

ANALISIS SENTIMEN TENTANG IMPLEMENTASI HAK ASASI MANUSIA DI INDONESIA PADA MEDIA SOSIAL TWITTER MENGGUNAKAN NAÏVE BAYES DAN SUPPORT VECTOR MACHINE

Magdalena Teofila Usniaty¹, Yudi Setyawan^{2*}, Rokhana Dwi Bekti³
Jurusan Statistika, Fakultas Sains Terapan, Institut Sains & Teknologi AKPRIND Yogyakarta
Email : setyawan@akprind.ac.id
*corresponding author

Abstract. *Human rights are a set of rights inherent in the essence and existence of human beings that must be respected, upheld and protected. Several cases of human rights violations in Indonesia are often trending topics on twitter, therefore in this study a sentiment analysis was conducted about the implementation of human rights in Indonesia on social media twitter. This analysis was conducted to classify tweets containing public sentiment regarding the implementation of human rights in Indonesia. The classification methods used in this research are Naïve Bayes Classifier and Support Vector Machine and TF.IDF to make weighting. The source of the data in this research was taken from twitter with keywords #hakasasimanusia and #HAM using crawling method. The results of the analysis showed that the percentage of negative sentiment was 81.2% and positive sentiment was 18.8%. The best method in classifying the implementation of human rights in Indonesia is Support Vector Machine, this is from the greater accuracy value of the Support Vector Machine which is 87% compared to the accuracy of Naïve Bayes Classifier which is 81%. Suggestion for improving accuracy in this research is to conduct thorough data preprocessing, such as adding vocabulary for word normalization and feature reduction through stopword removal.*

Keywords: *Sentiment Analysis, Human Rights, Naïve Bayes Classifier, Support Vector Machine*

Abstrak. Hak asasi manusia adalah seperangkat hak yang melekat pada hakikat dan keberadaan manusia yang wajib dihormati, dijunjung tinggi dan dilindungi. Beberapa kasus pelanggaran HAM di Indonesia sering menjadi trending topik *twitter*, oleh karena itu pada penelitian ini dilakukan analisis sentimen tentang implementasi HAM di Indonesia pada media sosial *twitter*. Analisis ini dilakukan untuk mengklasifikasikan *tweet* yang berisi sentimen masyarakat mengenai implementasi HAM di Indonesia. Metode klasifikasi yang digunakan dalam penelitian ini adalah *Naïve Bayes Classifier* dan *Support Vector Machine* dengan menggunakan pembobotan TF.IDF. Sumber data dalam penelitian ini diambil dari *twitter* menggunakan metode *crawling* dengan kata kunci #hakasasimanusia dan #HAM. Hasil analisis menunjukkan bahwa persentase sentimen negatif sebesar 81.2% dan sentimen positif sebesar 18.8%. Metode terbaik dalam melakukan klasifikasi implementasi HAM di Indonesia adalah *Support Vector Machine*, dilihat dari nilai akurasi *Support Vector Machine* yang lebih besar yakni sebesar 87% dibandingkan nilai akurasi *Naïve Bayes Classifier* yakni sebesar 81%. Saran yang diberikan dalam penelitian ini agar meningkatkan nilai akurasi yakni pada tahap pengolahan data dilakukan secara teliti, misalnya menambahkan kosa kata untuk normalisasi kata dan pengurangan fitur yaitu pada tahap *stopword removal*.

