



ANALISIS SENTIMEN *TWITTER* UNTUK MENGETAHUI KESAN MASYARAKAT TENTANG PELAKSANAAN POMPROV JAWA TIMUR TAHUN 2022 DENGAN PERBANDINGAN METODE *NAÏVE BAYES CLASSIFIER* DAN *DECISION TREE* BERBASIS *SMOTE*

Mas'ud Hermansyah

Sistem dan Teknologi Informasi, masudhermansyah@stie-mandala.ac.id, Institut Teknologi dan Sains Mandala

ABSTRACT

Sentiment analysis is a method used to understand, extract, and automatically process text data to get the sentiment contained in an opinion. Sentiment analysis will be used to process comments made by the community or supporters of each participant of POMPROV East Java 2022 through various media, including Twitter, regarding the progress or results of POMPROV East Java 2022. The number of comments, the authors use data mining methods and algorithms to process the comment data to get information about the POMPROV East Java 2022 event. The Naïve Bayes Classifier and Decision Tree classification algorithms are used as tools to classify comments expressed by users. Based on the results of experiments that have been carried out four times according to the number of data splits and twice based on the algorithm used, it can be concluded that the use of the SMOTE algorithm can increase the accuracy of the various data split compositions used. The best results of the Naïve Bayes Classifier method are found in the 7:3 data distribution which increases the accuracy by 14.52% and the Decision Tree method in the 9:1 data division increases the accuracy by 9.45%.

Keywords: *Sentiment analysis, Naïve Bayes Clasifier, Decision Tree, SMOTE*

ABTRAK

Analisis sentimen adalah metode yang digunakan untuk memahami, mengekstrak, dan secara otomatis memproses data teks untuk mendapatkan sentimen masyarakat dalam menuliskan opininya. Analisis sentimen akan digunakan untuk memproses komentar yang dibuat oleh komunitas atau pendukung dari setiap peserta POMPROV Jatim 2022 melalui berbagai media, termasuk Twitter, mengenai progres atau hasil POMPROV Jatim 2022. Banyaknya komentar, maka penulis menggunakan metode dan algoritma *data mining* untuk mengolah data komentar tersebut untuk mendapatkan informasi tentang *event* POMPROV Jawa Timur 2022. Algoritma klasifikasi *Naïve Bayes Clasifier* dan *Decision Tree* digunakan sebagai alat bantu mengklasifikasi komentar pengguna. Hasil empat percobaan yang dilakukan sesuai dengan banyaknya *split data* dan dua kali berdasarkan algoritma yang digunakan, dapat disimpulkan bahwa penggunaan algoritma *SMOTE* dapat meningkatkan *accuracy* dari berbagai komposisi *split data* yang digunakan. Hasil terbaik metode *Naïve Bayes Clasifier* terdapat pada pembagian data 7:3 yang meningkatkan *accuracy*-nya sebesar 14,52% dan metode *Decision Tree* pada pembagian data 9:1 meningkatkan *accuracy*-nya sebesar sebesar 9,45%.

Kata Kunci: Analisis sentimen, *Naïve Bayes Clasifier, Decision Tree, SMOTE*

1. PENDAHULUAN

Pekan Olahraga Mahasiswa Jawa Timur (POMPROV) 2022 akan diselenggarakan dengan sistem kompetisi olahraga bagi para mahasiswa yang ada di wilayah Provinsi Jawa Timur. POMPROV Jawa Timur 2022 merupakan ajang olahraga *multi event*. POMPROV Jawa Timur 2022 merupakan ajang tahunan olahraga pada tingkat mahasiswa komunitas yang merupakan bagian dari sejarah dan komitmen anak bangsa dalam membangun dunia olahraga bangsa. POMPROV berperan dalam mempromosikan dan menggali benih-benih yang baik, khususnya mahasiswa senior di perguruan tinggi Jawa Timur. Sama seperti tahun sebelumnya pada PORPROV VII Jawa Timur tahun 2022 ini diselenggarakan pada 4 kota berbeda yaitu Jember, Lumajang, Bondowoso, dan Situbondo.

