

Tinjauan Pustaka Sistematis: Penerapan Data Mining Metode Klasifikasi Untuk Menganalisa Penyalahgunaan Sosial Media

Bangkit Indarmawan Nugroho¹, Zaenal Ma'arif², Zaenul Arif³

¹Sistem Informasi STMIK TEGAL

^{2,3}Program Studi Teknik Informatika STMIK TEGAL

Email: ¹efbeterang@gmail.com, ²zaenalmaarif21@gmail.com, ³zendhunter@gmail.com

Abstrak

Data mining adalah proses analitis yang dirancang untuk mengeksplorasi sejumlah besar data untuk pengetahuan yang berharga, konsisten, dan tersembunyi. Salah satu metode yang digunakan dalam data mining adalah klasifikasi. Metode klasifikasi dapat digunakan untuk membuat prediksi menggunakan data terpadat dari Igracias. Algoritma klasifikasi yang digunakan dalam penelitian ini adalah pohon keputusan. Media sosial adalah tempat orang berbagi tentang diri mereka sendiri dan perspektif mereka tentang kehidupan. Menyadari bahwa banyak informasi dapat digali dari media sosial, beberapa penelitian telah dilakukan dengan menggunakan media sosial sebagai objek. Ini juga mencakup penelitian tentang penyalahgunaan media sosial. Dengan kemajuan informasi yang pesat, orang biasa dipermudah oleh teknologi yang dapat memperburuk keadaan. Pesan lelucon dan ujaran kebencian dapat menyebar dengan sangat mudah karena kehadiran media sosial seperti Facebook, Twitter, atau aplikasi ponsel seperti WhatsApp, Line. Ketika media komunikasi dan informasi tidak dapat difilter dari awal ini adalah penerapan data mining ke metode klasifikasi menggunakan Dari 2017 hingga 2021, penyalahgunaan di media sosial tetap lazim. Berdasarkan pilihan literatur dari artikel pilihan, Twitter adalah media sosial yang paling umum dari eksploitasi media sosial, dan taksonomi yang banyak digunakan untuk eksploitasi media sosial adalah Naive Bayes, SVM, dan Desison.Tree.

Kata Kunci: *data mining, klasifikasi, pohon keputusan, media sosial, taksonomi*

I. PENDAHULUAN

Data mining adalah proses analitis yang dirancang untuk mengeksplorasi sejumlah besar data untuk pengetahuan yang berharga, konsisten, dan tersembunyi. Salah satu metode yang digunakan dalam data mining adalah klasifikasi. Metode klasifikasi dapat digunakan untuk membuat prediksi menggunakan data terpadat dari Igracias. Algoritma klasifikasi yang digunakan dalam penelitian ini adalah *decision tree* (Rachmawaty & Febriani, 2022).

Media sosial adalah tempat orang berbagi tentang diri mereka sendiri dan perspektif mereka tentang kehidupan. Menyadari potensi untuk mengekstrak banyak informasi dari media sosial, beberapa penelitian telah dilakukan dengan menggunakan media sosial sebagai objek. (Yuliani et al., 2020). Termasuk penelitian tentang penyalahgunaan media sosial. Dengan kemajuan informasi yang pesat, orang biasa dipermudah oleh teknologi yang dapat memperburuk keadaan. Pesan lelucon dan ujaran kebencian dapat menyebar dengan sangat mudah karena kehadiran media sosial seperti Facebook, Twitter, atau aplikasi ponsel seperti WhatsApp, Line. Jika media komunikasi dan informasi tidak dapat disaring terlebih dahulu (Palupi, 2019).

Dengan menerapkan *Systematic Literatur Review*, tujuan yang ingin dicapai dalam paper ini untuk menganalisis data mining dengan menggunakan teknik klasifikasi dalam penyalahgunaan media sosial dengan melakukan *review* secara mendalam dan mengevaluasi penelusuran melalui literatur terpilih menggunakan parameter tertentu dan subjek yang terpilih akan diproses untuk menjawab pertanyaan penelitian, serta diharapkan mampu memberikan bukti pengaruh dan gambaran bahwa dalam data mining teknik klasifikasi dapat dijadikan sebagai sarana pengetahuan serta sebagai indikator paling kuat yang dapat dijadikan sebagai tolak ukur dalam penelitian bidang ilmiah.