Kata Kunci : *Analisis Sentimen, HAM, Naïve Bayes Classifier, Support Vector Machine*

1. Pendahuluan

Berdasarkan Undang-Undang Nomor 39 Tahun 1999 Tentang Hak Asasi Manusia, pengertian hak asasi manusia adalah seperangkat hak yang melekat pada hakikat dan keberadaan manusia sebagai makhluk Tuhan Yang Maha Esa dan merupakan anugerah-Nya yang wajib dihormati, dijunjung tinggi dan dilindungi oleh negara, hukum, pemerintah, dan setiap orang demi kehormatan serta perlindungan harkat dan martabat manusia. Di Indonesia sendiri sering terjadi masalah yang berkaitan dengan hak asasi manusia yang menarik perhatian publik di media sosial *Twitter*. Beberapa masalah tersebut antara lain, kerangkeng manusia di rumah Bupati Langkat yang terjadi pada bulan Januari tahun 2022, kasus ini diketahui saat Bupati Langkat terjaring operasi tangkap tangan dari pihak KPK, kerangkeng manusia ini digunakan untuk rehabilitasi narkoba, namun belum ada ijin terkait tempat rehabilitasi narkoba tersebut, ditemukan 26 bentuk

penyiksaan, kekerasan, dan perlakuan merendahkan martabat manusia yang dilakukan terhadap penghuni kerangkeng dan diduga terdapat keterlibatan dari oknum TNI-POLRI. Kasus lainnya yakni penembakan Brigadir J oleh Irjen Ferdy Sambo yang terjadi pada tanggal 8 Juli 2022 di Kompleks Rumah Dinas Polri, Jalan Duren Tiga Utara, Jakarta Selatan. Kasus ini banyak menarik perhatian publik hingga menjadi trending topik di sosial media *Twitter* dan semakin menarik banyak perhatian dikarenakan adanya rekayasa skenario yang dibuat oleh Ferdi Sambo. Berdasarkan penyelidikan, Komnas HAM menyatakan bahwa Sambo telah melakukan pelanggaran HAM berupa penghilangan hak untuk hidup dan hak memperoleh keadilan.

Analisis sentimen merupakan analisis yang digunakan untuk mengidentifikasi sebuah sentimen yang dimuat dalam sebuah teks dengan tujuan untuk melihat apakah teks tersebut menunjukkan sentimen positif atau negatif terhadap suatu subjek (Nasukawa & Yi, 2003). Tujuannya adalah untuk memperoleh opini dari pengguna yang terdapat pada platform tersebut. Pada penelitian ini digunakan metode *Naive Bayes Classifier* dan *Support Vector Machine (SVM)*. *Naive Bayes Classifier* merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasikan data uji pada kategori yang paling tepat, sedangkan *Support Vector Machine* adalah metode pada *machine learning* yang dapat digunakan untuk menganalisis data dan mengurutkannya ke dalam salah satu dari dua kategori. Informasi mengenai sentimen masyarakat terhadap hak asasi manusia sangat penting karena peran masyarakat penting dalam penegakan HAM dan tanpa adanya peran masyarakat maka HAM akan berjalan sia-sia. Partisipasi dan peran masyarakat juga diatur dalam UU No. 39 Tahun 1999. Pembentukan undang-undang tersebut merupakan perwujudan tanggung jawab dari masyarakat Indonesia.