Received September 8, 2022; Revised September 30, 2022; Accepted Oktober 30, 2022

Twitter adalah salah satu media komunikasi paling populer di dunia. Hal ini dibuktikan dengan meningkatnya jumlah pengguna *twitter* yang tercatat di seluruh dunia.. *Twitter* memiliki jumlah pengguna aktif sebesar 313 juta per bulan pada tahun 2016 dan sebagian besar pengguna mengakses *Twitter* melalui perangkat *mobile*, yaitu sebesar 82 persen. Jumlah pengguna yang banyak juga menimbulkan *tweet* yang banyak dari pengguna. Pengguna memberikan berita dan komentar terbaru tentang topik utama dunia saat ini, apa yang saat ini menjadi topik utama dan dikomentari oleh pengguna mengarah pada *trending* topik di *twitter*. [1].

Terdapat beberapa penelitian terdahulu yang melakukan analisis sentimen terhadap data ulasan *online* di *Twitter*, antara lain analisis sentimen untuk mengetahui kesan *Player Game Mobile Legends* menggunakan *Naïve Bayes Classifier* [2], dengan nilai *accuracy* sebesar 80%, *precision* sebesar 76%, dan *recall* sebesar 90,4%. Penelitian lain tentang analisis sentimen tentang opini terhadap performa timnas sepak bola indonesia pada *Twitter* [1] juga menggunakan metode *Naïve Bayes Classifier* dengan menggunakan tiga kali pengujian didapatkan *accuracy* tertinggi yaitu 87%, *precision* positif tertinggi yaitu 93%, *precision* negatif tertinggi yaitu 85%, *recall* positif tertinggi yaitu 75%, dan *recall* negatif tertinggi yaitu 98% pada presentase 90% data *training* dan 10% data *testing*. Penelitian lain mengkombinasikan metode *Modified Term Frequency Scheme* dan *Naïve Bayes Classifier* untuk analisa sentimen pada ulasan pengguna Aplikasi Mandiri *Online* yang memberikan hasil pengujian yang dilakukan *5-fold* mendapatkan nilai akurasi 83%, *recall* 86%, *precision* 76%, *f-measure* 77,70% dengan *fold* ke-3 merupakan *fold* terbaik yang menghasilkan akurasi 85%, *recall* 84,50%, *precision* 81,34%, *f-measure* 82,88% [3]. Pengkombinasian metode klasifikasi juga dilakukan untuk menganalisis sentimen transportasi KRL *Commuter Line* yang memberikan hasil bahwa memiliki perbedaan pada metode *Naive Bayes Classifier* menghasilkan akurasi sebesar 80%, *precision* 66,67%. Pada metode KNN menghasilkan akurasi sebesar 80%, *precision* 100%, dan pada metode *Decision Tree* menghasilkan akurasi sebesar 100%, *precision* 100% [4].

Analisis sentimen adalah sebuah cara yang digunakan untuk mengolah komentar-komentar yang diberikan oleh masyarakat atau pendukung dari masing-masing peserta POMPROV Jawa Timur 2022 melalui berbagai media, salah satunya *Twitter* mengenai berjalannya ataupun sebuah hasil dari POMPROV Jawa Timur 2022. Dengan banyaknya komentar tersebut, penggunaan *data mining* menjadi salah satu cara untuk memproses data komentar-komentar tersebut agar mendapatkan suatu informasi mengetahui tentang berjalannya acara POMPROV Jawa Timur 2022. Algoritma klasifikasi digunakan sebagai alat bantu diantara komentar yang diutarakan oleh *user*. *Naïve Bayes Classifier* dan *Decision Tree* sendiri merupakan metode yang sering digunakan dalam penelitian analisis sentimen dan masing-masing metode tersebut memiliki tingkat akurasi yang berbeda-beda, tergantung dengan kasus atau data yang diberikan dalam perhitungan dan pengklasifikasiannya.

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentimen adalah kombinasi dari penambahan data dan penambahan teks, atau metode yang digunakan untuk memproses pendapat yang berbeda dari konsumen atau pakar melalui berbagai media tentang suatu produk, layanan, atau agensi. Analisis sentimen merupakan metode yang digunakan untuk memahami, mengekstrak, dan secara otomatis memproses data teks untuk mendapatkan sentimen yang terkandung dalam suatu opini. Analisis sentimen terdiri dari tiga jenis opini: positif, negatif, dan netral. Analisis sentimen dapat digunakan untuk menguji reaksi yang diutarakan oleh masyarakat melalui *platform* media sosial.

Saat melakukan survei atau survei berdasarkan analisis sentimen, kita dapat mempelajari apa tanggapan atau reaksi pengguna terhadap topik yang sedang dibahas. [5].