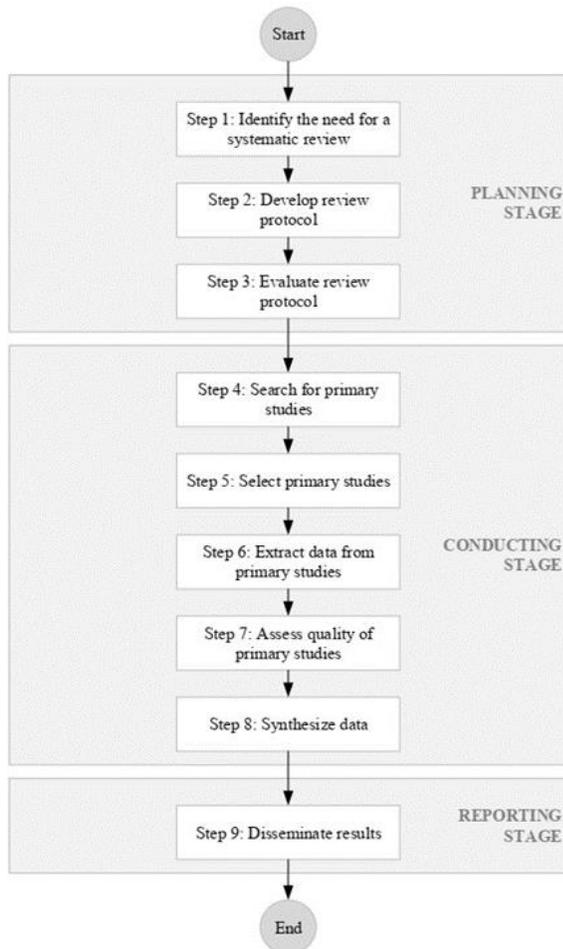
Dalam penelitian ini, metode pencarian paper dilakukan menggunakan teknik spontan dari *keyword* tertentu menggunakan 3 perpustakaan digital yang masih populer selama 5 tahun terakhir. Penelitian ini hanya membatasi untuk paper dari jurnal yang selanjutnya akan dikaji sehingga menghasilkan penelitian terbaru terkait data mining teknik klasifikasi penyalahgunaan media sosial.

II. METODE PENELITIAN

Systematic Literature Review

Tinjauan Pustaka Sistematis (*Systematic Literature Review/SLR*) adalah sebuah metode penelitian yang bertujuan untuk mengidentifikasi dan mengevaluasi hasil penelitian dengan teknik terbaik berdasarkan prosedur yang spesifik dari hasil perbandingan (Teknologi, n.d.). Pada penelitian ini, langkah-langkah yang digunakan mengikuti kaidah penelitian serta mengadopsi prosedur dari penelitian (Wahono, 2015).

Systematic Literature Review diproses dalam tiga langkah, yakni: perencanaan, pelaksanaan, dan pelaporan tinjauan pustaka.



Gambar 1. Menunjukkan tahapan pada metode SLR (Kitchenham et al., 2009)

Pertanyaan Penelitian

Pertanyaan Penelitian (*Research Question/RQ*) merupakan identifikasi pertama dalam metode SLR. RQ memiliki tujuan yakni melakukan teknik pencarian serta meneliti literatur yang mendalam secara otomatis dengan berorientasi pada objek tujuan. Untuk merancang RQ yang baik maka digunakan pendekatan PICOC (Kitchenham et al., 2009).

Tabel 1 menyajikan ringkasan untuk RQ dari PICOC. Tabel 2.2 menyajikan pertanyaan penelitian yang bertujuan untuk membahas lebih lanjut terkait penelitian ini.

Tabel 1. Ringkasan PICOC

<i>Population</i>	Data Mining, Klasifikasi, Sosial Media
<i>Intervention</i>	Penerapan data mining metode klasifikasi untuk penyalahgunaan sosial media
<i>Comparison</i>	-
<i>Outcomes</i>	Penyalahgunaan sosial media dengan menggunakan metode klasifikasi
<i>Context</i>	Penyalahgunaan sosial media

Tabel 2. Pertanyaan Penelitian dan Tujuan

ID	Pertanyaan Penelitian	Tujuan
RQ1	Topik Penelitian apa saja yang banyak digunakan dalam penelitian data mining untuk menganalisis penyalahgunaan media sosial?	Mengidentifikasi asi topik yang sering digunakan dalam penerapan data mining untuk menganalisis penyalahgunaan media sosial
RQ2	Sosial media apa saja yang banyak digunakan dalam penelitian data mining untuk menganalisis penyalahgunaan media sosial?	Mengidentifikasi asi sosial media yang sering digunakan dalam penerapan data mining untuk menganalisis penyalahgunaan media sosial
RQ3	Metode apa saja yang banyak digunakan dalam penelitian data mining untuk menganalisis penyalahgunaan media sosial?	Mengidentifikasi asi metode yang sering digunakan dalam penerapan data mining untuk menganalisis penyalahgunaan media sosial

