Systematic Literature Review: Klasifikasi Ujaran Kebencian Sosisal Media Dengan Algoritma Naïve Bayes

Aang Alim Murtopo*[1], Rito Cipta Sigitta H[2], Penulis ketiga[3], Yustira[4]
[1,4] Proram Studi Teknik Informatika, STMIK YMI Tegal, [2]Teknik Informatika Universitas Peradaban,, [3] Program Studi Sistem Informasi, STMIK YMI Tegal
E-mail: *[1]**aang.alim@gmail.com**, [2]ritocipta@peradaban.ac.id, [3]sarif_surorejo@yahoo.com, [4]yustira1997@gmail.com

Abstrak

Pada zaman modern ini pertumbuhan internet semakin pesat terutama penggunaan media sosial sehingga opini publik semakin luas dan bebas ditunjukan di berbagai media sosial. Kebebasan berpendapat yang dimiliki setiap orang malah akan disalahgunakan untuk ujaran kebencian. Untuk membuat analisis sentiment klasifikasi kata ujaran kebencian ini menggunakan Algoritma Naïve Bayes. Dengan menggunakan Tinjauan Pustaka sistematis (SLR) adalah strategi untuk melakukan tinjauan pustaka yang menemukan, meneliti, dan menilai semua temuan pada topik penelitian tertentu untuk memberikan jawaaban atas pertanyaan penelitian yang diajukan sebelumnya. Tinjauan pustaka ini bertujuan untuk menganalisis dan mengidentifikasi penelitian klasifikasi ujaran kebencian menggunakan metode Naïve Bayes antara tahun 2017-2022. Maka mendapatkan hasil analisis dan indentifikasi yaitu akurasi paling signifikan didapat pada kombinasi metode pada algoritma Naïve Bayes menggunakan seleksi fitur information gain yaitu sebesar 98%. Untuk tahun publikasi dan media sosial yang paling banyak membahas tentang klasifikasi kata ujaran kebencian pada algoritma Naïve Bayes yaitu tahun 2018 dan media social twitter. Jadi tinjauan pustaka sistematis ini menghasilkan analisis tentang tahun, kombinasi metode dan media sosial yang paling signifikan membahas tentang klasifikasi ujaran kebencian dengan Algoritma Naive Bayes.

Kata Kunci*— Ujaran Kebencian, Klasifikasi, Tinjauan Pustaka Sistematis, Naïve Bayes*

1. PENDAHULUAN

Perkembangan teknologi internet saat ini mempengaruhi perubahan di beberapa bidang kehidupan manusia, seperti pada bidang pendidikan, perdagangan, pemerintahan hingga komunikasi. Pada zaman modern ini pertumbuhan internet semakin pesat terutama penggunaan media sosial sehingga opini publik semakin luas dan bebas ditunjukan di berbagai media sosial. Kebebasan berpendapat yang dimiliki setiap orang malah akan disalahgunakan untuk ujaran kebencian.

Pada Januari 2022, di Indonesia ada 191 juta pengguna aktif media sosial, menurut studi We Are Social. Dari 170 juta tahun sebelumnya, jumlah itu naik 12,35% [1]. Twitter merupakan salah satu platform media sosial yang sering digunakan untuk menyebarkan ujaran kebencian. Di twitter, banyak sekali bentuk ujaran kebenciam yang didasari oleh motif SARA (suku, agama, ras, dan antargolongan) [2]. Hal tersebut didasarkan pada pertimbangan bahwa twitter menjadi media sosial yang dimanfaatkan oleh publik untuk mengartikulasikan ide, menyampaikan kritik, menyebarluaskan informasi secara instan, dan berdebat dengan sesama warganet [3].

Untuk membuat analisis sentiment ujaran kebencian diperlukan pemilihan klasifikasi kata yang akan digunakan. Metode klasifikasi naïve bayes digunakan dalam pembuatan penelitian ini. Naïve bayes adalah metode sederhana yang dikembangkan dengan memeriksa kondisi eksisting dan potensi setiap kondisi, berdasarkan aturan Bayes [4].

