

## Ensemble Machine Learning To Detect Sarcasm In English On Twitter Social Media

## Ensemble Machine Learning Untuk Mendeteksi Sarkasme Dalam Bahasa Inggris Pada Sosial Media Twiter

Muhammad Arginanta Kafi Sambada<sup>\*1)</sup>, Mochammad Alfian Rosid<sup>2)</sup>, Arif Senja Fitriani<sup>3)</sup> Suprianto<sup>4)</sup>

<sup>1)2)3)4)</sup>Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

\*Email Penulis Korespondensi: [argkaf@umsida.ac.id](mailto:argkaf@umsida.ac.id), [alfanrosid@umsida.ac.id](mailto:alfanrosid@umsida.ac.id), [asfjim@umsida.ac.id](mailto:asfjim@umsida.ac.id), [suprianto@umsida.ac.id](mailto:suprianto@umsida.ac.id)

**Abstract.** *Sarcasm on social media is often used to mock or hurt someone using language or words that appear positive but have a negative meaning. This poses a challenge in sentiment analysis on social media because it's difficult to detect sarcasm even for humans. Therefore, sentiment analysis is conducted to classify sarcastic sentences that contain positive sentiment but have a negative meaning. The aim of this study is to compare the performance of four machine learning methods, namely Logistic Regression, Naive Bayes, Decision Tree, and Support Vector Machine, in detecting sarcastic sentences in news headlines. The dataset used contains news headlines in English. The results show that the Support Vector Machine method has the best performance with a model of 20 that has an accuracy score of 80% and an f1-score of 80%, compared to the Logistic Regression, Naive Bayes, and Decision Tree methods. Therefore, it can be concluded that the Support Vector Machine method is the best solution for detecting sarcasm.*

**Keywords :** *Sarcasm; Social media; Sentiment analysis; Sarcasm detection; Machine learning;*

**Abstrak.** *Sarkasme pada media sosial seringkali digunakan untuk menyindir atau menyakiti hati seseorang dengan menggunakan bahasa atau kata yang seolah-olah mengandung kata positif tetapi maknanya negatif. Hal ini menimbulkan kendala pada analisis sentimen di media sosial, karena sulit untuk mendeteksi sarkasme secara otomatis bahkan oleh manusia. Oleh karena itu, analisis sentimen dilakukan untuk mengklasifikasikan kalimat sarkasme yang mengandung sentimen positif namun memiliki makna negatif. Tujuan Penelitian ini adalah untuk membandingkan kinerja empat metode machine learning, yaitu Logistic Regression, Naive Bayes, Decision Tree, dan Support Vector Machine dalam mendeteksi kalimat sarkasme pada judul berita. Dataset yang digunakan yakni berisi judul berita dalam bahasa inggris. Hasil penelitian adalah menunjukkan bahwa metode Support Vector Machine memiliki kinerja terbaik dengan model 20 yang memiliki nilai akurasi 80% dan f1-score 80%, dibandingkan dengan metode Logistic Regression, Naive Bayes, dan Decision Tree. Oleh karena itu penjelasan diatas dapat disimpulkan bahwa metode Support Vector Machine adalah solusi terbaik untuk mendeteksi sarkasme.*

**Kata Kunci :** *Sarkasme; Media Sosial; Analisis sentimen; Deteksi sarkasme; Machine Learning;*

### 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.

Pada penelitian ini dilakukan analisis dari kumpulan informasi dengan penelitian yang serupa. Penelitian yang serupa akan ditelusuri keutamaan dan kekurangan dalam penelitian tersebut dengan melihat dari hasil yang terdahulu. Penelitian untuk mendeteksi sarkasme mulai banyak dipelajari dalam beberapa tahun terakhir. Penelitian yang dilakukan[5] menggunakan fitur interjeksi dan unigram sebagai fitur utama pendeteksi kalimat sarkasme serta 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.

