

cek plagiasi argy

by Argy Sambadda

Submission date: 16-Feb-2023 02:53PM (UTC+0700)

Submission ID: 2015536230

File name: Ilmiah_191080200230_Muhammad_Arginanta_Kafi_Sambada_Revisi1.docx (192.4K)

Word count: 3146

Character count: 19786

ENSEMBLE MACHINE LEARNING UNTUK MENDETEKSI SARKASME DALAM BAHASA INGGRIS DI MEDIA SOSIAL TWITTER

Muhammad Arginanta Kafi Sambada¹⁾, Mochammad Alfian Rosid^{*2)}, Arif Senja Fitriani³⁾

¹⁾Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

^{*}Email Penulis Korespondensi: alfanrosid@umsida.ac.id

Abstract. *Text classification refers to the categorization of documents based on their content and textual characteristics. This can be done by grouping notes into predetermined classes. Sarcastic content in text can pose a challenge in sentiment analysis. Sentiment analysis is one of the research fields in text mining that is used to classify text documents, including tweets. Tweets can be classified according to their sentiment, either positive or negative. Often, the emotion of a tweet cannot be accurately determined when it contains sarcasm. Sarcasm occurs when someone provides information that means the opposite of what is said, and it is often difficult to recognize automatically, even by humans. In this study, we will classify English text data on Twitter by identifying sarcasm in public opinion on Twitter. Therefore, sarcasm detection becomes a solution to overcome challenges in sentiment analysis by applying an ensemble approach to machine learning methods. The machine learning methods used include Logistic Regression, Naïve Bayes, Decision Trees, and Support Vector Machines. The best result of this study, which uses the Support Vector Machine algorithm, obtained good accuracy results with 80% accuracy and an f1-score of 80%.*

Keywords : *sentiment analysis; ensembles; Twitter English; Logistics Regression; Naïve Bayes, Decision Tree; SVM;*

Abstrak. *Pengklasifikasian teks menunjukkan pembagian dokumen berdasarkan konten dan karakter teks. Pembagian ini dapat dilakukan dengan mengelompokkan catatan ke dalam kelas yang telah ditentukan. Konten sarkastis yang terdapat dalam teks menjadi halangan dalam analisis perasaan teks. Menggunakan analisis perasaan merupakan salah satu cabang riset pembuatan teks yang digunakan untuk mengelompokkan dokumen teks, termasuk tweet. Tweet dapat diklasifikasikan menurut sentimennya, yaitu sentimen positif dan negatif. Seringkali, emosi sebuah tweet tidak dapat ditentukan secara akurat ketika mengandung sarkasme. Sarkasme terjadi ketika seseorang memberikan informasi yang berarti sebaliknya dari apa yang dikatakan, dan seringkali sulit untuk dikenali secara otomatis bahkan oleh manusia. Dalam penelitian ini, kami akan melakukan klasifikasi data teks dalam bahasa Inggris di Twitter dengan mengidentifikasi sarkasme dalam opini publik di Twitter. Dengan demikian dalam penelitian ini, pendeteksian sarkasme menjadi solusi untuk mengatasi kendala yang ada pada analisis sentimen dengan mengaplikasikan pendekatan ensemble pada metode machine learning. Metode machine learning yang akan digunakan meliputi Regresi Logistik, Naive Bayes, Pohon Keputusan, dan Mesin Vektor Pendukung. Hasil terbaik dari penelitian ini menggunakan algoritma Support Vector Machine dengan hasil akurasi yang baik mendapatkan akurasi 80% dan f1-score sebesar 80%.*

Kata Kunci : *analisis sentimen; ensemble; Twitter Bahasa Inggris; Logistic Regression; Naïve Bayes, Decision Tree; SVM;*

I. PENDAHULUAN

Klasifikasi teks adalah proses pengelompokan dokumen berdasarkan isi dan karakteristik teksnya. Proses ini dapat dilakukan dengan membagi teks menjadi kelas atau kategori tertentu. Menurut KBBI, sarkasme adalah penggunaan makna kata yang kasar untuk menyakiti perasaan orang lain atau membual diri sendiri. Sarkasme mengubah arti kalimat atau teks dengan cara yang bertentangan dengan isi teks dan dapat membingungkan arti teks. Tentu saja mendeteksi keberadaan sarkasme memerlukan perhatian pada semantic kata dan memilih algoritme yang sesuai untuk masalah ironi. Dengan demikian studi kasus sarkasme adalah topik hangat bagi peneliti sarkasme[1].

