

Studi Komparasi Algoritma SVM Dan Random Forest Pada Analisis Sentimen Komentar Youtube BTS

Anisa Nur Syafia^{1*}, Muhammad Fikri Hidayattullah², Wirmanto Sutеды³

^{1,3}Program Studi Teknik Komputer, Universitas Pendidikan Indonesia

²Program Studi Teknik Informatika, Politeknik Harapan Bersama

^{1,3}Jl. Pendidikan No.15, Kabupaten Bandung, 40625, Indonesia

²Jalan Mataram No.9, Kota Tegal, 52147, Indonesia

email: ¹anisasyaf@upi.edu, ²fikri@poltektegal.ac.id, ³wirmanto.suteddy@upi.edu

Abstract – Sentiment analysis of YouTube boy group BTS comments uses the NLP approach to detect emotional patterns based on two category labels, namely positive and negative. With NLP, positive or negative polarity in an entity can be allocated as well as predicted high and low performance from various classification sentiments. The machine learning algorithms used to measure the accuracy of sentiment analysis developed are the Support Vector Machine and Random Forest algorithms. The steps taken start from the data collection obtained from the BTS YouTube Comment dataset and then go through the data preprocessing stage. Then proceed to the feature extraction stage by converting text into digital vectors or Bag of Words (BOW) and classified using machine learning algorithms until the evaluation stage. From the results comparison of the evaluated algorithms, the accuracy value between the two algorithms is 96% for training data and 85% for data testing using the SVM algorithm, while for the Random Forest algorithm it is 82% for training data and 80% for data testing. This shows that the SVM algorithm produces a higher accuracy value than the Random Forest for sentiment analysis of YouTube boy group BTS comments.

Abstrak – Analisis sentimen komentar youtube boy group BTS menggunakan pendekatan NLP untuk mendeteksi pola emosi berdasarkan dua label kategori yaitu positif dan negatif. Dengan NLP, polaritas positif atau negatif pada suatu entitas dapat dialokasikan juga diprediksi kinerja tinggi dan rendah dari berbagai pengklasifikasian sentimen. Algoritma *machine learning* yang digunakan untuk mengukur tingkat akurasi analisis sentiment yang dikembangkan yaitu algoritma *Support Vector Machine* dan *Random Forest*. Tahapan yang dilakukan dimulai dari data *collection* yang didapat dari *BTS Youtube Comment dataset* lalu melalui tahap data *preprocessing*. Kemudian dilanjutkan ke tahap *feature extraction* dengan mengubah teks menjadi vektor digital atau *Bag of Words (BOW)* dan diklasifikasikan menggunakan algoritma *machine learning* hingga tahap evaluasi. Dari hasil komparasi algoritma yang dievaluasi, didapat nilai akurasi antara kedua algoritma tersebut. Untuk algoritma SVM didapat 96% akurasi data latih dan 85% akurasi data uji, sedangkan untuk algoritma *Random Forest* didapat 82% untuk akurasi data latih dan 80% untuk data uji. Hal ini menunjukkan bahwa algoritma SVM menghasilkan nilai akurasi yang lebih tinggi dibandingkan *Random Forest* pada analisis sentimen komentar *youtube boy group* BTS.

*) penulis korespondensi: Anisa Nur Syafia

Email: anisasyaf@upi.edu

Kata Kunci – NLP, SVM, *random forest*, *machine learning*, *sentiment analysis*, BTS