2.2. Twitter

Twitter adalah sebuah *platform* media sosial yang mempunyai layanan *microblogging* yang dimana mengizinkan penggunaannya untuk mengirim pesan secara *real-time*. Pesan di *Twitter* terkenal dengan nama *tweet* [6]. Penggunaan media social *twitter* sangat populer dikalangan masyarakat, tidak terbatas umur hampir seluruh orang kenal, memakai bahkan aktif pada *twitter* [5]. Para pengguna *Twitter* dapat mengirimkan 140 karakter untuk dipost di *Twitter* tersebut, dari 140 karakter ini bisa berupa pujian ataupun komentar pada suatu hal yang dibahas dan juga karena keterbatasan karakter yaitu 140 yang bisa dituliskan, sehingga *tweet* seringkali mengandung bahasa singkatan maupun kesalahan pengejaan. *Twitter* juga mampu

menjadi wadah atau tempat berbagi ide ataupun informasi yang singkat dengan jumlah yang tidak terbatas, sehingga dari berbagai informasi inilah *Twitter* menjadi sumber yang digunakan pada *text classification* [7].

2.3. SMOTE

Metode ini diusulkan pertama kali pada tahun 2002 oleh Chawla, dimana kelas minoritas di-oversamplingkan dengan membuat "*data training sintetis*". SMOTE memodifikasi set data yang tidak seimbang dengan membuat data sintetis baru dari kelas minoritas dengan tujuan meningkatkan kinerja metode klasifikasi. *Overfitting* dapat terjadi di SMOTE dengan data kelas minoritas direplikasi [9]. Data latih sintetik didasarkan pada nilai k tetangga terdekat dalam model algoritma k -nn dan menghasilkan data latih sintetik skala numerik yang berbeda dari data numerik kategoris diukur berdasarkan kedekatan dengan jarak *Euclidean*, sedangkan data kategorikal mempertimbangkan nilai modulusnya. [8].

2.4. Naïve Bayes Classifier

Naive Bayes Classifier merupakan teknik pembelajaran algoritma *data mining* yang menggunakan teknik probabilistik dan statistik. Klasifikasi *Naive Bayes Classifier* memiliki dua proses penting: pembelajaran (*training*) dan pengujian (*testing*) [10]. Persamaan Matematika *Naive Bayes Classifier* sesuai persamaan (1):

$$P(C_j | W_i) = \frac{P(W_i | C_j) P(C_j)}{P(W_i)} \quad (1)$$

Di mana $P(C_j|W_i)$ adalah peluang munculnya kategori j ketika kata i muncul, $P(W_i|C_j)$ adalah peluang kata i termasuk dalam kategori j , $P(C_j)$ adalah peluang munculnya kategori j Probabilitas kejadian, $P(W_i)$, adalah probabilitas kemunculan kategori j .

Terdapat berbagai cara untuk menghitung $P(W_i|C_j)$, tetapi yang paling sederhana adalah melalui persamaan (2). N_{ic} pada persamaan (2) adalah banyaknya data pelatihan dengan *atribut* W_i kategori C_j , N_c merupakan jumlah data kategori C , dan V adalah jumlah kategori. Pada pengolahan pengujian dalam algoritma *Naive Bayes Classifier* menggunakan persamaan (3).

$$P(C_j | W_i) = \frac{N_{ic} + 1}{N_c + |V|} \quad (2)$$

$$P(d) = P(C) \prod_{i=1}^n P(W_i | C) \quad (3)$$

2.5. Decision Tree

Decision Tree merupakan model prediksi yang menerapkan struktur pohon untuk menemukan dan mengambil keputusan yang menyelesaikan suatu masalah dengan mempertimbangkan beberapa kondisi yang ada dalam suatu masalah tersebut. Algoritma ini membutuhkan nilai *entropi* dan *gain*. *Entropi* adalah parameter yang mengukur jumlah keragaman atau perbedaan dalam kumpulan data. *Gain* adalah perolehan informasi sebagai ukuran *validitas atribut* [11]. Persamaan Matematika *Decision Tree* sesuai persamaan (4) [9]:

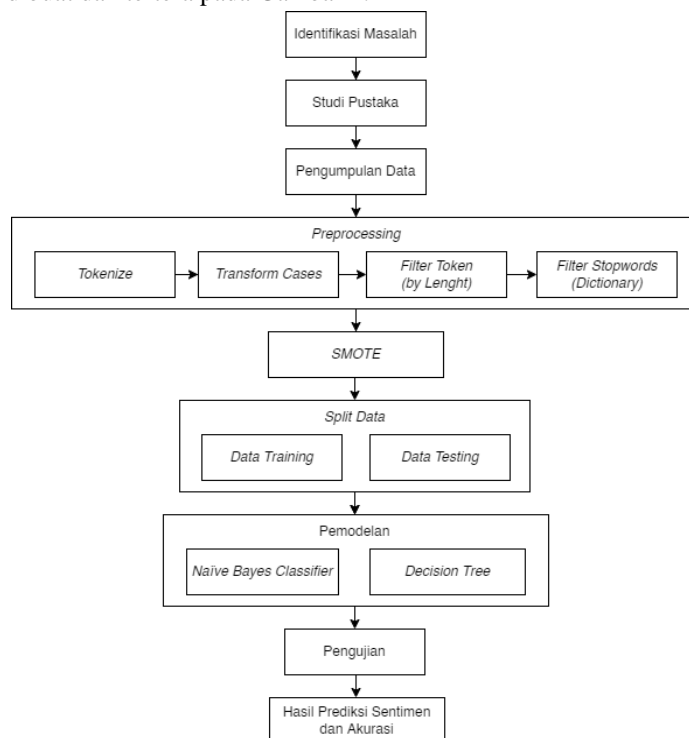
$$Entropy(S) = \sum_{i=0}^n -p_i * \log^2 p_i \quad (4)$$

$$Gain(S, A) = Entropy(S) - \sum_i^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (5)$$

3. METODOLOGI PENELITIAN

Tahap metodologi penelitian dimulai dari identifikasi masalah penelitian, memberikan tujuan pemecahan masalah, studi pustaka untuk meninjau penelitian-penelitian dengan topik yang sama, pengumpulan data dilanjutkan pemrosesan teks mulai dari *preprocessing*, pemodelan dengan metode hingga didapatkan luaran yang diharapkan yaitu sentimen ulasan masyarakat tentang pelaksanaan POMPROV Jawa Timur Tahun

2022 dengan kelas negatif, netral atau positif. Berikut akan dijelaskan mengenai gambaran metodologi penelitian yang akan dibuat dan tertera pada Gambar 1.



Gambar 1. Metodologi Penelitian

Berikut adalah penjelasan dari Alur Metode Penelitian :

1) Identifikasi Masalah

Mengamati dan menemukan permasalahan yang terjadi selama pelaksanaan POMPROV Jawa Timur 2022 dilihat dari komentar masyarakat di *Twitter*.

2) Studi Pustaka

Hal ini bertujuan untuk menemukan teori mana yang digunakan untuk memecahkan masalah yang diteliti dan untuk memberikan peneliti dengan dasar referensi yang kuat.

3) Pengumpulan Data

Proses pengumpulan data dan *input* pada sistem ini dimulai dengan *web scraping* data. Proses *web scraping* dilakukan dengan menggunakan bahasa pemrograman *Python* yang salah satu *tools* yang tersedia untuk *web data scraping*. Keluaran dari sistem ini berupa komentar masyarakat dengan nilai keluaran berupa sentimen positif, negatif, dan netral yang dikategorikan oleh ahli bahasa.

4) *Preprocessing*

Preprocessing adalah salah satu tahap kunci data dalam proses penambangan data. Data yang digunakan untuk proses penambangan tidak selalu dalam kondisi baik untuk diproses. Data dapat memiliki berbagai masalah yang dapat mempengaruhi hasil dari proses penambangan itu sendiri, termasuk: nilai yang hilang, data yang berlebihan, *outlier*, atau format data yang tidak kompatibel. Oleh karena itu, diperlukan tahapan *preprocessing* untuk mengatasi permasalahan tersebut. *Preprocessing* merupakan salah satu tahapan untuk menghilangkan masalah yang dapat mempengaruhi hasil pengolahan data. Pada tahap ini terdapat tahapan yang di lalui diantaranya :

a. *Tokenize*

Pemotongan *string* berdasarkan setiap kata yang membentuk dokumen. Setelah memisahkan setiap kata dari dokumen, *case folding* dilakukan untuk menghilangkan angka, tanda baca, *glyph*, dan karakter lain yang tidak digunakan oleh pengolah kata. Hasil *case folding* harus melewati tahapan *cleaning* yakni proses menghilangkan komponen seperti *tag HTML*, *link*, *url*, *hashtag (#)*, *username (@)*, dan *alamat website* [3].

