

**PENERAPAN METODE *INFORMATION GAIN* DAN
SUPPORT VECTOR MACHINE UNTUK IDENTIFIKASI
KONTEN KASAR PADA *TWEETS***

TUGAS AKHIR

**Kevin Aprilion
1116006**



**PROGRAM STUDI INFORMATIKA
INSTITUT TEKNOLOGI HARAPAN BANGSA
BANDUNG
2022**

**PENERAPAN METODE *INFORMATION GAIN* DAN
SUPPORT VECTOR MACHINE UNTUK IDENTIFIKASI
KONTEN KASAR PADA *TWEETS***

TUGAS AKHIR

**Diajukan sebagai salah satu syarat untuk memperoleh
gelar sarjana dalam bidang Informatika**

**Kevin Aprilion
1116006**



INSTITUT
TEKNOLOGI
HARAPAN
BANGSA

Veritas vos liberabit

**PROGRAM STUDI INFORMATIKA
INSTITUT TEKNOLOGI HARAPAN BANGSA
BANDUNG
2022**

ABSTRAK

Nama : Kevin Aprilion
Program Studi : Informatika
Judul : Penerapan Metode *Information Gain* dan *Support Vector Machine* untuk Identifikasi Konten Kasar pada *Tweets*

Dalam penggunaan jejaring sosial apalagi Twitter sering sekali menemukan tulisan yang mengandung kata-kata kasar dan bersifat ofensif dalam Bahasa Indonesia. Penelitian ini akan menguji akurasi metode *Information Gain*, metode TF-IDF dan menguji parameter pada metode SVM untuk mengetahui kombinasi metode dan parameter terbaik untuk identifikasi. Pada penelitian ini memiliki beberapa indikator utama yaitu menggunakan *stemming* atau tidak menggunakan *stemming* pada saat *preprocessing* jenis pengambilan fitur, parameter C dan *gamma*, dan jenis *kernel*. Data yang digunakan didapat dari website penyedia dataset, berupa kumpulan data Twitter yang terdapat kata kasar. Berdasarkan pengujian, hasil akurasi tertinggi didapatkan dengan nilai 93.1402% dengan *f-measure* 90.7787% hasil ini didapat dengan menggunakan dataset Twitter yang tidak dilakukan pelabelan manual metode yang gunakan adalah *Information Gain* dengan *stemming*. Dari penelitian ini, dapat disimpulkan bahwa menggunakan *stemming* dan pengambilan fitur dapat mempengaruhi nilai akurasi dalam pendeteksian kata kasar tersebut. Faktor seperti jenis *kernel* dan parameter SVM juga menjadi salah satu hal yang mempengaruhi hasil pendeteksian kata kasar tersebut.

Kata kunci: Identifikasi Kata Kasar, Pembobotan Kata, TF-IDF, *Information Gain* (IG), *Support Vector Machine* (SVM).

ABSTRACT

Name : Kevin Apriliom
Department : Informatics
Title : *Application of Information Gain and Methods Support Vector Machine for Identification of Abusive Content on Tweets*

In the use of social networks, especially Twitter, they often find posts that contain harsh and offensive words in Indonesian. This study will test the accuracy of Information Gain, the TF-IDF method and test the parameters on the SVM method to find out the best combination and parameters to find out. This study has several main indicators, namely using stemming or not using stemming when preprocessing the type of feature capture, parameters C and gamma, and the type of kernel. The data used is obtained from the website of the dataset provider, in the form of a Twitter data collection containing harsh words. Based on the test, the highest accuracy result was obtained with a value of 93.14% with f-measure 90.78% the results obtained using the Twitter dataset that were not manually labeled the method used was Information Gain with stemming . From this research, it can be said that using stemming and feature retrieval can affect the accuracy value in detecting the abusive words. Factors such as kernel and SVM parameters are also one of the things that affect the results of the detection of these abusive words.