I. PENDAHULUAN

Dengan meningkatnya minat masyarakat di Indonesia terhadap penggunaan *youtube*, banyak masyarakat bebas mengeluarkan pendapatnya terhadap video *youtube* tersebut. Pendapat atau komentar yang diberikan berisi pujian, masukan, bahkan kritikan berdasarkan isi konten. Dengan mengartikan komentar, kita bisa menilai apakah konten tersebut merupakan konten positif ataupun negatif berdasarkan makna dari komentar yang diberikan [1]. BTS merupakan grup laki-laki yang berasal dari Korea dan sangat populer di beberapa negara, termasuk Indonesia [2]. Hal ini bisa dilihat berdasarkan seberapa seringnya BTS *trending* di *youtube*, bukan hanya musiknya, namun wawancara, *vlog*, maupun video kegiatan lainnya yang berkaitan dengan BTS selalu *trending*. Namun, segala sesuatu yang *trending* belum tentu berisi konten positif, apalagi BTS merupakan sosok idola yang menjadi panutan bagi penggemarnya. Maka dari itu penulis melakukan analisis sentimen pada komentar *youtube* konten BTS karena analisis sentiment dapat menganalisis teks bahasa alami untuk mendeteksi emosi dengan pola emosi untuk membuat keputusan [3]. Hal ini digunakan untuk mengetahui seberapa banyaknya komentar positif dan negatif dalam konten BTS untuk dijadikan pertimbangan apakah BTS merupakan idola yang memberikan pengaruh positif berdasarkan karya musiknya atau tidak. Sehingga rumusan masalah dalam penelitian ini diantaranya bagaimana cara menganalisis komentar positif dan negatif pada komentar youtube BTS, bagaimana cara mengklasifikasikan komentar positif dan negatif konten youtube BTS berdasarkan tingkat akurasinya dan bagaimana cara mengetahui frekuensi kemunculan kata positif dan negatif pada komentar youtube BTS.

Tujuan dari penelitian ini untuk menerapkan pendekatan *Natural Language Processing (NLP)* dalam analisis konten youtube BTS. Kontribusi dalam penelitian ini terdapat di kontribusi masalah yaitu menentukan komentar positif dan negatif pada konten BTS menggunakan metode NLP dengan mengalokasikan polaritas positif atau negatif pada suatu entitas juga memprediksi kinerja tinggi dan rendah dari berbagai pengklasifikasian sentimen [4]. Proses yang dilakukan ialah pengumpulan data yang dilanjutkan dengan proses pengolahan data mentah yang telah dikumpulkan dari youtube lalu tahap *preprocessing* untuk mengatasi data

mentah yang bervariasi dilanjutkan dengan membagi data menjadi data latih dan data uji [5]. Selanjutnya dilakukan perhitungan TF-IDF yang nantinya menghasilkan nilai untuk dianalisis sehingga bisa dilakukan proses akhir yaitu klasifikasi menggunakan SVM dan *Random Forest*. Hasil yang didapat dari penelitian ini untuk menambah pengetahuan dan pemahaman mendalam terkait pemrosesan *Natural Language Processing* pada sentimen analisis, membantu masyarakat mengetahui banyaknya komentar positif dan negative, memberi informasi terkait kata yang paling sering muncul pada konten youtube BTS, juga menambah pengetahuan untuk selalu berkomentar baik di sosial media.

II. PENELITIAN TERKAIT

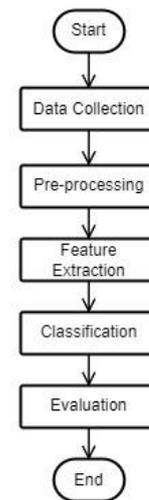
Penelitian mengenai analisis sentimen telah banyak dilakukan oleh peneliti di berbagai bidang, diantaranya oleh Khanvilkar dkk. [6] menggunakan algoritma *Random Forest* dan SVM untuk analisis sentimen produk rekomendasi. Darwis dkk. [7] juga melakukan analisis sentimen terhadap komentar twitter menggunakan algoritma SVM, data komentar tersebut kemudian digunakan untuk melakukan prediksi pada data *testing* dan menghasilkan nilai akurasi sebesar 82%. Alita dkk. [8] menggunakan algoritma *Random Forest* untuk mendeteksi sarkasme pada analisis sentimen. Perbandingan analisis sentimen menggunakan algoritma *Random Forest* dan SVM dilakukan oleh [9] pada data komentar *twitter* yang mengandung kata kunci PSBB. Perbandingan lain dilakukan oleh [10] pada analisis sentimen pengguna layanan BPJS menggunakan algoritma KNN, *Decision Tree*, dan *Naïve Bayes*. Selain itu, studi komparasi terhadap berbagai arsitektur dengan pendekatan *Deep Learning* dilakukan oleh [11].

