Jurnal Komputer dan Informatika
Vol 19 No 1, April 2024: hlm 64-72

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

KLASIFIKASI *TWEET* YANG MENGANDUNG UJARAN KEBENCIAN DENGAN *XGBOOST* DAN *LOGISTIC REGRESSION*

Dhiwa Aqsha
Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara
Jl. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia
e-mail:*dhiwa.535210087@stu.untar.ac.id*

ABSTRAK

Saat ini, masyarakat menggunakan media sosial sebagai platform untuk mengungkapkan pendapat mereka. Ada berbagai metode yang dapat digunakan untuk menyuarakan pandangan, baik yang bersifat positif maupun negatif.Jumlah pengguna media sosial yang tinggi juga memberikan peluang lebih besar untuk munculnya konten yang mengandung ujaran kebencian, termasuk di platform seperti Twitter. Twitter adalah platform media sosial yang memfasilitasi pengguna untuk menyampaikan perasaan dan pendapat melalui cuitan, termasuk cuitan yang berpotensi mengandung ujaran kebencian. Ujaran kebencian merupakan tindakan komunikasi yang mencakup provokasi, hasutan, atau hinaan terhadap individu atau kelompok berdasarkan faktor-faktor seperti suku, agama, ras, kewarganegaraan, dan lainnya. Untuk memperoleh informasi dan mengklasifikasikan teks, diperlukan analisis sentimen. Dalam konteks penelitian ini, analisis sentimen merupakan suatu proses klasifikasi dokumen teks ke dalam dua kelas, yakni kelas sentimen negatif dan positif.Dalam studi ini, kami membandingkan dua metode klasifikasi yang berbeda, yaitu *Logistic Regression* dan *Extreme Gradient Boosting* (XGBoost). Penelitian ini menggunakan data latih sebanyak 31962 data dan data uji sebanyak 17197 data. Penelitian ini berhasil mendapatkan model Logistic Regression terbaik dengan angka akurasi sebesar 95.74%. Sementara XGBoost menunjukkan hasil akurasi yang tidak kalah tinggi sebesar 94.97%. Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma logistic regression merupakan metode yang paling efektif untuk melakukan klasifikasi ujaran kebencian pada teks Twitter. Hal ini dibuktikan dengan hasil akurasi yang diperoleh, yaitu sebesar 95.74%.

Kata kunci: Akurasi, Ujaran Kebencian, Twitter, *Logistic Regression, Extreme Gradient Boosting*

ABSTRACT

Nowadays, people use social media as a platform to express their opinions. There are various methods that can be used to voice views, both positive and negative. The high number of social media users also provides greater opportunities for content containing hate speech to appear, including on platforms such as Twitter. Twitter is a social media platform that facilitates users to convey feelings and opinions through tweets, including tweets that have the potential to contain hate speech. Hate speech is an act of communication that includes provocation, incitement, or insults against individuals or groups based on factors such as ethnicity, religion, race, nationality, and others. To obtain information and classify text, sentiment analysis is needed. In the context of this research, sentiment analysis is a process of classifying text documents into two classes, namely negative and positive sentiment classes. In this study, we compare two different classification methods, namely Logistic Regression and Extreme Gradient Boosting (XGBoost). This research used 31962 training data and 17197 test data. This research succeeded in obtaining the best Logistic Regression model with an accuracy rate of 95.74%. Meanwhile, XGBoost shows no less high accuracy results of 94.97%. Based on the results of the research that has been carried out, it can be concluded that the logistic regression algorithm is the most effective method for classifying hate speech in Twitter text. This is proven by the accuracy results obtained, namely 95.74%.

Keywords: *Accuracy, Hate Speech, Twitter, Logistic Regression, Extreme Gradient Boosting*

1 PENDAHULUAN

Pada masa sekarang, penggunaan social media telah menjadi bagian tak terpisahkan dari aktivitas harian kita. Media sosial bukan hanya merupakan alat komunikasi, tetapi juga sarana untuk berbagi informasi dan menyuarakan pendapat. Dampak dari penggunaan media sosial dapat beragam, membawa pengaruh negatif maupun positif yang mungkin berdampak pada kehidupan seseorang[1].