Strategi Penelitian

Tinjauan pustaka memiliki beberapa bagian tertentu pada proses pencarian, yakni harus menentukan *digital library*, memilih *keyword* tertentu, menerapkan *keyword* yang sudah ditentukan, mengoreksi *keyword* dan mengumpulkan beberapa literatur dari *digital library*.

Sebelum memulai pencarian, tahap awal yang harus dilakukan adalah melakukan penentuan atau pemilihan *database* yang sesuai untuk menemukan paper yang relevan. *Database* dari *digital library* yang digunakan adalah ResearchGet dan Google Scholar..

Menurut(Wahono, 2015), pemilihan pencarian menggunakan *keyword* tertentu dilakukan beberapa langkah sebagai berikut:

1. Identifikasi pencarian berdasarkan kata kunci dengan menggunakan ringkasan PICOC terutama pada *Population* dan *Intervention*.
2. Identifikasi pencarian menggunakan pertanyaan penelitian.
3. Identifikasi pencarian kata kunci menggunakan abstrak dan *keyword* serta judul yang selaras.
4. Identifikasi persamaan kata, lawan kata, dan pergantian kata berdasarkan kata kunci pencarian.
5. Menggunakan kata kunci sebagai pencarian lanjutan dengan menerapkan Boolean AND dan OR.

Kata kunci yang digunakan untuk pencarian:

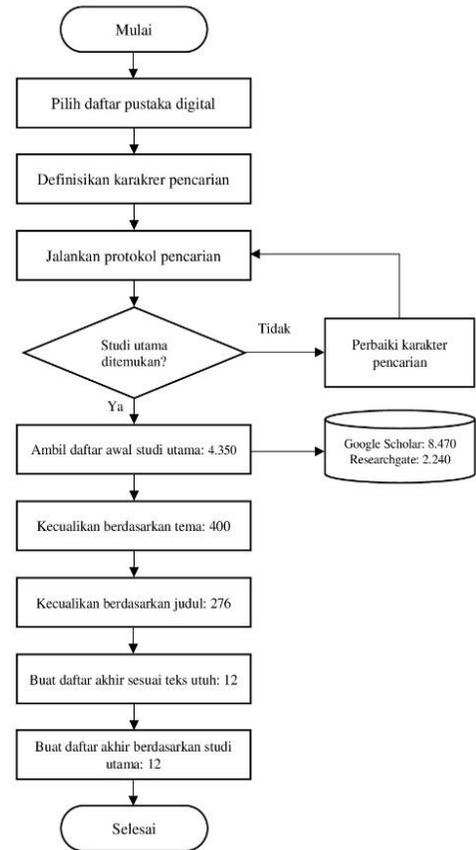
(Data Mining OR DM*) AND (Klasifikasi OR Clasifier*)
 AND (Sosial Media*)

Penyeleksian Kajian

Menggunakan kriteria inklusi dan kriteria eksklusi dengan tujuan untuk mendapatkan literatur *primer* yang kemudian akan ditinjau. Tabel 3 menyajikan rangkuman kriteria yang akan dikaji.

Tabel 3. Rangkuman Kriteria

Kriteria Inklusi	Kajian termasuk ke dalam Data Science atau Computer Science
	Diutamakan berbahasa inggris
Kriteria Eksklusi	Kajian tidak menggunakan validasi
	Kajian tanpa menggunakan teks penuh



Gambar 2. Hasil Pencarian dan Seleksi Kajian

III. HASIL DAN PEMBAHASAN

Evaluasi Literatur

Setelah dilakukan tahap pemilihan literatur dengan berdasarkan jurnal yang paling aktif dalam mempublikasikan paper terkait sosial media dengan kualitas terbaik, maka tahap selanjutnya adalah mengevaluasi literatur berdasarkan karakteristik melalui metode serta melakukan perbandingan dari database yang paling sering mempublikasikan jurnal terkait klasifikasi penyalahgunaan sosial media. Pada tahap ini diperoleh 12 paper final yang memenuhi syarat dari kriteria inklusi dan dan eksklusi. Hal ini tersaji dalam Tabel 4 berikut ini