Penelitian lainnya dilakukan oleh [12] yang melakukan improvisasi terhadap penggunaan algoritma *Random Forest* menggunakan dataset UCI. Pajri dkk. [13] menggunakan algoritma KNN berbasis *Particle Swarm Optimization* untuk analisis sentimen terhadap Tokopedia. Pada kasus analisis sentimen komentar *youtube*, Saputra dkk. [14] menggunakan algoritma *Naïve Bayes* untuk analisis sentimen komentar *youtube* pada video pelayanan pemerintah. Begitupun Munthe dkk. [15] melakukan analisis sentimen pada saluran *youtube food vlogger* menggunakan algoritma *Naïve Bayes*. Udayana dkk. [16] juga menggunakan algoritma *Naïve Bayes* dan *Laplacian Smoothing* pada analisis sentimen komentar *youtube* vaksin nasional.

III. METODE PENELITIAN

Analisis sentimen adalah salah satu teknik yang aktif digunakan secara luas dalam *Machine Learning*. Analisis sentiment menggunakan pemrosesan bahasa alami, analisis teks, komputasi linguistik sehingga dapat memahami makna data tekstual lalu diolah untuk mendapatkan informasi dari data tersebut. Sentimen dalam penelitian ini dibagi menjadi dua kategori label sentimen yaitu positif dan negatif.

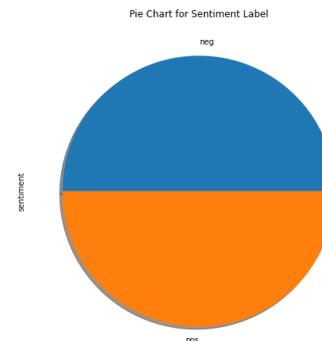
Analisis sentiment dalam penelitian ini menggunakan pendekatan dua algoritma *Machine Learning* yaitu *Support Vector Machine* dan *Random Forest*. Kedua algoritma *machine learning* ini efisien untuk melakukan analisis sentimen. Alur proses penelitian digambarkan dapat dilihat di Gambar 1.



Gambar 1. Alur Penelitian

A. Data Collection

Data yang digunakan merupakan kumpulan komentar yang didapat dari video populer yang mengandung kata kunci BTS dan diambil dari tanggal 1 Januari 2021 hingga 30 Januari 2021 berupa *dataset* dengan nama *BTS Youtube Comment* [17]. *Dataset* terdiri dari 13 kolom diantaranya *query, url, title, upload_date, channel, views, likes, dislikes, comment_count, comment_text, comment_author, comment_date, comment_likes*. Jumlah data yang diambil dalam penelitian ini berjumlah 2213 untuk data berlabel positif dan 2213 data label negatif.



Gambar 2. Jumlah Data

B. Data Pre-processing

Selanjutnya dilakukan tahap *pre-processing* data. Tahap ini merupakan teknik data *mining* dapat mentransformasikan data mentah menjadi format dapat dimengerti untuk menghilangkan komentar yang tidak mengandung opini, seperti nilai yang datanya tidak ada, *noisy* data, dan data yang redundansi [18]. Tahapan *pre-processing* yang diimplementasikan pada penelitian ini diantaranya *remove punctuation* juga *removing all numbers* yang dilanjutkan menggunakan tahap:

- 1) *Stopwords Removal*, ialah suatu proses untuk menghilangkan kata-kata yang tidak memiliki arti dalam *text mining* atau kategori *stopword*;
- 2) *Tokenizing*, merupakan proses membagi dokumen menjadi token untuk mengidentifikasi kata kunci

yang dipisahkan dengan spasi, dan membuang karakter tanda baca [19]. Sehingga, dokumen dapat dipisahkan tiap katanya;

- 3) *Case Folding*, proses ini dilakukan untuk mengubah semua huruf besar seluruh dokumen menjadi kecil atau *lowercase*. Proses tersebut memanfaatkan fitur *transform cases*;
- 4) *Stemming* menggunakan *porter stemmer* untuk mengubah atau menghilangkan semua kata imbuhan yang ada dalam komentar menjadi kata dasar [20].