Pada jurnal penelitian sebelumnya berkaitan dengan penerapan analisis sentimen ujaran kebencian dengan metode Naïve Bayes diantaranya adalah menggunakan teknik Naïve Bayes berbasis N-Gram dengan pemilihan fitur Information Gain untuk mengklasifikasikan ujaran kebencian di Twitter. 250 buah data diidentifikasi sebagai ujaran kebencian dan 250 buah sebagai bukan ujaran kebencian untuk penelitian ini, menggunakan rasio data latih sebesar 80% dan rasio data uji sebesar 20%. Nilai presisi 92%, nilai recall 79,31%, dan nilai f-measure 85,18%. Hasil akurasi terbaik yang dihasilkan menggunakan fitur Unigram dan tanpa menggunakan pemilihan fitur Information Gain adalah 84% [2]. Peneliti selanjutnya, yang menggunakan metode Naïve Bayes dan seleksi fitur Information Gain dengan normalisasi kata. 250 tweet kebencian dalam bahasa Indonesia dimasukkan dalam analisis, dengan perbandingan data latih sebesar 80% dan data uji sebesar 20%. Dengan ambang batas sebesar 20%, 40%, 60%, 80%, dan 90%. Hasil akurasi terbaik penelitian tersebut adalah 98% dengan 100% nilai precision, 96,15% nilai recall, dan 98,03% nilai f-measure. Hasil ini dicapai dengan menggunakan langkah pre-processing normalisasi kata dan pemilihan fitur Information Gain dengan threshold 80%. Penelitian ini mampu meningkatkan akurasi hasil menjadi lebih baik berdasarkan analisis tes dan hasil [5].

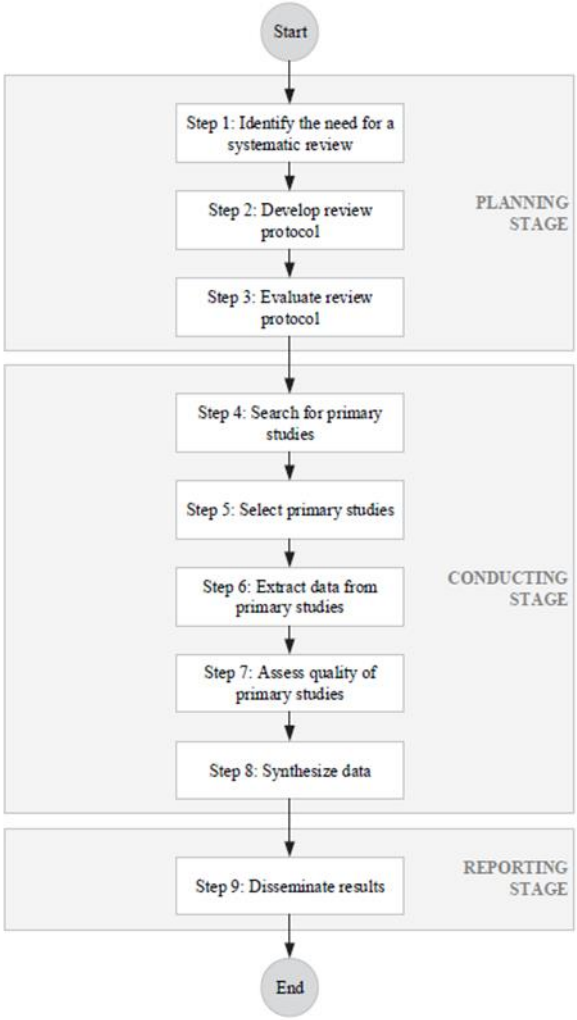
Penelitian selanjutnya yang menggunakan metode Naive Bayes untuk mengklasifikasikan kategori ujaran kebencian dengan hasil yang diperoleh menunjukkan bahwa sentiment irrelevant sebanyak 11,3% dengan 573 data, 35,4% sentiment negatif dengan 1786 data, 26,7% sentiment netral sebanyak 1350 data dan 26,6% sentiment positif sebanyak 1343 data. Sentimen negatif memperoleh skor tertinggi dengan nilai sebesar 35,4% [6]. Berikutnya sebuah aplikasi situs web dikembangkan yang akan membantu pemerintah dan institusi terkait dalam menemukan kata ujaran kebencian di postingan media sosial twitter menggunakan metode machine learning berupa naïve bayes. Hasil dari penelitian ini menunjukkan bahwa sistem pendeteksi ujaran ancaman pada twitter yang di buat mendapatkan akurasi sebesar 66%, precision 64%, recall 63%, dan F1 score sebesar 63% [7].