Beberapa penelitian tentang analisis sentimen antara lain penelitian yang dilakukan oleh Tuhuteru dan Iriani (2018). Dalam penelitian ini peneliti melakukan pengujian pada *K-Fold Cross Validation* yang memiliki pengaruh dalam meningkatkan akurasi metode klasifikasi *Naive Bayes Classifier* dan *Support Vector Machine*, sehingga nilai akurasi terbaik didapatkan berdasarkan nilai *fold* yang paling baik yakni 2 *fold*. Kelemahan pada penelitian ini adalah kurangnya variasi dalam pembagian data *training* dan *testing* sehingga nilai akurasi dari hasil klasifikasi kurang bervariasi dan menyebabkan tidak terdapat perbandingan nilai akurasi yang lebih baik. Penelitian yang dilakukan oleh Hasry dan Alita (2022). Penelitian ini membahas tentang pendapat Masyarakat mengenai kebijakan-kebijakan yang telah dibuat oleh pemerintah pusat dan daerah tentang PPKM di tahun 2021. Kelebihan pada penelitian ini adalah peneliti melakukan pengujian pada *K-Fold Cross Validation* yang diujikan sebanyak 1-10 kali untuk meningkatkan akurasi metode klasifikasi *naive Bayes Classifier* dan *Support Vector Machine*. Kelemahan pada penelitian ini adalah kurangnya variasi dalam pembagian data *training* dan *testing* sehingga nilai akurasi dari hasil klasifikasi kurang bervariasi hal ini menyebabkan tidak terdapat perbandingan nilai akurasi yang lebih baik. Penelitian yang dilakukan oleh Ramadhan dan Setiawan (2019). Penelitian ini membahas tentang sentimen penonton terhadap kepuasan program acara televisi di salah satu saluran televisi swasta yaitu SCTV (Surya Citra Televisi). Kelebihan pada penelitian ini adalah peneliti melakukan pengujian pengujian akurasi menggunakan *n-gram* untuk meningkatkan akurasi metode klasifikasi *Naive Bayes Classifier* dan *Support Vector Machine*. Kelemahan pada penelitian ini adalah peneliti tidak melakukan normalisasi data pada tahap *praprocessing* data sehingga mengurangi konsistensi dan keseragaman kata. Penelitian yang dilakukan oleh Fitriana, Utami, dan Fatta (2021). Penelitian ini membahas mengenai sentimen masyarakat tentang vaksin *Covid-19*. Kelebihan pada penelitian ini adalah peneliti melakukan pengujian pada *K-Fold Cross Validation* yang diujikan sebanyak 5 kali perulangan untuk meningkatkan akurasi metode klasifikasi *Naive Bayes Classifier* dan *Support Vector Machine*. Kelemahan pada penelitian ini adalah kurangnya variasi dalam pembagian data *training* dan *testing* sehingga nilai akurasi dari hasil klasifikasi kurang bervariasi hal ini menyebabkan tidak terdapat perbandingan nilai akurasi yang lebih baik. Pada penelitian sebelumnya, peneliti hanya melakukan satu kali pengacakan dan menggunakan satu proporsi data dalam melakukan pembagian data *training* dan *testing* sehingga tidak terdapat rata-

rata nilai akurasi terbaik. Sedangkan pada penelitian ini menggunakan 30 *set seed* dalam melakukan pengacakan dan dua proporsi data yakni 80:20 dan 70:30 pada saat pembagian data *training* dan *testing* untuk mendapatkan nilai rata-rata akurasi terbaik.

2. Metode

2.1 Desain Penelitian

Metode penelitian yang digunakan pada penelitian ini adalah deskriptif kuantitatif. Pemilihan deskriptif kuantitatif pada penelitian ini didasarkan pada penelitian yang ingin mengkaji sentimen tentang implementasi HAM di Indonesia pada media sosial Twitter. Pada penelitian ini beberapa tahapan rancangan penelitian yang dilakukan yakni :

1. Identifikasi topik penelitian dan tinjauan literatur untuk memahami konteks dan solusi yang sudah ada.
2. Perumusan masalah yang akan diidentifikasi.
3. Pengumpulan data.
4. *Pra-processing* data meliputi pembersihan dan persiapan data.
5. Pelabelan data.
6. Analisis deskriptif.
7. Analisis data dan Interpretasi hasil.
8. Penulisan laporan.
9. Presentasi dan diskusi hasil penelitian.
10. Publikasi hasil penelitian.

2.2 Objek Penelitian

Objek penelitian mencakup kalimat *tweet* yang mengandung opini, pandangan, atau sentimen terkait implementasi HAM. Kalimat *tweet* ini didapatkan dari hasil *Crawling* data menggunakan *Software Python*.

