

KLASIFIKASI KOMENTAR TERHADAP KRIMINAL SAAT MALAM HARI PADA MEDIA SOSIAL *YOUTUBE* MENGGUNAKAN *NAÏVE BAYES*

Vanness Bee ¹⁾, Muhammad Alfa Rizki ²⁾, Nur Rachmat ³⁾

^{1,2,3)} Informatika, Fakultas Ilmu Komputer dan Rekayasa, Universitas Multi Data Palembang
Email : vannes.204@mhs.mdp.ac.id ¹⁾, fairiz04@mhs.mdp.ac.id ²⁾, rachmat.nur91@mdp.ac.id ³⁾

Diterima : 12 Juni 2024 ; Disetujui : 24 Juli 2024 ; Dipublikasikan : 31 Juli 2024

ABSTRAK

Penelitian ini bertujuan untuk mengklasifikasikan komentar mengenai kejahatan kriminal pada malam hari dari media sosial di Youtube dengan menggunakan metode Naïve Bayes. Data komentar dikumpulkan dari video yang telah membahas kejadian kejahatan kriminal pada malam hari, secara khusus yaitu perampokan dengan kata lain adalah begal. Pada Proses Pemrosesan data meliputi normalisasi teks, tokenisasi, pemfilteran, dan stemming untuk mempersiapkan data sebelum dianalisis oleh peneliti. Setelah prapemrosesan, metode Naïve Bayes diterapkan untuk melatih dan menguji pada model klasifikasi. Hasil penelitian telah menunjukkan bahwa metode ini mampu mengklasifikasikan komentar dengan tingkat akurasi sebesar 0.703. Visualisasi data menunjukkan bahwa sentimen positif lebih dominan daripada sentimen negatif dalam komentar yang telah dianalisis. Hasil pada Laporan klasifikasi menunjukkan bahwa precision, recall, dan f1-score masing-masing sebesar 0.64, 0.65, dan 0.64 untuk sentimen negatif, serta 0.75, 0.74, dan 0.75 untuk sentimen positif. Akurasi keseluruhan model adalah 0.70, dengan macro average dan weighted average untuk precision, recall, dan f1-score masing-masing sebesar 0.70. Pada Penelitian ini memberikan pandangan mengenai tanggapan dari masyarakat terhadap kejahatan kriminal pada malam hari serta dapat digunakan untuk mengembangkan strategi yang lebih efektif sehingga untuk mengurangi kejadian perampokan dan kejahatan lainnya.

Kata Kunci : Analisis Sentimen, Klasifikasi, Media Sosial, Naïve Bayes.

ABSTRACT

This study aims to classify comments regarding nighttime criminal activities from social media on YouTube using the Naïve Bayes method. The comment data were collected from videos that have discussed criminal incidents at night, specifically robberies, in other words, muggings. In the data preprocessing process, text normalization, tokenization, filtering, and stemming were conducted to prepare the data before analysis by the researcher. After preprocessing, the Naïve Bayes method was applied to train and test the classification model. The research results have shown that this method can classify comments with an accuracy rate of 0.703. Data visualization indicates that positive sentiment is more dominant than negative sentiment in the analyzed comments. The results in the classification report show that precision, recall, and f1-score are 0.64, 0.65, and 0.64 respectively for negative sentiment, and 0.75, 0.74, and 0.75 for positive sentiment. The overall model accuracy is 0.70, with macro average and weighted average for precision, recall, and f1-score all at 0.70. This study provides insights into public responses to nighttime criminal activities and can be used to develop more effective strategies to reduce robberies and other crimes.

Keywords : Classification, Naïve Bayes, Sentimen Analysis, Social Media.

1. PENDAHULUAN

Kriminal merupakan segala macam bentuk tindakan dan perbuatan yang merugikan secara ekonomis dan psikologis yang melanggar hukum yang berlaku dalam Negara Indonesia serta norma-norma sosial dan agama [1].