Banyak penelitian yang menerapkan berbagai metode klasifikasi untuk menganalisis ujaran kebencian pada media sosial. Tinjauan Pustaka ini bertujuan untuk menganalisis dan mengidentifikasi penelitian klasifikasi ujaran kebencian menggunakan metode Naïve Bayes antara tahun 2017-2022..

2. METODE PENELITIAN

2.1. Tahapan Review

Penelitian ini merupakan Systematic Literature Review (SLR). Tinjauan Pustaka Sistematis adalah strategi untuk melakukan tinjauan pustaka yang menemukan, meneliti, dan menilai semua temuan pada topik penelitian tertentu untuk memberikan jawaban atas pertanyaan penelitian yang diajukan sebelumnya[8]. Tinjauan sistematis literatur telah dilakukan berdasarkan pedoman awal yang digunakan oleh peneliti [9]. Tahap perencanaan, pelaksanaan, dan pelaporan tinjauan pustaka adalah tiga langkah yang membentuk tinjauan pustaka sistematis (SLR) [10]. Seperti terlihat pada Gambar 1, berikut alur penelitian yang dipakai pada penelitian SLR ini.



Gambar 1. Alur penelitian [9].

2.1.1. Gambar dan Tabel

2.2. Research questions

Research questions (RQ) atau pertanyaan penelitian, digunakan untuk mengumpulkan data spesifik dari setiap studi yang diselesaikan [11]. RQ dirancang dengan bantuan Population, Intervention, Comparison, Outcomes, dan Context (PICOC) menurut [9]. Tabel 1 menunjukkan struktur (PICOC) dari pertanyaan penelitian.

Tabel 1. Rangkuman PICOC

Populasi	*Naïve Bayes Classifier*, *Naïve Bayes*, ujaran kebencian
Intervensi	Ujaran kebencian klasifikasi *Naïve bayes*, *hate speech* klasifikasi *Naïve bayes*, ujaran kebencian
Perbandingan	-

Hasil	Klasifikasi ujaran kebencian pada sosial media dengan algoritma *Naïve bayes*
Konteks	Literatur ujaran kebencian dengan Algoritma *Naïve Bayes*

Pada Tabel 2 berisi pertanyaan penelitian dan motivasi yang dibuat dalam tinjauan pustaka ini:

Tabel 2. Research Questions

ID	*Research questions*	Motivasi
RQ1	Jurnal tahun berapakah yang paling banyak mempublikasikan algoritma *naive bayes*?	Identifikasi rahun publikasi jurnal yang membahass algoritma *Naïve Bayes*.
RQ2	Jurnal manakah yang paling signifikan membahas algoritma *Naïve Bayes*?	Indentifikasi jurnal signifikan yang membahas algoritma *Naïve Bayes*.
RQ3	Metode apa yang banyak digunakan pada algoritma *Naïve Bayes?*	Identifikasi metode yang banyak digunakan pada algoritma *Naïve Bayes.*
RQ4	*Media social* apa yang paling banyak dipakai pada ulasan ujaran kebencian?	Indentifikasi *media social* yang paling banyak dipakai dalam jurnal.

