

Analisis Sentimen *Hate Speech* Pada Berita Lombok Menggunakan Metode *Support Vector Machine*

M. Imam Dinata^{a,1,*}, Hady Rasikhun^{b,2}, Muhammad Rizkillah^{c,3}

^a Sistem dan Teknik Informatika, Universitas Muhammadiyah Mataram, Indonesia

¹ imam.dinata@ummat.ac.id; ² hady.rasikhun@ummat.ac.id; ³ rizkillah@ummat.ac.id

* Penulis Korespondensi

ABSTRAK

Berita *online* telah menjadi kebutuhan mendasar di kalangan masyarakat umum. Berita *online* juga berpotensi dapat merugikan masyarakat *online* atau *netizen* melalui tidak terkendali komentar atau opini mengenai penyebaran berita yang bersifat negatif. *Support Vector Machine* (SVM) merupakan salah satu metode pada *machine learning* yang dapat menentukan tingkat sentimen dari komentar negatif atau positif. Metode SVM terdapat beberapa tahapan yaitu *cleansing*, *tokenizing*, *emojis*, *stopword*, dan *stemming*. Penelitian ini bertujuan menganalisis sentimen *hate speech* pada berita di Lombok dengan metode SVM. Hasil pengujian pada data latih menggunakan metode SVM memberikan tingkat akurasi sebesar 81%.

Riwayat Artikel

Diterima 10 Juni 2024
Diperbaiki 25 Juni 2024
Diterbitkan 30 Juni 2024

Kata Kunci

Berita Online
Analisis Sentimen
SVM



This is an open-access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license

1. Pendahuluan

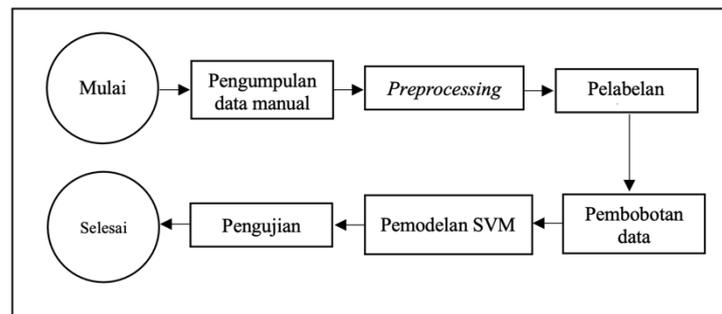
Dengan berkembangnya teknologi informasi dan komunikasi, berita *online* telah menjadi kebutuhan mendasar dalam menyajikan berita di kalangan masyarakat umum [1]. Terutama dalam pemberitaan secara *online* mengenai wisata, peristiwa, ataupun *event* yang terdapat di pulau Lombok dengan tujuan upaya meningkatkan daya tarik tersendiri bagi para wisatawan [2]. Disisi lain, berita *online* juga berpotensi dapat merugikan masyarakat *online* atau *netizen* berupa tidak terkendali komentar atau opini mengenai penyebaran berita negatif terutama membahas sebuah peristiwa penyerangan dan kerusakan di desa Batu Layar, Lombok.

Beberapa peneliti telah melakukan penelitian terkait komentar dari *netizen* yaitu analisis sentimen *hate speech* pada portal berita *online* menggunakan *support vector machine* (SVM) dengan tingkat akurasi sebesar 53.88% [3]. Peneliti kedua membahas analisis sentimen pada ulasan aplikasi WeTv sebagai upaya meningkatkan layanan kepada pengguna menggunakan *K-Nearest Neighbor* dengan tingkat akurasi sebesar 70% [4]. Peneliti ketiga membahas analisis sentimen klasifikasi GoJek pada sosial media Twitter dengan validasi silang 10 kali lipat dan kerjel linier serta kerjel RBF [5]. Hasil pelabelan data secara manual, memperoleh hasil akurasi sebesar 79,19% [5]. Penelitian keempat membahas analisis sentimen pinjaman *online* pada media Twitter menggunakan SVM dengan validasi silang dan menghasilkan akurasi sebesar 62,99% [6]. Peneliti kelima juga membahas analisis sentimen terhadap data Twitter tentang KPK Republik Indonesia menggunakan metode *confusion matrix* dengan menerapkan metode SVM yang menghasilkan akurasi sebesar 82% [7].