Hasil data *pre-processing punctuation removal* dan *case folding* dapat dilihat di Tabel 1.

TABEL 1.
DATA PRE-PROCESSING 1

Comment	Punctuation Removal	Case Folding
BTS never Fail to make GREAT SONG	BTS never Fail to make GREAT SONG	bts never fail to make great song
Omg this video made my day love it!!!!	Omg this video made my day love it	omg this video made my day love it
Don't you guys also think that when Jimin shouted on the boat, it sounded like the shout of TOM from TOM AND JERRY	Dont you guys also think that when Jimin shouted on the boat it sounded like the shout of TOM from TOM AND JERRY	dont you guys also think that when jimin shouted on the boat it sounded like the shout of tom from tom and jerry
i can stop laughing to jin 😊	i can stop laughing to jin	i can stop laughing to jin
Aww! So cute and funny!	Aww So cute and funny!	aww so cute and funny

Selanjutnya dilakukan tahap *preprocessing stopwords removal* dan *stemming* yang hasilnya dapat dilihat di Tabel 2.

TABEL 2.
DATA PRE-PROCESSING 2

Stopword Removal	Stemming	Label
bts never fail make great song	bt never fail make great song	neg
omg video made day love	omg video made day love	pos
guys also think jimin shouted boat sounded like shout	guy also think jimin shout boat sound like shout	pos
stop laughing jin	stop laugh jin	pos
aww cute funny!	aww cute funni	pos

C. Feature Extraction

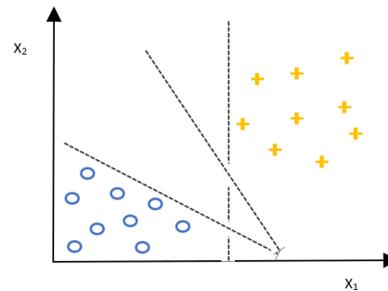
Dalam ekstraksi fitur, sebuah kalimat atau dokumen dibagi menjadi kata-kata untuk membangun matriks fitur berupa baris dan setiap kata membentuk fitur sebagai kolom, dan nilainya adalah jumlah frekuensi kata. Kemudian matriks fitur diteruskan ke tiap *classifier* dan terdapat evaluasi kinerja. Untuk itu perlu dicari transformasi representatif yang mengubah teks menjadi vektor digital atau *Bag of Words* (BOW). BOW merupakan representasi dari kata-kata unik dari ulasan teks yang diidentifikasi menggunakan penandaan

Parts of Speech (PoS) kemudian direpresentasikan menggunakan vektor [21]. Sehingga selanjutnya dilakukan perubahan ulasan menjadi representasi numerik untuk algoritma *machine learning* dengan membuat BOW menggunakan *count vectorizer* dan *tf-idf vectorizer*. *Count vectorizer* mengimplementasikan tokenisasi dan penghitungan kemunculan dalam satu kelas dengan *output* representasi *sparse* matriks. Sedangkan *tfidf vectorizer* mengimplementasikan tokenisasi dan penghitungan berbobot *tf-idf* dalam satu kelas [22].

D. Classification

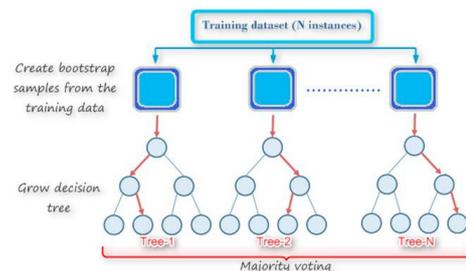
Tahap selanjutnya yaitu klasifikasi dengan pendekatan algoritma SVM dan *Random Forest* menggunakan data *training* dan mengujinya pada data *testing* untuk memberikan polaritas sentiment [23].