Twitter ialah salah satu platform media sosial yang sangat terkenal dan digunakan secara luas. Sekitar 19,5 juta penduduk Indonesia menggunakan Twitter, menjadikan Indonesia sebagai negara kelima dengan jumlah pengguna Twitter terbanyak dan aktif[2]. Karena media sosial beroperasi sebagai saluran komunikasi publik yang terbuka dan transparan, sifatnya dapat memicu masyarakat untuk cenderung mengeluarkan pernyataan kebencian di bagian komentar media sosial[3].

Ujaran kebencian merujuk pada penggunaan bahasa untuk menyatakan perasaan kebencian terhadap suatu kelompok, dengan tujuan merendahkan, menghina, atau mempermalukan anggota kelompok tersebut. Ujaran kebencian seringkali kita dapatkan dalam keseharian kita, baik saat berinteraksi langsung dengan orang yang berada di sekitar kita ataupun ketika berselancar di dunia maya, terlebih di social media. Ujaran kebencian memiliki potensi untuk menciptakan perpecahan di antara kelompok-kelompok dan mengancam persatuan bangsa Indonesia. Oleh karena itu, penanganan yang cepat diperlukan sebelum masalah ini berkembang menjadi lebih luas dan signifikan. Klasifikasi ujaran kebencian menjadi suatu kebutuhan untuk mengurangi penyebaran pesan-pesan berisi ujaran kebencian, terutama di platform media sosial seperti Twitter. Proses klasifikasi ujaran kebencian secara manual mengharuskan penggunaan waktu dan sumber daya yang besar. Oleh karena itu, penting untuk mengimplementasikan klasifikasi secara otomatis agar proses tersebut menjadi lebih efisien. Dengan adanya klasifikasi otomatis, diharapkan dapat mempermudah proses identifikasi ujaran kebencian, dan solusi ini dapat diadopsi oleh pihak-pihak yang memerlukan. Dengan melibatkan sejumlah besar tweet yang mengandung ujaran kebencian, penelitian ini akan menggunakan data tersebut sebagai sumber informasi utama. Data tersebut akan diolah secara cermat untuk mencapai tujuan penelitian dengan hasil yang diinginkan.

Machine learning adalah salah satu bagian dari Artificial Intelligence yang berfokus pada pengembangan sistem yang dapat terus belajar dari data dan meningkatkan akurasinya seiring waktu. Salah satu sub-bidang dari machine learning, yaitu text analytics, melibatkan penggunaan algoritmaalgoritma yang mampu mengenali atau mengelompokkan objek-objek teks. Teknik text analytics ini dapat digunakan untuk mengatasi permasalahan ujaran kebencian di media sosial dengan kemampuannya dalam mendeteksi *cyber bullying*, penggunaan bahasa kasar, dan cyber hate [4].

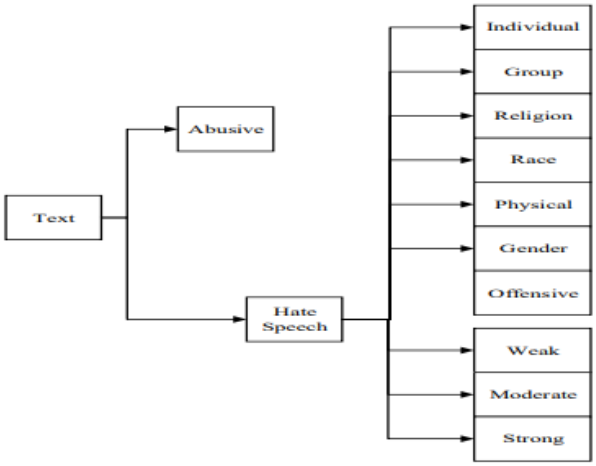
Peneliti akan membandingkan dua algoritma pengajaran mesin, *Logistic Regression* dan XGBoost, berdasarkan masalah yang ada untuk melakukan klasifikasi ujaran kebencian pada cuitan Twitter. Selain itu, tujuan penelitian ini adalah untuk menguji seberapa baik kedua model tersebut mendeteksi ujaran kebencian. Diharapkan bahwa penelitian ini akan menghasilkan model terbaik untuk tugas klasifikasi. Diharapkan model ini dapat digunakan sebagai referensi untuk penelitian penelitian lanjutan.