Metode SVM telah berhasil diterapkan secara luas dalam berbagai konteks penelitian. SVM adalah alat yang efektif dalam mengklasifikasi dan menganalisis data kompleks seperti berita atau opini dari *netizen*. Penelitian ini bertujuan menganalisis sentimen *hate speech* dari *netizen* pada portal berita di Lombok menggunakan SVM. Penelitian ini menjanjikan kontribusi yang signifikan dalam meningkatkan pemahaman dan pengambilan keputusan terkait isu penting di Lombok.

2. Metode

Subjek penelitian ini yaitu analisis sentimen *hate speech* terhadap berita Lombok dengan menggunakan algoritma *Support Vector Machine* (SVM). *Dataset* yang digunakan sebanyak 500 komentar yang didapatkan dari berita Lombok yang membahas mengenai penyerangan dan kerusakan di desa Batu Layar. Setiap komentar dilakukan analisis dengan menggunakan SVM untuk menentukan tingkat sentimen dari komentar yaitu komentar negatif atau positif. Dataset tersebut terlebih dahulu harus dihilangkan kata penghubung (*stopword*) untuk dapat digunakan sebagai data *input* pada proses *document preprocessing* [3]. Tahapan penelitian meliputi pengambilan data, *preprocessing*, pelabelan, pembobotan data, pemodelan SVM, pengujian, dan akurasi [8].



Gambar 1 Tahap Penelitian

2.1 Pengambilan Data

Pengambilan data pada penelitian ini dilakukan secara manual dengan memilih dan memilah komentar sebanyak 500 data dengan *keyword* pencarian yang digunakan yaitu peristiwa penyerangan dan kerusakan di desa Batu Layar Lombok. Setiap data diolah dan diklasifikasi menjadi komentar positif dan negatif.

2.2 Preprocessing

Data yang didapatkan dari komentar masih bersifat data *unstruktur*, maka dilakukan proses *preprocessing* [9]. Tahap ini akan menghilangkan data-data yang bersifat duplikat, *noise*, dan lainnya.

a. *Cleansing*

Pada tahap *cleansing* dilakukan proses membersihkan data yang baru diambil dari media sosial (facebook, twitter, Instagram, dan lain). Dalam proses ini, kalimat dan karakter yang tidak diperlukan seperti hashtag, URL, @, retweet, dan lainnya harus dibersihkan.

b. *Tokenizing*

Pada tahap *tokenizing* dilakukan proses memotong kalimat menjadi kata-kata yang dibagi-bagi atau terpisahkan, sehingga menjadi kata satu-satu.

c. *Emojis*

Pada tahap *emojis* dilakukan proses untuk menghilangkan *emotion* atau emot yang ada pada di media sosial sehingga kalimat tidak ada berisikan tanda emot.

d. *Stopword*

Pada tahap *stopword* dilakukan proses untuk mengambil kata-kata penting dari hasil *tokenizing*. Proses ini akan menghilangkan kata-kata yang tidak memiliki makna misal: “di”, “lain”, “atau” dan lain-lain.

e. *Stemming*

Pada tahap *Stemming* dilakukan proses pencarian dari hasil *stopword* dengan membuang kata-kata berimbuhan seperti “me-kan”, ”ber-kan” dan lain-lain.

2.3 Pelabelan Data

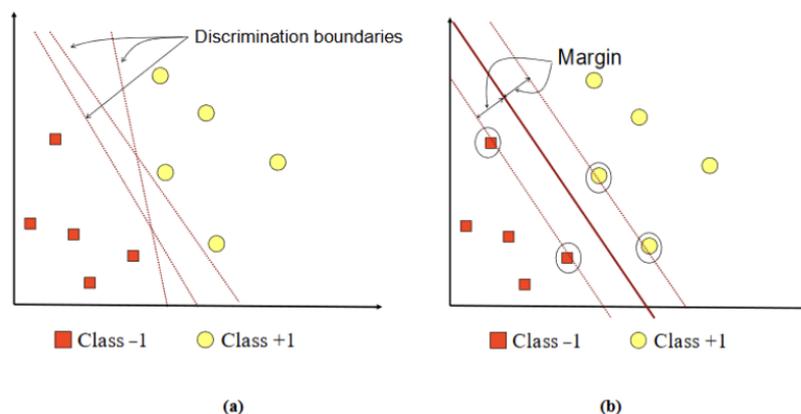
Pada proses pelabelan ini, data dari sosial media yang didapatkan selanjutnya dibagi menjadi ke dalam kelas *hate speech* atau tidak. Penilaian parameter tidak *hate speech* yaitu tidak mengandung indikasi kalimat *hatespeech* [10].