Support Vector Machine termasuk algoritma *machine learning* yang dapat dipakai pada kasus klasifikasi juga regresi. Namun, algoritma ini lebih banyak digunakan pada studi kasus klasifikasi. SVM merupakan salah satu metode yang sering digunakan untuk mendeteksi polaritas data tekstual. Ilustrasi SVM digambarkan oleh Gambar 3.



Gambar 3. Ilustrasi SVM

Random forest classifier adalah metode klasifikasi yang diumpamakan sebagai sekelompok pohon keputusan dengan disertai data latih dan fitur acak independen yang memiliki beragam fitur berbeda [24]. *Random forest* merupakan algoritma yang mampu mengklasifikasikan sejumlah besar data dengan akurat yang hasil akhirnya didapat dari penentuan *node* akar dan berakhir dengan beberapa *node* daun. Algoritma *Random Forest* direpresentasikan oleh Gambar 4.



Gambar 4. Random Forest Algorithm

E. Evaluation

Selanjutnya dilakukan evaluasi pengujian untuk mengukur performa algoritma yang digunakan dengan metrik klasifikasi diantaranya *Accuracy*, *Precision*, *Recall* dan *F-Measure*. Evaluasi pengujian didasarkan pada tinjauan positif dan negatif yang dihitung dengan pendekatan yang diusulkan. Notasi TP menunjukkan *True Positives*, FP menunjukkan

False Positives, TN menunjukkan True Negatives dan FN menunjukkan False Negative.

Accuracy didefinisikan sebagai rasio prediksi benar dengan total semua prediksi keseluruhan data. Accuracy dihitung berdasarkan:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

Precision ialah nilai yang didapat dari perbandingan antara true positive dan banyaknya prediksi data yang dinilai positif, dengan rumus sebagai berikut:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Recall ialah nilai yang didapat dari perbandingan true positive dan banyaknya data yang benar positif. Recall disebut juga True Positive Rate (TPR) atau sensitivity dengan rumus sebagai berikut:

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

F-Measure merupakan perbandingan nilai rata-rata antara presisi dan recall yang dirumuskan sebagai berikut:

$$F - Measure = 2 * \frac{Recall * Precision}{Recall + Precision} \tag{4}$$

IV. HASIL DAN PEMBAHASAN

Analisis sentimen BTS youtube comment dijalankan pada komputer dengan processor Intel Pentium CPU 5405U @2.30GHz, RAM 4096 MB, dan sistem operasi Windows 10 Home Single Language 64-bit. Penelitian ini menghasilkan evaluasi dari model machine learning yang telah dibuat.

A. Hasil

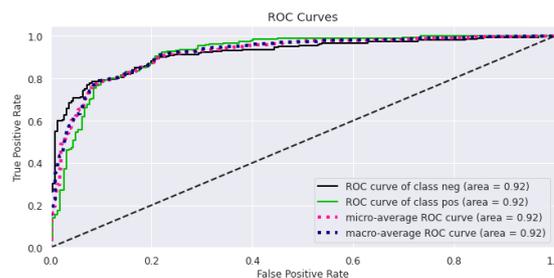
Evaluasi nilai akurasi dengan algoritma SVM menggunakan 3 kernel klasifikasi diantaranya Radial Basis Function (RBF), Linear dan Polynomial. Penelitian ini dilakukan untuk membandingkan data antar kelas berdasarkan ukuran kesamaan menggunakan kernel trick sehingga memudahkan pengklasifikasian data. Data yang digunakan berjumlah 2213 untuk data berlabel positif dan 2213 data label negatif yang merupakan kumpulan komentar dari video youtube yang mengandung kata kunci BTS dan diambil dari tanggal 1 Januari 2021 hingga 30 Januari 2021.