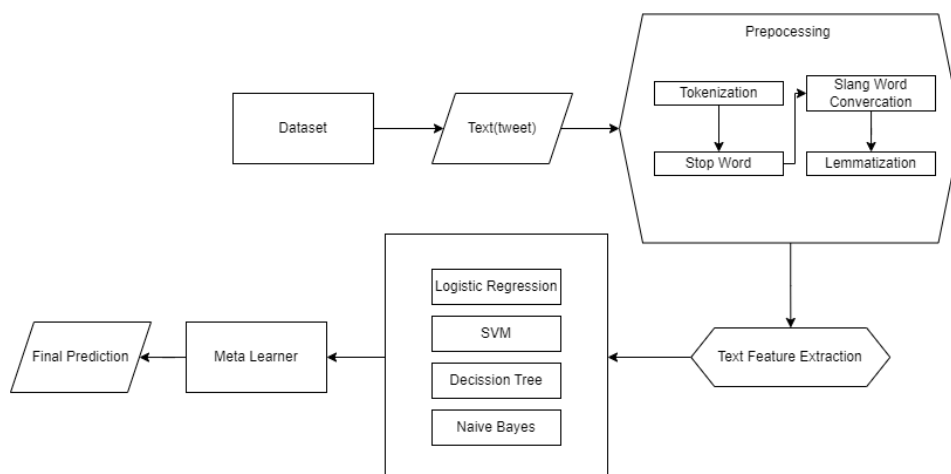
Penelitian selanjutnya dilakukan[8] menyelidiki tweet sentimen negatif dengan kehadiran hiperbola untuk deteksi sarkasme. Enam ribu enam ratus tweet sentimen negatif yang telah diproses sebelumnya yang terdiri dari #Chinesevirus, #Kungflu, #COVID19, #Hantavirus, dan #Coronavirus dikumpulkan untuk deteksi 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%. 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. Tahapan penelitian diawali dari pengumpulan data, Pre-Processing, Feature Extraction, Proses Klasifikasi menggunakan Logistic Regression, Naive Bayes, Decision Tree, dan Support Vector Machine , kemudian dilakukan prediksi akhir dengan Meta Learning. Tujuannya adalah untuk menemukan metode ensemble machine learning yang memiliki kinerja terbaik dan dapat menjadi acuan untuk pembuatan model pendeteksian sarkasme.

## II. METODOLOGI

Berikut ini merupakan arsitektur sistem untuk ensemble machine learning yang digunakan pada penelitian.



Gambar 1. Arsitektur Sistem untuk Ensemble Machine Learning

### 2.1 Dataset

Dataset diambil dari situs web yaitu Kaggle. Data yang diperoleh tersebut adalah data berita utama. Kumpulan data berita utama memiliki kesalahan tata bahasa yang minimal, kalimat yang unik dan data yang sangat bersih. Data tersebut menggunakan Natural Language Toolkit (NLTK) untuk tokenisasi, lemmatisasi dan pra pemrosesan. Dataset headline berita tersebut terdiri dari 26.709 headline berita.

Tabel 1. Contoh Dataset

| No. | headline   | sarcastic |
|-----|--|-----------|
| 1.  | former versace store clerk sues over secret 'black code' for minority shoppers       | 0         |
| 2.  | the 'roseanne' revival catches up to our thorny political mood, for better and worse | 0         |
| 3.  | mom starting to fear son's web series closest thing she will have to grandchild      | 0         |
| 4.  | boehner just wants wife to listen, not come up with alternative debt-reduction ideas | 1         |
| 5.  | j.k. rowling wishes snape happy birthday in the most magical way                     | 0         |
| 6.  | advancing the world's women  | 0         |
| 7.  | the fascinating case for eating lab-grown meat                                       | 0         |
| 8.  | this ceo will send your kids to school, if you work for his company                  | 0         |
| 9.  | top snake handler leaves sinking huckabee campaign                                   | 1         |
| 10. | friday's morning email: inside trump's presser for the ages                          | 0         |

## 2.2 Preprocessing Text

Proses preprocessing text di terapkan untuk data yang akan di gunakan dalam proses analisa sentimen, dimana data yang kita proses akan kita ambil informasi yang terkandung didalamnya dalam hal sentimen penulisnya yaitu negaitf atu positif. Guna memudahkan dalam mengelola data maka data perlu kita berikan analisa sentimen secara manual dengan membaca maksud dari kalimat yang ada dalam sentimen tersebut, sehingga dapat diberikan penilaian bahwa sentimen tersebut merupakan setimen negatif atau positif [5]. Pada penelitian ini di proses yang dilakukan dalam tahapan *preprocessing* adalah sebagai berikut:

### 1. *Tokenizing*

*Tokenizing* adalah proses pemotongan sebuah dokumen menjadi bagian-bagian, yang disebut dengan token. Pada saat bersamaan tokenizing juga berfungsi untuk membuang beberapa karakter tertentu yang dianggap sebagai tanda baca.

### 2. *Stopword Removal*

*Stopword removal* adalah proses penghilangan kata-kata yang tidak berkontribusi banyak pada isi dokumen. Kata-kata yang termasuk ke dalam *stopword* dihilangkan karena memberikan pengaruh yang tidak baik dalam proses text mining seperti kata-kata “and”, “I”, “you”, “with”, “she”, “he”, dan lain-lain.