2 TINJAUAN LITERATUR

Dari penjelasan tersebut, penelitian ini bertujuan untuk mengembangkan model Machine Learning dengan menerapkan metode Logistic Regression dan XGBoost untuk mengklasifikasikan ujaran kebencian dalam teks yang berasal dari Twitter. Harapannya, penelitian ini dapat menghasilkan model terbaik dengan kinerja optimal dalam tugas klasifikasi, yang nantinya dapat menjadi landasan untuk pengembangan lebih lanjut dalam penelitian-penelitian berikutnya.

Dengan menggunakan algoritma *Support Vector Machine* (SVM), penelitian sebelumnya telah membahas ujaran kebencian dalam bentuk multilabel pada teks Twitter [9]. Dalam proses pengujian, penelitian ini menggunakan desain Klasifikasi *Hierarchical Multi-Label* (HMC). Tujuan implementasi HMC adalah untuk mengevaluasi kombinasi model yang paling akurat dengan menggunakan lima skenario labeling yang dibangun berdasarkan hierarki yang dihasilkan dari label label yang memiliki karakteristik yang sebanding. Gambar 1 menunjukkan detail hierarki label pada dataset yang digunakan.

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)



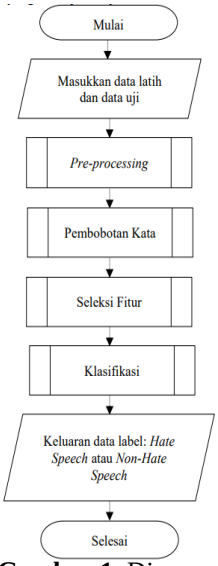
Gambar 1. Hirarki label pada dataset multilabel ujaran kebencian bahasa Indonesia

Mengurangi jumlah label dari 12 menjadi 9, penelitian yang dijelaskan dalam literatur [5] mencapai tingkat akurasi tertinggi sebesar 68,43%. Label-label yang digunakan termasuk "*Religion*", "*Abusive*", "*Individual*", "*Group*", "*Hate Speech*", "*Race*", "*Gender*", "*Offensive*" dan "*Physical*". Sementara itu, label-label yang dikurangi termasuk "*Strong*", "*Weak*", dan "*Moderate*".

Penelitian lain yang membandingkan beberapa metode dalam klasifikasi teks [6] menunjukkan bahwa *Support Vector Machine* (SVM) adalah metode yang efektif untuk klasifikasi teks. Hasil uji coba dari metode SVM pada klasifikasi teks yang terbagi menjadi empat kategori menunjukkan nilai ratarata tingkat akurasi sebesar 90,72%, nilai F1 sebesar 88,97%, dan tingkat recall sebesar 86,37%. Dengan nilai-nilai tersebut, SVM dianggap memiliki kestabilan performa yang lebih baik jika dibandingkan dengan metode-metode lainnya[6]. Kelebihan dari algoritma ini termasuk kemampuannya dalam menangani volume data yang besar, terutama dalam konteks klasifikasi teks.

3 METODE PENELITIAN

Gambar 2 menunjukkan alur tahapan penyelesaian masalah penelitian dalam diagram alir sistem. Data uji dan data latihan dimasukkan dalam tahap awal, yang akan digunakan oleh sistem untuk proses klasifikasi berikutnya. Penulis menggunakan *dataset* yang diambil dari https://www.kaggle.com/datasets/arkhoshghalb/twitter-sentiment-analysis-hatred-speech dengan j



Gambar 1 Diagram Alir

Judul dataset “Twitter Sentiment Analysis”. Dataset tersebut terbagi menjadi dataset latih dan dataset uji. Dataset latih berjumlah 31962 sampel dan data uji berjumlah 17197 sampel.

	id	label	tweet
0	1	0	@user when a father is dysfunctional and is s...
1	2	0	@user @user thanks for #lyft credit i can't us...
2	3	0	bihday your majesty
3	4	0	#model i love u take with u all the time in ...
4	5	0	factsguide: society now #motivation
...	...	...	...
31957	31958	0	ate @user isz that youuu?ð□□□ð□□□ð□□□ð□□□ð□□□ð...
31958	31959	0	to see nina turner on the airwaves trying to...
31959	31960	0	listening to sad songs on a monday morning otw...
31960	31961	1	@user #sikh #temple vandalised in in #calgary,...
31961	31962	0	thank you @user for you follow