Keyword: *Identification of Abusive Word, Word Weight, TF-IDF, Information Gain (IG), Support Vector Machine (SVM).*

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan yang Maha Esa karena atas rahmat dan bimbingan-Nya penulis dapat menyelesaikan Tugas Akhir yang berjudul "PENERAPAN METODE *INFORMATION GAIN* DAN *SUPPORT VECTOR MACHINE* UNTUK IDENTIFIKASI KONTEN KASAR PADA *TWEETS* ". Laporan ini disusun sebagai salah satu syarat kelulusan di Institut Teknologi Harapan Bangsa. Pada kesempatan ini penulis menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Tuhan Yang Maha Esa, karena oleh bimbingan-Nya penulis selalu mendapat pengharapan untuk menyelesaikan tugas akhir ini.
2. Ibu Ken Ratri Retno Wardani, S.Kom., M.T., selaku pembimbing I Tugas Akhir yang senantiasa memberi berbagai saran dan ilmu serta selalu memberikan dukungan kepada penulis selama proses pembuatan tugas akhir.
3. Ibu Ir. Inge Martina, M.T., selaku penguji I dalam Tugas Akhir Terima kasih untuk semua ilmu dan masukan serta dukungan yang telah diberikan kepada penulis dalam menyelesaikan Laporan Tugas Akhir
4. Seluruh dosen dan staff Departemen Teknik Informatika ITHB yang telah membantu dalam menyelesaikan Laporan Tugas Akhir ini.
5. Segenap jajaran staf dan karyawan ITHB yang turut membantu kelancaran dalam menyelesaikan Laporan Tugas Akhir ini.
6. Teman-teman yang telah menyediakan waktunya untuk memberikan dukungan, semangat, dan masukan kepada penulis dalam menyelesaikan
7. Keluarga tercinta yang selalu menyediakan waktu untuk memberikan doa, semangat dan dukungan yang tak habis-habisnya kepada penulis untuk menyelesaikan Laporan Tugas Akhir ini. Terima kasih untuk nasib, masukan, perhatian, teguran dan kasih sayang yang diberikan hingga saat ini.
8. Pihak-pihak lain yang penulis tidak bisa sebutkan satu-satu, yang selalu membantu penulis selama di ITHB.

Penulis menyadari bahwa laporan ini masih jauh dari sempurna karena keterbatasan waktu dan pengetahuan yang dimiliki oleh penulis. Oleh karena itu, kritik dan saran

untuk membangun kesempurnaan tugas akhir ini sangat diharapkan. Semoga tugas akhir ini dapat membantu pihak-pihak yang membutuhkannya.

Bandung, Juli 2022

Hormat penulis,

A handwritten signature in black ink, consisting of stylized, overlapping loops and strokes, likely representing the name Kevin Aprilion.

Kevin Aprilion

DAFTAR ISI

ABSTRAK	iv
ABSTRACT	v
KATA PENGANTAR	vi
DAFTAR ISI	viii
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN	xv
BAB 1 PENDAHULUAN	1-1
1.1 Latar Belakang	1-1
1.2 Rumusan Masalah	1-2
1.3 Tujuan Penelitian	1-2
1.4 Batasan Masalah	1-3
1.5 Kontribusi Penelitian	1-3
1.6 Metodologi Penelitian	1-3
1.7 Sistematika Pembahasan	1-4
BAB 2 LANDASAN TEORI	2-1
2.1 Tinjauan Pustaka	2-1
2.1.1 <i>Natural Language Processing</i>	2-1
2.1.2 Bahasa Kasar atau Umpatan dalam Bahasa Indonesia	2-1
2.1.3 <i>Text preprocessing</i>	2-4
2.1.3.1 <i>Case Folding</i>	2-4
2.1.3.2 <i>Remove Emoticon</i>	2-4
2.1.3.3 <i>Remove Unnecessary Word</i>	2-5
2.1.3.4 <i>Remove Non AlphaNumeric</i>	2-5
2.1.3.5 <i>Replace Slang Word</i>	2-5
2.1.3.6 <i>Stemming</i>	2-6
2.1.3.7 <i>Stop Word Removal</i>	2-6
2.1.3.8 <i>Unigram</i>	2-6