Analisis sentimen adalah bagian dari penelitian text mining yang bertujuan untuk mengklasifikasikan dokumen teks, termasuk tweet. Tweet dapat dikelompokkan menurut sentimen mereka, yaitu positif dan negatif. Sayangnya, seringkali sulit untuk menentukan emosi sebuah tweet dengan akurat jika tweet tersebut mengandung sarkasme. Ironi adalah bentuk ironi khusus yang terjadi ketika seseorang menyampaikan informasi yang tersirat, seringkali berarti kebalikan dari apa yang dikatakan. Sarkasme sulit dianalisis secara otomatis, bahkan oleh manusia. Analisis sentimen mendeteksi polaritas berdasarkan nilai setiap kata, sedangkan deteksi sarkasme juga memperhitungkan intonasi atau gerakan wajah saat orang tersebut berbicara[2]. Sayangnya, informasi mengenai intonasi atau gerakan wajah masih belum tersedia. Karena itu, deteksi ironi masih menjadi permasalahan yang sulit dalam analisis sentimen, termasuk analisis sentimen pada tweet. Berdasarkan penelitian yang dilakukan oleh Aisyah dan rekan-rekannya[3], sebanyak 21% komentar pada akun instagram politikus di Indonesia menggunakan sarkasme. Penelitian juga dilakukan oleh

Seindoufer[4]. dilakukan perbandingan dengan menggunakan 4 metode klasifikasi pada pendeteksian sarkasme, hasil akurasi tertinggi sebesar 83% diperoleh dengan algoritma Random Forest.

Pada penelitian yang telah dijelaskan diatas masih menggunakan machine learning dan deep learning, belum ada penelitian yang menggunakan ensemble machine learning untuk mendeteksi sarkasme. Sehingga ada peluang untuk mengembangkan deteksi sarkasme menggunakan metode ensemble machine learning.

Dalam kajian ini, kami akan melakukan klasifikasi terhadap teks dalam bahasa Inggris yang terdapat pada platform Twitter dengan mencari tahu opini publik yang terkandung dalam sarkasme. Kami akan melakukan pendeteksian sarkasme [12] melalui pendekatan metode ensemble machine learning. Metode machine learning yang akan digunakan termasuk Logistic Regression, Naive Bayes, Decision Tree, dan Support Vector Machine. Tujuannya adalah untuk menemukan metode ensemble machine learning yang memiliki kinerja terbaik dan dapat menjadi acuan untuk pembuatan model pendeteksian sarkasme.

II. LITERATUR REVIEW

Dalam penelitian ini dilakukan analisis terhadap sekumpulan informasi dengan penelitian sejenis. Pencarian serupa akan [7] mengeksplorasi pro dan kontra dari pencarian ini dengan melihat hasil sebelumnya. Penelitian yang dilakukan [5] menggunakan fitur interjeksi dan unigram sebagai fitur utama untuk mendeteksi kalimat sarkasme serta [7] membandingkan 2 metode klasifikasi yaitu Naive Bayes dan Support Vector Machine dengan kernel polinomial. Hasil akurasi terbaik adalah metode Naive Bayes dengan akurasi terbaik yang diperoleh mencapai lebih dari 91%.

Penelitian selanjutnya dilakukan [6] menggunakan pendekatan berdasarkan Metode BERT, Pertama, kami menyempurnakan model AraBERTv02 untuk tugas deteksi sarkasme. Kemudian, kami menggunakan model Sentence-BERT yang dilatih dengan pembelajaran kontrasif untuk mengekstrak embedding tweet yang representatif. Akhirnya, terinspirasi oleh bagaimana otak manusia memahami permukaan dan makna implisit dari kicauan sarkastik. Dan hasil yang diperoleh adalah 2,36% lebih rendah dari tempat pertama yang menegaskan kemampuan model gabungan yang digunakan dalam mendeteksi sarkasme.

Penelitian selanjutnya berkaitan pada Deteksi sarkasme pada pendekatan berbasis fitur menggunakan model pembelajaran mesin yang diawasi [7] melakukan Pendekatan yang ada untuk deteksi sarkasme otomatis terutama bergantung pada isyarat leksikal dan linguistik. Penelitian ini mengusulkan penerapan sistem yang kuat dan efisien untuk mendeteksi sarkasme guna meningkatkan akurasi analisis sentimen. Dan hasil menunjukkan bahwa Pohon Keputusan (91,84%) dan Hutan Acak (91,90%) mengungguli dalam hal akurasi dibandingkan dengan algoritme pembelajaran mesin terawasi lainnya untuk pemilihan fitur yang tepat.