2.3 Metode Pengumpulan Data

Dalam penelitian ini data yang digunakan adalah data sekunder. Metode yang digunakan dalam pengumpulan adalah metode *crawling Twitter* yang menggunakan *Software Python*. *Crawling Twitter* adalah proses mengumpulkan informasi dari *Twitter* secara otomatis. Kata kunci dalam *Crawling Twitter* pada penelitian ini adalah #hakasimanusia dan #HAM, dengan rentang pada tanggal 01 bulan Januari tahun 2022 hingga 06 Juni tahun 2023.

2.4 Variabel

Variabel yang digunakan dalam penelitian ini antara lain *Tweet* yakni kalimat *tweet* yang telah melewati tahap *pre-processing* data, *Polarity* yakni klasifikasi pada *tweet* yang telah diberikan label positif, netral dan negatif dan *Score* yakni nilai klasifikasi yang dihitung berdasarkan kamus kata positif dan negatif, hasil skoring dibagi menjadi dua kelas yakni kelas positif jika memiliki score > 0 dan negatif jika memiliki score < 0 .

2.5 Metode Analisis

Dalam penelitian ini, dilakukan serangkaian langkah sebagai berikut.

1. Pertama, dilakukan pengambilan data dari *Twitter* dengan teknik *crawling*. Selanjutnya, data hasil *crawling* tersebut diolah dan diproses untuk persiapan analisis selanjutnya.
2. Setelah itu, dilakukan pelabelan pada data hasil *crawling* untuk memberikan label sentimen pada setiap data.
3. Selanjutnya, data dibagi menjadi dua bagian, yaitu data *training* dan data *testing*, untuk melatih dan menguji model yang akan digunakan.
4. Selanjutnya, dilakukan pembobotan menggunakan metode TF-IDF untuk memberikan bobot pada setiap kata dalam teks.
5. Kemudian, dilakukan klasifikasi menggunakan metode *Naive Bayes* dan *Support Vector Machine* untuk memprediksi sentimen dari teks.

6. Hasil klasifikasi tersebut kemudian dievaluasi menggunakan *Confusion Matrix* untuk mengukur akurasi dan performa model.
7. Langkah terakhir adalah melakukan visualisasi hasil akhir dari analisis. Visualisasi ini dapat membantu dalam memahami hasil analisis secara lebih intuitif dan memberikan gambaran yang jelas tentang sentimen yang terkait dengan teks yang dianalisis.

3. Hasil dan Pembahasan

3.1 Pengumpulan Data

Pada penelitian ini pengumpulan data *Twitter* tentang implementasi HAM di Indonesia menggunakan metode *crawling* dan *Software* yang digunakan adalah *Software Python*. Data hasil *crawling* didapatkan sebanyak 2171 data, yang diambil dari rentang tanggal 01 Januari 2022 hingga 06 Juni 2023 dengan kata kuncinya adalah #hakasasimanusia dan #HAM.

3.2 *Pra-Processing* Data

3.2.1 Pembersihan Data

Pada proses ini dilakukan penghapusan *tweet* yang berupa artikel dan berita menggunakan *Excel*. Selanjutnya dilakukan penghapusan karakter-karakter yang tidak diperlukan seperti *hashtag*, alamat situs (<http://www/situs.com>), username (@username), tanda baca seperti (.,?![]/%;;<>()), angka(1,2,3,4,5,6,7,8,9,0), data duplikat, kata berulang dalam satu kalimat seperti hak-hak menjadi hak, dan karakter-karakter lain selain alphabet, hal ini dilakukan untuk mengurangi *noise* pada saat melakukan analisis data. Setelah semua langkah pembersihan dilakukan maka data akhir adalah sebanyak 515 data yang akan digunakan dalam analisis.

3.2.2 *Case folding*

Proses ini dilakukan perubahan huruf besar menjadi huruf kecil, manfaat *case folding* dalam analisis sentimen adalah untuk menghindari ambiguitas dalam teks, sehingga mesin dapat memproses kata-kata dengan benar.