2.3. Research questions

Proses pencarian data pada SLR ini memiliki beberapa tahapan, yaitu memilih perpustakaan digital, menentukan kata kunci pencarian, melakukan pencarian kata kunci, memperbaiki kata kunci pencarian, dan mengambil kata kunci pada perpustakaan digital. Dari hasil pencarian pada perpustakaan digital yang dilakukan, artikel yang di cari antara tahun 2017-2022 dengan database yang digunakan yaitu Google Scholar.
Kata kunci dicari dengan langkah sebagai berikut [10]:

1. Mengidentifikasi kata kunci pencarian dari PICOC.
2. Mengidentifikasi kata kunci dari research questions (RQ) atau pertanyaan penelitian.
3. Mengidentifikasi dari judul, abstrak dan kata kunci yang relevan.
4. Mengidentifikasi persamaan kata, lawan kata dan kata aternatif dari kata kunci pencarian.
5. Menggunakan string pencarian yang menggunakan istilah Boolean AND dan OR.

Kata kunci yang di gunakan:
(Ujaran Kebencian OR Ujar Kebencian OR Hate Speech*) AND (Klasifikasi OR Classifier OR classification*) AND (Naïve Bayes Classifier OR NBC*) AND (Naïve Bayes*).

2.4. Study Selection

Studi utama yang akan diteliti dipilih dengan memakai kriteria inklusi dan eksklusi. Kriteria ini ditunjukan pada Tabel 3.

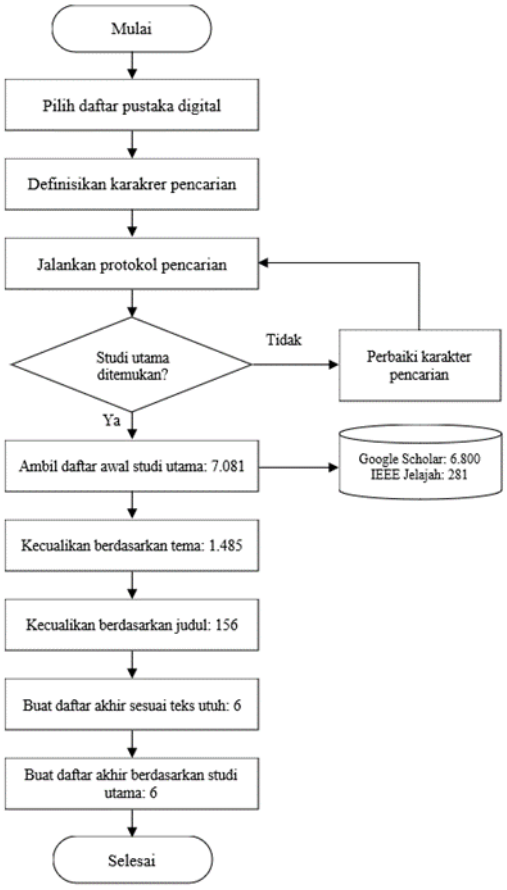
Tabel 3. Kriteria inklusi dan kriteria eksklusi

Kriteria inklusi	Studi yang membahas akurasi penerapan klasifikasi *Naïve Bayes* pada *tweet* ujaran kebencian.

	Studi yang membahas kolaborasi metode yang membahas akurasi atau klasifikasi dengan *Naïve bayes.*
Kriteria eksklusi	Studi yang membahas selain konteks *naïve bayes.* Studi yang dilakukan dibawah tahun 2017.

Seleksi studi utama dalam penelitian ini memakai kriteria inklusi dan eksklusi dimana hasil inklusi yang didapat yaitu penelitian yang membahas akurasi penerapan klasifikasi naïve bayes pada kata ujaran kebencian, peneliti yang membahas kolaborasi metode yang digunakan pada akurasi dan klasifikasi ujaran kebencian, dan penelitian yang diambil yang paling lengkap dan baru. Sedangkan kriteria eksklusi yang di gunakan yaitu penelitian yang membahas selain konteks naïve bayes dan penelitian yang dilakukan dibawah tahun 2017.

Pada Gambar 2, proses pemilihan studi utama yaitu dilakukan dengan mengambil daftar awal, mengecualikan berdasarkan judul atau abstrak dan mengecualikan berdasarkan teks utuh.



Gambar 2. Proses Pemilihan Studi Utama

2.5. Ektrasi data

Data studi utama yang dipilih akan di ekstraksi untuk dapat menjawab Reserch Question pada penelitian ini. Proses ekstraksi data pada 6 artikel utama yang dipilih, kemudian dibagi dalam beberapa properti seperti ditunjukan oleh tabel 4.