Aksi tindakan kejahatan kriminalitas yang sering terjadi ialah pencurian, pembunuhan, perampokan, pemerkosaan, penculikan, penipuan, korupsi. Banyaknya tindakan kejahatan yang terjadi pada belakangan ini yang telah menjadi pusat perhatian masyarakat di berbagai media sosial seperti *Youtube*, *Instagram*, *Twitter*, dan sebagainya, yang telah menjadi sarana utama bagi masyarakat untuk membagikan pengalaman pribadinya, menyampaikan kekhawatiran, serta dapat membahas kejadian-kejadian kriminal terbaru agar dapat membantu dalam menyebarkan informasi dengan cepat dan luas.

Fitur komentar merupakan salah satu aspek penting yang menjadi ciri interaktif bagi media sosial. [2] Pada penelitian ini bertujuan untuk mengklasifikasikan komentar-komentar terkait kriminalitas pada malam hari yang diambil dari platform *Youtube* menggunakan metode *Naïve Bayes*. *Youtube* Merupakan media sosial yang digunakan untuk mengupload video, menonton berbagai video, dan juga bisa berbagi video yang dimana video itu bisa dilihat oleh semua orang [3]. Aplikasi *Youtube* adalah yang paling sering digunakan di Indonesia dan akan terus meningkat hingga 2020 [4].

Dalam penelitian ini, pada klasifikasi komentar-komentar dalam media sosial di *youtube* yang berkaitan dengan kejadian kriminal pada malam hari, terutama perampokan. Penggunaan metode *Naïve Bayes* dalam analisis komentar ini didasarkan untuk menangani data teks dan kemampuannya agar dapat memberikan hasil yang akurat walaupun dengan jumlah data latih yang relatif kecil. Metode ini bekerja dengan cara menghitung probabilitas berdasarkan kemunculan kata-kata dalam komentar, yang kemudian digunakan untuk mengklasifikasikan sentimen komentar menjadi positif, netral, dan negatif.

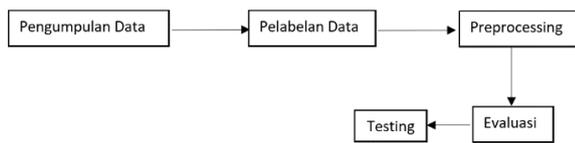
Hasil dari penelitian ini menunjukkan bahwa metode *Naive Bayes* mampu mengklasifikasikan komentar-komentar dengan tingkat akurasi yang memadai. Visualisasi data

mengungkapkan bahwa sentimen positif lebih dominan daripada sentiment negatif dalam komentar yang dianalisis. Penelitian ini juga memberikan bahwa pentingnya analisis sentimen dalam konteks keamanan publik untuk menyelesaikan masalah keamanan di dunia nyata.

Terdapat beberapa penelitian yang telah membahas Analisa sentiment antara lain yaitu, Analisis Sentimen Pengguna Twitter terhadap pemilihan Gubernur DKI Jakarta dengan Metode *Naïve Bayesian Classification* dan *Support Vector Machine*. Hasil dari penelitian ini hasil nilai akurasi yang didapat menggunakan metode *naïve bayes* yaitu *accuracy* 85,77%; *precision* 85.90%, *recall* 85,77, *F-Measure* 85.67%. Sedangkan hasil menggunakan metode *Support Vector Machine* diperoleh lebih baik yaitu *accuracy* 87.80%, *precision* 98.48%, *recall* 87,80%, *F-Measure* 92.64% [5].

2. METODE

Dalam penelitian ini, menggunakan jenis penelitian kuantitatif. Metode penelitian kuantitatif adalah sebuah metode penelitian yang di dalamnya menggunakan banyak angka. [6] Metode yang digunakan yaitu *Naïve bayes*, Metode *Naïve Bayes* merupakan metode yang hanya memerlukan jumlah data latih (training data) yang kecil untuk memilih parameter yang dibutuhkan selama proses klasifikasi. [7] *Naïve Bayes* merupakan salah satu algoritma klasifikasi yang sederhana namun memiliki kemampuan dan akurasi tinggi.[8] *Naive bayes* memiliki 2 tahapan yaitu tahap pelatihan dan tahap klasifikasi. Tahap pelatihan, tahap klasifikasi dilakukan dengan menggunakan jumlah kemunculan kata pada kumpulan data dari setiap kelas. Metode penelitian ini bertujuan untuk menentukan cara yang tepat untuk urutan langkah-langkah yang harus dilakukan dalam memecahkan sebuah permasalahan yang berguna mencapai tujuan tertentu yang ingin diraih. Gambar 1 menunjukkan langkah-langkah yang dilakukan dalam penelitian ini.



