

# Analisis Komparasi Algoritma *Machine Learning* untuk *Sentiment Analysis* (Studi Kasus: Komentar YouTube “Kekerasan Seksual”)

Chandra Ayunda Apta Soemedhy<sup>1</sup>, Nora Trivetisia<sup>2</sup>, Nawang Anggita Winanti<sup>3</sup>, Dwi Puspa Martiyaningsih<sup>4</sup>, Tri Wulandari Utami<sup>5</sup>, Sudianto Sudianto<sup>6\*)</sup>

<sup>1,2,3,4,5,6</sup>Jurusan Teknik Informatika, Fakultas Informatika, Institut Teknologi Telkom, Purwokerto  
<sup>1,2,3,4,5,6</sup>Jalan D.I. Panjaitan No. 128, Purwokerto, Banyumas, Jawa Tengah, Indonesia  
email: 19102301@ittelkom-pwt.ac.id<sup>1</sup>, sudianto@ittelkom-pwt.ac.id<sup>6</sup>

**Abstract** – Cases of sexual violence in the last decade have been rampant in Indonesia. Cases of sexual violence are increasingly exposed, along with the increasing use of social media. One of them is violence against women. Cases of sexual violence often cause various kinds of stigma in the community, so a study is needed to determine how the community responds to these cases. This study aims to determine the public's response to cases of sexual harassment using sentiment analysis. The data used in this study were sourced from YouTube comments titled "Kasus Bunuh Diri NW: Bripda Randy Tersangka, Penanganan Polisi Dikritik | Narasi Newsroom" The method used is machine learning algorithms such as the SVM algorithm, Naive Bayes, and Random Forest. The three algorithms are then compared to determine which algorithm is the best in analyzing public sentiment on cases of sexual harassment. As a result of a comparison of the three machine learning algorithms, Random Forest obtained the best accuracy rate of 78% compared to the other two algorithms in conducting sentiment analysis on YouTube comments about sexual harassment topics.

**Keywords** – Algorithm comparison, Naïve Bayes, Random Forest, sexual harassment, SVM

**Abstrak** – Kasus kekerasan seksual dalam dekade terakhir marak terjadi di Indonesia. Kasus kekerasan seksual semakin terekspos seiring dengan meningkatnya penggunaan media sosial. Salah satunya kekerasan yang terjadi pada kaum perempuan. Kasus kekerasan seksual acap menimbulkan berbagai macam stigma dari masyarakat sehingga diperlukannya sebuah penelitian untuk mengetahui bagaimana tanggapan masyarakat terhadap kasus tersebut. Penelitian ini bertujuan untuk mengetahui tanggapan masyarakat terhadap kasus pelecehan seksual menggunakan analisis sentimen. Data yang digunakan dalam penelitian ini bersumber dari komentar YouTube dengan judul “Kasus Bunuh Diri NW: Bripda Randy Tersangka, Penanganan Polisi Dikritik | Narasi Newsroom”. Metode yang dipakai yaitu algoritme pembelajaran mesin seperti algoritme SVM, Naive Bayes, dan Random Forest. Ketiga algoritme tersebut kemudian dibandingkan untuk mengetahui algoritme mana yang paling baik dalam menganalisis sentiment masyarakat terhadap kasus pelecehan seksual. Hasil dari perbandingan ketiga algoritme Pembelajaran mesin, Random Forest memperoleh tingkat akurasi terbaik sebesar 78% dibandingkan kedua algoritme lainnya dalam melakukan analisis sentimen terhadap komentar YouTube tentang bahasan pelecehan seksual.