2.1.4	TF-IDF	2-6
2.1.5	<i>Information Gain</i>	2-7
2.1.6	<i>Support Vector Machine</i>	2-8
2.1.7	Evaluation Metrics	2-12
2.2	Tinjauan Studi	2-13
2.2.1	Pembahasan Penelitian Terkait	2-15
2.2.2	Pustaka Pendukung	2-16
2.2.2.1	Pustaka Pandas	2-17
2.2.2.2	Pustaka Regular <i>Expression</i> (RegEx)	2-17
2.2.2.3	Pustaka Sastrawi	2-18
2.2.2.4	Pustaka Scikit-learn	2-18
2.3	Tinjauan Objek	2-19
2.3.1	Twitter	2-19
2.3.2	Dataset	2-20
2.3.2.1	Dataset Twitter	2-20
2.3.2.2	Kamus <i>New_Kamusalay</i>	2-20
2.3.2.3	Kamus Stopwords	2-20
2.3.2.4	Kamus <i>Abusive</i>	2-20
BAB 3	ANALISIS DAN PERANCANGAN	3-1
3.1	Analisis Masalah	3-1
3.2	Kerangka Pemikiran	3-2
3.3	Analisis Urutan Proses Global	3-2
3.3.1	Preprocessing	3-3
3.3.2	Proses <i>Training</i>	3-3
3.3.3	Proses <i>Testing</i>	3-4
3.4	Data Sampling	3-5
3.4.1	Dataset Asli	3-5
3.4.2	Dataset Pelabelan Manual	3-5
3.5	Analisis Kasus	3-6
3.5.1	<i>Tweet preprocessing</i>	3-6
3.5.1.1	<i>Case folding</i>	3-6
3.5.1.2	<i>Remove Emoticon</i>	3-6
3.5.1.3	<i>Remove Unnecessary Word</i>	3-7
3.5.1.4	<i>Remove Non AlphaNumeric</i>	3-7
3.5.1.5	<i>Replace Slang Word</i>	3-7
3.5.1.6	<i>Stemming</i>	3-7
3.5.1.7	<i>Stop Word Removal</i>	3-8

3.5.2	<i>TF-IDF</i>	3-8
3.5.3	<i>Information Gain</i>	3-9
3.5.4	<i>Support Vector Machine</i>	3-10
3.5.5	<i>Evaluation Matrix</i>	3-14
BAB 4	IMPLEMENTASI DAN PENGUJIAN	4-1
4.1	Lingkungan Implementasi	4-1
4.1.1	Spesifikasi Perangkat Keras	4-1
4.1.2	Lingkungan Perangkat Lunak	4-1
4.2	Daftar <i>Class</i> dan <i>Method</i>	4-1
4.2.1	<i>Class Preprocessing</i>	4-1
4.2.2	<i>Class Training</i>	4-3
4.2.3	<i>Class Testing</i>	4-4
4.2.4	Tampilan antarmuka aplikasi	4-4
4.3	Implementasi Perangkat Lunak	4-5
4.3.1	Pengambilan Data	4-5
4.3.2	<i>Tweet Preprocessing</i>	4-5
4.3.3	<i>TF-IDF</i>	4-6
4.3.4	<i>Information Gain</i>	4-7
4.3.5	<i>Training Data</i>	4-7
4.3.6	<i>Testing Data</i>	4-7
4.3.7	Skenario Uji	4-8
4.4	Pengujian	4-9
4.4.1	Ekstraksi Fitur <i>TF-IDF</i>	4-10
4.4.1.1	Dataset Asli Menggunakan <i>Stemming</i>	4-10
4.4.1.2	Dataset Asli Tidak Menggunakan <i>Stemming</i>	4-12
4.4.1.3	Dataset Pelabelan Manual Menggunakan <i>Stemming</i>	4-13
4.4.1.4	Dataset Pelabelan Manual Tidak Menggunakan <i>Stemming</i>	4-15
4.4.2	Seleksi Fitur <i>Information Gain</i>	4-16
4.4.2.1	Dataset Asli Menggunakan <i>Stemming</i>	4-16
4.4.2.2	Dataset Asli Tidak Menggunakan <i>Stemming</i>	4-18
4.4.2.3	Dataset Pelabelan Manual Menggunakan <i>Stemming</i>	4-19
4.4.2.4	Dataset Pelabelan Manual Tidak Menggunakan <i>Stemming</i>	4-21
4.4.3	Analisis Indikator	4-22
4.4.3.1	Menggunakan <i>Stemming</i> dan Tidak Menggunakan <i>Stemming</i>	4-23