Penelitian selanjutnya dilakukan [8] menyelidiki tweet sentimen negatif dengan kehadiran hiperbola untuk deteksi sarkasme. Enam ribu enam ratus tweet sentimen negatif yang telah diproses sebelumnya yang terdiri dari #Chinesevirus, #Kungflu, #COVID19, #Hantavirus, dan #Coronavirus dikumpulkan untuk deteksi sarkasme. Eksperimen dan analisis yang dilakukan dalam penelitian ini menyimpulkan bahwa hiperbola ada dalam kumpulan data yang tidak bias yang juga membantu meningkatkan deteksi sarkasme. Dan hasil metode yang diusulkan dengan kata memanjang mencapai akurasi dan F-score masing-masing 78,74% dan 71%.

Penelitian selanjutnya dilakukan [5] melakukan analisis sentiment dimana sarkasme sulit diidentifikasi dalam teks karena arti kata-kata yang diungkapkan oleh seseorang berlawanan dengan maksud sebenarnya dari orang tersebut. Dan hasil eksperimen menunjukkan bahwa kombinasi fastText embeddings dan BiGRU sebagai pengklasifikasi menghasilkan performa terbaik, dengan akurasi sebesar 93,85%.

2.1 Analisis Sentimen

Analisis sentimen adalah bidang studi yang mempelajari bagaimana menganalisis pandangan, perasaan, evaluasi, sikap, dan emosi orang terhadap suatu objek seperti produk, layanan, organisasi, individu, isu, peristiwa, topik dan lain-lain. Analisis sentimen, juga disebut dengan istilah opinion mining, berfokus utama pada pandangan yang mengungkapkan atau menyiratkan sentimen positif atau negatif, tetapi juga memperhatikan pandangan yang bersifat netral [6].

2.2 Sarkasme

Sarkasme adalah bentuk ekspresi emosi atau tanggapan negatif yang disampaikan dengan menggunakan kata-kata yang positif bahkan berlebihan. Dalam bentuk teks, sarkasme seringkali sulit dipahami sehingga dapat menimbulkan kesalahpahaman tentang polaritas suatu teks [7].

2.3 Ensemble Learning

Ensemble learning adalah teknik pembelajaran data yang menggabungkan beberapa algoritma atau model untuk memperoleh hasil yang lebih akurat [8]. Beberapa metode pembelajaran ensemble yang bisa digunakan adalah *voting*, *bagging*, *boosting*, dan *stacking*. Dalam penelitian ini, penulis menggunakan Voting. Itu metode klasifikasi voting tersedia di *scikit-learn* framework. Metode tersebut menghasilkan sebuah output yang akan digabungkan dengan aturan rata-rata voting probabilitas. Kemudian kelas dengan tinggi skor rata-rata akan dipilih untuk hasil akhir[9].

2.4 Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah sebuah sistem pembelajaran mesin yang bertindak sebagai sebuah klasifikasi. Ini menggunakan fungsi-fungsi linear sebagai hipotesis dalam sebuah ruang fitur (feature space) yang memiliki dimensi tinggi. SVM dilatih menggunakan algoritma pembelajaran yang didasarkan pada teori optimasi dan memasukkan bias pembelajaran yang berasal dari teori statistik.

Support Vector Machine (SVM) adalah metode untuk melakukan prediksi baik untuk klasifikasi maupun regresi. SVM berdasarkan pada prinsip bahwa masalah linier dapat dipisahkan, namun juga dapat mengatasi masalah non-linier dengan menambahkan konsep kernel ke dalam ruang fitur berdimensi tinggi. Dalam ruang ini, dicari garis pemis (hyperlane) yang dapat memaksimalkan jarak antara layer data [9].

2.5 Decision Tree

Algoritma Decision Tree merupakan salah satu algoritma untuk metode data mining yang sering diterapkan sebagai solusi untuk mengklasifikasikan sebuah masalah. Decision Tree sendiri menggunakan metode supervised machine learning yaitu proses pembelajaran dimana data baru diklasifikasikan berdasarkan training samples yang sudah ada[10]. Pohon Keputusan atau decision tree adalah implementasi dari teknik pengelompokan yang melibatkan proses penentuan suatu fungsi yang memetakan setiap himpunan perilaku ke dalam satu kelas yang telah didefinisikan sebelumnya.