b. *Transform Cases*

Fitur konversi huruf besar memungkinkan untuk secara otomatis mengonversi semua huruf dalam teks menjadi huruf kecil atau huruf besar secara keseluruhan. Dalam penelitian ini, sebagian besar

teks berupa opini dan sebagian besar huruf kecil, sehingga semua huruf diubah menjadi huruf kecil semuanya. [12].

c. *Filter Token (by Length)*

Penyaringan token (berdasarkan panjang) adalah proses mengekstraksi kata-kata signifikan dari hasil kata. Kata-kata dengan panjang tertentu dihapus. Dalam penelitian ini peneliti menggunakan kata dengan minimal 4 huruf dan maksimal 25 huruf.

d. *Filter Stop Words (Dictionary)*

Filter stop words (kamus) adalah proses menghilangkan kata-kata yang sering muncul tetapi tidak mempengaruhi ekstraksi sentimen dari ulasan. Kata-kata yang mengandung kata kerja dan interogati [13].

Tabel 1. Hasil *Preprocessing* yang Menjadi Masukan untuk *Term Frequency*

No.	<i>Preprocessing</i>	Hasil
1.	<i>Twweet Awal</i>	@andisuwiji23 : Jelang Penutupan Pomprov Jatim 2022, Bojonegoro Tambah 2 Perak & 2 Perunggu #PorprovJatim2022
2.	<i>Tokenize</i>	Jelang Penutupan Pomprov Jatim 2022 Bojonegoro Tambah 2 Perak 2 Perunggu
3.	<i>Transform Cases</i>	jelang penutupan pomprov jatim 2022 bojonegoro tambah 2 perak 2 perunggu
4.	<i>Filter Token (by Length)</i>	jelang penutupan pomprov jatim bojonegoro tambah perak perunggu
5.	<i>Filter Stopwords (Dictionary)</i>	penutupan pomprov jatim bojonegoro tambah perak perunggu

5) *SMOTE*

SMOTE memodifikasi *set data* yang tidak seimbang dengan membuat data sintetis baru dari kelas minoritas dengan tujuan meningkatkan kinerja metode klasifikasi. [9].

6) *Split Data*

Setelah proses pelabelan, data dibagi menjadi data latih dan data uji. Untuk masing-masing pembagian data data latih dan uji ini penulis membagi data dengan presentase 6:4, 7:3, 8:2, dan 9:1. Pada tahap ini juga di uji pada presentase oembagian yang mana yang memberikan akurasi paling tinggi.

7) *Pemodelan*

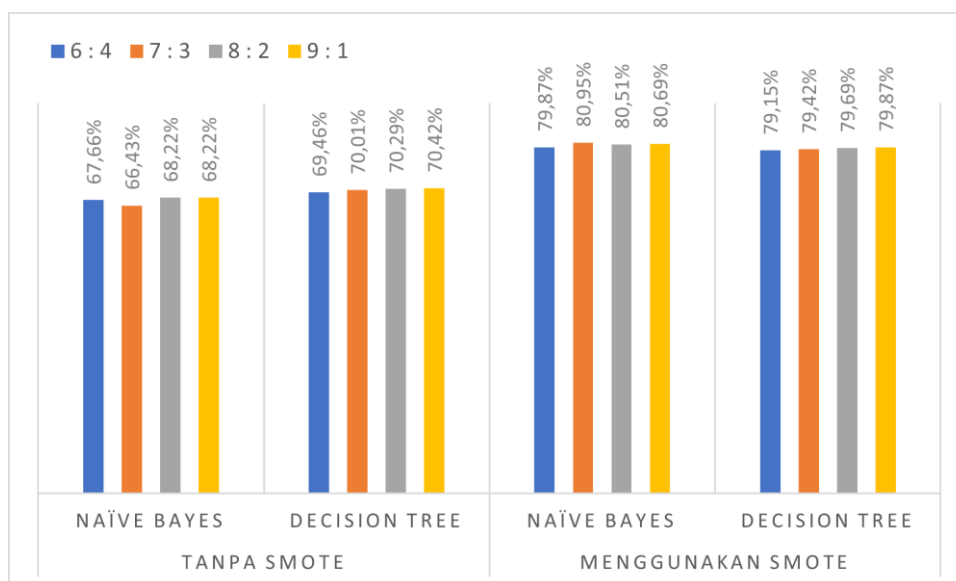
Pada tahap pemodelan dan analisis komparasi klasifikasi yang digunakan adalah algoritma *Naïve Bayes Classifier* dan *Decision Tree*.