Tabel 4. Daftar Paper Literatur Terpilih dan Perbandingan Database

No	Referensi	Judul	Topik	Metode	Hasil	Media
1	Salim, Yulita Budi, Lutfi (2020).(Sali m & Budi, 2020)	Analisis Sentimen terhadap Komentar Negatif di Media Sosial Facebook dengan Metode	Cyberbullying	Naive Bayes	Simpulkan bahwa data yang telah diproses sebelumnya dapat digunakan untuk menghapus kata-kata seperti huruf dan tanda	Facebo ok

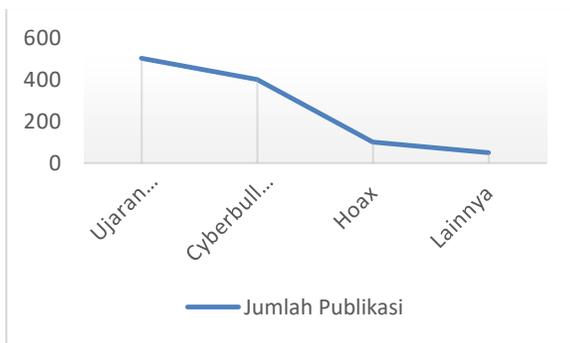
		Klasifikasi Naïve Bayes			baca yang tidak perlu, mengubah kata-kata prochem menjadi kata-kata standar, dan menghapus kata-kata umum yang digunakan sebagai komentar.	
2	Baihaqi, Wiga Maulana Pinilih, Muliastari Rohmah, Miftakhul (2020)(Baihaqi et al., 2020)	Kombinasi k-means dan support vector machine (svm) untuk k-means and support vector machine (svm) combination to predict sara elements on tweet	Cyberbullying	SVM	Berhasil menganalisis tweet positif dengan sara dan tweet negatif dengan sara, namun terbukti dengan rendahnya hasil klasifikasi oleh algoritma svm, tweet dengan sara dan tweet dengan sara hasil pengelompokan tweet yang tidak memberikan hasil yang baik.	Twitter
3	Maulana, Fajar Agus Ernawati, Iin(Maulana et al., 2020)	Analisa sentimen cyberbullying di jejaring sosial twitter dengan algoritma naïve bayes	Cyberbullying	Naive bayes	Dapat digunakan untuk mengklasifikasikan tweet yang mengandung unsur cyberbullying atau tidak mengandung unsur cyberbullying tetapi harus didukung dengan data training yang sesuai. Keduanya untuk data pelatihan yang dilabeli mengandung cyberbullying.	Twitter
4	Rani Kurnia Putri dkk. (2021) (Studi et al., 2021)	Support vector machine untuk identifikasi berita hoax terkait virus	Hoax	SVM	Kumpulan hoax dan non-hoax yang sebelumnya diverifikasi oleh tim MAFINDO berdasarkan diskusi anggota dan hasil pencarian fakta. Hasilnya dipublikasikan di laman turnbackhoax.id.	Facebook
5	Auliasari, Karina Prasetya, Renaldi Primaswara(2021)(Auliasari, Prasetya Karina & Industri, 2021)	Analisis sentiment cyberbullying pada sosial media twitter menggunakan metode support vector machine	Cyberbullying	SVM	Berdasarkan hasil pengujian black box, kami menunjukkan bahwa sistem yang kami bangun mampu mengonversi dan mengklasifikasikan konten tweet yang berisi: Elemen selain cyberbullying di dalamnya	Twitter
7	Ulfah, Aniq Noviciatie Anam, M Khairul	Analisis Sentimen Hate Speech Pada Portal Berita Online	Hate Speech	SVM	SVM berlaku saat menganalisis komentar tentang isu politik yang mengandung ujaran	Facebook