2.6 Logistic Regression

Logistic Regression adalah sebuah teknik dalam bidang belajar mesin statistik yang diterapkan untuk melakukan klasifikasi. Ia termasuk dalam metode belajar bimbingan. Performa Logistic Regression sangat baik saat memproses data dalam jumlah besar, sehingga ia menjadi metode yang paling sering digunakan dalam aktivitas penambangan data.

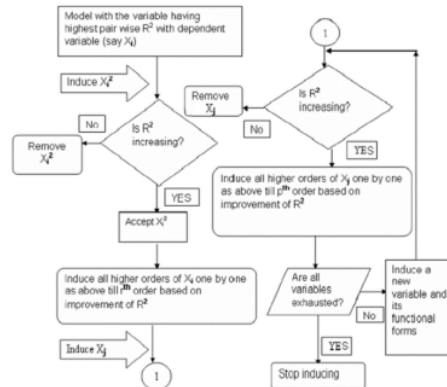
Logistic Regression adalah metode yang mempertemukan hubungan antara output sebagai klasifikasi biner dan variabel independen menggunakan probabilitas. Metode ini memprediksi nilai dari variabel dependen. Fungsi matematis dari logistic regression dinyatakan dengan Persamaan 1, di mana $\sigma(\cdot)$ adalah fungsi melalui dari sigmoid activation, seperti yang ditunjukkan dalam Persamaan 2. Persamaan ketiga menyatakan kedua persamaan tersebut.

$$\hat{p} = h\theta(x) = \sigma(\theta^T \cdot x) \quad (1)$$

$$\sigma(t) = \frac{1}{1 + \exp(-t)} \quad (2)$$

$$\hat{y} = f(x) = f(x) = \begin{cases} 0, & \text{jika } \hat{y} < 0.5 \\ 1, & \text{jika } \hat{y} \geq 0.5 \end{cases} \quad (3)$$

Selain untuk melakukan pengklasifikasi biner, logistic regression juga dapat melakukan multinomial regression, multi-kelas logistic regression atau pengklasifikasi entropy maksimum[11].



Gambar 3. Diagram intruksi pada Logistic Regression [12]

Proses pembelajaran Logistic Regression membutuhkan data latihan untuk membentuk model yang nantinya dapat digunakan untuk klasifikasi. Dalam penelitian yang telah dilakukan sebelumnya, telah membandingkan Logistic Regression dengan algoritma Naive Bayes dan SVM dalam klasifikasi biner. Beberapa studi juga menunjukkan bahwa Logistic Regression memperlihatkan hasil yang baik dalam mengklasifikasi teks.

2.7 Naïve Bayes

Naive Bayes Classifier adalah teknik klasifikasi yang didasarkan pada Teorema Bayes dan sangat cocok untuk menangani jumlah data yang besar. Ini menggunakan model statistik untuk menghitung probabilitas suatu kelas memiliki setiap set atribut dan menentukan kelas yang paling sesuai. Dalam metode ini, semua atribut dianggap saling bebas satu sama lain [13][14].

2.8 Confusion Matrix

Confusion Matrix adalah sebuah tabel yang sering digunakan untuk mendeskripsikan performansi model klasifikasi pada kumpulan data testing dengan nilai actual yang sudah diketahui. Tabel confusion matrix multiclass yang diterapkan pada penelitian ini adalah sebagai berikut:

		Predicted Class	
		Positive	Negative
Actual Class	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Gambar 4. Confusion Matrix

Pada Gambar 4, terdapat label prediksi yang dipengaruhi oleh kumpulan data. TP adalah singkatan dari True Positive, yaitu situasi di mana nilai prediksi dan nilai sebenarnya sama-sama benar. Dalam matrix multi-class, masing-masingnya adalah hanya TP yang dicantumkan, karena FN (False Negative) ditentukan dari semua baris setiap label dan FP (False Positive) ditentukan dari jumlah setiap label. Label dan TN (True Negative) mengacu pada situasi di mana nilai prediksi tidak ada dan nilai sebenarnya salah. Kinerja dalam mendeteksi penggunaan kalimat abusive diukur menggunakan beberapa parameter seperti precision, recall, dan F1 Score. Persamaan untuk parameter ini diberikan dalam persamaan (1-3)[15]. Precision adalah perbandingan antara data yang terklasifikasi dengan benar dan seluruh data yang terklasifikasi dengan benar.