3.2.3 Normalisasi data

Normalisasi data adalah mengubah kata menjadi kata baku dan mengubah kata-kata gaul atau kata singkatan menjadi kata standar, mengubah kata tidak menjadi kata tak agar kata tersebut tidak dihapus saat proses *stopword removal*.

3.2.4 *Tokenizing*

Proses ini digunakan untuk menghapus *emoticon* atau tanda lainnya yang bukan merupakan huruf dan membagi suatu kalimat menjadi kata tunggal.

3.2.5 *Stopword removal*

Proses ini digunakan untuk menghapus kata-kata yang tidak relevan, dan kata-kata yang mempunyai makna tersendiri yang tidak terkait dengan sentiment.

3.2.6 *Stemming*

Proses ini digunakan untuk mencari kata dasar dengan menghilangkan imbuhan dari kata tersebut

3.3 Pelabelan Data

Pelabelan data menggunakan *Software Rstudio*, pelabelan dilakukan dengan menghitung skor, untuk perhitungannya adalah skor kata positif dikurangi skor kata negatif dalam setiap ulasan, dimana jika $score\ tweet > 0$ diklasifikasikan sebagai sentimen positif, $score\ tweet < 0$ diklasifikasikan sebagai sentimen negatif. Pada analisis ini sentimen netral tidak digunakan karena dianggap kurang memberikan manfaat bagi pihak pemerintah Indonesia. Pelabelan ini menggunakan kamus *lexicon* positif dan negatif. Kata dalam kamus positif seperti kata apresiasi, baik, bagus, bijak, bijaksana, lindung, sayang, mantap. Kata dalam kamus negatif seperti kata agresif, adzab, benci, bencana, ganggu, jembut, keji

3.7 Hasil Klasifikasi Naive Bayes Classifier

Algoritma *Naive Bayes Multinomial* adalah salah satu variasi dari metode Naive Bayes yang umum digunakan dalam klasifikasi teks. Algoritma ini cocok digunakan saat fitur yang digunakan dalam klasifikasi adalah variabel diskrit seperti jumlah kata yang muncul. Data inputan yang memiliki representasi vektor digunakan pada analisis *Naive Bayes Classifier* ini adalah data yang didapatkan dari proses pembobotan TF. Berikut adalah penjelasan langkah-langkah dalam algoritma *Naive Bayes Classifier* menggunakan *library scikit-learn* di *Python* pada proporsi data 80:20 dan *Set Seed 2*

1. Langkah pertama adalah melakukan persiapan data
 - a) Data teks *training* dan *testing* yang telah diubah menjadi fitur numerik menggunakan pembobotan TF-IDF akan digunakan sebagai input untuk algoritma *Naive Bayes*.
2. Langkah selanjutnya adalah inialisasi dan pelatihan model
 - a) Import kelas ``MultinomialNB`` dari *library`sklearn.naive_bayes`*.
 - b) Membuat objek model menggunakan ``MultinomialNB()``.
 - c) Memanggil metode ``fit()`` pada model dengan memberikan data *training* yang sudah diubah menjadi fitur numerik menggunakan metode TF-IDF dan label yang sesuai.
 - d) Pada tahap ini, dengan menggunakan data *training* maka model akan mempelajari distribusi probabilitas dari setiap fitur dalam kelas target (positif atau negatif).
3. Langkah terakhir adalah melakukan prediksi
 - a) Setelah model dilatih, kita dapat menggunakan metode ``predict()`` pada objek model untuk melakukan prediksi pada data *testing*.
 - b) Input untuk prediksi harus berupa data *testing* yang telah diubah menjadi fitur numerik hasil dari pembobotan TF-IDF.

3.8 Hasil Klasifikasi Support Vector Machine

Pada penelitian ini digunakan *Support Vector Machine* linear. Data inputan yang memiliki representasi vektor digunakan pada penelitian ini adalah data yang didapatkan dari proses pembobotan. Tujuan dilakukan *training* pada klasifikasi *Support Vector Machine* adalah untuk menghasilkan nilai atau pola yang akan digunakan pada proses *testing Support Vector Machine* yang akan memberikan label sentimen pada data *tweet*.