Gambar 1. Kerangka Penelitian

2.1 Pengumpulan Data

Dataset diambil berdasarkan kategori dari berita kriminal, yaitu begal. Dataset yang digunakan dalam penelitian ini berupa komentar-komentar yang diambil dari begal berita yang ada di *Youtube* dengan sebanyak 1419 dataset menggunakan *Netlytic*.

2.2 Pelabelan Data

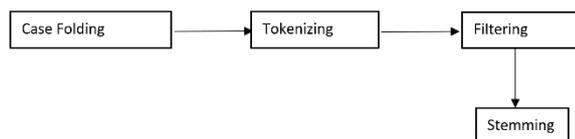
Data yang dikumpulkan oleh peneliti akan dilaberi label sentimen seperti positif, netral, atau negatif. Peneliti telah memberikan label sentimen di setiap komentar komentar yang telah dikumpulkan oleh peneliti yang ada pada Tabel 1.

Tabel 1. Pelabelan Data

Komentar	Sentimen
Tembakan apa itu kok berkali-kali gak lumpuh	Netral
Untuk mengurangi kepadatan penduduk itu gampang. Musnahkan saja para preman, pembegal, perampok, pencuri, koruptor, dan para penjahat berat lainnya	Positif
Wafat gak? Kalau wafat seneng deh	Negatif

2.3 Preprocessing

Preprocessing data melibatkan serangkaian proses untuk mengubah data mentah menjadi format yang lebih sesuai untuk analisis. Proses *preprocessing* ditunjukkan pada Gambar 2.



Gambar 2. Tahapan Preprocessing

2.2.1 Case Folding

Pada tahap ini dilakukan mengubah semua teks dalam dataset menjadi huruf kecil untuk menghindari duplikasi yang disebabkan

oleh perbedaan kapitalisasi yang ditunjukkan pada Tabel 2.

Table 2. Case Folding

Sebelum <i>Case Folding</i>	<i>Case</i>	Setelah <i>Case Folding</i>	<i>Case</i>
AKU YANG TEMBAK NYA. BOLEH NGGA		aku yang tembak begal boleh ngga	
Makasih pakpol		makasih pakpol	

2.2.2 Tokenizing

Memisahkan teks menjadi token atau kata-kata individu, sambil menghilangkan karakter non-alfabetik. Hasil *tokenizing* dapat dilihat pada Tabel 3.

Tabel 3. Tokenizing

Sebelum <i>Tokenizing</i>	Sesudah <i>Tokenizing</i>
Tepat pak, tembak aja begal meresahkan, Pak Polisi	tepat, pak, tembak, aja, begal, makin, meresahkan, bravo, pak, polisi
wanted level bintang satu	wanted, level, bintang, satu

2.2.3 Filtering

Memilih kata-kata yang relevan dengan menggunakan *wordlist* yang telah ditentukan, menghilangkan kata-kata umum atau *stopwords* yang tidak memberikan informasi signifikan. Tabel 4 menunjukkan hasil *filtering*.

Tabel 4. Filtering

Kalimat Asli	Kalimat Setelah <i>Filtering</i>
Harus ada hukuman yang berat	Hukuman berat
Tindakan tegas sangat diperlukan	Tindakan tegas diperlukan
Biar jadi pelajaran bagi yang lain	Biar pelajaran lain

2.2.4 Stemming

Mengurangi kata-kata ke bentuk dasar mereka, sehingga variasi kata yang memiliki makna yang sama dapat dianalisis sebagai satu entitas. Hasil dari proses *stemming* dapat dilihat pada Tabel 5.