$$Precision = \frac{TP}{(TP+FP)} \tag{8}$$

Recall adalah perbandingan antara data yang terklasifikasi dengan benar dan jumlah data yang terdapat dalam kelas tersebut. Rumus untuk setiap perhitungan adalah sebagai berikut.

$$Recall = \frac{TP}{(TP+FN)} \quad (9)$$

F1 Score adalah rata-rata bobot antara nilai precision dan nilai recall.

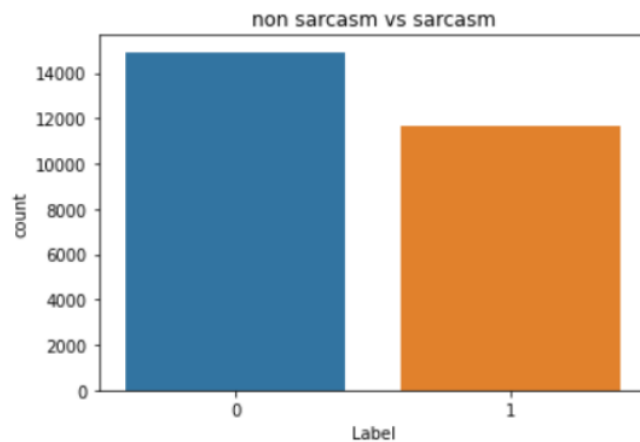
$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (10)$$

III. HASIL DAN PEMBAHASAN

Dataset yang dipakai pada survei ini terdiri dari 26709 data, diantaranya ada 14985 judul yang bermakna sarkasme dan 11724 judul berita tanpa makna sarkasme. Label sentiment yang digunakan hanya 2, yaitu 1 untuk data positif dan 0 untuk data negatif. Distribusi dataset dapat dilihat pada Tabel 1 dan Gambar 1.

Tabel 1. Distribusi Data

Kelas	Jumlah Data.
Sarkasme	14985
Tidak Sarkasme	11724
Total	26709



Gambar 3.1 Diagram Penyebaran Dataset

Selama preprocessing kumpulan data dilakukan *punctuation removal*, *Tokenizing*, *Stopword Removal* dan *Lematization*. Pada tahap menghilangkan stopwords dan tanda baca, semua kata pada data akan diubah menjadi huruf kecil. Kata-kata yang tidak memiliki makna seperti "and", "then", "also", "a", "an", dan lain-lain akan dihapus, serta tanda baca yang ada akan juga dihilangkan. Hasil dari proses menghilangkan stopwords dan tanda baca akan terlihat pada Tabel 2.

Tabel 2. Proses *Stopword Removal* dan *Punctuation Removal*

Sebelum	Sesudah

10 former versace store clerk sues over secret 'black code' for minority shoppers	Former Versace store employee files lawsuit over secret black code for minority shoppers.
5 the 'roseanne' revival catches up to our thorny political mood, for better and worse	1 the roseanne revival catches up to our thorny political mood for better and worse
3 mom starting to fear son's web series closest thing she will have to grandchild	mom starting to fear son s web series closest thing she will have to grandchild
3 boehner just wants wife to listen, not come up with alternative debt- reduction ideas	boehner just wants wife to listen not come up with alternative debt reduction ideas
1 j.k. rowling wishes snape happy birthday in the most magical way	1 j k rowling wishes snape happy birthday in the most magical way

Setelah melakukan penghapusan kata-kata tak berarti dan penghapusan tanda baca, dataset kemudian dibagi menjadi token-token dan hasilnya ditunjukkan pada Tabel 3.

