapat sepuluh metode yang digunakan untuk membangun model dalam melakukan klasifikasi terhadap emosi. Lima belas metode dengan penggunaan terbanyak ditunjukkan pada Gambar 5.



Gambar 5. Grafik Distribusi Metode Pada Klasifikasi Emosi

Diantaranya sepuluh metode tersebut, metode dengan jumlah penggunaan terbanyak adalah SVM (Support Vector Machine). Pada Gambar 6 ditunjukkan lima metode teratas, yaitu SVM, Naive Bayes, BiLSTM, CNN, dan Random Forest. Kelima metode tersebut digunakan pada lebih dari setengah dari literatur yang dianalisis pada *systematic literature review* ini.



Gambar 6. Grafik Penggunaan Lima Metode Teratas pada Literatur

C. Metode Terbaik dalam Klasifikasi Emosi pada Data Teksual

Dalam *systematic literature review* ini, terdapat cukup banyak metode yang digunakan untuk membangun model dalam melakukan klasifikasi emosi. Walaupun terdapat beberapa literatur yang membandingkan performa antar metode seperti [11], namun tidak dapat dinyatakan secara jelas bahwa metode tersebut lebih baik diantaranya metode yang lainnya ketika dilihat secara individual. Akan tetapi, berdasarkan grafik pada Gambar 4 dan Gambar 5, dapat dilihat bahwa beberapa metode populer yang digunakan adalah SVM, Naive Bayes, Random Forest, serta Neural Network seperti BiLSTM dan CNN.

Naive Bayes dan SVM dalam *systematic literature review* ini merupakan dua metode *machine learning* yang paling populer dalam melakukan klasifikasi emosi pada data teksual. Naive Bayes sendiri didasari pada teorema Bayes. Kelebihan dari Naive Bayes adalah metode ini bekerja sangat baik pada data teks dan mudah untuk diimplementasikan serta cenderung cepat jika dibandingkan dengan algoritma yang lain [12]. Akan tetapi Naive Bayes juga memiliki beberapa kekurangan, salah satunya adalah Naive Bayes memiliki asumsi yang kuat terhadap distribusi data [13]. Sehingga, diperlukan dataset dengan distribusi yang imbang agar dapat membangun model yang baik.

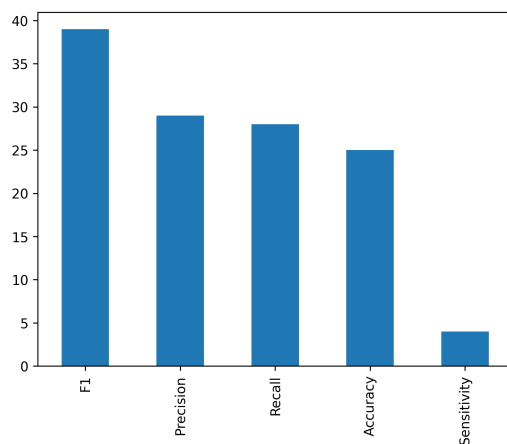
SVM atau Support Vector Machine adalah klasifikasi dan regresi dengan menggunakan konsep kernel dalam dimensi tinggi untuk memecahkan masalah non-linear [14]. SVM beberapa kali menunjukkan performa yang lebih baik dibandingkan model lainnya seperti pada [11] dan [15]. SVM memiliki kelebihan untuk membangun *hyperlane* yang bersifat non-linear sehingga dapat digunakan untuk memaksimalkan jarak margin antar kelas. Kekurangan dari SVM salah satunya adalah kurangnya transparansi dalam hasil klasifikasi yang disebabkan karena tingginya dimensi pada data teks [12].