### 3. *Slangword*

*Slangword* merupakan perubahan kata berbahasa Indonesia yang tidak baku di tweet dan tidak ada dalam kamus, kata tersebut akan di translasikan ke kata terdekat dengan menggunakan kamus yang dibuat dengan melihat pola kemunculan kata-kata baku.

### 4. *Lemmatization*

*Lemmatization* adalah proses mengubah kata-kata dalam teks ke dalam bentuk dasar atau kata dasar yang beragam. Tujuannya adalah untuk mengurangi variasi bentuk kata sehingga analisis teks dapat dilakukan dengan lebih efisien dan akurat.

Selanjutnya dilakukan proses *Feature extraction* atau ekstrasi fitur yang merupakan proses merubah sebuah data di proyeksikan ke dalam fitur baru dengan ruang dimensi yang lebih rendah [6]. Terakhir adalah melakukan tahapan klasifikasi machine learning pada proses analisis sentimen dengan menggunakan metode Logistic Regression, Naive Bayes, Decision Tree, dan Support Vector Machine.

## 2.3 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 [7]. 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 pemisah (hyperlane) yang dapat memaksimalkan jarak antara layer data [9].

## 2.4 Decision Tree

Decision Tree (pohon keputusan) adalah metode yang umum digunakan dalam data mining dan machine learning untuk klasifikasi dan prediksi. Metode ini membangun model prediktif dengan membagi data menjadi subset yang lebih kecil dengan menentukan fitur terbaik yang memisahkan data tersebut dengan cara yang paling efisien. Decision Tree sendiri menggunakan metode supervised machine learning yaitu proses pembelajaran dimana data baru diklasifikasikan berdasarkan training samples yang sudah ada[10]. Model yang dihasilkan dari metode Decision Tree berbentuk seperti pohon, dengan node yang mewakili keputusan dan cabang yang mewakili hasil dari keputusan tersebut.

Pada setiap tahapan, algoritma Decision Tree menggunakan berbagai metrik untuk memilih fitur terbaik, yaitu metrik yang dapat mengukur seberapa baik fitur tersebut memisahkan data. Beberapa metrik yang umum digunakan dalam Decision Tree antara lain Gini Index, Information Gain, dan Chi-Square. Dalam praktiknya, Decision Tree dapat digunakan untuk berbagai masalah, seperti klasifikasi dan regresi. Dalam klasifikasi, model Decision Tree memprediksi kelas target dengan mengambil keputusan berdasarkan fitur yang dipilih pada setiap node, sedangkan

dalam regresi, model Decision Tree memprediksi nilai target berdasarkan rata-rata nilai target pada subset data yang ada di setiap daun.

Dalam pengembangan model Decision Tree, terdapat beberapa teknik seperti Pruning, Boosting dan Random Forest yang dapat digunakan untuk meningkatkan kinerja model dan mencegah overfitting.

## 2.5 Logistic Regression

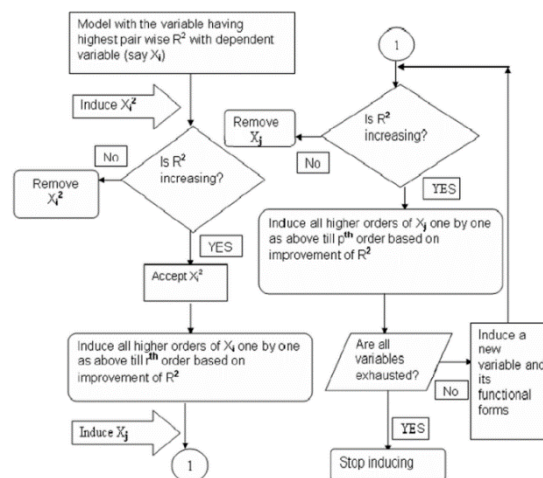
Logistic Regression adalah metode statistik yang digunakan untuk menganalisis dan memodelkan hubungan antara variabel dependen dan satu atau lebih variabel independen. Variabel dependen dalam regresi logistik adalah variabel biner, yang artinya hanya bisa memiliki dua nilai (misalnya, 0 atau 1, benar atau salah, ya atau tidak).