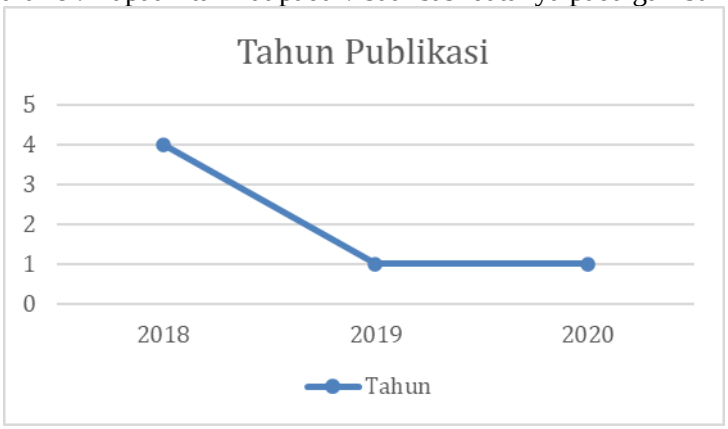
Tabel 4. Ekstraksi Data

Properti	Pertanyaan Penelitian
Tahun publikasi jurnal	RQ1
Jurnal yang paling signifikan, metode yang dipakai dalam algoritma dan sosial media yang dipakai	RQ2, RQ3, RQ4

3. HASIL DAN PEMBAHASAN

3.1. Tahun publikasi jurnal

Pada studi utama yang didapat sebanyak 6 artikel dipilih berdasarkan kata kunci, tahun jurnal yang paling banyak membahas tentang tentang klasifikasi ujaran kebencian dengan algoritma Naïve Bayes adalah tahun 2018 sebanyak 4 artikel, 2019 sebanyak 1 artikel dan tahun 2020 sebanyak 1 artikel. Dapat kita lihat pada visualisasi datanya pada gambar 3 dibawah ini.



Gambar 3. Diagram tahun publikasi

3.2. Riview dan Analisis Literature

Berdarkan dari proses pemilihan studi utama yang menggunakan kriteria inklusi dan eksklusi maka data akhir literatur yang didapat sebanyak 6 judul. Pada tabel 5 dapat kita lihat:

Tabel 5. Riview dan analisis perbandingan literatur

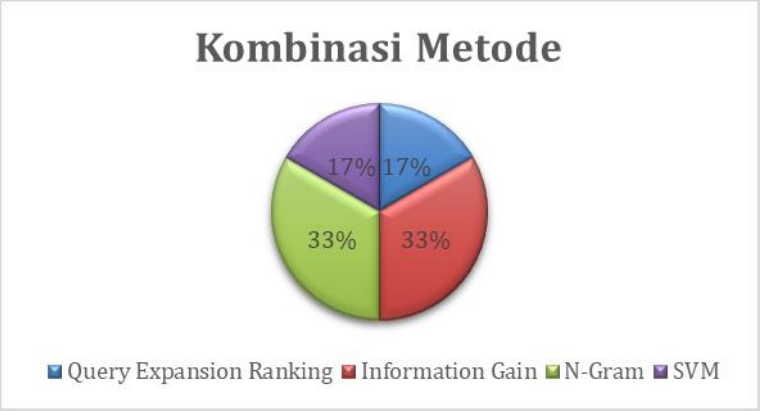
No	Referensi	Judul	Kombinasi metode	Media sosial	Akurasi	*Database*
1	Shima Fanissa, M. Ali Fauzi, Sigit Adinugroho (2018) [4]	Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode *Naive Bayes* dan Seleksi Fitur	*Query Expansion Ranking*	*Web TripAdvisor*	86,5%	*Google Scholar*