Tabel 5. Stemming

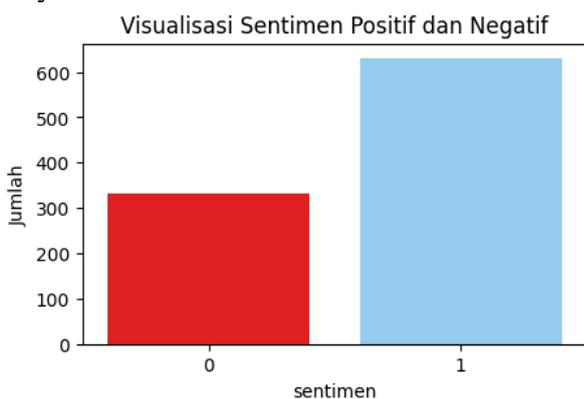
Kalimat Asli	Kalimat Setelah Stemming
Jangan kasih ampun para penjahat	Jangan kasih ampun penjahat
Polisi harus bertindak tegas	Polisi harus tindak tegas
Harus ada hukuman yang berat	Harus ada hukum berat

2.3 Evaluasi Model

Penelitian ini menggunakan metode *Naïve Bayes* dan 1419 dataset yang bersumber dari komentar di Youtube. Pembagian data latih: data uji sebesar 85%:15% atau sebanyak 1206 data latih dan 213 data uji.

3. HASIL DAN PEMBAHASAN

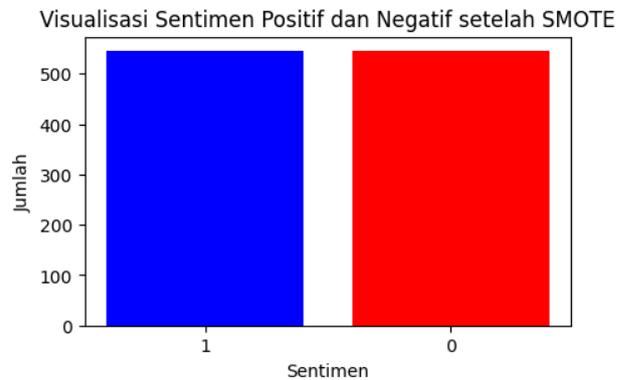
Setelah dilakukan *pre-processing* dan pembagian data latih dan data uji, dilakukan *oversampling* menggunakan SMOTE pada data *training*. Visualisasi sentimen sebelum dan sesudah SMOTE dapat dilihat pada Gambar 3 dan Gambar 4. Berdasarkan Gambar 3, gambaran positif lebih mendominasi dibandingkan sentimen negatif. Hal ini dapat dilihat dari tinggi kolom merah (sentimen positif) yang lebih unggul dari pada kolom biru muda (sentimen negatif). Meskipun interpretasi ini tidak memberikan persentase pasti karena sumbu y tidak memiliki label, namun gambaran umum menunjukkan bahwa opini positif lebih banyak dalam data tersebut.



Gambar 3. Visualisasi Sentiment Positif dan Negatif Sebelum SMOTE

Selanjutnya Gambar 4 menunjukkan dominasi sentimen positif setelah SMOTE, dengan kolom merah (positif) lebih tinggi daripada biru (negatif). Meskipun persentase pasti tidak tersedia, SMOTE terbukti

meningkatkan jumlah data sentimen minoritas (positif), menyeimbangkan distribusi, dan berpotensi meningkatkan performa model klasifikasi sentimen.



Gambar 4. Visualisasi Sentiment Positif dan Negatif setelah SMOTE

Setelah itu, dilakukan pemodelan menggunakan metode *Naïve Bayes* menggunakan data setelah SMOTE. Hasil pemodelan dapat dilihat pada Tabel 6.