Metode Logistic Regression mengestimasi probabilitas variabel dependen memiliki nilai tertentu (misalnya, 1) yang diberikan sekumpulan variabel independen. Hal ini dilakukan dengan menggunakan fungsi logistik (juga dikenal sebagai fungsi sigmoid), yang memetakan masukan bernilai riil menjadi nilai antara 0 dan 1. Analisis Logistic Regression didasarkan pada suatu fungsi yang disebut fungsi Logistic Regression yang di tulis;

$$x = \frac{\exp(Q_0 + Q_1x_{1i} + Q_1x_{2i} + \dots + Q_px_{pi})}{1 + \exp(Q_0 + Q_1x_{1i} + Q_1x_{2i} + \dots + Q_px_{pi})} \quad (1)$$

Nilai ( $x$ ) merupakan peluang kejadian sukses dengan variabel bebas  $x$  sebanyak  $p$ . Dalam persamaan ini, nilai ( $x$ ) merupakan fungsi yang non linier, sehingga perlu dilakukan transformasi logit agar diperoleh fungsi linier, sehingga kita dapat melihat hubungan antara variabel bebas dan variabel terikat [8].

Logistic Regression adalah metode statistik yang digunakan untuk memodelkan hubungan antara variabel dependen biner dan satu atau lebih variabel independen. Ini mengestimasi probabilitas variabel dependen memiliki nilai tertentu menggunakan fungsi logistik, dan koefisien dari variabel independen diestimasi menggunakan estimasi maksimum kemungkinan. Selain untuk melakukan pengklasifikasi biner, logistic regression juga dapat melakukan multinomial regression, multi-kelas logistic regression atau pengklasifikasi entropy maksimum[9].



Gambar 2. Diagram intruksi pada Logistic Regression

Sumber [10]

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.

## 2.6 Naïve Bayes

Naive Bayes adalah metode klasifikasi yang berdasarkan pada teorema Bayes. Metode ini cukup populer dan sering digunakan pada bidang Natural Language Processing (NLP) dan pengenalan teks. Dalam NLP, metode ini dapat digunakan untuk klasifikasi teks, seperti klasifikasi spam atau non-spam email, klasifikasi dokumen, atau klasifikasi sentimen [11]. Rumus dasar dari teorema Bayes adalah sebagai berikut:

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \quad (2)$$

Persamaan (1) menunjukkan  $Y$  suatu kelas spesifik,  $X$  merupakan data pada kelas yang belum diketahui, sedangkan  $P(Y|X)$  adalah probabilitas hipotesis berdasarkan kondisi, sedangkan  $P(Y)$  dan  $P(X|Y)$  merupakan probabilitas sebelumnya dari kelas berdasarkan kondisi hipotesis, sedangkan  $P(X)$  merupakan probabilitas  $Y$  [12].

## 2.7 Ensemble Learning

Ensemble learning adalah teknik pembelajaran data yang menggabungkan beberapa algoritma atau model untuk memperoleh hasil yang lebih akurat [13]. 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 [14].

## 2.8 Confusion Matrix

Confusion Matrix adalah sebuah tabel yang sering digunakan untuk mendeskripsikan peformansi 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 3. 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, masalahnya 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.

$$Precision = \frac{TP}{(TP+FP)} \quad (3)$$

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

$$Recall = \frac{TP}{(TP+FN)} \quad (4)$$

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

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (5)$$

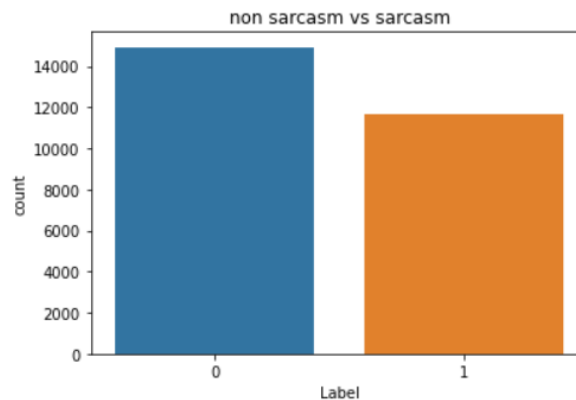
## III. HASIL DAN PEMBAHASAN

Dataset yang digunakan dalam penelitian ini terdiri dari 26709 data tweet yang diambil dari platform twitter. Tweet-tweet ini telah dianotasi oleh empat annotator yang terlatih untuk mendeteksi keberadaan sarcasm. Setiap tweet diberi label "sarkastik" atau "tidak sarkastik" berdasarkan konsensus dari tiga anotator. contoh data dapat dilihat pada

Tabel 1. Dataset terdiri dari 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 2 dan Gambar 1.

Tabel 2. Distribusi Data

| Kelas          | Jumlah Data. |
|----------------|--------------|
| Sarkasme       | 14985        |
| Tidak Sarkasme | 11724        |
| Total          | 26709        |



Gambar 4. 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 3.