Tabel 3. Proses *Tokenizing*

Sebelum	Sesudah
1 courtroom sketch artist has clear manga influences	'courtroom', 'sketch', 'artist', 'has', 'clear', 'manga', 'influences'
1 top snake handler leaves sinking huckabee campaign	'top', 'snake', 'handler', 'leaves', 'sinking', 'huckabee', 'campaign'
1 friday s morning email inside trump s presser for the ages	5 'friday', 's', 'morning', 'email', 'inside', 'trump', 's', 'presser', 'for', 'the', 'ages'
airline passengers tackle man who rushes cockpit in bomb threat	5 'airline', 'passengers', 'tackle', 'man', 'who', 'rushes', 'cockpit', 'in', 'bomb', 'threat'
facebook reportedly working on healthcare features and apps	1 'facebook', 'reportedly', 'working', 'on', 'healthcare', 'features', 'and', 'apps'

Penelitian ini melakukan pemodelan untuk mendeteksi sarkasme dengan menggunakan metode Logistic Regression, Support Vector Machine, Decision Tree, dan Naive Bayes dengan pendekatan ensemble. Model dari Logistic Regression, Support Vector Machine, Decision Tree, dan Naive Bayes dibuat dengan cara yang sama sehingga hasilnya dapat dibandingkan secara adil dan diklasifikasi pada pendekatan ensemble. Proses pengujian dilakukan untuk memperoleh hasil akurasi terbaik, menggunakan ukuran akurasi, presisi, recall, dan f1-score. Dataset dibagi menjadi data training dan data testing, dengan 21262 data training dan 5316 data testing. Percobaan dilakukan dengan mengubah dimensi dari 20, 50, dan 80. Hasil percobaan pada dimensi 20 dari dapat dilihat pada Tabel 4. Hasil percobaan untuk dimensi 50 pada dapat dilihat pada Tabel 5. Hasil percobaan untuk dimensi 80 dari dapat dilihat pada Tabel 6.

Tabel 4. Hasil Percobaan dengan Dimensi 20

No	Metode	Dimensi	Akurasi	Precision	Recall	F1
1	Logistik Regression	20	0,78	0,78	0,78	0,78
2	Naïve Bayes	20	0,76	0,83	0,76	0,77
3	Support Vecktor Machine	20	0,80	0,80	0,80	0,80
4	Decision Tree	20	0,63	0,89	0,63	0,70
5	Ensemble	20	0,75	0,80	0,75	0,76

Tabel 5. Hasil Percobaan dengan Dimensi 50

No	Metode	Dimensi	Akurasi	Precision	Recall	F1
1	Logistik Regression	50	0,77	0,77	0,77	0,77
2	Naïve Bayes	50	0,74	0,84	0,74	0,76
3	Support Vecktor Machine	50	0,78	0,79	0,78	0,79
4	Decision Tree	50	0,63	0,89	0,63	0,70
5	Ensemble	50	0,74	0,80	0,74	0,75

Tabel 6. Hasil Percobaan dengan Dimensi 80

No	Metode	Dimensi	Akurasi	Precision	Recall	F1
1	Logistik Regression	80	0,74	0,74	0,74	0,74
2	Naïve Bayes	80	0,72	0,82	0,72	0,74
3	Support Vecktor Machine	80	0,75	0,75	0,75	0,75
4	Decision Tree	80	0,63	0,88	0,63	0,70
5	Ensemble	80	0,71	0,77	0,71	0,73

Dari beberapa percobaan diatas, hasil percobaan terbaik diperoleh pada dimensi 20 dengan metode Support Vector Machine dengan nilai akurasi 80% dan f1-score 80% diikuti dimensi 50 dengan nilai akurasi 78% dan f1-score 79% diikuti dimensi 80 dengan nilai akurasi 75% dan f1-score 75% yang dapat dilihat pada Tabel 7.

Tabel 7. Perbandingan dari Dimensi 20, 50, 80 pada metode Support Vector Machine

Metode	Dimensi	Akurasi	Precision	Recall	F1
Support Vecktor Machine	20	0,80	0,80	0,80	0,80
	50	0,78	0,79	0,78	0,79
	80	0,75	0,75	0,75	0,75

IV. KESIMPULAN DAN SARAN

Kesimpulan dari hasil penelitian ini menunjukkan bahwa metode Support Vector Machine dengan dimensi 20 adalah metode terbaik untuk deteksi sarkasme. Ini karena tingkat akurasi dan f1-score yang paling tinggi dibandingkan metode lain. Dimensi 20 terbukti tepat untuk menganalisis sentiment data yang digunakan. Dalam penelitian selanjutnya, akan diterapkan teknik ensemble machine learning pada dataset yang berbeda.