**Kata Kunci** – Analisis sentimen, Naive Bayes, pelecehan seksual, Random Forest, SVM

\*) penulis korespondensi: sudianto  
Email: sudianto@ittelkom-pwt.ac.id

## I. PENDAHULUAN

Kekerasan sering terjadi di Indonesia menimpa orang tak berdaya seperti anak-anak, perempuan, dan lansia. Kekerasan banyak jenisnya, salah satunya kekerasan seksual pada perempuan. Kekerasan pada kalangan tertentu, khususnya perempuan menjadi ketakutan tersendiri, apalagi bagi perempuan yang sering melakukan aktifitas di luar rumah maupun yang berdiam diri di dalam rumah [1]. Berdasarkan catatan tahunan Komisi Nasional Perempuan, data kasus kekerasan seksual terhadap perempuan sepanjang 2019 sebanyak 431.471 kasus. Ditambah lagi pada bulan Oktober 2020 terdapat 4.200 kasus [2]. Selain itu, peran media sosial yang mengekspos korban daripada pelaku melalui *headline* berita membuat stigma bahwa kejadian yang menimpa korban merupakan kesalahan korban sendiri, karena tidak bisa menjaga diri dari segi pergaulan maupun cara berpakaian. Fenomena tersebut membuat korban lainnya menjadi bungkam. Namun, dari sisi lain terdapat juga sebagian masyarakat memberikan dukungan terhadap korban dengan menyatakan bahwa korban tidak bersalah. Oleh karena itu, untuk mengetahui tanggapan masyarakat mengenai kasus pelecehan seksual pada platform media sosial, maka diperlukan analisis sentimen pada media sosial yaitu bersumber dari YouTube.

Analisis sentimen adalah komputasi dari sentimen, opini, dan emosi terhadap suatu objek yang ditumpahkan dalam bentuk teks. Analisis ini digunakan untuk mengekstrak atribut dan komponen yang ada didalamnya serta untuk menentukan komentar yang terkategori bermakna negatif atau positif [2]. Analisis sentimen dapat dilakukan dengan beberapa algoritme Pembelajaran mesin diantaranya NV (Naive Bayes) [3] [4] [5], RF (Random Forest) [6] [7], SVM (Support Vector Machine) [8] [9], LR (Logistic Regression) [10], KNN (K-Nearest Neighbors) [11], dan masih banyak lagi. Adapun untuk analisis sentimen pada penelitian ini menggunakan algoritme Naive Bayes [12], Random Forest [13], dan SVM [14] yang memiliki akurasi yang baik untuk melakukan analisis sentimen. Dari analisis sentimen yang dilakukan, diharapkan mampu untuk mengetahui berapa jumlah persentase masyarakat yang mendukung dan tidak mendukung pelecehan seksual khususnya pada korban. Selain itu, dengan menerapkan ketiga algoritme, diharapkan dapat mengetahui algoritme mana yang memiliki akurasi terbaik untuk analisis sentimen terhadap komentar YouTube tentang pelecehan seksual.

## II. PENELITIAN TERKAIT

Beberapa penelitian yang terkait pada analisis sentimen yaitu penelitian oleh Rabathy dan Komala [15] dengan judul “Pelecehan Seksual Di Ruang Publik” menjelaskan bahwa pelecehan seksual merupakan masalah sensitif yang membutuhkan tahapan atau cara khusus untuk mengatasinya. Ditambah lagi, kejelasan hukum bagi pelaku seharusnya mulai diterapkan untuk memberikan efek jera kepada pelaku. Masyarakat sekitar juga diharuskan untuk meningkatkan tingkat kewaspadaan serta membuka pikiran mereka sehingga tidak hanya korban pelecehan seksual yang disalahkan. Hal tersebut dapat membantu para korban untuk memulihkan trauma dari pelecehan seksual yang mereka alami.

Penelitian selanjutnya oleh Nasution dan Hayaty [14] memberikan hasil yang lebih baik untuk algoritme K-Nearest Neighbor yang berjudul “Perbandingan Algoritme Analisis Sentimen KNN dan SVM untuk Presiden Amerika Serikat”. Proses lama 0,0160 detik tanpa KFold dan verifikasi timbal balik dengan KFold 0,1505 detik. Sebaliknya, untuk akurasi metode SVM lebih baik, dimana hasil tanpa KFold Cross Validation 88,76% dan dengan skor 89,70% dengan KFold Cross Validation.