Gambar 2 Format Dataset

Variabel dependen pada penelitian ini yaitu Kolom "label" yang berisi data mengenai ujaran kebencian. Dalam konteks ini dinyatakan sebagai "1" jika terdapat ujaran kebencian yang terdeteksi dan "0" jika tidak ada ujaran kebencian yang terdeteksi pada tweet tertentu berdasarkan data yang tercatat dalam dataset.

Setelah tahap penginputan data uji dan data latih, langkah berikutnya adalah melakukan pre-processing terhadap data tersebut. *Pre-processing teks* adalah langkah awal sebelum data diolah lebih lanjut, dan hasilnya akan digunakan dalam proses pembuatan sistem. Tahap pre-processing teks melibatkan serangkaian langkah untuk mengubah data agar lebih mudah diproses oleh sistem, dan juga untuk menghilangkan data yang tidak relevan. Proses ini sangat penting dalam analisis teks, terutama untuk media sosial yang cenderung memiliki teks tidak formal, tidak terstruktur, dan tingkat kebisingan yang tinggi (Mujilahwati, 2016).Tahapan pre-processing teks yang akan digunakan melibatkan *Cleaning*, *Case Folding*, Tokenisasi, Penghapusan Stopword, dan Stemming. Setelah melalui tahapan ini, data akan siap untuk digunakan dalam proses seleksi fitur dan klasifikasi, dengan keluaran berupa label data: *Hate Speech* atau *Non-Hate Speech*.

Langkah berikutnya adalah pembobotan kata. Proses pembobotan kata dilakukan dengan tujuan untuk mengidentifikasi jumlah kemunculan kata-kata yang telah diperoleh dari proses pre-processing pada dokumen yang dimiliki. Dengan pembobotan kata, kita dapat menilai seberapa sering kata-kata tertentu muncul dalam data. Setelah pembobotan kata, dilakukan seleksi fitur pada fitur-fitur yang telah diperoleh dari proses sebelumnya. Tujuannya adalah untuk mendapatkan fitur yang memiliki pengaruh signifikan pada proses klasifikasi yang akan dijalankan. Proses klasifikasi adalah tahapan akhir dari penelitian ini, di mana data uji dimasukkan untuk memperoleh hasil klasifikasi kelas. Hasil akhir dari proses ini berupa label kelas, yakni hate speech atau non-hate speech, berdasarkan karakteristik data tersebut. Keseluruhan tahapan dianggap selesai setelah diperoleh keluaran berupa label kelas tersebut.

Untuk membangun sistem klasifikasi ujaran kebencian dengan menggunakan metode *Logistic Regression* dan *Extreme Gradient Boosting* dengan seleksi fitur, diperlukan dasar teori yang kuat sebagai landasan bagi penelitian ini. Integrasi teori-teori ini akan membantu memperkuat dasar penelitian dan memberikan kerangka kerja yang kokoh untuk pengembangan sistem klasifikasi.

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

3.1 *Logistic Regression*

Algoritma *Logistic Regression* merupakan algoritma machine learning untuk melakukan klasifikasi yang digunakan untuk memprediksi hasil biner, seperti 1(ya) atau 2(tidak), berdasarkan pengamatan sebelumnya dari dataset. Model *logistic regression* memprediksi variabel data dependen melalui analisis korelasi antara satu atau lebih variabel independen yang ada. Algoritma ini digunakan untuk pengolahan data dan dapat diartikan sebagai urutan logis dalam pengambilan keputusan untuk memecahkan masalah. Algoritma Logistic Regression merupakan salah satu jenis algoritma supervised learning yang digunakan dalam berbagai konteks, seperti klasifikasi dan regresi. Beberapa contoh penggunaan *Logistic Regression* adalah dalam analisis keputusan investasi, analisis masalah bank yang tidak membayar pinjaman, dan analisis keputusan lainnya. Algoritma *Logistic Regression* juga dapat digunakan dalam memprediksi risiko diabetes dan dalam deteksi kebakaran hutan dan asap [7].