REFERENSI

- [1] Y. V. Aritonang, D. P. Napitupulu, M. H. Sinaga, and J. Amalia, "Pengaruh Hyperparameter pada Fasttext terhadap Performa Model Deteksi Sarkasme Berbasis Bi-LSTM," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 3, pp. 2612–2625, 2022, doi: 10.35957/jatisi.v9i3.1331.
- [2] Y. Yunitasari, A. Musdholifah, and A. K. Sari, "Sarcasm Detection For Sentiment Analysis in Indonesian Tweets," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 13, no. 1, p. 53, 2019, doi: 10.22146/ijccs.41136.
- [3] A. Muhaddisi, B. N. Prastowo, D. Utami, and K. Putri, "Sentiment Analysis With Sarcasm Detection On P olitician ' s Instagram," vol. 15, no. 4, pp. 349–358, 2021.
- [4] V. Govindan and V. Balakrishnan, "A machine learning approach in analysing the effect of hyperboles using negative sentiment tweets for sarcasm detection," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 8, pp. 5110–5120, 2022, doi: 10.1016/j.jksuci.2022.01.008.
- [5] U. Riyanto, "ANALISIS PERBANDINGAN ALGORITMA NAIVE BAYES DAN SUPPORT VECTOR MACHINE DALAM MENGLASIFIKASIKAN JUMLAH PEMBACA," pp. 62–72, 2018.
- [6] M. G. Undap, V. P. Rantung, and P. T. D. Rompas, "Analisis Sentimen Situs Pembajak Artikel Penelitian Menggunakan Metode Lexicon-Based," 2021.
- [7] S. Christina, "Sarcasm in Sentiment Analysis of Indonesian Text : A Literature Review," vol. 13, no. 2, pp. 54–59, 2019.
- [8] A. Subekti, "Analisis Sentiment pada Ulasan Film Dengan Optimasi Ensemble Learning," vol. 7, no. 1, pp. 5–8, 2020.
- [9] M. Ma, A. Prayogo, P. Subarkah, and F. Nida, "Sentiment analysis of customer satisfaction levels on smartphone products using Ensemble Learning," vol. 14, no. 3, pp. 339–347, 2022.
- [10] Y. A. Wijaya *et al.*, "Analisa Klasifikasi menggunakan Algoritma Decision Tree pada Data Log Firewall Jurnal Sistem Informasi dan Manajemen," vol. 9, no. 3, 2021.
- [11] M. Shandy, T. Putra, and Y. Azhar, "Perbandingan Model Logistic Regression dan Artificial Neural Network pada Prediksi Pembatalan Hotel," vol. 6, no. 1, pp. 29–37, 2021.
- [12] R. Rahmanda and D. S. Informasi, "Rancang bangun aplikasi berbasis microservice untuk klasifikasi sentimen. studi kasus: pt. yesboss group indonesia (kata.ai)," 2018.
- [13] A. Setiawan, L. W. Santoso, R. Adipranata, U. K. Petra, and J. Siwalankerto, "Klasifikasi

- Artikel Berita Bahasa Indonesia Dengan Naive Bayes Classifier,” pp. 3–8.
- [14] E. I. Setiawan *et al.*, “Deteksi Validitas Berita pada Media Sosial Twitter dengan Algoritma Naive Bayes,” pp. 55–60.
- [15] J. Nasional, S. Informasi, M. Kamil, T. Endra, and E. Tju, “Naïve Bayes dan Confusion Matrix untuk Efisiensi Analisa Intrusion Detection System Alert,” vol. 02, pp. 81–88, 2022.

cek plagiasi argy

ORIGINALITY REPORT

19%

SIMILARITY INDEX

19%

INTERNET SOURCES

6%

PUBLICATIONS

9%

STUDENT PAPERS

PRIMARY SOURCES

1	gist.github.com Internet Source	3%
2	openlibrarypublications.telkomuniversity.ac.id Internet Source	2%
3	www.theonion.com Internet Source	2%
4	media.neliti.com Internet Source	2%
5	practicaldatascience.co.uk Internet Source	2%
6	123dok.com Internet Source	2%
7	jisai.mercubuana-yogya.ac.id Internet Source	1%
8	Submitted to Universitas Lancang Kuning Student Paper	1%
9	prosiding.unimus.ac.id Internet Source	1%

10 www.huffingtonpost.com.au 1 %
Internet Source

11 Submitted to UIN Sultan Syarif Kasim Riau 1 %
Student Paper

12 link.springer.com 1 %
Internet Source

13 lib.ui.ac.id 1 %
Internet Source

Exclude quotes Off

Exclude matches < 15 words

Exclude bibliography On