4.4.3.2	Seleksi Fitur <i>Information Gain</i> dan Ekstraksi Fitur TF-IDF	4-23
4.4.3.3	Tipe <i>Kernel</i> , Parameter C dan <i>Gamma</i> di SVM . . .	4-24
4.4.4	Analisis <i>library</i> Sastrawi	4-24
4.5	Analisis Kesalahan	4-25
4.5.1	Analisis Kesalahan <i>Preprocessing</i>	4-25
4.5.2	Analisis Pemilihan Tipe <i>Kernel</i> dan Akurasi Terkecil . . .	4-26
4.5.3	Analisis Kesamaan Akurasi yang Identik	4-26
BAB 5	KESIMPULAN DAN SARAN	5-1
5.1	Kesimpulan	5-1
5.2	Saran	5-1
BAB A	DAFTAR PELABELAN MANUAL	A-1
BAB B	PENGUJIAN SVM DENGAN TF-IDF	B-14
BAB C	PENGUJIAN SVM DENGAN IG	C-21

DAFTAR TABEL

2.1	Tabel <i>Confusion Matrix</i> untuk <i>Information gain</i>	2-7
2.2	Persamaan <i>kernel Support Vector Machine</i>	2-10
2.3	<i>State of the Art</i>	2-14
2.4	Tabel pustaka Pandas	2-17
2.5	Pustaka Regular Expression	2-18
2.6	Pustaka sastrawi	2-18
2.7	Pustaka scikit-learn	2-19
3.1	Asumsi kemunculan kata	3-9
3.2	Contoh perhitungan bobot TF, IDF pada data latih	3-9
3.3	Contoh relasi kata takut antara <i>tweet</i>	3-10
3.4	Tabel perhitungan <i>Confusion Matrix</i> untuk <i>Information Gain</i>	3-10
3.5	Data pelatihan	3-11
3.6	Skenario pengujian parameter dan <i>kernel SVM</i>	3-12
3.7	Hasil persamaan linear	3-13
3.8	Data uji	3-13
3.9	Hasil evaluasi matrik	3-14
4.1	Daftar attribute <i>class preprocessing</i>	4-2
4.2	Daftar <i>method class preprocessing</i>	4-2
4.3	Daftar attribute <i>class training</i>	4-3
4.4	Daftar <i>method class training</i>	4-3
4.5	Daftar attribute <i>class training</i>	4-4
4.6	Daftar <i>method class testing</i>	4-4
4.7	Jumlah dataset yang dipakai	4-8
4.8	Skenario pengujian parameter dan <i>kernel SVM</i>	4-8
4.9	Hasil pengujian skenario menggunakan TF-IDF	4-22
4.10	Hasil pengujian skenario menggunakan <i>Information Gain</i>	4-23
A-1	Daftar Perubahan Dataset	A-1
B-1	Hasil Pengujian TF-IDF dengan Dataset Asli Menggunakan Stemming	B-14
B-1	Hasil Pengujian TF-IDF dengan Dataset Asli Menggunakan Stemming	B-15