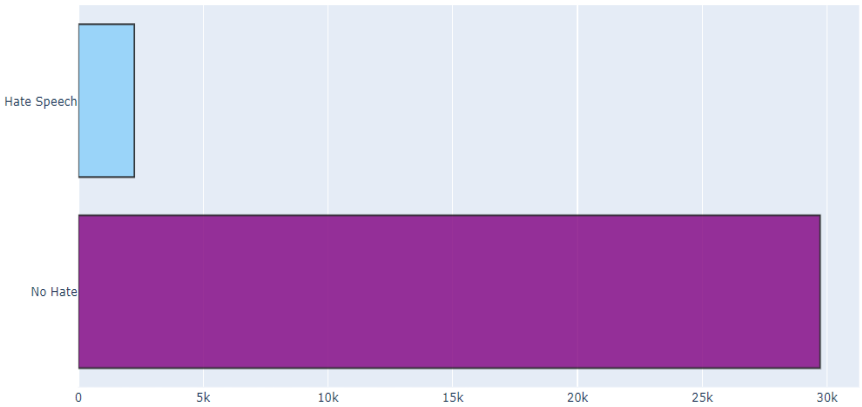
3.2 XGBoost

Algoritma XGBoost (*Extreme Gradient Boosting*) merupakan salah satu algoritma machine learning berdasarkan pohon keputusan sebagai *classifier* dengan *Gradient Boosting* sebagai intinya. Perbedaan antara *Gradient Boosting* dengan XGBoost, tidak seperti *Gradient Boosting*, proses penambahan "*weak learner*"pada XGBoost tidak terjadi secara berurutan, tetapi secara multi-threaded. Dalam hal ini, pemanfaatan inti CPU digunakan secara tepat (*efisien*), sehingga kecepatan dan kinerja algoritma akan lebih baik [8].

XGBoost merupakan suatu implementasi dari *Gradient Boosting Method* (GBM) yang lebih efisien dan scalable. Keunggulan utama XGBoost terletak pada kemampuannya menangani berbagai fungsi seperti regresi, klasifikasi, dan ranking dengan tingkat efisiensi yang tinggi. Pencapaian tersebut membuat *Extreme Gradient Boosting* menjadi pilihan yang sangat populer dalam berbagai kompetisi machine learning. *Extreme Gradient Boosting* pertama kali diperkenalkan pada *Higgs Boson Competition*, di mana metode ini adalah metode yang paling banyak digunakan oleh pesertacompetition. Keberhasilan ini melanjutkan popularitasnya, dan *Extreme Gradient Boosting* juga merupakan metode yang dominan pada kompetisi machine learning yang diadakan oleh Kaggle pada tahun 2015. Kombinasi performa tinggi dan fleksibilitasnya menjadikan XGBoost sebagai alat yang efektif dalam berbagai tantangan analisis data.

4 HASIL DAN PEMBAHASAN

Pada bagian ini kami akan membahas hasil dari penelitian mengenai pendeteksi asap menggunakan metode *Logistic Regression, dan Extreme Gradient Boosting* (*XGBoost).* Seperti dijelaskan pada bagian di atas jumlah sampel data yang digunakan sebagai bahan Analisa sebanyak 31,962. Pada penelitian kali ini setelah pengidentifikasian maka didapatkanlah hasil pada Gambar 3.



Gambar 3. Jumlah *hate speech* yang terdeteksi

Terlihat dari Gambar 3 bahwa dari 31,962 data dari kelas "label" didapatkan hasil sebanyak 29,720 sampel data memenuhi kondisi 0 yang berarti tidak mendeteksi adanya ujaran kebencian pada tweet, sedangkan sebanyak 2,242 sampel data memenuhi kondisi 1 yang berarti mendeteksi terdapat ujaran kebencian pada tweet tersebut

4.1 Pengukuran Akurasi dengan *Confusion Matrix*

Di mana kinerja klasifikasi dipengaruhi oleh akurasi klasifikasi, metode klasifikasi *Confusion Matrix* didasarkan pada hasil klasifikasi sebelumnya. *Matrix confusion* memberikan informasi yang membandingkan hasil klasifikasi yang dibuat oleh sistem (model) dengan hasil klasifikasi yang sebenarnya. Pentingnya confusion matrix adalah bahwa itu akan menunjukkan seberapa baik model yang telah dibuat sebelumnya melalui pengukuran akurasi sebelumnya untuk menentukan seberapa akurat model tersebut. Kinerja model klasifikasi pada set data uji yang nilai sebenarnya diketahui digambarkan dengan matriks kekacauan. Untuk menghitung ketepatan, *Confusion Matrix* digunakan. Tabel 1 Menampilkan Matriks Konflik [9].