8) *Hasil Prediksi dan Akurasi*

Hasil prediksi dan akurasi menggunakan metode *confusion matrix*. *Confusion matrix* adalah tabel yang menyatakan klasifikasi jumlah data uji yang benar dan jumlah data uji yang salah.

4. HASIL DAN PEMBAHASAN

Penelitian didukung dengan tools *Rapidminer 9.10*. Penelitian ini menggunakan *dataset* dari *twitter* mengenai pelaksanaan Pomprov Jatim 2022 berjumlah 727 ulasan yang terdiri dari 470 ulasan positif, 84 ulasan negatif, dan 173 ulasan netral. Pemberian label data ulasan dilakukan oleh pakar bahasa, yaitu guru Bahasa Indonesia di sebuah sekolah. Pengujian dilakukan dengan menguji coba model usulan terhadap dataset. Model usulan yang diuji meliputi *Naïve Bayes Classifier*, *Decision Tree* dan optimalisasi dataset tidak seimbang dengan menggunakan *Synthetic Minority Oversampling Technique (SMOTE)*. Pengujian terhadap algoritma tersebut dilakukan secara bergantian, dimulai dari *split data* dan penggunaan algoritmanya untuk mendapatkan informasi *accuray* tertinggi. Hasil perhitungan diperoleh probabilitas sebagai berikut:



Gambar 2. Tabel Perbandingan Hasil *Accuracy* metode *Naive Bayes Classifier* dan *Decision Tree* dengan *Naive Bayes Classifier + SMOTE* dan *Decision Tree + SMOTE*

Dari hasil klasifikasi dengan mengintegrasikan *SMOTE* dari respon masyarakat tentang pelaksanaan Pomprov Jatim 2022, diketahui *SMOTE* dapat meningkatkan hasil akurasi dari algoritma *Naive Bayes Classifier* dan *Decision Tree*. *Naive Bayes Classifier* dengan *SMOTE* memiliki nilai *accuracy* terbaik di ukuran *split data* 7:3 sebesar 80,95% meningkat 14,52% dari pengujian *Naive Bayes Classifier* sebelumnya sebesar 66,43%. Sedangkan *Decision Tree* dengan *SMOTE* memiliki nilai *accuracy* terbaik di ukuran *split data* 9:1 sebesar 79,87% meningkat 9,45% dari pengujian *Decision Tree* sebelumnya sebesar 70,42%.

5. KESIMPULAN DAN SARAN

Dalam penelitian ini, kami mengoptimalkan pengklasifikasi *Naive Bayes Classifier* dan *Decision Tree* yang dikombinasikan dengan algoritma *SMOTE* (*Synthetic Minority Oversampling Technique*) untuk mendapatkan hasil akurasi tertinggi pada kumpulan data komentar publik tentang event POMPROV di Jawa Timur 2022. Berdasarkan hasil eksperimen yang telah dilakukan sebanyak empat kali sesuai dengan banyaknya *split data* dan dua kali berdasarkan algoritma yang digunakan, dapat disimpulkan bahwa penggunaan algoritma *SMOTE* dapat meningkatkan *accuracy* dari berbagai komposisi *split data* yang digunakan. Hasil terbaik metode *Naive Bayes Classifier* terdapat pada pembagian data 7:3 yang meningkatkan *accuracy*-nya sebesar 14,52% dan metode *Decision Tree* pada pembagian data 9:1 meningkatkan *accuracy*-nya sebesar 9,45%.