B-2	Hasil Pengujian TF-IDF dengan Dataset Asli Menggunakan Tidak Stemming	.B-15
B-2	Hasil Pengujian TF-IDF dengan Dataset Asli Menggunakan Tidak Stemming	.B-16
B-2	Hasil Pengujian TF-IDF dengan Dataset Asli Menggunakan Tidak Stemming	.B-17
B-3	Hasil Pengujian TF-IDF dengan Dataset Pelabelan Manual Menggunakan Stemming	.B-17
B-3	Hasil Pengujian TF-IDF dengan Dataset Pelabelan Manual Menggunakan Stemming	.B-18
B-4	Hasil Pengujian TF-IDF dengan Dataset Pelabelan Manual Tidak Menggunakan Stemming	.B-18
B-4	Hasil Pengujian TF-IDF dengan Dataset Pelabelan Manual Tidak Menggunakan Stemming	.B-19
B-4	Hasil Pengujian TF-IDF dengan Dataset Pelabelan Manual Tidak Menggunakan Stemming	.B-20
C-1	Hasil Pengujian IG dengan Dataset Asli Menggunakan Stemming .	.C-21
C-1	Hasil Pengujian IG dengan Dataset Asli Menggunakan Stemming .	.C-22
C-2	Hasil Pengujian IG dengan Dataset Asli Tidak Menggunakan Stemming	.C-22
C-2	Hasil Pengujian IG dengan Dataset Asli Tidak Menggunakan Stemming	.C-23
C-2	Hasil Pengujian IG dengan Dataset Asli Tidak Menggunakan Stemming	.C-24
C-3	Hasil Pengujian IG dengan Dataset Pelabelan Manual Menggunakan Stemming	.C-24
C-3	Hasil Pengujian IG dengan Dataset Pelabelan Manual Menggunakan Stemming	.C-25
C-4	Hasil Pengujian IG dengan Dataset Pelabelan Manual Tidak Menggunakan Stemming	.C-25
C-4	Hasil Pengujian IG dengan Dataset Pelabelan Manual Tidak Menggunakan Stemming	.C-26
C-4	Hasil Pengujian IG dengan Dataset Pelabelan Manual Tidak Menggunakan Stemming	.C-27

DAFTAR GAMBAR

2.1	Daftar Emoji [5]	2-5
2.2	<i>Hyperlane Class -1 dan +1</i> [9]	2-8
2.3	Ilustrasi <i>kernel</i> linear	2-10
2.4	Ilustrasi <i>kernel</i> Polynomial	2-11
2.5	Ilustrasi <i>kernel</i> RBF	2-11
2.6	Ilustrasi <i>kernel sigmoid</i>	2-12
2.7	<i>Confusion matrix</i>	2-12
3.1	Kerangka pemikiran	3-2
3.2	<i>Flowchart preprocessing</i>	3-3
3.3	<i>Flowchart training</i>	3-4
3.4	<i>Flowchart testing</i>	3-4
3.5	Contoh data sampling	3-5
3.6	Data masukan	3-11
4.1	Tampilan antarmuka awal	4-4
4.2	Tampilan antarmuka akhir	4-5
4.3	Hasil Akurasi Dataset Asli Menggunakan <i>Stemming</i> Pengujian <i>kernel</i> dan Parameter tanpa menggunakan <i>Information Gain</i>	4-10
4.4	Hasil <i>f-measure</i> Dataset Asli Menggunakan <i>Stemming</i> Pengujian <i>kernel</i> dan Parameter tanpa menggunakan <i>Information Gain</i>	4-11
4.5	Hasil Akurasi Dataset Asli Tidak Menggunakan <i>Stemming</i> Pengujian <i>kernel</i> dan Parameter tanpa menggunakan <i>Information</i> <i>Gain</i>	4-12
4.6	Hasil <i>f-measure</i> Dataset Asli Tidak Menggunakan <i>Stemming</i> Pengujian <i>Kernel</i> dan Parameter tanpa menggunakan <i>Information</i> <i>Gain</i>	4-12
4.7	Hasil Akurasi Dataset Pelabelan Manual Menggunakan <i>Stemming</i> Pengujian <i>Kernel</i> dan Parameter tanpa menggunakan <i>Information</i> <i>Gain</i>	4-13
4.8	Hasil <i>f-measure</i> Dataset Pelabelan Manual Menggunakan <i>Stemming</i> Pengujian <i>Kernel</i> dan Parameter tanpa menggunakan <i>Information</i> <i>Gain</i>	4-14
4.9	Hasil Akurasi Dataset Pelabelan Manual Tidak Menggunakan <i>Stemming</i> Pengujian <i>Kernel</i> dan Parameter tanpa menggunakan <i>Information Gain</i>	4-15