	(2020) (Ulfah et al., 2020)	Menggunakan <i>Support Vector Machine</i> (SVM)			kebencian dengan skor akurasi 53,88%.	
8	Musdar, Izmy Alwiah Angriani, Husni (2022)(Musdar & Husni, 2022)	Implementasi Teori <i>Naive Bayes</i> dalam Klasifikasi Ujaran Kebencian di Facebook	Ujaran Kebencian	<i>Naive Bayes</i>	Tingkat akurasi klasifikasi ujaran kebencian menggunakan metode <i>Naive Bayes</i> mencapai angka 83%.	Facebook
9	Umar Syahid Aulia Rahman dkk (2020)(Syahid et al., 2020).	Implementasi <i>Multinomial Naive Bayes</i> Untuk Klasifikasi Ujaran Kebencian Pada Dataset Kicauan (Twitter) Bahasa Indonesia	Ujaran Kebencian	<i>Naive Bayes</i>	Proses pembuatan kumpulan data yang terfokus Sejumlah masalah membuat penerapan <i>Multinomial Naive Bayes</i> pada kumpulan data kurang berhasil.	Twitter
10	Chalida, Muftia Didik, M (2019) (Chalida & Didik, 2019)	Analisis sentimen ujaran kebencian pemilihan presiden 2019 menggunakan algoritma <i>Naive Bayes</i>	Ujaran Kebencian	<i>Naive Bayes</i>	Hasil untuk masing-masing kota dengan sentimen ujaran kebencian tertinggi adalah Jakarta dengan 33,8%, Bandung dengan sentimen negatif 65,4%, Surabaya dengan sentimen positif 37,2%, dan Yogyakarta dengan sentimen negatif 51,8%. Semarang dengan sentimen negatif 61,7%.	Twitter
11	Fauzi Ihsan, Iwan Iskandar dkk (2021) (Ihsan et al., 2021)	Algoritme <i>decision tree</i> untuk mendeteksi ujaran kebencian dan bahasa kasar multilabel pada Twitter berbahasa Indonesia <i>Decision tree algorithm for multi-label hate speech and abusive language</i>	Ujaran Kebencian	<i>decision tree</i>	Meningkatkan hasil akurasi klasifikasi dari algoritme pohon keputusan saat mengklasifikasikan teks Twitter untuk mendeteksi beberapa kelas yang berbeda.	Twitter
12	Basuki, Setio Maghfiroh, Siti Azhar, Yufis (2020) (Basuki et al., 2020)	Klasifikasi Tweets Tindak Kejahatan Berbahasa Indonesia Menggunakan <i>Naive Bayes</i>	<i>Cybercrime</i>	<i>Naive Bayes</i>	Data Tweet diperoleh dengan mengimplementasikan library Twitter 4jt dengan menambahkan kunci konsumen, rahasia konsumen, kunci token dan rahasia token yang disediakan oleh Twitter.	Twitter

Objek Penelitian Berpengaruh

Berdasarkan publikasi jurnal paling aktif yang mempublikasikan paper terkait penerapan data mining untuk

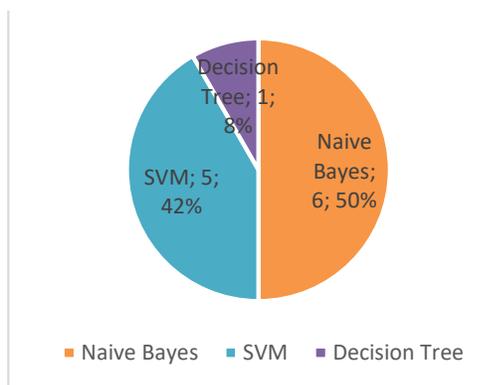
menganalisis penyalahgunaan sosial media adalah *cyberbullying* sebanyak 500 artikel, *Hoax* sebanyak 100 artikel, *Ujaran Kebencian* sebanyak 400 artikel. Gambar 3 menyajikan statistik distribusi publikasi artikel dari literatur terpilih.



Gambar 3. Distribusi Publikasi Paper dari Literatur Terpilih

Metode Berpengaruh

Berdasarkan perbandingan dari literature terpilih, gambar 4 menunjukkan bahwa sebagian besar metode penerapan data mining untuk menganalisis penyalahgunaan sosial media didominasi oleh *Naive Bayes*, *SVM*, dan *Decision tree*.



Gambar 4. Perbandingan Metode Berpengaruh

IV. KESIMPULAN

Setelah melakukan penelitian, dapat disimpulkan bahwa: penerapan data mining penyalahgunaan media sosial metode klasifikasi masih banyak digunakan dalam rentang waktu 2017-2021. Berdasarkan pemilihan literatur dari artikel terpilih, sosial media yang banyak digunakan terhadap penyalahgunaan sosial media adalah Twitter dengan metode klasifikasi yang banyak digunakan terhadap penyalahgunaan sosial media adalah *Naive Bayes*, *SVM* dan *Decision Tree*