TABEL 3.
EVALUASI SVM

SVM	Training Acc	Testing Acc	Precision		Recall		F1-Score	
			Pos	Neg	Pos	Neg	Pos	Neg
Linear	90	85	83	88	88	82	86	85
RBF	96	83	82	85	85	81	83	83
Poly	96	82	89	78	74	91	81	84

Berdasarkan hasil evaluasi yang telah dilakukan, SVM menggunakan kernel linear menunjukkan akurasi testing data yang paling tinggi yaitu sebesar 85%. Dan untuk data training akurasi yang paling tinggi di kernel RBF dan Polynomial karena memiliki akurasi yang sama yaitu sebesar 96%. Hal ini

bersesuaian dengan fungsi kernel linear yang cocok digunakan untuk data yang dipisahkan secara linear.



Gambar 5. Kurva ROC SVM

ROC merupakan kurva untuk menunjukkan probabilitas dan AUC yang mewakili derajat atau nilai keterpisahan. Kurva ini dapat memberi tahu banyaknya model yang mampu untuk membedakan kelas. Berdasarkan model pembeda antar kelas SVM, kurva untuk kelas negatif berada di area 0,92 begitupun kurva di kelas positif.

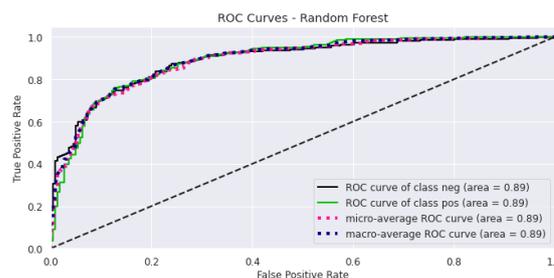
Selanjutnya dilakukan evaluasi keakuratan menggunakan algoritma Random Forest, untuk perbandingannya dilakukan percobaan pada rasio pembagian data training sebesar 90% dan data testing 10%. Perbandingan data tersebut akan digunakan untuk melatih dan menguji model. Data splitting dilakukan untuk melihat akurasi tiap teknik validasi dan membagi data komentar secara acak menjadi dua bagian, sebagian digunakan untuk data latih dan data lainnya digunakan pada data uji

TABEL 4.

EVALUASI RANDOM FOREST

Training Acc	Testing Acc	Precision		Recall		F1-Score	
		Pos	Neg	Pos	Neg	Pos	Neg
82	80	76	86	89	72	82	79

Berdasarkan hasil evaluasi yang telah dilakukan pada algoritma Random Forest, akurasi data mengalami penurunan dibandingkan SVM yaitu 82% untuk data latih dan 80% untuk data uji. Nilai akurasi yang didapat menggunakan parameter jumlah tree dan kedalaman tree terbaik sehingga memiliki pengaruh penting untuk hasil akurasi.



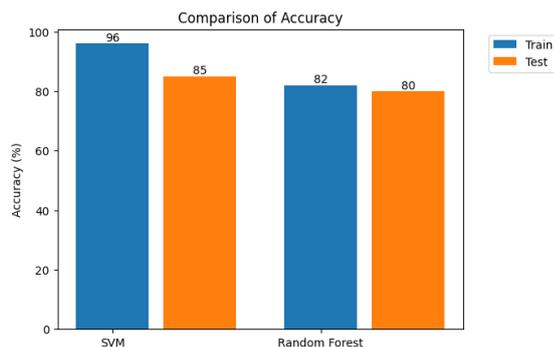
Gambar 6. Kurva ROC Random Forest

Berdasarkan model pembeda antar kelas SVM, kurva untuk kelas negatif berada di area 0,89 begitupun kurva di kelas positif berada di nilai 0,89.

B. Pembahasan

Setelah didapatkan hasil dari percobaan training dan testing menggunakan SVM dan Random Forest, maka dilakukan komparasi hasil akurasi berdasarkan kedua

algoritma tersebut. Perbandingan nilai akurasi dapat dilihat di Gambar 7.



Gambar 7. Perbandingan Akurasi

Berdasarkan hasil komparasi nilai akurasi, algoritma SVM merupakan algoritma dengan akurasi terbaik untuk sentimen analisis dataset BTS *youtube comment* karena mendapat hasil *training* 96% dan *testing* 85%. Hal ini menunjukkan performa yang cukup baik karena dalam penelitian sebelumnya akurasi yang didapat menggunakan algoritma SVM sebesar 89.17%. Sehingga dapat disimpulkan bahwa penelitian ini menunjukkan peningkatan dari penelitian sebelumnya.