Neural Network atau Deep Neural Network adalah model yang meniru bagaimana otak manusia bekerja [16]. Arsitektur dari Neural Network terbentuk dari lapisan yang saling terhubung yang terdiri dari tiga tingkatan yang berbeda, yaitu *input layer*, *hidden layer*, dan *output layer* [17]. Secara umum terdapat dua jenis arsitektur dari Neural Network, yaitu *feed-forward network* dan *recurrent network*. Contoh dari *feed-forward network* adalah CNN sementara contoh dari *recurrent network* adalah RNN, LSTM, dan GRU. BERT atau juga menggunakan basis Neural Network untuk membangun *pre-trained natural language* model [18]. Neural Network memiliki kelebihan dalam mengolah data dengan fitur yang kompleks, namun memiliki kelemahan dalam komputasi yang berat dan membutuhkan data yang banyak [12].

Menentukan model yang memiliki kinerja terbaik harus disesuaikan dengan teknik dan data yang tepat. Setiap model akan memberikan hasil yang berbeda dengan perlakuan yang berbeda dan jenis dataset yang berbeda. Oleh karena itu diperlukan pemahaman terhadap setiap algoritma untuk dapat menentukan model yang paling cocok untuk digunakan dalam permasalahan klasifikasi emosi.

D. Metrik Evaluasi Terbanyak Digunakan dalam Klasifikasi Emosi pada Data Teksual

Performa dari sebuah model dapat diketahui dengan melakukan evaluasi pada model. Kinerja dari sebuah model direpresentasikan oleh metrik tertentu. Metrik-metrik evaluasi ini memberikan informasi yang berbeda-beda sesuai dengan informasi yang ingin ditunjukkan oleh metrik tersebut [12]. Dalam *systematic literature review* ini, metrik yang paling umum digunakan adalah *accuracy*, *precision*, *recall*, dan *F-measure* yang didasarkan pada *confusion matrix*. *Confusion matrix* sendiri adalah sebuah matriks yang menggambarkan hasil prediksi atau klasifikasi dari model (predicted classification) dan klasifikasi yang sebenarnya (actual classification). Gambar 7 menunjukkan sebaran penggunaan metrik untuk evaluasi.



Gambar 7. Grafik Distribusi Metrik Evaluasi

Penggunaan metrik evaluasi yang paling banyak adalah metrik *F-measure* atau *F1-score*. Metrik *F-measure* digunakan untuk mengetahui *harmonic-mean* dengan menghitung agregat antara metrik *precision* dan *recall* [19].

IV. KESIMPULAN

Penelitian ini bertujuan untuk mengidentifikasi dan menganalisis dataset, metode, dan juga metrik evaluasi yang digunakan dalam klasifikasi teks emosi pada data tekstual berdasarkan penelitian yang dilakukan mulai dari tahun 2013 sampai tahun 2022. Metode yang digunakan pada penelitian ini adalah *systematic literature review*. Berdasarkan desain inklusi dan eksklusi dalam pemilihan literatur, didapatkan sebanyak 50 literatur yang digunakan dalam ekstraksi dan sintesis data. Analisis dari data tersebut menunjukkan bahwa dari 50 literatur, terdapat 36 literatur yang menggunakan dataset publik sementara 14 literatur menggunakan dataset privat. Pada metode dalam pengembangan model untuk melakukan klasifikasi, terdapat 15 metode yang paling banyak digunakan, diantaranya adalah SVM, Naive, Bayes, BiLSTM, CNN, Random Forest, Decision Tree, LSTM, KNN, Logistic Regression, BERT, RNN, GRU, ANN, Multilayer Perceptron, dan Pointwise Mutual Information. Model SVM dan Naive Bayes merupakan yang paling populer diantara model lainnya. Dalam menentukan model, diperlukan pemahaman dan analisis terhadap permasalahan agar mendapatkan model yang optimal untuk klasifikasi emosi. Dalam melakukan evaluasi terhadap model, metrik *F-measure* atau *F1-score* adalah metrik yang paling banyak digunakan dibandingkan dengan metrik lainnya.