Berikut adalah tahapan lengkap algoritma *Support Vector Machine* pada proporsi data 80:20 dan *Set Seed 2* dengan kernel linear:

1. Inialisasi Awal
 - a) *Import library* yang diperlukan yaitu *sklearn*.
 - b) Definisikan *parameter* seperti nilai α , C , ϵ , γ , λ , *ittermax* seperti berikut
 - 1) Parameter α awal menggunakan nilai *default* dari algoritma *Support Vector Machine* yakni $\alpha_i = 0$
 - 2) Parameter C ditentukan melalui metode *cross validation* untuk mengetahui nilai C terbaik. Nilai C yang diujikan adalah (1, 1,2,3,4,5,6,7,8,9, 10, 100), Setelah dilakukan pengujian didapatkan nilai C terbaik adalah 2 dilihat dari nilai skor rata-rata pengujian terbesar.
 - 3) Parameter ϵ menggunakan *default* dari algoritma *Support Vector Machine* itu sendiri, yakni $\epsilon = 0.001$
 - 4) Parameter γ menggunakan nilai *default* dari algoritma *Support Vector Machine* yakni $\gamma = scale$, artinya bahwa Nilai γ akan dihitung secara otomatis berdasarkan skala data input.

- 5) Parameter *Itermax* juga menggunakan *default* dari algoritma *Support Vector Machine* itu sendiri, yakni *ittermax* = -1, Nilai -1 menandakan bahwa tidak ada batasan pada jumlah iterasi. Dalam hal ini, algoritma akan terus melakukan iterasi hingga kondisi berhenti yang ditentukan tercapai, seperti mencapai toleransi kesalahan atau $C = 2$.
2. Memasukkan Data TF-IDF
 - a) Persiapkan dataset yang terdiri dari matriks TF-IDF (variabel x) dan label kelas (variabel y).
 - b) Pisahkan data menjadi data *training* dan data *testing* yakni 80:20 dan 70:30 dengan menggunakan *set seed* = 2.
3. Menghitung Matriks Kernel Linear:
Gunakan persamaan $K(x_i, x_j^T) = x_i \cdot x_j^T$ untuk menghitung matriks kernel linear (K) dari data *training* (x_{train}) dan data *testing* (x_{test})
4. Menghitung Matriks D_{ij}
Dari matriks kernel linear (K), hitung matriks D_{ij} , menggunakan rumus perhitungan matriks $D_{ij} = y_i y_j (K(x_i, x_j) + \lambda^2)$, nilai λ disini menggunakan nilai C .
5. Iterasi Pertama:
 - a) Dalam iterasi pertama, cari nilai *error* (E_i) untuk setiap data *training* menggunakan rumus sebagai berikut
$$E_i = \sum_{i,j=1}^n \alpha_i D_{ij}$$
 - b) Hitung nilai delta alpha ($\delta\alpha_i$) dengan menggunakan rumus sebagai berikut
$$\delta\alpha_i = \min\{\max[\gamma(1 - E_i), -\alpha_i], C - \alpha_i\}$$
 - c) Menghitung nilai α_i baru menggunakan rumus sebagai berikut
$$\alpha_i = \alpha_i + \delta\alpha_i$$
6. Mencari Nilai w
Dalam *Support Vector Machine* dengan kernel linear, nilai w dapat dihitung menggunakan rumus sebagai berikut
$$w = \sum_{i=1}^n \alpha_i y_i x_i$$
7. Mencari Nilai Bias
Dalam *Support Vector Machine* dengan kernel linear, nilai bias (b) dapat dihitung menggunakan rumus sebagai berikut
$$b = -\frac{1}{2} w \cdot x^+ + w \cdot x^-$$