Dalam penelitian selanjutnya, metode klasifikasi pada penelitian ini dapat digunakan untuk dataset yang berbeda untuk melihat perbandingan akurasi antara kedua algoritma tersebut. Selain itu bisa diterapkan metode *Deep Learning* untuk pemrosesan datanya untuk dilihat perbandingan nilai akurasinya

V. KESIMPULAN

Hasil evaluasi algoritma *Support Vector Machine* dan *Random Forest* untuk menganalisis sentimen komentar *youtube boygroup* BTS menunjukkan peningkatan nilai akurasi dibandingkan penelitian terkait sebelumnya. Akurasi *training* dengan algoritma SVM menggunakan *kernel RBF* dan *Polynomial* menghasilkan nilai paling tinggi dibandingkan *kernel Linear* yaitu sebesar 96%, sedangkan untuk akurasi *testing*, *kernel Linear* menunjukkan angka lebih tinggi yaitu sebesar 85%. Selain itu, berdasarkan perbandingan pembagian data rasio 90:10 menggunakan algoritma *Random Forest*, nilai akurasi pada percobaan *training* yaitu sebesar 82%, sedangkan nilai akurasi percobaan *testing* sebesar 80%.

Berdasarkan hasil komparasi algoritma SVM dan *Random Forest*, SVM menghasilkan nilai yang lebih tinggi untuk percobaan *training* dan *testing* karena SVM mampu mencari *hyperplane* atau batas keputusan terbaik dengan membagi tiap komentar kedalam dua label sentimen komentar yaitu positif dan negatif. Sedangkan hasil pengujian menggunakan *Random Forest* lebih rendah dibandingkan SVM dikarenakan pembagian parameter *tree* dan kedalaman *tree* yang tidak terbagi dengan sempurna. Sehingga dapat disimpulkan bahwa algoritma SVM merupakan algoritma terbaik untuk melakukan analisis sentimen pada komentar *youtube boygroup* BTS.

DAFTAR PUSTAKA

[1] P. A. Permatasari, Linawati and L. Jasa, "Survei Tentang Analisis Sentimen Pada Media Sosial," *Majalah Ilmiah Teknologi Elektro*, vol. 20, 2021.

[2] M. I. Alrasid and M. I. Gautama, "Daya Tarik BTS Sebagai Brand Ambassador Tokopedia dalam Membentuk Perilaku Mahasiswa Berbelanja Online," *Jurnal Perspektif: Jurnal Kajian Sosiologi dan Pendidikan*, vol. 4, no. 4, pp. 555-566, 2021.

[3] K. Nimala and R. Jebakumar, "Sentiment topic emotion model on students feedback for educational benefits and practices," *Behav. Inf. Technol.*, vol. 40, p. 311-319, 2019.

[4] M. R. Hasan, M. Maliha and M. Arifuzzaman, "Sentiment analysis with NLP on Twitter data," *International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2)*, pp. 1-4, 2019.

[5] A. Onan, "Mining opinions from instructor evaluation reviews: A deep learning," *Comput. Appl. Eng. Educ.*, vol. 28, p. 117-138, 2020.

[6] G. Khanvilkar and D. Vora, "Product Recommendation using Sentiment Analysis of Reviews: A Random Forest Approach," *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 8, no. 2S2, pp. 146-152, 2019.

[7] D. Darwis, E. S. Pratiwi and A. F. O. Pasaribu, "PENERAPAN ALGORITMA SVM UNTUK ANALISIS SENTIMEN PADA DATA TWITTER KOMISI PEMBERANTASAN KORUPSI REPUBLIK INDONESIA," *Jurnal Ilmiah Edutic*, vol. 7, no. 1, 2020.

[8] D. Alita and R. Auliya, "Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random," *Jurnal Komputasi*, vol. 8, no. 2, pp. 50-58, 2020.