Table 1 *Confussion Matrix*

Kelas	Terklarifikasi Positif	Terklarifikasi Negatif
Positif	TP(True Positive)	FN(False Negative)
Negatif	FP(False Positive)	TN(True Negative)

Confusion Matrix dapat dievaluasi kinerjanya dengan menggunakan nilai FP (*False Positive*), TP (*True Positive*), TN (*True Negative*), dan FN (*False Negative*), *Confusion Matrix* dapat dievaluasi kinerjanya. *True Positive* adalah data positif yang berhasil diprediksi dengan benar sebagai positif; *True Negative* adalah data negatif yang salah diprediksi sebagai positif; dan *False Negative* adalah data positif yang salah diprediksi sebagai negatif. Menghitung akurasi dapat menggunakan persamaan:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

4.1.1 Hasil Akurasi Algoritma *Logistic Regression*

Klasifikasi data dengan algoritma *Logistic Regression* menghasilkan data yang disajikan dalam Tabel 3 *Confussion Matrix Logistic Regression*

Table 2. *Confussion Matrix Logistic Regression*

	True 0	True 1	class recall
pred. 0	5892	23	99%
pred. 1	249	229	26%
Class Precision	94%	96%	

Akurasi mencapai 95.74%, dengan presisi kelas nol (prediksi negatif) sebesar 99% dan presisi kelas satu (prediksi positif) mencapai 26%. Akurasi dihitung menggunakan persamaan 1, dengan nilai *true positive* sebanyak 5892, *true negative* sebanyak 229, *false negative* sebanyak 23, dan *false positive* 249. Dengan menggunakan perhitungan ini, akurasi dapat dikonfirmasi sebagai berikut:

$$Accuracy = \frac{5892 + 229}{5892 + 229 + 23 + 249} = 95.74\%$$

Performance Vector adalah bentuk deskripsi dari tabel hasil analisis yang diperoleh dalam penelitian yang dilakukan. Terdapat nilai *True Positive* sebanyak 5892 , yang menggambarkan data positif yang diprediksi dengan benar. Selanjutnya, terdapat nilai F*alse Positive* sebanyak 249, yang mengindikasikan data negatif namun salah prediksi sebagai data positif. Nilai *False Negative* sebanyak 23 mencerminkan data positif yang salah prediksi sebagai data negatif. Terakhir, terdapat

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

nilai *True Negative* sebanyak 229, menunjukkan data negatif yang diprediksi dengan benar. Tabel 6 menampilkan *performance vector* dari algoritma *Logistic Regression*

Table 3 performance vector Logistic Regression

Performance Vektor Accuracy: 95.74% Confussion Matrix		
True	0	1
0	5892	23
1	249	229

4.1.2 Hasil Akurasi Algoritma *Extreme Gradient Boosting (XGBoost)*

Klasifikasi data dengan algoritma *Extreme Gradient Boosting* menghasilkan data yang disajikan dalam Tabel 5 *Confussion Matrix Extreme Gradient Boosting (XGBoost)*

Table 4 Confussion *Extreme Gradient Boosting*

	True 0	**True 1**	**class recall**
pred. 0	5899	16	99%
pred. 1	305	173	36%
Class Precision	95%	92%	