Berdasarkan perhitungan nilai b didapatkan hasil sebagai berikut
$$b = -0.45549376$$
8. Prediksi Data *Testing*
 - a) Gunakan nilai alpha, matriks kernel linear (K), nilai w , dan nilai bias untuk memprediksi kelas dari data *testing*.
 - b) Prediksi kelas data *testing* menggunakan persamaan berikut
$$f(x) = w \cdot x^T + b$$
 - c) Jika nilai tersebut positif, klasifikasikan sebagai kelas positif. Jika nilai tersebut negatif, klasifikasikan sebagai kelas negatif.

3.9 Evaluasi

Pada tahap ini dilakukan evaluasi hasil prediksi klasifikasi menggunakan *Naive Bayes Classifier* dan *Support Vector Machine*. Dari hasil klasifikasi untuk melihat seberapa akurat hasil klasifikasi menggunakan metode tersebut maka kita lakukan pengujian akurasi menggunakan *confusion matrix* dan didapatkan bahwa rata-rata akurasi hasil klasifikasi menggunakan *Naive Bayes Classifier* pada data 80:20 adalah sebesar 81%, rata-rata nilai presisi adalah sebesar 67%, rata-rata nilai *recall* adalah sebesar 82% dan rata-rata nilai skor F1 adalah 73%. Pada data 70:30 rata-rata nilai akurasi adalah sebesar 81%, rata-rata nilai presisi adalah

sebesar 67%, rata-rata nilai *recall* adalah sebesar 81% dan rata-rata nilai skor F1 adalah 73%. Sedangkan rata-rata nilai akurasi hasil klasifikasi menggunakan metode *Support Vector Machine* pada data 80:20 adalah sebesar 87%, rata-rata nilai presisi adalah sebesar 86%, rata-rata nilai *recall* adalah sebesar 87% dan rata-rata nilai skor F1 adalah 85%. Pada data 70:30 rata-rata nilai akurasinya adalah sebesar 87%, rata-rata nilai presisi adalah 86%, rata-rata nilai *recall* adalah 87% dan rata-rata skor F1 adalah sebesar 85%. Berdasarkan perbandingan nilai akurasi pada hasil klasifikasi menggunakan metode *Naive Bayes Classifier* dan *Support Vector Machine*, diketahui bahwa metode terbaik dalam melakukan klasifikasi sentimen tentang hak asasi manusia di Indonesia pada media sosial *twitter* adalah metode *Support Vector Machine* yang memiliki nilai rata-rata akurasi tertinggi adalah sebesar 87%.

4. Kesimpulan

Berdasarkan analisis sentimen masyarakat terhadap implementasi hak asasi manusia di Indonesia didapatkan bahwa persentase sentimen negatif masyarakat terhadap implementasi HAM di Indonesia adalah sebanyak 81.2%, sedangkan sentimen positif sebanyak 18.8%. Pada sentimen negatif kata yang paling banyak muncul adalah langgar, dengan total sebanyak 98 kata yang muncul. Pada sentimen positif kata yang paling banyak muncul adalah kata lindung, dengan total sebanyak 13 kata yang muncul. Sentimen negatif yang dominan ini dapat memberikan indikasi adanya ketidakpuasan atau keprihatinan masyarakat terhadap implementasi hak asasi manusia di negara Indonesia pada periode tanggal 01 Januari 2022 hingga 06 Juni 2023. Berdasarkan rata-rata nilai akurasi pada hasil *confusion matrix* didapatkan bahwa metode *Support Vector Machine* lebih baik dalam melakukan klasifikasi implementasi hak asasi manusia di Indonesia karena memiliki rata-rata nilai akurasi lebih besar yakni sebesar 87% dibandingkan menggunakan metode *Naive Bayes Classifier* yang memiliki nilai akurasi sebesar 81%.