4.10	Hasil <i>f-measure</i> Dataset Pelabelan Manual Tidak Menggunakan <i>Stemming</i> Pengujian <i>Kernel</i> dan Parameter tanpa menggunakan <i>Information Gain</i>	4-15
4.11	Hasil Akurasi Dataset Asli Menggunakan <i>Stemming</i> Pengujian <i>Kernel</i> dan Parameter menggunakan <i>Information Gain</i>	4-16
4.12	Hasil <i>f-measure</i> Dataset Asli Menggunakan <i>Stemming</i> Pengujian <i>Kernel</i> dan Parameter menggunakan <i>Information Gain</i>	4-17
4.13	Hasil Akurasi Dataset Asli Tidak Menggunakan <i>Stemming</i> Pengujian <i>Kernel</i> dan Parameter menggunakan <i>Information Gain</i> . .	4-18
4.14	Hasil <i>f-measure</i> Dataset Asli Tidak Menggunakan <i>Stemming</i> Pengujian <i>Kernel</i> dan Parameter menggunakan <i>Information Gain</i> . .	4-18
4.15	Hasil Akurasi Dataset Pelabelan Manual Menggunakan <i>Stemming</i> Pengujian <i>Kernel</i> dan Parameter menggunakan <i>Information Gain</i> . .	4-19
4.16	Hasil <i>f-measure</i> Dataset Pelabelan Manual Menggunakan <i>Stemming</i> Pengujian <i>Kernel</i> dan Parameter menggunakan <i>Information Gain</i> . .	4-20
4.17	Hasil Akurasi Dataset Pelabelan Manual Tidak Menggunakan <i>Stemming</i> Pengujian <i>Kernel</i> dan Parameter menggunakan <i>Information Gain</i>	4-21
4.18	Hasil <i>f-measure</i> Dataset Pelabelan Manual Tidak Menggunakan <i>Stemming</i> Pengujian <i>Kernel</i> dan Parameter menggunakan <i>Information Gain</i>	4-21
4.19	Contoh data <i>tweet</i> yang gagal di <i>preprocessing</i>	4-26

DAFTAR LAMPIRAN

LAMPIRAN A	A-1
LAMPIRAN B	B-14
LAMPIRAN C	C-21

DAFTAR REFERENSI

- [1] Agus Tri Haryanto, "Pengguna Aktif Medsos RI 170 Juta, Bisa Main 3 Jam Sehari", Detik , 2021. [Online]. Available: <https://inet.detik.com/cyberlife/d-5407834/pengguna-aktif-medsos-ri-170-juta-bisa-main-3-jam-sehari>. [Accessed: 05 April 2021].
- [2] R F Hidayattullah, A A F Yusuf, K P Juwairi, R A N Nayoan, "Identifikasi Konten Kasar pada *Tweet* Bahasa Indonesia", Journal Komputasional (JLK), Vol. 2, No. 1, Maret 2019.
- [3] R Kusumawati, A D 'arofah and P A Pramana, "*Comparison Performance of Naive Bayes Classifier and Support Vector Machine Algorithm for Twitter's Classification of Tokopedia Services*", *Journal of Physics: Conference Series*, 1320 (2019) 012016 doi:10.1088/1742-6596/1320/1/012016.
- [4] Xiaoamei Zou, Jing Yang. 2017 "*Microblog Sentiment Analysis Using Social and Topic Context*". PLoS One 13(2):e0191163, <https://doi.org/10.1371/nejournal.po.0191163>.
- [5] Tim Whitlock, "Emoji Unicode Tables, timwhitlock , 2021. [Online]. Available: <https://apps.timwhitlock.info/emoji/tables/unicode>. [Accessed: 24 April 2022].
- [6] Muhammad Okky Ibrohim and Indra Budi, "*Indonesian Abusive and Hate Speech Twitter Text*", Kaggle , 2019. [Online]. Available: <https://www.kaggle.com/datasets/c7af46bd64825107fff61076eeec0df32f905f0e9b7f43b7c361b> [Accessed: 05 April 2021].
- [7] Oswin Rahadiyan Hartono, "*Indonesian Stoplist*", Kaggle , 2019. [Online]. Available: <https://www.kaggle.com/datasets/c7af46bd64825107fff61076eeec0df32f905f0e9b7f43b7c361b> [Accessed: 05 April 2021].
- [8] Reza Maulana, Panny Agustia R., Windi Irmayani, Dedi Saputra, Wanty Eka Jayanti, "*Improved Accuracy of Sentiment Analysis Movie Review Using Support Vector Machine Based Information Gain*". *Journal of Physics: Conference Series*, 1641 (2020) 012060 doi:10.1088/1742-6596/1641/1/012060.