2.4 Pembobotan Data

Tahap ini yaitu melakukan pembobotan setiap kata setelah tahap *pre-processing*. Di tahap ini setiap kata akan diubah ke dalam bentuk numerik karena proses klasifikasi hanya dapat dilakukan dalam bentuk numerik [10]. Proses ini bertujuan menghasilkan label kata atau sentimen kata yang memiliki nilai atau bobot pada sebuah *term* berdasarkan tingkat kepentingan pada dokumen [11].

2.5 Support Vector Machine

Dalam proses pemodelan dengan algoritma *Support Vector Machine* (SVM) untuk mengetahui pola data *training*, proses klasifikasi ditentukan menjadi 2 kelas yaitu *hate speech* dan tidak [11]. *Dataset* terbagi menjadi data *training* dan data *testing*.



Gambar 2 Ilustrasi Support Vector Machine

3. Hasil dan Pembahasan

3.1 Pengumpulan Data

Pada proses pengambilan data, hal utama dilakukan yaitu mengambil 500 data komentar dari Facebook yang membahas mengenai peristiwa penyerangan dan kerusakan di desa Batu Layar, Lombok. Data terkumpul dilakukan penghapusan duplikasi data untuk menghilangkan komentar yang duplikat, dan disimpan dalam jenis format *csv* atau *excel*. Sampel data yang diambil secara manual sebagai berikut:

Khairul	Kabarnya anak kandung Kades Rembitan
Niaa	gubuk mna itu pak.. Yg sbnernya.. Dn smga di hkum. Seberat2 nya.
Ayundia	sudah fix dari polisi kalau yang nyerang itu dari ketara & rambitan
Jun	oke bro , udah aman kalo ketare karna udah hadir pak kades ketare buat jelasin.
Willy	Maap bro bisa di jelaskan yg menyerang desa motong ini dari desa mana? Klo bukan dari de
Sukarna	Mestinya admin hati² upprove berita yg menyebutkan nama desa atau tempat kejadiannya b
Andini	banyak org menonton video...tapi tidak membaca judul/caption...memang pernah di sebutkn
Abhi Alghina	kayaknya bukan admin yg nyebut tapi komentator netizen.🙏🙏

Gambar 3 Hasil Pengambilan Data Manual

3.2 Pelabelan Data

Pelabelan data dilakukan untuk mengetahui data yang akan digunakan untuk data uji dan data latih. Sampel dari proses pelabelan data sebagai berikut:

Putra	Itu bukan dari desa ketare pastikan dulu alamat kronologi	Tidak
Septy	Apapun bentuk masalah nya , emang harus bawa2 massa	Tidak
Rofix	bukan ketare boss..tapi rembitan. edit postingan endah.	Hate Speech
Akara	bukan orang ketare cook cook. Sok tau kamu itu	Hate Speech
Darwis	SEHUBUNGAN SG HAL TSB UTK MENJAGA KONDUSIFITA	Tidak
Didix	🙏🙏🙏 dri peristiwa itu apa korban dpt santunan, mereka	Tidak
Herman	ndk boleh arip kepada anjing anjing yg sudah melukai ora	Hate Speech
Khairul	Kalau memang sampai segitunya lagi menyerang kampur	Hate Speech
Ampenan	Banyak cctv yg terpasang di setiap lampu merah,pertoko2	Tidak

Gambar 4 Hasil Data Dilabelkan

3.3 Proses Cleansing

Setelah dilakukan pelabelan dilakukan proses *cleansing* sebagai pembersihan data [12]. Pada proses ini terdapat data baru, maka harus dihilangkan atau dilakukan proses penghapusan terkait hastg, url, @, retweet dan lain-lain. Sampel dari proses *cleansing* data sebagai berikut:

Data Mentah Jadi miris lihatnya saya asli dgn Lombok merasa sedih lihat teman ² ribut saling bacok wsesama org lombok lagi. #mari kita saling kasih mengasihi dan saling menyayangi. Sesuai dgn nama LOMBOK itu sendiri yakni Polos bender.
Cleansing Jadi miris lihatnya saya asli dgn Lombok merasa sedih lihat teman ² ribut saling bacok wsesama org lombok lagi

Tabel 1 Proses Cleansing

3.4 Preprocessing

Text *preprocessing* merupakan proses mengubah data yang belum terstruktur menjadi data yang terstruktur [12]. Pada proses ini terdapat beberapa tahapan yaitu *tokenizing*, *casefolding*, *emotion*, *stopword*, dan *stemming*. Sampel dari *preprocessing* data sebagai berikut:

Casefolding kayaknya bukan admin yg nyebut tapi komentator netizen 🙏🙏
Tokenizing ['kayaknya' 'bukan' 'admin' 'yg' 'nyebut' 'tapi' 'komentator' 'netizen' '🙏🙏']
Emotion ['kayaknya' 'bukan' 'admin' 'yg' 'nyebut' 'tapi' 'komentator' 'netizen']
Stopword ['kayaknya' 'bukan' 'admin' 'nyebut' 'komentator' 'netizen']
Stemming ['kayak' 'bukan' 'admin' 'komentator' 'netizen']

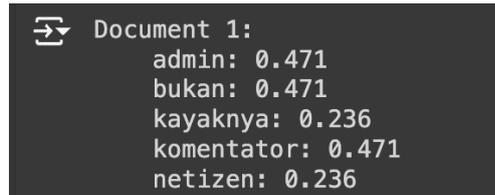
Tabel 2 Proses Preprocessing

3.5 Pembobotan Data TDF-IDF

Metode TF-IDF merupakan metode untuk melakukan perhitungan pembobotan pada setiap kata yang paling sering digunakan pada *information retrieval*. Metode ini diproses dengan cara menghitung nilai *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) pada setiap token (kata) di setiap dokumen. Metode ini memiliki rumus [3] :

$$W_{dt} = t_{fdt} \times IDF_t \tag{1}$$

- d : dokumen ke-d
- t : kata ke-t dari kata kunci
- W : bobot dokumen ke-d terhadap kata ke-t
- f : banyaknya kata yang dicari pada sebuah dokumen
- IDF : *Inversed Document Frequency*

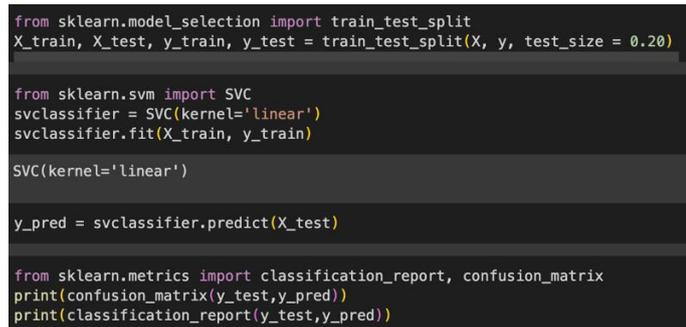


```
Document 1:  
admin: 0.471  
bukan: 0.471  
kayaknya: 0.236  
komentator: 0.471  
netizen: 0.236
```

Gambar 5 Contoh Hasil TF-IDF

3.6 Pemodelan SVM

Pada proses perhitungan dengan metode *Support Vector Machine* dilakukan dengan Google Collab, dimana saat melakukan pengujian model dilakukan membagi data komentar positif dan komentar negatif. Data yang digunakan sebanyak 400 data latih dan 100 data uji.



```
from sklearn.model_selection import train_test_split  
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.20)  
  
from sklearn.svm import SVC  
svclassifier = SVC(kernel='linear')  
svclassifier.fit(X_train, y_train)  
  
SVC(kernel='linear')  
  
y_pred = svclassifier.predict(X_test)  
  
from sklearn.metrics import classification_report, confusion_matrix  
print(confusion_matrix(y_test, y_pred))  
print(classification_report(y_test, y_pred))
```

Gambar 6 Contoh Pemodelan SVM Pada Google Collab

3.7 Hasil Akurasi

Hasil dari pengujian data latih menggunakan data pelatihan pada metode SVM memberikan hasil tingkat keakuratan, dengan nilai akurasi sebesar 81%. Hasil lain mengidentifikasi *F-1 Score* dari hasil perhitungan dengan menggunakan metode SVM yaitu sebesar 70%.

	precision	recall	f1-score	support
0	0.46	0.26	0.33	90
1	0.92	0.95	0.94	285
2	0.83	0.85	0.84	125
micro avg	0.81	0.81	0.81	500
macro avg	0.74	0.69	0.70	500
weighted avg	0.88	0.81	0.81	500

SVM, Accuracy Score: 0.8132822271535202

Gambar 7 Hasil Perhitungan dengan Menggunakan Metode SVM

4. Kesimpulan

Hasil implementasi metode *Support Vector Machine* (SVM) dapat menyelesaikan prediksi pada analisis sentimen *hate speech* berita Lombok dengan jumlah total 500 data. Data dibagi menjadi 400 data latih dan 100 data uji. Setiap data dikategori menjadi *hate speech* dan tidak dan diproses untuk mendapatkan tingkat akurasi sebesar 81%. Keberlanjutan penelitian sebagai upaya meningkatkan nilai akurasi disarankan untuk menambahkan jumlah data dan menambahkan *learning rate*.