Penelitian yang dilakukan oleh Baskoro, *et al* [13] dengan judul “Analisis Sentimen Pelanggan Hotel di Purwokerto (Studi Kasus Ulasan Pelanggan Pada Situs (TRIPADVISOR) Menggunakan metode Random Forest dan TF-IDF”. Memiliki hasil “Akurasi model mencapai 87,23% dengan tahapan mengoreksi secara manual komentar negatif pada rating 30 dari rating 10 - 50 dan preprocessing pembenaran/konversi kata *slang word* sesuai kata dalam KBBI dan penambahan daftar *stopword* yang ditambah *stopword* buatan sendiri. Akurasi ini didapat tanpa proses *stemming*.

Selanjutnya adalah penelitian yang dilakukan oleh Fitriyyah, *et al* [2] berjudul “Analisis Sentimen Tweet Terhadap Calon Presiden Indonesia 2019 dengan Menggunakan metode Naive Bayes”. Diperoleh hasil dengan menguji tiga dan dua untuk masing-masing pasangan calon (Paslon). Pada uji tiga kelas, pasangan calon 01 dan pasangan calon 02 memperoleh hasil 64,6 dan 58%. Pada pengujian dua kelas, pasangan calon 01 dan pasangan calon 02, diperoleh hasil akurasi sebesar 77,7% dan 88%. Dari hasil tersebut dapat dilihat bahwa Naive Bayes memiliki performa yang cukup baik.

Berdasarkan penelitian terkait, terdapat tiga metode terbaik yang dihasilkan dari masing - masing jurnal. Metode tersebut adalah SVM, Naive Bayes, dan Random Forest. Sehingga pada penelitian ini, menggunakan ketiga metode untuk menguji algoritme yang terbaik pada analisis sentimen terhadap kasus pelecehan seksual. Selain itu, dari pengujian ketiga algoritme, diharapkan mengetahui kecenderungan stigma masyarakat berbasis kecerdasan buatan pada platform media sosial.

## III. METODE PENELITIAN

### A. Subjek Penelitian

Penelitian yang dilakukan adalah analisis sentimen terhadap media sosial YouTube dengan studi kasus kekerasan seksual yang menghasilkan kesimpulan mengenai tanggapan masyarakat terhadap kasus tersebut. Data yang dipakai merupakan dataset menggunakan format CSV yang diambil menurut channel YouTube Narasi menggunakan judul “Kasus

Bunuh Diri NW: Bripda Randy Tersangka, Penanganan Polisi Dikritik | Narasi Newsroom”.

### B. Algoritme Pembelajaran Mesin

#### 1. Support Vector Machine

Support Vector Machine (SVM) merupakan jenis algoritme pembelajaran mesin (supervised learning) [9]. SVM melakukan hasil dari proses pelatihan sebagai prediksi kelas. Pada proses algoritme SVM terdapat pengelompokan data menjadi dua atau lebih. Pengelompokan data untuk menemukan permukaan keputusan berupa garis pemisah yang disebut dengan *hyperlane*. Tujuan algoritme SVM adalah sebagai penemu garis pemisah *hyperlane* yang optimal, dengan adanya pengelompokan maka garis pemisah digunakan untuk memisahkan masing-masing area kelas positif, negatif, dan netral [8].

Tahap-tahap dalam penghitungan menggunakan algoritme SVM yaitu [16]:

- Menetapkan kata paling banyak keluar pada setiap komentar sebagai data yang akan digunakan.
- Menetapkan variabel awal sebagai penentu nilai  $\alpha=0.5$ ,  $C=1$ ,  $\lambda=0.5$ ,  $\gamma=0.5$ , dan  $\epsilon=0.001$ .
- Menentukan matriks dengan rumus (1):

$$D_{ij} = y_i y_j (K(x_i \cdot x_j) + \frac{2}{\lambda}) \quad (1)$$

Keterangan:

$D_{ij}$  : elemen matiks data ke-ij

$Y_i$  : kelas atau label data ke-i

$Y_j$  : kelas atau label data ke-j

$\lambda$  : turunan batas teoritis

- Untuk data ke  $n=1,2,3$  gunakan persamaan (2),(3),(4) berikut ini:

$$E_i = \quad (2)$$

$$\sum a_i n_j = 1 \quad D_{ij} \quad (3)$$

$$\delta a_i = \min \{ \max[y(1 - E_i), -\alpha_i], C - \alpha_i \} \quad (3)$$

$$\alpha_i = \alpha_i + \delta a_i \quad (4)$$

Keterangan:

$E_i$  : nilai error data ke-i

$Y$  : tingkat pembelajaran

$\text{Max}_{(i)} D_{ij}$  = nilai maksimum diagonal matriks hessian

- Mencari nilai simpangan (b) dengan menggunakan persamaan (5)

$$b = -\frac{1}{2} [w \cdot x + w \cdot x -] \quad (5)$$

- Pengujian pada data yang dipakai
- Perhitungan keputusan dengan ketentuan (6):

$$h(x) = \begin{cases} +1, & \text{if } w \cdot x + b \geq 0 \\ -1, & \text{if } w \cdot x + b < 0 \end{cases} \quad (6)$$

Perhitungan Hasil Jika hasil  $\geq 0$ , nilai tanda  $h(x)+1$  termasuk dalam tipe kelas yang bernilai positif. Jika hasil perhitungan keputusan adalah  $\leq 0$ , nilai tanda  $h(x)$  adalah  $-1$  dan termasuk dalam kelas nilai negatif. Keputusan dihitung dengan menerapkan persamaan (7)(8).

$$h(x) = w \cdot x + b \quad (7)$$

atau

(8)

$$\sum \alpha_i y_i K(x, x_i) m_i = 1 + b$$

2. Random Forest

Random Forest merupakan metode Pembelajaran mesin dengan konsep *supervised* untuk klasifikasi. Algoritme Random Forest mengkombinasikan prediksi berdasarkan *Multiple Decision Tree* [7]. Algoritme RF memiliki tujuan untuk membangun pohon keputusan yang terdiri dari simpul-simpul akar yang merupakan akar dari pohon keputusan. Lalu ada simpul cabang internal dengan satu input dan setidaknya dua output. Akhirnya, ada simpul daun. Ini adalah simpul terakhir tanpa jalan keluar dan hanya satu pintu masuk [6]. Dimana pohon keputusan diperoleh dari atribut dan data secara random sesuai dengan kebutuhan. Perhitungan nilai perolehan informasi dengan rumus matematika (9) [17]:

$$Entropy(Y) = \sum_v p(Y) \log_2 p(Y) \tag{9}$$

Y adalah himpunan pada kasus dengan  $p(c|Y)$  adalah proporsi nilai Y pada kelas c.

Information Gain (Y, a)

$$= Entropy(Y) - \sum_{v \in Values(a)} \frac{|Y_v|}{|Y|} Entropy(Y_v) \tag{10}$$

Nilai (a) adalah nilai peluang yang terdapat pada himpunan kasus a.  $Y_v$  adalah anak dari kelas Y, dan kelas v berasosiasi dengan kelas a.  $Y_a$ , nilai totalnya sama dengan a dalam kumpulan data dan mewakili tipe nilai atribut dari database yang digunakan.

3. Naive Bayes

Naive Bayes merupakan salah satu Algoritme klasifikasi yang digunakan untuk menentukan dan menghitung probabilitas anggota dari suatu *class*. Naive Bayes menggunakan model independen, dimana data berdiri sendiri sehingga tidak berkaitan dengan atribut yang lain [18]. Algoritme ini melatih data yang kemudian diproses untuk membentuk pengetahuan berupa nilai probabilitas [16]. Kelebihan dari Algoritme ini adalah dapat memproses data yang bersifat diskrit maupun *continue* [19]. Algoritme Naive Bayes memiliki persamaan umum, dapat dilihat pada persamaan berikut(11).