		Query Expansion Ranking				
2	Muhammad Hakiem, Mochammad Ali Fauzi, Indriati (2019) [2]	Klasifikasi Ujaran Kebencian pada *Twitter* Menggunakan Metode *Naïve Bayes* Berbasis *N-Gram* Dengan Seleksi Fitur *Information Gain*	*Fitur Information Gain*	*Twitter*	84%	*Google Scholar*
3	Ivan, Yuita Arum Sari, Putra Pandu Adikara (2019) [5]	Klasifikasi *Hate Speech* Berbahasa Indonesia di *Twitter* Menggunakan *Naive Bayes* dan Seleksi Fitur *Information Gain* dengan Normalisasi Kata	*Fitur Information Gain*	*Twitter*	98%	*Google Scholar*
4	Luh Putu Ary Sri Tjahyanti (2020) [12]	PENDETEKSIAN BAHASA KASAR (*ABUSIVE LANGUAGE*) DAN UJARAN KEBENCIAN (*HATE SPEECH*) DARI KOMENTAR DI JEJARING SOSIAL	*N-Gram* dengan kata *Unigram*	*Twitter*	87,26%	*Google Scholar*
5	Kevin Antariksa, Y. Sigit Purnomo	Klasifikasi Ujaran Kebencian	*N-Gram*	*Twitter*	80%	*Google Scholar*

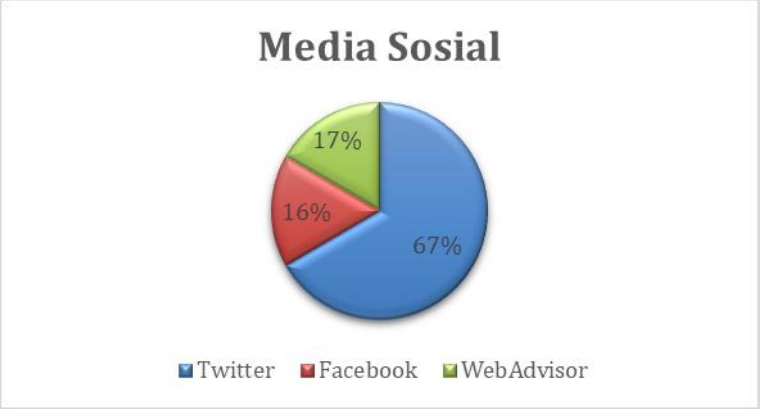
	WP., Dra. Ernawati (2019) [13]	pada Cuitan dalam Bahasa Indonesia *Twitter*				
6	Asogwa D.C, Chukwuneke C.I, Ngene C.C, Anigbogu G.N (2019) [14]	*Hate Speech Classification Using SVM and Naive BAYES*	SVM	*Facebook*	72%	*IEEE*

3.3. Hasil Analisis

Berdasarkan hasil riview dan analisis perbandingan yang dibuat, maka dapat dibuat visualisasi dari artikel ujaran kebencian klasifikasi Naïve bayes yang paling signifikan atau akurat dengan kombinasi metode dan media sosial yang paling banyak dipakai pada klasifikasi ujaran kebencian. Pada Gambar 4 dan Gambar 5 dapat kita lihat diagram berikut.



Gambar 4. Persentase metode yang digunakan



Gambar 5. Persentase sosial media yang digunakan

4. KESIMPULAN

Berdasarkan *literature riview* yang telah diakukan maka jumlah studi utama yang didapat sebanyak 6 artikel dengan menggunakan proses inklusi dan eksklusi agar mendapat hasil yang akurat dari penelitian tentang klasifikasi kata ujaran kebencian menggunakan algoritma *naïve bayes*. Maka hasil penelitian yaitu akurasi paling signifikan didapat pada kombinasi metode pada algoritma *Naïve Bayes* menggunakan metode seleksi fitur *information gain* yaitu sebesar 98%. Untuk tahun publikasi yang paling banyak membahas tentang algoritma *Naïve Bayes* yaitu tahun 2018, dan media social yang banyak digunakan pada algoritma *Naïve Bayes* klasifikasi ujaran kebencian yaitu *twitter*.

DAFTAR PUSTAKA

[1] M. I. Mahdi, "Pengguna Media Sosial di Indonesia Capai 191 Juta pada 2022," *Dataindonesia.id*, 2022. .