Akurasi mencapai 94.97%, dengan presisi kelas nol (prediksi negatif) sebesar 99% dan presisi kelas satu (prediksi positif) mencapai 36%. Akurasi dihitung menggunakan persamaan 1, dengan nilai true positive sebanyak 5899, true negative sebanyak 173, false negative sebanyak 16, dan false positive 305. Dengan menggunakan perhitungan ini, akurasi dapat dikonfirmasi sebagai berikut:

$$Accuracy = \frac{5899 + 173}{5899 + 173 + 16 + 305} = 94.97\%$$

Performance Vector adalah bentuk deskripsi dari tabel hasil analisis yang diperoleh dari penelitian yang telah dilakukan. Terdapat nilai *True Positive* sebanyak 5899, yang menggambarkan data positif yang diprediksi dengan benar. Selanjutnya, terdapat nilai *False Positive* sebanyak 305, yang mengindikasikan data negatif namun salah prediksi sebagai data positif. Nilai *False Negative* sebanyak 16 mencerminkan data positif yang salah prediksi sebagai data negatif. Terakhir, terdapat nilai *True Negative* sebanyak 173, menunjukkan data negatif yang diprediksi dengan benar. Tabel menampilkan performance vector dari algoritma *Logistic Regression.*

Table 5 performance vector Extreme Gradient Boosting

Performance Vektor Accuracy: 95.74% Confussion Matrix		
True	0	1
0	5899	23
1	305	173

5 KESIMPULAN

Tujuan dilakukan penelitian ini untuk mengetahui hasil perbandingan tingkat keakuratan dari metode penelitian yang digunakan yaitu *Logistic Regression, dan Extreme Gradient Boosting.* Dilihat dari *Class Recall* dan *Class Precision* metode yang menghasilkan tingkat keakuratan yang paling tinggi adalah *Logistic Regression* yaitu sebesar 95.74%. Metode klasifikasi *Extreme Gradient Boosting pada* penelitian ini cukup baik digunakan karena menghasilkan tingkat akurasi 94.97%, namun untuk mendapatkan hasil akurasi yang lebih maksimal untuk penelitian selanjutnya bisa menggunakan metode yang lain.

DAFTAR PUSTAKA

[1] H. Nurrahmi and D. Nurjanah, "Indonesian Twitter Cyberbullying Detection using Text Classification and User Credibility," 2018 International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia, 2018, pp. 543-548, doi: 10.1109/ICOIACT.2018.8350758.

[2] A. Pravina, I. Cholissodin, and P. Adikara, "Tampilan Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM)," *Ub.ac.id*, 2023. https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/4793/2232 (accessed Dec. 05, 2023).

[3] D. J. Ningrum, S. Suryadi, and D. E. Chandra Wardhana, "KAJIAN UJARAN KEBENCIAN DI MEDIA SOSIAL," *Jurnal Ilmiah KORPUS*, vol. 2, no. 3, pp. 241–252, Feb. 2019, doi: https://doi.org/10.33369/jik.v2i3.6779.

[4] D. Indonesia, "Pengguna Twitter di Indonesia Capai 18,45 Juta pada 2022," *Dataindonesia.id*, Aug. 10, 2022. https://dataindonesia.id/internet/detail/pengguna-twitter-di-indonesia-capai-1845-juta-pada-2022

[5] F. A. Prabowo, M. O. Ibrohim and I. Budi, "Hierarchical Multi-label Classification to Identify Hate Speech and Abusive Language on Indonesian Twitter," 2019 6th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE), Semarang, Indonesia, 2019, pp. 1-5, doi: 10.1109/ICITACEE.2019.8904425.

[6] W. Wahyudi, S. Adriko, M. Harits, and D. Hapsari, "Perbandingan Kinerja Algoritma Klasifikasi Naive Bayes, k-Nearest Neighbor dan Logistic Regression pada Dataset Multiclass," *Seminar Nasional Teknik Elektro, Sistem Informasi, dan Teknik Informatika FTETI*, vol. 1, no. 1, pp. 380–385, Mar. 2023.

[7] Khusni Mubarok, T. Wibowo, and S. Wibowo, "Kaji Awal Pendeteksi Api Menggunakan Kamera dengan Program Machine Learning," *Prosiding The 13th Industrial Research Workshop and National Seminar*, pp. 639–643, Jul. 2022.

[8] B. Wijaya and V. Mawardi, "PENDETEKSI UJARAN KEBENCIAN PADA PLATFORM MEDIA SOSIAL TWITTER MENGGUNAKAN SUPPORT VECTOR MACHINE," *Jurnal Serina Sains, Teknik dan Kedokteran*, vol. 1, no. 1, pp. 11–17, Feb. 2023.