(11)

$$P(X) = \frac{P(C).P(C)}{P(x)}$$

Keterangan

- X : Data yang belum memiliki *class*.
- C : Asumsi data variabel X dari kelas yang spesifik
- P(C|X): Probabilitas asumsi C sesuai dengan kondisi variabel X
- P(C) : Probabilitas sebelum C
- P(X|C): Probabilitas variabel X berdasar keadaan C
- P(X) : Probabilitas dari variabel X

Persamaan 11 dapat menganalisis variabel sehingga Naive Bayes dapat mengklasifikasikan data dengan baik dan benar [20] [21].

B. Metode Pengumpulan Data

Metode pengumpulan data yang digunakan dalam

penelitian ini adalah Teknik *crawling*. *Crawling* sendiri merupakan suatu program yang otomatis dapat mengumpulkan data ataupun informasi yang diperlukan [22]. Data dikumpulkan dengan aplikasi Python dengan mengakses API YouTube yang disediakan oleh Google yang kemudian diberi label sesuai dengan kebutuhan [17].

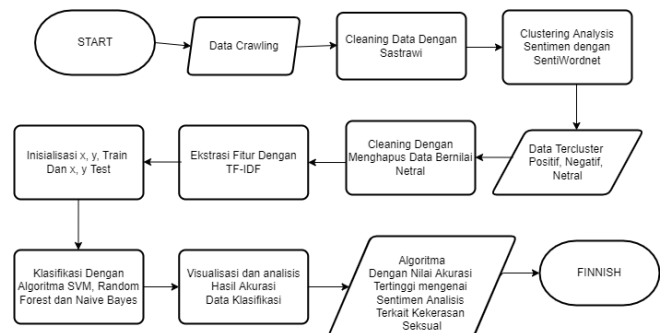
C. Alat dan Bahan

Penelitian ini menggunakan alat dan bahan untuk analisis sentimen diantaranya berupa *hardware* dan *software*:

- 1) Smartphone
- 2) Laptop
- 3) Google Colab
- 4) Data *crawling* komentar YouTube
- 5) VSCod (Visual Studio Code)

D. Alur Metode

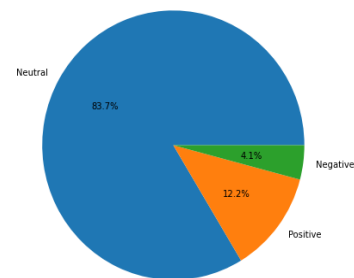
Alur yang dilakukan dalam penelitian ini dengan mengcrawling data dari komentar YouTube yang kemudian di *cleaning*. Setelah itu data yang sudah di *cleaning* akan di cluster dan di *split* menjadi data *train* dan data *test*. Setelah pengolahan data selesai, data diklasifikasikan menggunakan tiga Algoritme yaitu SVM, Random Forest, dan Naive Bayes. Pada ketiga Algoritme tersebut, akan dianalisis hasil akhirnya untuk menentukan algoritma mana yang terbaik seperti pada Gbr. 1.



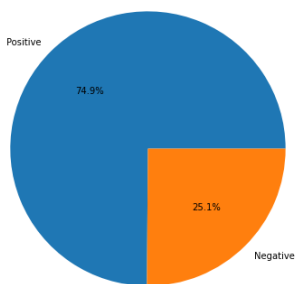
Gbr. 1 Alur Metode

VII. HASIL DAN PEMBAHASAN

Data hasil *crawling* dilakukan *cleaning* menggunakan Sastrawi dan SentiWordNet untuk menghapus simbol dan karakter yang tidak diperlukan. Kemudian mengelompokkan data dan membaginya menjadi tiga kelas: positif, negatif, dan netral. Gbr. 2 menunjukkan hasil dari data yang sudah dikelompokkan menjadi tiga kelas. Setelah pembagian kelas selesai, kelas netral dihilangkan atau dihapus.



Gbr. 2 Data dalam tiga kelas (positif, negatif, dan netral)



Gbr. 3 Data dalam dua kelas (positif dan negatif)

Gbr. 3 menunjukkan hasil dari penghapusan kelas netral pada data yang telah terkluster. Gbr. 3 dapat dilihat bahwa data dengan kelas negatif sebesar 25,1% dan kelas positif sebesar 74,9%. Data tersebut akan ditampilkan menjadi WordCloud untuk mengetahui kata positif dan negatif yang sering muncul seperti pada Gbr. 4 dan Gbr. 5.