- [9] Oman Somantril, Dyah Apriliani , ”*Support Vector Machine Berbasis Feature Selection untuk Sentiment Analysis Kepuasan Pelanggan Terhadap Pelayanan Warung Dan Restoran Kuliner Kota Tegal*”. Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), Vol.5, No.5 Oktober 2018, hlm.537-548 p-ISSN: 2355-7699.
- [10] Christevan Destitus, Wella, Suryasari, ”*Support Vector Machine VS Information Gain: Analisis Sentimen Cyberbullying Twitter Indonesia*”. ULTIMA InfoSys, Vol. XI, No. 2 Desember 2020, ISSN 2085-4579.
- [11] Gerardus Blokdyk, *Information gain ratio Standard Requirements*, 2022.
- [12] L. Derczynski, ”*Complementarity, F-score, and NLP Evaluation*”, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’ 16)*, pp. 261-266, 2016.
- [13] M. O. Ibrohim and I. Budi, ”*Multi-label Hate Speech and Abusive Language Detection in Indonesia Twitter*”, ALW3: 3rd Workshop on Abusive Language Online, 46-57. 2019.
- [14] Tom M. Mitchell, *Machine Learning (McGraw-Hill International Editions Computer Science Series)*, 1997.
- [15] D. Jurafsky and J. H. Martin. *Speech and Language Processing. Prentice Hall, 2nd edition*, 2008.
- [16] C. Stocan and R. Stoean, *Support vector machines and evolutionary algorithms for classification*. Cham: Springer, 2014.
- [17] C. Manning, P. Raghavan and H. Schutze, *Introduction to information retrieval*. Cambridge: Cambridge University Press, 200.
- [18] H. Lin and C. Lin, ”*A Study on Sigmoid Kernels for SVM and the Training of non-PSD Kernels by SMO-type Methods*”, National Taiwan University, 2003.
- [19] M. S. Kaari. ”*Sentiment analysis for hate speech detection on social media: TF-IDF weighted N-Grams based approach*”. (Thesis). Strathmore University. 2017. Retrieved from <http://suplus.strathmore.edu/handle/11071/5657>.
- [20] Alan Davies, *An Introduction to Applied Linguistics: From Practice to Theory, 2nd edition*, 2007.

- [21] P. Jackson, and I. Moulinier. *Natural Language Processing for Online Applications Text Retrieval, Extraction and Categorization*, 2004.
- [22] H. Lane, C. Howard ,and H. M. Hapke. *Natural Language Processing in Action*, Manning Publications Co., 2019.
- [23] Fauzi Ihsan, Iwan Iskandar, N. S. Harahap, Surya Agustian, "Algoritme decision tree untuk mendeteksi ujaran kebencian dan bahasa kasar multilabel pada Twitter berbahasa Indonesia ", Jurnal Teknologi dan Sistem Komputer, ISSN: 2338-0403.
- [24] Faisal Rahutomo , Pramana Yoga Saputra, Miftahul Agtamas Fidyawan, "IMPLEMENTASI TWITTER SENTIMENT ANALYSIS UNTUK REVIEW FILM MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE", Jurnal Informatika Polinema, ISSN: 2407-070X.
- [25] Paul Pu Liang , Terrance Liu, Anna Cai , Michal Muszynski1 , Ryo Ishii, Nicholas Allen, Randy Auerbach, David Brent, Ruslan Salakhutdinov , Louis-Philippe Morency, "*Learning Language and Multimodal Privacy-Preserving Markers of Mood from Mobile Data*", *59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, ACL-IJCNLP 2021.