Tabel 3. Proses *Stopword Removal* dan *Punctuation Removal*

| Sebelum  | Sesudah   |
|--|---|
| ormer versace store clerk sues over secret 'black code' for minority shoppers        | Former Versace store employee files lawsuit over secret black code for minority shoppers. |
| the 'roseanne' revival catches up to our thorny political mood, for better and worse | the roseanne revival catches up to our thorny political mood for better and worse         |
| 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           |
| 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       |
| j.k. rowling wishes snape happy birthday in the most magical way                     | 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 4.

Tabel 4. Proses *Tokenizing*

| Sebelum   | Sesudah  |
|---|--|
| courtroom sketch artist has clear manga influences              | 'courtroom', 'sketch', 'artist', 'has', 'clear', 'manga', 'influences'                       |
| top snake handler leaves sinking huckabee campaign              | 'top', 'snake', 'handler', 'leaves', 'sinking', 'huckabee', 'campaign'                       |
| friday s morning email inside trump s presser for the ages      | 'friday', 's', 'morning', 'email', 'inside', 'trump', 's', 'presser', 'for', 'the', 'ages'   |
| airline passengers tackle man who rushes cockpit in bomb threat | 'airline', 'passengers', 'tackle', 'man', 'who', 'rushes', 'cockpit', 'in', 'bomb', 'threat' |
| facebook reportedly working on healthcare features and apps     | '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 6. Hasil percobaan untuk dimensi 80 dari dapat dilihat pada Tabel 7.

Tabel 5. Hasil Percobaan dengan Model 20

| No | Metode                  | Dimensi | Akurasi | Precision | Recall | F1   |
|----|-------------------------|---------|---------|-----------|--------|------|
| 1  | Logistik Regression     | 20      | 0,78    | 0,78      | 0,78   | 0,78 |
| 2  | Naive 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 6. Hasil Percobaan dengan Model 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 7. Hasil Percobaan dengan Model 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 8.

Tabel 8. Perbandingan dari Dimensi 20, 50, 80 pada Model 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

Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa model deteksi sarcasm yang dikembangkan dengan menggunakan algoritma Support Vector Machine (SVM) memiliki akurasi yang cukup baik. Model ini dapat digunakan untuk mendeteksi keberadaan sarcasm pada teks dengan tingkat keberhasilan sebesar 80% pada dimensi 20. Meskipun model deteksi sarcasm yang dikembangkan sudah cukup baik, masih terdapat beberapa tantangan dalam mengembangkan model yang lebih akurat dan mampu mendeteksi berbagai jenis sarcasm yang berbeda. Oleh karena

itu, disarankan untuk melanjutkan penelitian ini dengan mencoba algoritma deteksi sarcasm lain yang lebih canggih, melakukan eksperimen dengan dataset yang lebih besar, dan mempertimbangkan faktor konteks dalam deteksi sarcasm. Selain itu, disarankan pula untuk mempertimbangkan penggunaan teknologi pemrosesan bahasa alami (natural language processing) untuk meningkatkan akurasi model deteksi sarcasm.

## REFERENSI

- [1] Y. V. Aritionang, 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 Politician '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] F. Ugm and F. Ugm, "Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine," vol. 8, no. 1, pp. 91–100, 2014.
- [6] A. F. Hidayatullah *et al.*, "Analisis sentimen dan klasifikasi kategori terhadap tokoh publik pada twitter," vol. 2014, no. semnasIF, pp. 115–122, 2014.
- [7] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 147, 2021, doi: 10.25126/jtiik.0813944.
- [8] A. Syahadati, N. C. Lengkong, O. Safitri, S. Machsus, Y. R. Putra, and R. Nooraeni, "ANALISIS SENTIMEN PENERAPAN PSBB DI DKI JAKARTA DAN DAMPAKNYA TERHADAP PERGERAKAN IHSG," vol. 15, no. 1, pp. 20–25, 2021.
- [9] 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.
- [10] R. Rahmanda and D. S. Informasi, "Rancang bangun aplikasi berbasis microservice untuk klasifikasi sentimen. studi kasus: pt. yesboss group indonesia (kata.ai)," 2018.
- [11] 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.
- [12] U. Verawardina, F. Edi, and R. Watrianthos, "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naive Bayes," vol. 5, pp. 157–163, 2021, doi: 10.30865/mib.v5i1.2604.
- [13] A. Subekti, "Analisis Sentiment pada Ulasan Film Dengan Optimasi Ensemble Learning," vol. 7, no. 1, pp. 5–8, 2020.
- [14] 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.
- [15] J. Nasional, S. Informasi, M. Kamil, T. Endra, and E. Tju, "Naive Bayes dan Confusion Matrix untuk Efisiensi Analisa Intrusion Detection System Alert," vol. 02, pp. 81–88, 2022.

### **Conflict of Interest Statement:**

*The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.*