[9] M. I. Abidin, K. A. Notodiputro, and B. Sartono, "Improving Classification Model Performances using an Active Learning Method to Detect Hate Speech in Twitter," *Indonesian Journal of Statistics and Its Applications*, vol. 5, no. 1, pp. 26–38, Mar. 2021, doi: https://doi.org/10.29244/ijsa.v5i1p26-38.

[10] Y. S. Mahardhika and E. Zuliarso, "ANALISIS SENTIMEN TERHADAP PEMERINTAHAN JOKO WIDODO PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAIVES BAYES CLASSIFIER," *SINTAK*, vol. 2, Nov. 2018, Accessed: Nov. 06, 2023. [Online]. Available: https://www.unisbank.ac.id/ojs/index.php/sintak/article/view/6651

[11] Renaldo Yosia Rafael and Fransiskus Adikara, "PENGIMPLMENTASIAN ALGORITMA LONG SHORT-TERM MEMORY UNTUK MENDETEKSI UJARAN KEBENCIAN PADA APLIKASI TWITTER," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 2, pp. 551–560, May 2023, doi: https://doi.org/10.29100/jipi.v8i2.3490.

[12] K. N. Widyatnyana, I. W. Rasna, and I. B. Putrayasa, "UJARAN KEBENCIAN DI DALAM TWITTER #SEBELUM2024JOKOWILENGSER: KAJIAN CYBERPRAGMATICS," *repo.undiksha.ac.id*, Aug. 01, 2023. https://repo.undiksha.ac.id/15785/ (accessed Dec. 05, 2023).

[13] Metatags generator, "Klasifikasi Ujaran Kebencian pada Media Sosial Twitter Menggunakan Support Vector Machine | Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)," *www.jurnal.iaii.or.id*, vol. 5, no. 1, Mar. 2021, Accessed: Dec. 05, 2023. [Online]. Available: https://www.jurnal.iaii.or.id/index.php/RESTI/article/view/2700

[14] Muhammad Mishbahul Munir, Mochamad Ali Fauzi, and Rizal Setya Perdana, "Implementasi Metode Backpropagation Neural Network Berbasis Lexicon Based Features dan Bag Of Words untuk Identifikasi Ujaran Kebencian pada Twitter," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 10, pp. 3182–3191, 2018, Accessed: Dec. 05, 2023. [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/2573

[15] Muhammad Hakiem, Mochammad Ali Fauzi, and Indriati Indriati, "Klasifikasi Ujaran Kebencian pada Twitter Menggunakan Metode Naive Bayes Berbasis N-Gram Dengan Seleksi Fitur Information Gain," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 3, pp. 2443–2451, 2019, Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/4682

[16] M. Ridwan and A. Muzakir, "Model Klasifikasi Ujaran Kebencian pada Data Twitter dengan Menggunakan CNN-LSTM":, *Teknomatika*, vol. 12, no. 02, pp. 209–218, Sep. 2022, Accessed: Dec. 05, 2023. [Online]. Available: https://ojs.palcomtech.ac.id/index.php/teknomatika/article/view/604

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

[17] M. Munir, M. Fauzi, and R. Perdana, "Implementasi Metode Backpropagation Neural Network Berbasis Lexicon Based Features dan Bag Of Words untuk Identifikasi Ujaran Kebencian pada Twitter," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 10, pp. 3182–3191, Oct. 2018.

[18] A. P. J. Dwitama and S. Hidayat, "Identifikasi Ujaran Kebencian Multilabel Pada Teks Twitter Berbahasa Indonesia Menggunakan Convolution Neural Network," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 3, no. 2, p. 117, Dec. 2021, doi: *https://doi.org/10.30865/json.v3i2.3610*.

[19] A. P. J. Dwitama, "DETEKSI UJARAN KEBENCIAN PADA TWITTER BAHASA INDONESIA MENGGUNAKAN MACHINE LEARNING: REVIU LITERATUR," *Jurnal Sains, Nalar, dan Aplikasi Teknologi Informasi*, vol. 1, no. 1, Aug. 2021, doi: https://doi.org/10.20885/snati.v1i1.5.