Gbr. 4 WordCloud positif



Gbr. 5 WordCloud Negatif

Data tersebut kemudian dipecah menjadi data latih dan data uji menggunakan vektor TF-IDF. Kemudian data akan ditraining menggunakan tiga Algoritme klasifikasi, yaitu SVM, Random Forest, dan Naive Bayes. Untuk proses pertama menggunakan SVM dilakukan dengan mencari Kernel terbaik diantara Kernel *linear* dan *rbf*. Hasil dari pemilihan Kernel menunjukkan kernel *linear* merupakan kernel terbaik dengan nilai parameter C adalah 10 dan score sebesar 77,9%. Kemudian data yang sudah di *split* akan di *train* dan di *test* yang menghasilkan akurasi sebesar 77%. Hasil dari klasifikasi menggunakan Algoritme SVM dapat dilihat pada Gbr. 6.

	precision	recall	f1-score	support
0	1.00	0.17	0.29	12
1	0.76	1.00	0.86	31
accuracy			0.77	43
macro avg	0.88	0.58	0.57	43
weighted avg	0.82	0.77	0.70	43

Gbr. 6 Hasil akurasi dengan SVM

Selanjutnya pada Algoritme Random Forest, tahap awal yang dilakukan setelah data di *split* adalah menentukan *max\_depth* untuk mengetahui kedalaman maksimum pohon keputusan dimana hasilnya adalah bernilai 5. Setelah itu data ditrain dan di test menggunakan *max\_depth* yang ditentukan dengan hasil akhir akurasi pada Algoritme Random Forest adalah 78%. Hasil dari klasifikasi menggunakan Algoritme Random Forest dapat dilihat pada Gbr. 7.

	precision	recall	f1-score	support
0	1.00	0.33	0.50	3
1	0.75	1.00	0.86	6
accuracy			0.78	9
macro avg	0.88	0.67	0.68	9
weighted avg	0.83	0.78	0.74	9

Gbr. 7 Hasil akurasi dengan Random Forest

Pada Algoritme Naive Bayes, data yang telah di *split* kemudian di *train* dan di *test* langsung yang mendapatkan hasil akhir berupa akurasi sebesar 77%. Hasil dari klasifikasi menggunakan Algoritme Naive Bayes dapat dilihat pada Gbr. 8.

	precision	recall	f1-score	support
0	0.45	0.56	0.50	9
1	0.88	0.82	0.85	34
accuracy			0.77	43
macro avg	0.66	0.69	0.67	43
weighted avg	0.79	0.77	0.78	43

Gbr. 8 Hasil akurasi dengan Naive Bayes

Tabel I menunjukkan hasil analisis dari klasifikasi menggunakan tiga Algoritme, yaitu SVM, Random Forest, dan Naive Bayes.

TABEL I  
HASIL PERBANDINGAN TIGA ALGORITME

Algoritme	Waktu Komputasi (s)	Akurasi (%)
Naive Bayes	02,39	77
SVM	01,86	77
Random Forest	01,22	78

Dapat dilihat dari Tabel 1. Algoritme terbaik dalam klasifikasi untuk sentimen analisis adalah Random Forest dengan *max\_depth* bernilai 5 dan error yang rendah serta akurasi 78%. Random Forest memiliki akurasi terbaik karena Algoritme tersebut memproses data dengan pohon keputusan yang mempengaruhi peningkatan akurasi. Kemudian diikuti SVM dan Naive Bayes sebesar 77%. Naive Bayes memiliki akurasi rendah karena tidak bisa mengatasi apabila prediksi bernilai nol, sedangkan SVM kurang efektif apabila digunakan pada data sampel besar sehingga menghasilkan banyak error. Ketiga Algoritme tersebut memiliki waktu komputasi yang berbeda, dimana waktu komputasi tercepat adalah Algoritme Random Forest dengan 01,22 detik, diikuti SVM 01,86 detik, dan terakhir adalah Naive Bayes dengan 02,39 detik.