[2] M. Hakiem, M. A. Fauzi, and Indriati, "Klasifikasi Ujaran Kebencian pada Twitter Menggunakan Metode Naïve Bayes Berbasis N-Gram Dengan Seleksi Fitur Information Gain," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 2443–2451, 2019.

[3] A. Nurul F, N. Nurhadi, and S. Pranawa, "Konflik dan Ujaran Kebencian di Twitter (Studi Tentang Hashtag #2019TetapJokowi and #2019GantiPresiden Periode Januari-Februari 2019)," *Jupiis J. Pendidik. Ilmu-Ilmu Sos.*, vol. 12, no. 1, p. 132, 2020, doi: 10.24114/jupiis.v12i1.16083.

[4] Sh. Fanissa, M. A. Fauzi, and S. Adinugroho, "Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking Optimasi Sisa Bahan Baku Pada Industri Mebel Menggunakan Algoritma Genetika View project Automatic Essay Scoring View project," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 8, pp. 2766–2770, 2018.

[5] Y. A. S. Ivan and P. P. Adikara, "Klasifikasi Hate Speech Berbahasa Indonesia di Twitter Menggunakan Naive Bayes dan Seleksi Fitur Information Gain dengan Normalisasi Kata," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 5, pp. 4914–4922, 2019.

[6] M. Chalida and M. D. R. Wahyudi, "Analisis Sentimen Ujaran Kebencian Pemilihan Presiden 2019 Menggunakan Algoritma Naïve Bayes (Studi Kasus: Tweet #Pilpres2019 Di Kota Jakarta, Bandung, Semarang, Surabaya Dan Yogyakarta)," *Jnanaloka (Jurnal Open Access Yayasan Lentera Dua Indones.*, no. 2001, pp. 5–10, 2019.

[7] A. Rafi R, M. Nasrun, and R. Astuti N, "Deteksi Ujaran Ancaman Berbasis Website Pada Postingan Media Sosial Twitter Menggunakan Metode Naive Bayes," *e-Proceeding Eng.*, vol. 8, no. 1, p. 500, 2021.

[8] E. Suhendar, M. I. Komputer, F. T. Informasi, B. Luhur, L. Publik, and K. Data, "Tinjauan Sistematis : Implementasi Cloud Computing Terhadap Keamanan Layanan Publik," pp. 599–606.

[9] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering - A systematic literature review," *Inf. Softw. Technol.*, vol. 51, no. 1, pp. 7–15, 2009, doi: 10.1016/j.infsof.2008.09.009.

[10] R. S. Wahono, "A Systematic Literature Review of Software Defect Prediction: Research Trends, Datasets, Methods and Frameworks," *J. Softw. Eng.*, vol. 1, no. 1, pp. 1–16, 2015.

[11] M. Ahmad, S. Aftab, M. S. Bashir, and N. Hameed, "Sentiment analysis using SVM: A systematic literature review," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 2, pp. 182–188, 2018, doi: 10.14569/IJACSA.2018.090226.

[12] L. P. A. S. Tjahyanti, "Pendeteksian Bahasa Kasar (Abusive Language) Dan Ujaran Kebencian (Hate Speech) Dari Komentar Di Jejaring Sosial," *J. Chem. Inf. Model.*, vol. 07, no. 9, pp. 1689–1699, 2020.

[13] K. Antariksa, Y. S. Purnomo WP, and E. Ernawati, "Klasifikasi Ujaran Kebencian pada Cuitan dalam Bahasa Indonesia," *J. Buana Inform.*, vol. 10, no. 2, p. 164, 2019, doi: 10.24002/jbi.v10i2.2451.

[14] S. Ahammed, M. Rahman, M. H. Niloy, and S. M. M. H. Chowdhury, "Implementation of Machine Learning to Detect Hate Speech in Bangla Language," *Proc. 2019 8th Int. Conf. Syst. Model. Adv. Res. Trends, SMART 2019*, pp. 317–320, 2020, doi: 10.1109/SMART46866.2019.9117214.