Saran yang dituliskan untuk penelitian lanjut adalah perlunya penambahan jumlah dataset yang digunakan, menggunakan atau menambahkan metode klasifikasi lainnya, dan menggunakan media sosial selain twitter

DAFTAR PUSTAKA

- [1] M. Astiningrum, M. Haniah, and Y. rahmat yoga Pradana, "Analisis Sentimen Tentang Opini Terhadap Performa Timnas Sepak Bola Indonesia Pada Twitter," *Semin. Inform. Apl. Polinema*, p. 35–39, 2020.
- [2] D. Sinaga and C. Jatmoko, "Analisis Sentimen Untuk Mengetahui Kesan Player Game Mobile Legends Menggunakan Naive Bayes Classifier," *Semnas LPPM UMP*, vol. V, pp. 540–547, 2020.
- [3] E. P. Nirwandani, Indriati, and R. C. Wihandika, "Analisis Sentimen Pada Ulasan Pengguna Aplikasi Mandiri Online Menggunakan Metode Modified Term Frequency Scheme Dan Naive Bayes," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 3, pp. 1039–1047, 2021.
- [4] N. Tri Romadloni, I. Santoso, S. Budilaksono, and M. Ilmu Komputer STMIK Nusa Mandiri Jakarta, "Perbandingan Metode Naive Bayes, Knn Dan Decision Tree Terhadap Analisis Sentimen Transportasi Krl Commuter Line," *J. IKRA-ITH Inform.*, vol. 3, no. 2, pp. 1–9, 2019.
- [5] E. M. Sipayung, H. Maharani, and I. Zefanya, "Perancangan Sistem Analisis Sentimen Komentar

- Pelanggan Menggunakan Metode Naive Bayes Classifier,” *J. Sist. Inf. UNSRI*, vol. 8, no. 1, pp. 958–965, 2016.
- [6] D. Ayu Wulandari, R. Rohmat Saedudin, R. Andreswari, P. S. Studi, and S. Informasi, “Analisis Sentimen Media Sosial Twitter Terhadap Reaksi Masyarakat Pada Ruu Cipta Kerja Menggunakan Metode Klasifikasi Algoritma Naive Bayes,” *e-Proceeding Eng.*, vol. 8, no. 5, pp. 9007–9016, 2021.
- [7] H. Basri, “Peranmedia Sosial Twitter Dalam Interaksi Sosial Pelajar Sekolah Menengah Pertama Di Kota Pekanbaru (Studi Kasus Pelajar SMPN 1 Kota Pekanbaru),” *Strateg. Bertahan Hidup Petani Penggarap Di Jorong Sarilamak Nagari Sarilamak Kec. Harau Kabupaten Lima Puluh Kota*, vol. 4, no. 1, pp. 1–13, 2017.
- [8] A. A. Arifiyanti and E. D. Wahyuni, “Smote: Metode Penyeimbang Kelas Pada Klasifikasi Data Mining,” *SCAN - J. Teknol. Inf. dan Komun.*, vol. 15, no. 1, pp. 34–39, 2020, doi: 10.33005/scan.v15i1.1850.
- [9] C. Cahyaningtyas, Y. Nataliani, and I. R. Widiyari, “Analisis Sentimen Pada Rating Aplikasi Shopee Menggunakan Metode Decision Tree Berbasis SMOTE,” *Aiti J. Teknol. Inf.*, vol. 18, no. 2, pp. 173–184, 2021, doi: 10.24246/aiti.v18i2.173-184.
- [10] F. Afshoh, “Analisa Sentimen Menggunakan Naïve Bayes untuk Melihat Persepsi Masyarakat Terhadap Kenaikan Harga Jual Rokok pada Media Sosial Twitter,” *Inform. Progr. Stud. Komunikasi, Fak. Inform. D A N Surakarta, Univ. Muhammadiyah*, vol. 12, p. 17, 2017.
- [11] M. F. Asshiddiqi and K. M. Lhaksana, “Perbandingan Metode Decision Tree dan Support Vector Machine untuk Analisis Sentimen pada Instagram Mengenai Kinerja PSSI,” *e-Proceeding Eng.*, vol. 26, no. 10, pp. 44–58, 2020.
- [12] A. T. Jaka, “Preprocessing Text untuk Meminimalisir Kata yang Tidak Berarti dalam Proses Text Mining,” *J. Inform. UPGRIS*, vol. 1, pp. 1–9, 2015.
- [13] Lilyani Asri Utami, “Analisis Sentimen Opini Publik Berita Kebakaran Hutan Melalui Komparasi Algoritma Support Vector Machine Dan K-Nearest Neighbor Berbasis Particle Swarm Optimization,” *J. Pilar Nusa Mandiri*, vol. 13, no. 1, pp. 103–112, 2017.