V. DAFTAR PUSTAKA

- [1] Auliasari, Prasetya Karina, P. R., & Industri, F. T. (2021). ANALISIS SENTIMENT CYBERBULLYING PADA SOSIAL MEDIA TWITTER MENGGUNAKAN METODE SUPPORT VECTOR MACHINE. 5(2), 548–556.
- [2] Baihaqi, W. M., Pinilih, M., Rohmah, M., Informasi, T., Purwokerto, U. A., Informasi, S., Purwokerto, U. A., & Korespondensi, P. (2020). KOMBINASI K-MEANS DAN SUPPORT VECTOR MACHINE (SVM) UNTUK K-MEANS AND SUPPORT VECTOR MACHINE (SVM) COMBINATION TO PREDICT SARA ELEMENTS ON TWEET. 7(3), 501–510. <https://doi.org/10.25126/jtiik.202072126>
- [3] Basuki, S., Maghfiroh, S., & Azhar, Y. (2020). Klasifikasi Tweets Tindak Kejahatan Berbahasa Indonesia Menggunakan Naive Bayes. 2(7), 933–944.
- [4] Chalida, M., & Didik, M. (2019). Analisis sentimen ujaran kebencian pemilihan presiden 2019 menggunakan algoritma Naive Bayes.
- [5] Ihsan, F., Iskandar, I., Harahap, N. S., Agustian, S., Iskandar, I., Harahap, N. S., & Agustian, S. (2021). Algoritme decision tree untuk mendeteksi ujaran kebencian dan bahasa kasar multilabel pada Twitter berbahasa Indonesia Decision tree algorithm for multi-label hate speech and abusive language. 9(October), 199–204. <https://doi.org/10.14710/jtsiskom.2021.13907>
- [6] Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering – A systematic literature review. *Information and Software Technology*, 51(1), 7–15. <https://doi.org/10.1016/j.infsof.2008.09.009>
- [7] Maulana, F. A., Ernawati, I., Labu, P., & Selatan, J. (2020). Analisa sentimen cyberbullying di jejaring sosial twitter dengan algoritma naive bayes. 529–538.
- [8] Musdar, A. A. I., & Husni. (2022). Implementasi Teori Naive Bayes dalam Klasifikasi Ujaran Kebencian di Facebook. 6(4), 666–671.
- [9] Palupi, R. (2019). Penyalahgunaan Media Sosial Sebagai Alat Propaganda. *Jurnal Komunikasi*, 10(1), 69–76. <https://doi.org/10.31294/jkom.v10i1.5364>
- [10] Rachmawaty, D., & Febriani, J. M. A. (2022). Penerapan Metode Klasifikasi Decision Tree Untuk Memprediksi Kelulusan Tepat Waktu. *Journal of Industrial Engineering and Technology*, 2(1), 61–74. <https://doi.org/10.24176/jointech.v2i1.7432>
- [11] Salim, Y., & Budi, L. et al. (2020). Analisis Sentimen terhadap Komentar Negatif di Media Sosial Facebook dengan Metode Klasifikasi Naive Bayes. 1(4), 259–265.
- [12] Studi, P., Matematika, P., Sains, F., Pgr, U., & Buana, A. (2021). SUPPORT VECTOR MACHINE UNTUK IDENTIFIKASI BERITA HOAX TERKAIT VIRUS. 6(3), 162–167.
- [13] Syahid, U., Rahman, A., Wibisono, Y., & Nugroho, E. P. (2020). Implementasi Multinomial Naive Bayes Untuk Klasifikasi Ujaran Kebencian Pada Dataset Kicauan (Twitter) Bahasa Indonesia. 3(2), 78–84.
- [14] Teknologi, M. K. (n.d.). *Jurnal iptek*. 17–24. <https://doi.org/10.31284/j.iptek.2019.v23i1>
- [15] Ulfah, A. N., Anam, M. K., Informatika, J. T., Speech, H., & Machine, S. V. (2020). Analisis Sentimen Hate Speech Pada Portal Berita Online Menggunakan Support Vector Machine (SVM). 7(1).
- [16] Wahono, S. R. (2015). A Systematic Literature Review of Software Defect Prediction : Research Trends , A Systematic Literature Review of Software Defect Prediction : Research Trends , Datasets , Methods and Frameworks. May.
- [17] Yuliani, E., Utami, E., & Raharjo, S. (2020). Klasifikasi Kepribadian Pengguna Media Sosial. *Jurnal Informa : Jurnal Penelitian Dan Pengabdian Masyarakat*, 6(1), 15–19. <https://doi.org/10.46808/informa.v6i1.167>