PERNYATAAN PENGHARGAAN

DAFTAR PUSTAKA

- [1] A. Dzedzickis, A. Kaklauskas, and V. Bucinskas, "Human Emotion Recognition: Review of Sensors and Methods," *Sensors*, vol. 20, no. 3, 2020
- [2] L. Schoneveld, A. Othmani, and H. Abdelkawy, "Leveraging Recent Advances in Deep Learning for Audio-Visual Emotion Recognition," *Pattern Recognition Letters*, vol. 46, pp. 1-7, 2021
- [3] F. A. Acheampong, C. Wenyu, and H. Nunoo-Mensah, "Text-based emotion detection: Advances, challenges, and opportunities," *Engineering Reports*, vol. 2, no. 7, 2020
- [4] Oberländer, L.A.M. and Klinger, R., "An analysis of annotated corpora for emotion classification in text," in *Proceedings of the 27th International Conference on Computational Linguistics, USA*, 2018
- [5] P. Nandwani and R. Verma, "A review on sentiment analysis and emotion detection from text," *Social Network Analysis and Mining*, vol. 11, no. 1, 2021
- [6] S. Zad, M. Heidari, J. H. J. Jones, and O. Uzuner, "Emotion Detection of Textual Data: An Interdisciplinary Survey," *2021 IEEE World AI IoT Congress (AIIoT)*, USA, 2021
- [7] Keele, S., "Guidelines for performing systematic literature reviews in software engineering", In: *Technical report*, Ver. 2.3 EBSE Technical Report, UK, 2007
- [8] V. Torres-Carrion, C. S. Gonzalez-Gonzalez, S. Aciar, and G. Rodriguez-Morales, "Methodology for systematic literature review applied to engineering and education," *2018 IEEE Global Engineering Education Conference (EDUCON)*, Spain, 2018
- [9] Wahono RS, "A systematic literature review of software defect prediction", *Journal of software engineering*, vol. 1, no. 1, pp. 1-6, 2015
- [10] Xiao, Y. and Watson, M., "Guidance on conducting a systematic literature review". *Journal of planning education and research*, vol. 39, no. 1, pp. 93-112, 2019
- [11] T. Parvin, O. Sharif, and M. M. Hoque, "Multi-class Textual Emotion Categorization using Ensemble of Convolutional and Recurrent Neural Network," *SN Computer Science*, vol. 3, no. 1, 2021
- [12] K. Kowsari, K. Jafari Meimandi, M. Heidarysafa, S. Mendu, L. Barnes,

- and D. Brown, "Text Classification Algorithms: A Survey," *Information*, vol. 10, no. 4, p. 150, 2019
- [13] Y. Wang, R. Khardon, and P. Protopapas, "Nonparametric Bayesian estimation of periodic light curves," *The Astrophysical Journal*, vol. 756, no. 1, pp. 67–67, 2012
- [14] Y. Azhar, A. K. Firdausy, and P. J. Amelia, "Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit Stroke," *SINTECH (Science and Information Technology) Journal*, vol. 5, no. 2, pp. 191–197, 2022
- [15] Ardiada, I.M.D., Sudarma, M. and Giriantari, D., 2019. "Text Mining pada Sosial Media untuk Mendeteksi Emosi Pengguna Menggunakan Metode Support Vector Machine dan K-Nearest Neighbour". *Maj. Ilm. Teknol. Elektro*, vol. 18, no. 1, 2019
- [16] Fudholi, D.H., "Klasifikasi Emosi pada Teks dengan Menggunakan Metode Deep Learning". *Syntax Literate; Jurnal Ilmiah Indonesia*, vol. 6, no. 1, pp.546-553, 2021
- [17] Osinga, D., *Deep learning cookbook: practical recipes to get started quickly*, Sebastopol: O'Reilly Media, Inc., 2018
- [18] Devlin, J., Chang, M.W., Lee, K. and Toutanova, K., "Bert: Pre-training of deep bidirectional transformers for language understanding", *arXiv (Cornell University)*, 2018
- [19] M. Grandini, Enrico Bagli, and Giorgio Visani, "Metrics for Multi-Class Classification: an Overview," *arXiv (Cornell University)*, 2020