Daftar Pustaka

- [1] Huda, I.A., "Perkembangan Teknologi Informasi dan Komunikasi (TIK) Terhadap Kualitas Pembelajaran Di Sekolah Dasar". Jurnal Pendidikan Dan Konseling (JPDK), 2(1), pp.121-125,2020,doi: <https://doi.org/10.31004/jpdk.v1i2.622>.
- [2] Suardana, I.K. "Peran Media Online Firstlomboktour.com Dalam Memasarkan Pariwisata Budaya Di Pulau Lombok. Jurnal ekonomi vol 1 no 2 (2022), 2022, doi: <https://doi.org/10.53977/jw.v1i2.710>.

- [3] Aniq Noviciatie Ulfah*1 , M. Khairul Anam2.2020. Analisis Sentimen Hate Speech Pada Portal Berita Online Menggunakan Support Vector Machine (SVM).Jurnal Teknik Informatika dan Sistem Informasi.Vol. 7, No. 1, April 2020, Hal. 1-10. DOI: <https://doi.org/10.35957/jatisi.v7i1.196>
- [4] Faridhotun, N. Haerani, E. Candra, R.M. 2023. Analisis Sentimen Ulasan Aplikasi WeTV Untuk Peningkatan Layanan Menggunakan Metode K-Nearest Neighbor. Journal Of Information System Research (JOSH). doi: <https://doi.org/10.47065/josh.v4i3.3349>
- [5] N. Fitriyah, B. Warsito, D. Asih, dan I. Maruddani, “Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (SVM),” JURNAL GAUSSIAN, vol. 9, no. 3, hlm. 376–390, 2020.
- [6] D. S. Utami dan A. Erfina, 2021. Analisis Sentimen Pinjaman Online Di Twitter Menggunakan Algoritma Support Vector Machine (SVM). vol. 1 (2021): meningkatkan inovasi dan daya saing di era pandemi melalui riset pada bidang sistem informasi dan manajemen informatika
- [7] D. Darwis, E. Shintya Pratiwi, A. Ferico, dan O. Pasaribu.2020.penerapan algoritma svm untuk analisis sentimen pada data twitter komisi pemberantasan korupsi republik indonesia.Jurnal Ilmiah Edutic Pendidikan dan Informatika.DOI:10.21107/edutic.v7i1.8779
- [8] G. Song, “Sentiment analysis of Japanese text and vocabulary learning based on natural language processing and SVM,” J Ambient Intell Humaniz Comput, 2021, doi: 10.1007/s12652-021-03040-z.
- [9] R. Sandoval-Almazan dan D. Valle-Cruz, “Sentiment Analysis of Facebook Users Reacting to Political Campaign Posts,” Digital Government: Research and Practice, vol. 1, no. 2, Apr 2020, doi: 10.1145/3382735.
- [10]J. Putri, A. S. Syafira, M. E. Purbaya, dan D. Purnomo, “Analisis Sentimen ECommerce Lazada pada Jejaring Sosial Twitter Menggunakan Algoritma Support Vector Machine,” Jurnal TRINISTIK: Jurnal Teknik Industri, Bisnis Digital, dan Teknik Logistik, vol. 1, no. 1, hlm. 16–21, Mar 2022, doi:10.20895/trinistik.v1i1.447.
- [11]H. C. Husada dan A. S. Paramita, “Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM),” Teknika, vol. 10, no. 1, hlm. 18–26, Feb 2021, doi: 10.34148/teknika.v10i1.311.
- [12]P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, dan W. Gata, “Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi,” Jurnal Teknoinfo, vol. 14, no. 2, hlm. 115, Jul 2020, doi: 10.33365/jti.v14i2.679.
- [13]N. Yadav, O. Kudale, A. Rao, S. Gupta, and A. Shitole, “Twitter sentiment analysis using supervised machine learning,” in Intelligent Data Communication Technologies and Internet of Things. Singapore: Springer, 2021,pp. 631–642.
- [14]M. Z. Ali, K. Javed, E. U. Haq, and A. Tariq, “Sentiment and emotion classification of epidemic related bilingual data from social media,” 2021, arXiv:2105.01468. [Online].Available: <http://arxiv.org/abs/2105.01468>
- [15]M. E. Basiri, S. Nemat, M. Abdar, E. Cambria, and U. R. Acharya, “ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis,” Future Gener. Comput. Syst., vol. 115, pp. 279–294, Feb. 2021.