#### IV. KESIMPULAN

Kesimpulan dari penelitian ini adalah sebagai berikut:

- Opini masyarakat memiliki kecenderungan positif sebesar 74,9% dan negatif sebesar 25,1%. Hal tersebut

membuktikan bahwa masyarakat cenderung untuk membela korban, tetapi masih banyak juga yang menyalahkan dilihat dari nilai presentase negatif yang masih diatas 20%.

2. Algoritme dengan akurasi terbaik adalah Random Forest yaitu sebesar 78%, kemudian diikuti Naïve Bayes dan SVM sebesar 77%. Sedangkan untuk proses komputasi tercepat yaitu Random Forest dengan 01,22 detik, diikuti SVM dengan 01,86 detik, dan yang terakhir Naïve Bayes dengan 02,39 detik.

Adapun saran yang dapat dilakukan bagi penelitian selanjutnya yaitu sebagai berikut:

1. Pada penelitian selanjutnya, ditambahkan *tunning parameter* untuk peningkatan akurasi pada tiap Algoritme.
2. Penelitian selanjutnya disarankan menggunakan dimensi data yang lebih besar untuk memastikan hasil yang diperoleh lebih akurat.

#### UCAPAN TERIMA KASIH

Terima kasih kepada Fakultas Informatika, Institut Teknologi Telkom Purwokerto yang telah memberikan wadah pembelajaran Pemrosesan Bahasa Alami, dan juga kepada pihak - pihak terkait yang telah memberikan dukungan dan kontribusi pada penelitian ini sehingga terlaksana.