Tabel 6. Hasil Pemodelan

	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
Negatif	0.64	0.65	0.64	60
Positif	0.75	0.74	0.75	85
<i>Accuracy</i>	-	-	0.70	145
<i>Macro Avg</i>	0.69	0.70	0.70	145
<i>Weighted Avg</i>	0.70	0.70	0.70	145

Pada Tabel 6, untuk sentimen negatif, diperoleh *precision*, *recall*, dan *f1-score* sebesar 0.64 atau 64%, 0.65 atau 65%, dan 0.64 atau 64%. Sedangkan untuk sentimen positif diperoleh *precision*, *recall*, dan *f1-score* sebesar 0.75 atau 75%, 0.74 atau 74%, dan 0.75 atau 75%. Secara keseluruhan, akurasi yang diperoleh sebesar 0.70 atau 70%. Hal ini menunjukkan bahwa model cukup baik dalam mengklasifikasikan sentimen secara umum.

4. PENUTUP

4.1. Kesimpulan

Algoritma *Naive Bayes* terbukti cukup baik dalam mengklasifikasi sentimen dengan akurasi 70%. Namun, terdapat perbedaan performa saat mengklasifikasikan sentimen positif dan negatif. Untuk sentimen negatif, diperoleh *precision*, *recall*, dan *f-1 score* sebesar 64%, 65%, dan 64% sedangkan untuk sentimen positif diperoleh *precision*, *recall*, dan *f1-score* sebesar 75%, 74%, dan 75%. Dapat disimpulkan pula bahwa model lebih baik dalam mengklasifikasikan sentimen positif dibandingkan negatif.

4.2. Saran

Untuk penelitian selanjutnya, diharapkan dapat menggunakan dataset yang bersumber dari media sosial lain yang sedang *hype* sekarang, seperti Tik-Tok. Kemudian, dapat digunakan juga metode klasifikasi lain dan membandingkannya dengan metode *Naive Bayes*.

5. DAFTAR PUSTAKA

- [1] S. M. Dewi, A. P. Windarto, I. S. Damanik, and H. Satria, "Analisa Metode K-Means pada Pengelompokan Kriminalitas Menurut Wilayah," *Semin. Nas. Sains Teknol. Inf.*, pp. 620–625, 2019.
- [2] A. N. Husna and P. Rianto, "Membaca Komentar di Media Sosial Sebagai Hiburan Reading Comments on Social Media for Entertainment," *J. Mhs. Komun. Cantrik*, vol. 1, pp. 29–40, 2021.
- [3] F. Nur Setiyana and A. Badu Kusuma, "Potensi Pemanfaatan Youtube Dalam Pembelajaran Matematika," *EduMatSains J. Pendidikan, Mat. dan Sains*, vol. 6, no. 1, pp. 71–90, 2021, doi: 10.33541/edumatsains.v6i1.2945.
- [4] K. D. DataBooks, "Salip Facebook, Youtube Paling Aktif di Indonesia," *DataBooks, Kata Data*, p. 2020, 2017.
- [5] E. Putra Nuansa, "Analisis Sentimen Pengguna Twitter Terhadap Pemilihan Gubernur Dki Jakarta Dengan Metode *Naive Bayesian Classification* Dan *Support Vector Machine*," *Inst. Teknol. Sepuluh Nop. Surabaya*, pp. 1–101, 2017.
- [6] M. M. Ali, T. Hariyati, M. Y. Pratiwi, and S. Afifah, "Metodologi Penelitian Kuantitatif dan Penerapannya dalam Penelitian," *Educ. Journal.2022*, vol. 2, no. 2, pp. 1–6, 2022.
- [7] A. Pebdika, R. Herdiana, and D. Solihudin, "Klasifikasi Menggunakan Metode *Naive Bayes* Untuk Menentukan Calon Penerima Pip," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 452–458, 2023, doi: 10.36040/jati.v7i1.6303.
- [8] E. Martantoh and N. Yanih, "Implementasi Metode *Naive Bayes* Untuk Klasifikasi Karakteristik Kepribadian Siswa Di Sekolah MTS Darussa'adah Menggunakan *Php Mysql*," *J. Teknol. Sist. Inf.*, vol. 3, no. 2, pp. 166–175, 2022, doi: 10.35957/jtsi.v3i2.2896.