[9] M. R. Adrian, M. P. Putra, M. H. Rafialdy and N. A. Rakhmawati, "Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB," *JURNAL INFORMATIKA UPGRIS*, vol. 7, no. 1, pp. 36-40, 2021.

[10] R. Puspita and A. Widodo, "Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS," *Jurnal Informatika Universitas Pamulang*, vol. 5, no. 4, pp. 646-654, 2020.

[11] W. Sutteddy, D. A. R. Agustini, A. Adiwilaga and D. A. Atmanto, "End-To-End Evaluation of Deep Learning Architectures for Offline Handwriting Writer Identification: A Comparative Study," *INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION*, vol. 7, no. 1, pp. 178-185, 2023.

[12] Y. Ying, ., L. Wang, H. Huang and Yang, "An Improved Random Forest Algorithm," *Journal of Physics: Conference Series*, 2020.

[13] D. Pajri, Y. Umaidah and T. N. Padilah, "K-Nearest Neighbor Berbasis Particle Swarm Optimization untuk Analisis Sentimen Terhadap Tokopedia," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 6, no. 2, pp. 242-253, 2020.

[14] P. Y. Saputra, D. H. Subhi and F. Z. A. Winatama, "IMPLEMENTASI SENTIMEN ANALISIS KOMENTAR CHANNEL VIDEO PELAYANAN PEMERINTAH DI YOUTUBE MENGGUNAKAN ALGORITMA NAÏVE BAYES," *Jurnal Informatika Polinema*, vol. 5, no. 3, pp. 209-213, 2019.

[15] M. P. Munthe, A. S. R. Ansori and R. R. Septiawan, "Analisis Sentimen Komentar Pada Saluran Youtube Food Vlogger Berbahasa Indonesia Menggunakan Algoritma Naïve Bayes," *e-Proceeding of Engineering*, vol. 8, no. 6, 2021.

[16] I. A. E. D. Udayana, I. G. I. Sudipa and Risaldi, "Sentimen Analisis Inisiatif Vaksin Nasional Menggunakan Naïve Bayes dan Laplacian Smoothing Pada Komentar Video Youtube," *JURNAL RESISTOR*, vol. 5, no. 2, 2022.

[17] "BTS YouTube Comments," 13 September 2022. [Online]. Available: <https://www.kaggle.com/datasets/seungguini/bts-youtube-comments>. [Accessed 2022].

[18] S. Juniarsih, E. F. Ripanti and E. E. Pratama, "Implementasi Naive Bayes Classifier pada Opinion Mining Berdasarkan Tweets Masyarakat Terkait Kinerja Presiden dalam Aspek Ekonomi," *Jurnal Sistem dan Teknologi Informasi (JustIN)*, vol. 8, no. 3, pp. 239-249, 2020.

[19] J. Joseph and J. R. Jeba, "Information Extraction using Tokenization And Clustering Methods," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 8, no. 4, pp. 3690-3692, 2019.

[20] A. Siswandi, A. Y. Permana and A. Emarilis, "Stemming Analysis Indonesian Language News Text..

[21] M. Geetha and D. K. Renuka, "Improving the performance of aspect based sentiment analysis using," *International Journal of Intelligent*

Networks, vol. 2, pp. 64-69, 2021.

- [22] Z. Kastrati, F. Dalipi, A. S. Imran, K. Pireva Nuci and M. A. Wani, "Sentiment analysis of students' feedback with NLP and deep learning: A systematic mapping study," *Applied Sciences*, vol. 11, no. 9, p. 3986, 2021.
- [23] G. Khanvilkar and P. D. Vora, "Sentiment Analysis for Product Recommendation Using Random Forest," *International Journal of Engineering & Technology*, vol. 7, no. 3, pp. 87-89, 2018.
- [24] I. Klyueva, "Improving Quality of the Multiclass SVM Classification Based on the Feature Engineering," *Proc. - 2019 1st Int. Conf. Control Syst. Math. Model. Autom. Energy Effic. SUMMA*, p. 491-494, 2019.