#### DAFTAR PUSTAKA

- [1] U. Z. NOVIANI P, R. Arifah, C. CECEP, and S. Humaedi, "Mengatasi Dan Mencegah Tindak Kekerasan Seksual Pada Perempuan Dengan Pelatihan Asertif," *Pros. Penelit. dan Pengabd. Kpd. Masy.*, vol. 5, no. 1, p. 48, 2018, doi: 10.24198/jppm.v5i1.16035.
- [2] S. N. J. Fitriyyah, N. Safriadi, and E. E. Pratama, "Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 3, p. 279, 2019, doi: 10.26418/jp.v5i3.34368.
- [3] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, 2018, doi: 10.33096/ilkom.v10i2.303.160-165.
- [4] A. A. Rahman and Y. I. Kurniawan, "Aplikasi Klasifikasi Penerima Kartu Indonesia Sehat Menggunakan," *Progr. Stud. Inform. Univ. Muhammadiyah Surakarta*, 2016.
- [5] C. Fadlan, S. Ningsih, and A. P. Windarto, "Penerapan Metode Naive Bayes Dalam Klasifikasi Kelayakan Keluarga Penerima Beras Rasta," *J. Tek. Inform. Musirawas*, vol. 3, no. 1, p. 1, 2018, doi: 10.32767/jutim.v3i1.286.
- [6] F. Komunikasi, U. M. Surakarta, J. A. Yani, and T. Pos, "Sistem Klasifikasi Variabel Tingkat Penerimaan Konsumen Terhadap Mobil Menggunakan Metode Random Forest," *J. Tek. Elektro*, vol. 9, no. 1, pp. 24–29, 2017, doi: 10.15294/jte.v9i1.10452.
- [7] S. Saadah and H. Salsabila, "Jurnal Politeknik Caltex Riau Prediksi Harga Bitcoin Menggunakan Metode Random Forest (Studi Kasus: Data Acak Pada Awal Masa Pandemi Covid-19)," *J. Komput. Terap.*, vol. 7, no. 1, pp. 24–32, 2021, [Online]. Available: <https://jurnal.pcr.ac.id/index.php/jkt/>.
- [8] A. M. Pravina, I. Cholissodin, and P. P. Adikara, "Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 2789–2797, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id>.
- [9] M. A. Maulana, A. Setyanto, and M. P. Kurniawan, "Analisis Sentimen Media Sosial Universitas Amikom," *Semin. Nas. Teknol. Inf. dan Multimed. 2018 Univ. AMIKOM Yogyakarta, 10 Februari 2018*, pp. 7–12, 2018.
- [10] M. Al Omari, M. Al-Hajj, N. Hammami, and A. Sabra, "Sentiment classifier: Logistic regression for Arabic services' reviews in Lebanon," *2019 Int. Conf. Comput. Inf. Sci. ICCIS 2019*, no. 2012, pp. 1–5, 2019, doi: 10.1109/ICCISci.2019.8716394.
- [11] N. O. F. Daeli and Adiwijaya, "Sentiment Analysis on Movie Reviews Using Information Gain and K-Nearest Neighbor," *J. Data Sci. Its Appl.*, vol. 3, no. 1, pp. 1–7, 2020, doi: 10.34818/JDSA.2020.3.22.
- [12] M. Wongkar and A. Angdresey, "Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter," *Proc. 2019 4th Int. Conf. Informatics Comput. ICIC 2019*, pp. 1–5, 2019, doi: 10.1109/ICIC47613.2019.8985884.
- [13] T. S. Kasus, P. Pada, and S. Tripadvisor, "Analisis Sentimen Pelanggan Hotel di," vol. 8106, pp. 21–29, 2021, doi: 10.20895/INISTA.V3.
- [14] M. R. A. Nasution and M. Hayaty, "Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter," *J. Inform.*, vol. 6, no. 2, pp. 226–235, 2019, doi: 10.31311/ji.v6i2.5129.
- [15] H. Dalam and P. Hukum, "Pelecehan Seksual Di Ruang Publik ( Street)," vol. 01, no. 02, pp. 56–65, 2016.
- [16] R. Tineges, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)," *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 650, 2020, doi: 10.30865/mib.v4i3.2181.
- [17] R. A. Maisal, A. N. Hidayanto, N. F. Ayuning Budi, Z. Abidin, and A. Purbasari, "Analysis of Sentiments on Indonesian YouTube Video Comments: Case Study of the Indonesian Government's Plan to Move the Capital City," *Proc. - 1st Int. Conf. Informatics, Multimedia, Cyber Inf. Syst. ICIMCIS 2019*, pp. 121–124, 2019, doi: 10.1109/ICIMCIS48181.2019.8985228.
- [18] F. Tempola, M. Muhammad, and A. Khairan, "Perbandingan Klasifikasi Antara KNN dan Naive Bayes pada Penentuan Status Gunung Berapi dengan K-Fold Cross Validation," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 5, p. 577, 2018, doi: 10.25126/jtiik.201855983.
- [19] H. Nalattissifa, W. Gata, S. Diantika, and K. Nisa, "Perbandingan Kinerja Algoritma Klasifikasi Naive Bayes, Support Vector Machine (SVM), dan Random Forest untuk Prediksi Ketidakhadiran di Tempat Kerja," *J. Inform. Univ. Pamulang*, vol. 5, no. 4, p. 578, 2021, doi: 10.32493/informatika.v5i4.7575.
- [20] Y. I. Kurniawan, "Perbandingan Algoritma Naive Bayes dan C.45 dalam Klasifikasi Data Mining," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, p. 455, 2018, doi: 10.25126/jtiik.201854803.
- [21] Yunitasari, H. S. Hopipah, and R. Mayasari, "Optimasi Backward Elimination untuk Klasifikasi Kepuasan Pelanggan Menggunakan Algoritme k-nearest neighbor (k-NN) and Naive Bayes," *Technomedia J.*, vol. 6, no. 1, pp. 99–110, 2021, doi: 10.33050/tmj.v6i1.1531.
- [22] F. A. Suharno and L. Listiyoko, "Aplikasi Berbasis Web dengan Metode Crawling sebagai Cara Pengumpulan Data untuk Mengambil Keputusan," *Semin. Nas. Rekayasa Teknol. Inf.*, no. November, pp. 105–109, 2018.