Ucapan Terima Kasih

Penulisan jurnal ini tidak terlepas dari bimbingan serta dukungan berbagai pihak, oleh karena itu penulis mengucapkan terima kasih kepada :

1. Bapak Drs. Yudi Setyawan, M.S., M.Sc sebagai Ketua Jurusan Statistika sekaligus sebagai dosen pembimbing satu Skripsi yang selalu membimbing dan mengarahkan dalam pelaksanaan hingga penyelesaian laporan ini.
2. Ibu Rokhana Dwi Bakti, S.Si., M.Si sebagai dosen pembimbing dua Skripsi yang berkenan mengarahkan dan memberikan bimbingan dalam pelaksanaan sampai dengan penyelesaian laporan ini.
3. Dosen-dosen Pengajar di Jurusan Statistika IST AKPRIND Yogyakarta yang telah memberikan ilmu dan bantuannya.
4. Kedua orang tua yang memberikan dukungan dan doa.

Daftar Pustaka

- adminlp2m. (2022, Februari 21). *Analisis Sentimen (Sentiment Analysis) : Definisi, Tipe Dan Cara Kerjanya*. (adminlp2m, Penyunt.) Diambil kembali dari LP2M Universitas Medan Area: <https://lp2m.uma.ac.id/2022/02/21/analisis-sentimen-sentiment-analysis-definisi-tipe-dan-cara-kerjanya/>
- Brownlee, J. (2019). How to Evaluate Machine Learning Models. Dalam *Machine Learning Mastery*.
- Cholissodin, I. (2016, November 07). *Analitik Data Tingkat Lanjut Klasifikasi Big Data*. Diambil kembali dari imamcs.lecture: <http://imamcs.lecture.ub.ac.id>

- Christian, H., Agus, M. P., & Suhartono, D. (2016, Desember). Single Document Automatic Text Summarization Using Term Frequency-Inverse Document Frequency (TF-IDF). *ComTech*, 7, 285-294.
- Fitriana, F., Utami, E., & Fatta, H. A. (2021, Mei). Analisis Sentimen Opini Terhadap Vaksin Covid-19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naive Bayes. *Jurnal Komputasi dan Informatika*, Vol. 5, 19-25. doi:<https://www.doi.org/%2010.31603/komtika.v5i1.5185>
- Hasry, C. F., & Alita, D. (2022, Juni). Penerapan Metode Naïve Bayes Classifier dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona di Twitter. *Jurnal Informatika dan Rekayasa Perangkat Lunak*, Vol. 3, 145-160.
- Leung, K. M. (2007, November 28). *Naive Bayesian Classifier*. Dipetik Oktober 25, 2022, dari <https://cse.engineering.nyu.edu>:
<https://cse.engineering.nyu.edu/~mleung/FRE7851/f07/naiveBayesianClassifier.pdf>
- Manusia, K. N. (1997). *Hak Asasi Manusia*. Jakarta: Gramedia Pustaka Utama.
- Nasukawa, T., & Yi, J. (2003). *Sentiment Analysis Capturing favorability using Natural Language Processing*. Yamato, Japan: Conference Paper. doi:DOI: 10.1145/945645.945658
- Nasution, L. M. (2017). Statistik Deskriptif. *Jurnal Hikmah*, 49-55.
- Normawati, D., & Prayogi, S. A. (2021, September). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, Volume 5, 697-711.
- Ramadhan, D. A., & Setiawan, E. B. (2019, Agustus). Analisis Sentimen Program Acara di Sctv Pada Twitter Menggunakan Metode Naive Bayes dan Support Vector Machine. *e-Proceeding of Engineering*, Vol.6, 9736-9743.
- Tuhuteru, H., & Iriani, A. (2018, September). Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine dan Naive Bayes Classifier. *Jurnal Informatika: Jurnal Pengembangan IT*, Vol. 03, 394-401. doi:10.30591/jpit.v3i3.977